2021 International Conference on Unmanned Aircraft Systems (ICUAS), 15-18 June 2021, Athens, Greece
DOI:10.1109/ICUAS51884.2021.9476765

A Review of Safe Online Learning for Nonlinear Control Systems*

Matthew Osborne[1], Hyo-Sang Shin[1], and Antonios Tsourdos[1]

***Abstract*— Learning for autonomous dynamic control systems that can adapt to unforeseen environmental changes are of great interest but the realisation of a practical and safe online learning algorithm is incredibly challenging. This paper highlights some of the main approaches for safe online learning of stabilisable nonlinear control systems with a focus on safety certification for stability. We categorise a non-exhaustive list of salient techniques, with a focus on traditional control theory as opposed to reinforcement learning and approximate dynamic programming. This paper also aims to provide a simplified overview of techniques as an introduction to the field. It is the first paper to our knowledge that compares key attributes and advantages of each technique in one paper.**

I. INTRODUCTION

Safe online learning of nonlinear dynamic systems is a significant challenge in the robotics and control systems community [1]–[7]. Although there is much interest in intelligent systems that can adapt to unforeseen environmental changes and system dynamics autonomously, attempts to realise a practical and safe online learning algorithm are incredibly challenging [8]–[10]. To learn by trial and error, there is the risk of entering unsafe states which is unacceptable for safety critical systems. Furthermore, traditional analytical techniques like Lyapunov theory used in the control systems field to guarantee stability become intractable for highly nonlinear systems which are prevalent in real world dynamics [11], [12]. The most successful adaptive control techniques rely on extensive simulation and testing to generate accurate models. However, the models are typically limited by the unknown or complex nonlinearities present in real-world environments, which reduces system performance [13]–[15]. State-of-the-art learning frameworks bridge the gap between traditional adaptive and optimal control techniques and reinforcement learning (RL) within the computer science field [16]–[20]. Traditional control theory is model-based whereas RL is classically model-free learning of discrete state-action spaces arising from dynamic programming and machine learning techniques. Nonlinear model predictive control (NMPC) uses a model to plan ahead at every time step over a time horizon to ensure tracking control, but online computation for highly dynamic systems has not yet been fully realised [21], [22]. RL for control systems like the actor-critic network are currently reliant on offline simulation using high capacity computing resources [23]. Models using RL are also dependant upon the skill of the designer in hand crafting a reward function [24] and there are also no guarantees that trained controllers operating in the real world will act safely in the presence of unseen system changes or environmental disturbances. Recently, adaptive control techniques that have combined model-free representations of system dynamics for learning unknown nonlinearities with traditional control theory have shown promising results for low-dimensional and dynamically modest scenarios [2], [25], [26]. One of the biggest challenges, however, for grid-based or tree-based algorithms such as reachability and funnels has been the *curse of dimensionality* for generating state space safety certificates of modelled behaviour and the tradeoff of conservatism in dynamics modelling vs tractable controller synthesis under safety constraints [2], [8], [27], [28]. To date, the most successful implementations of control system learning have either been via imitation learning [1], [29]–[32], a form of supervised learning, or been constrained to low-dimensional systems [11], [25], [33], [34] or quasi-linear stable systems [35]. Recent developments in contraction theory [11], [12], [36], [37], have alleviated some of the constraints in guaranteeing stability in differential form. Furthermore, adoption of advanced uncertainty estimation techniques [38]–[41] are being more widely adopted for intelligent safety-aware learning frameworks.

Nonlinear systems can be analytically or otherwise sampled to test for a region of the state space, called the region of attraction (RoA). This is a region bounded around an equilibrium point where all phase plane trajectories converge to the equilibrium point, ensuring safe operation and control of the system. The computation of the RoA is however a difficult and computationally expensive process [42] states that they are typically NP-hard due to the typically nonconvex and high dimensionality of the systems' state space dynamics. Solutions to calculate the safe regions for continuous time systems include polynomial candidate functions and sum-of-squares (SoS) techniques, continuous piecewise affine (CPA) candidate Lyapunov functions (LFs), sampling, state-space partitions and convex programming techniques [43]. The aim of the paper is to survey the current state-of-the-art techniques in safe online learning of nonlinear dynamic systems with a focus from a control systems background. To highlight the current challenges in terms of time complexity and online learning time frames for scaleability to more complex systems. The complexity and mathematical proofs typical of these techniques means that they are difficult to access for new researchers and this paper aims to provide a simplified overview as an introduction to the field.

The paper is structured as follows: Section 2 provides a generic control systems safe learning framework. Section 3

*This work was carried out under an Industrial CASE studentship jointly funded by the EPSRC and BAE Systems.

[1]Centre for Autonomous and Cyber-Physical Systems, School of Aerospace, Transport and Manufacturing, Cranfield University, MK430AL Cranfield, UK. matthew.osborne@cranfield.ac.uk

reviews the main approaches to safe online learning. Section 4 compares the benefits and limitations of each approach. Section 5 discusses future work and section 6 concludes the paper.

II. THE GENERIC CONTROL LEARNING FRAMEWORK

The generic safe learning control system framework, subject to environmental disturbances is shown in Fig. 1. A learning algorithm updates the control policy, $\pi_i(x,t)$ via optimisation of a cost function under dynamic constraints as data is collected online at discrete time steps k as $\{x_k, x_k^*, u_k, \dot{x}_k\}$ where x^* is the set point. The system is initialised with an a priori control policy π_0 and an approximation of the system dynamics $\hat{F}(.)$ which is improved over time as the system explores the state space and environmental disturbance bounds.

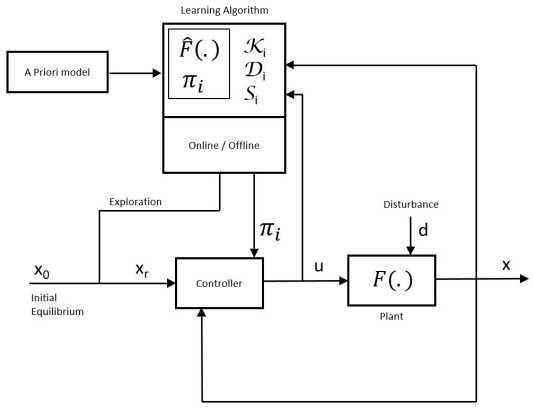

Fig. 1: Generic Safe Online Learning Control System Framework

An initial safe set, $\mathcal{S}_0$, defines where the learning algorithm can explore new states, in order to maximise learning about unknown states and to optimise the controller performance. This strategy of learning can take any form of stochastic search or Bayesian optimisation (BO) via an acquisition function. We point the reader to the following papers [2], [25], [44] for in depth set theory. We focus on system stability for control in this review, but the frameworks extend to trajectory planning and collision avoidance tasks [45], [46]. The state is assumed to be fully observable and the system equation has the following generic form:

$$\dot{x} = F(x) + B(x)\,u + B_d\left(d(x)\right) \tag{1}$$

where the state vector is $x \in \mathcal{X} \subset \mathbb{R}^n$ and control input $u \in \mathcal{U} \subset \mathbb{R}^m$, with disturbance $d \in \mathcal{D} \subset \mathbb{R}^{n_d}$. $d(x)$ is unknown but assumed bounded at each state by a compact set $\hat{\mathcal{D}}(x) \subseteq \mathcal{D}$. The environmental disturbances must be learned as the system explores the environment. The system dynamics are subject to constraints defined by a constraints set $x \in \mathcal{K} \subset \mathbb{R}^n$. Safe limits are those states and actions which keep the system within controllable states $\forall x(t) \forall u(t), t \geq 0$. The safe states are stored as the safe set denoted $\mathcal{S}_i$ which is grown over time incrementally in i steps as the system learns more about the dynamics and disturbance bounds. Model learning is performed on batches of trajectory data $\mathcal{D} = \{(x_k, u_k, F_k), k = 1:T\}$. A key feature of safety critical learning is that under uncertain conditions or in states and disturbances outside of the safe limits, the system will either fall back to a previous more cautious safe set or a more cautious policy until confidence has again been made in the stability of the system. This is achieved via control barrier functions (CBFs) described in [47]. They provide a method for pruning the state space of unsafe states by defining an acceptable set that is invariant under the control actions available to the system. Barrier functions, also known as safety functions, can be added to cost functions to avoid undesirable regions or can explicitly exclude regions of the state space as certificates, defining an unsafe set $\mathcal{G}_0$ and an initial safe set as $\mathcal{S}_0$ with a function $J(x) \geq 0 \;\forall\; x \in \mathcal{S}_0$ and $J(x) < 0 \;\forall x \in \mathcal{G}_0$, Fig. 2. Safe actions are then defined as preserving the inequality $\dot{J}(x,u) \geq -\gamma J(x)$, where $\gamma \geq 0$, by enforcing actions to remain inside the set at the boundary whilst allowing all actions inside the safe set. Control barrier functions are a generalisation of control Lyapunov functions (CLFs), where Lyapunov functions define level sets for stable trajectories and the function remains positive and centered at the origin, $\dot{V}(x,u) \leq -\alpha V(x)$. Figure 2 shows both the Lyapunov function and the barrier function forms in the same plot with the second iteration of a safe set $\mathcal{S}_\in$. Where a Lyapunov function is centered at zero the barrier function defines both a positive and negative discriminating function for safe and unsafe regions of the state space. Extensions of the generic safe framework include robust control techniques [33], [48], an extension of standard techniques to ensure that the controller performance is maintained when the environmental conditions deviate from the nominal baseline.

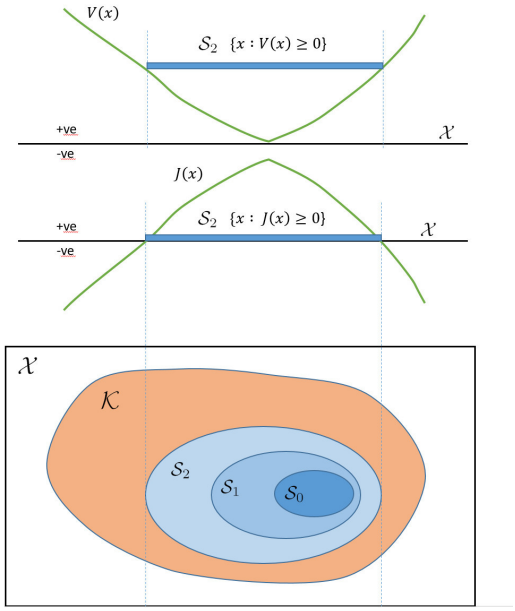

Fig. 2: Safe Set Iterations and Zero Level Set Barrier Safety Function

In Lyapunov's direct method for the function $V(x)$, stability is guaranteed if the derivative $\dot{V}(x) \leq 0$ and asymptotically stable if a strict inequality is found. Typically in control systems however exponetial stability is desired with additional convergence properties as shown by the following conditions [49], [50]:

$$V(x) > 0, \quad \forall x \in \mathcal{B} \setminus \{0\}, \quad V(0) = 0 \tag{2a}$$

$$\dot{V}(x) = \frac{\partial V}{\partial x} f(x) \leq -\alpha V(x), \quad \forall x \in \mathcal{B}, \setminus \{0\} \quad \dot{V}(0) = 0 \tag{2b}$$

where $\alpha > 0$. The stable region around the origin is described by a ball of states forming an open subset of $\mathbb{R}^n$ called set $\mathcal{B}$ where $\setminus$ means not included in the set. Global stability is guaranteed if $f(x)$ is continuous and $V(x)$ is continuously differentiable, $V(x) \to \infty$ whenever $\|x\| \to \infty$.

III. APPROACHES TO SAFE ONLINE LEARNING OF NONLINEAR SYSTEMS

Safe learning techniques can broadly be broken down into offline library-based techniques with online uncertainty handling and online trajectory optimisation techniques with safety filtering. Table I provides a subset of recent studies that consider robust learning under disturbances. We omit model-free RL techniques here that employ dynamic programming and deep learning and focus on model-based learning as we are interested in tractability and traditional control theory although there are significant merits in both fields [51]. Model-based techniques are significantly more sample efficient and faster at learning in general [51]. The techniques are typically hybrid models with nominal model-based and affine model-free uncertainty or learning component. Offline techniques to determine a library of safe states and actions include explicit Hamilton Jacobi Isaacs (HJI), [25], [52], [53], calcualtions over a discretised state space or precomputed robust trajectories in the form of funnel libraries, [54] or linearised CPA techniques that adopt linear quadratic regulator (LQR) parameter varying controllers, [55]. Online techniques adopt some form of MPC over a time horizon either directly via a nonlinear model and sequential quadratic programming (SQP), [56], [57] or as in the case of the control contraction metric (CCM) technique via optimisation of a stabilizing control trajectory using a Riemannian metric, [1], [11], [58]. Uncertainty aware (UA) techniques define algorithms that employ BO for exploration, [2], [25], [59].

TABLE I: Subset of Safe Learning Techniques

	LMPC	**RCCM**	**HJI**	**UA-BO**
Incremental Online	[56], [57]	[1], [11]	-	[2], [62]
	[7], [60]	[12], [58], [61]	-	[60], [63]
Library Based Offline	[54], [55]	[37]	[25], [52]	[25], [67]
	[64]	-	[65], [66]	-

In this section, we review some of the current state of the art techniques.

A. Linearisation Techniques

The fundamental linearised control framework is LQR with feedback control and is used widely for systems where nonlinearities can be approximated locally by linear equations. To accomodate approximations over the entire state space parameter varying techniques are used. Standard techniques to linearise system dynamics include Taylor series expansion [68] up to a suitable degree of accuracy or taking the Jacobian of the system at piecewise locations to prove stability via quadratic functions. Nonlinear dynamics that can be approximated by linear piecewise varying functions are termed Continuous Piecewise Affine (CPA) [55], [69], [70]. In CPA techniques the state space is partitioned into polyhedral regions where system constraints and optimal linear approximations are stored as a lookup table. The solution is generated offline due to the typical scale of the programs that employ mixed integer programming (MIP). These techniques are extensions of traditional gain scheduling or linear parameter varying (LPV) techniques. The generic robust linear control equation is:

$$\dot{x} = A(\alpha)x + Bu \tag{3}$$

where A and B are stabilisable and controllable system matrices via the control action $u^* = -Kx$. K is the feedback gain matrix and forms the closed-loop solution, $\dot{x} = (A - BK)x$. The cost or performance function $J(x, u)$ is quadratic in x and u via postive definite symmetric weighting terms Q and R, $J(x, u) = \frac{1}{2}\int_0^\infty (x^T Q x + u^T R u)dt$. Linear systems can be solved in real time via the algebraic Riccati equation (ARE), where $-Q = PA + A^T P - PBR^{-1}B^T P$, which generates a positive definite P matrix that certifies the system as stable over a region of the state space for a bounded uncertainty in α. The gain term K can then be calculated from the equation $K = R^{-1}B^T P$. The CLF, also termed the safety function or storage function, $V(x, t)$ is just the quadratic state equation weighted by the P matrix, $V(x, t) = x^T P x$. The linearised equation with uncertainty can also be converted from a nonconvex problem to a convex one where a congruence transformation can be used as described in [71].

B. Convex Optimisation

Convex optimisation makes use of the convex property of functions and can be solved efficiently by interior point methods or active set methods, [72]. One such function is the sum-of-squares (SoS) polynomial and has become popular for proving positivity of polynomial systems in control theory [68], [71], [73], [74]. A multivariate polynomial is a SoS if there are a set of polynomials $f_i(x), (i = 1..m)$ such that:

$$p(x) = \sum_{i=1}^{m} f_i^2(x) \tag{4}$$

The standard method involves enforcing symmetric positive definite matrix properties on Q for the quadratic equation $p(x) = x^T Q x$, which is itself a sum-of-square of the state

vector x. The technique is extended to nonlinear systems via the use of monomial functions of the state vector, $p(x) = m(x)^T Q m(x)$. The SoS formulation is therefore reduced to a linear matrix inequality (LMI) problem. Semi-definite programming (SDP) parsers and solvers such as SOSTOOLS [75], Yalmip/Mosek [76], [77] or SPOT [78] can be used to find solutions efficiently. In the case where positivity is not possible globally, which is typical in robotics systems, a local relaxation is used via the *Positivestellenzatz* or S-procedure whereby an extra polynomial term, $L(x)$ relaxes the function derivative positivity-constraint, (5c), [74].

Recently, inner approximations of the SoS matrix equations have been developed [68]. By pruning the matrix equations to prioritise the diagonal entries, a diagonally dominant algorithm DSoS and scaled diagonally dominant SDSoS algorithm have been shown to optimise solutions in faster time frames at the expense of more conservatism. DSoS and SDSoS are faster to process compared to SoS optimisation above 10 dimensions and significantly faster above 30 dimensional system optimisation. Currently the implementation is available in SPOT, [78]. Pre and post-processing of SoS has also been carried out by Löfberg, [73] in the Yalmip parser to improve optimisation times. [79] recently developed a sparse SoS (SSoS) formulation of SoS polynomials, which offers a middle ground between full SoS formulation and DSoS/SDSoS formulation and in some cases offers faster computation times as an SDP with less conservatism.

C. LQR-Trees and Funnels

Funnels can be thought of as the forward or backward propagation of RoAs in time. At discrete points in a state space trajectory, as long as the system is confined to a trajectory within the funnel, the system is verified as stable. Tedrake et al. [28], introduced the LQR-Tree algorithm for feedback motion planning that computes a tree of LQR stabilised trajectories with verified stable RoAs using SoS techniques. Majumdar et al [64] propose a library of funnel systems computed offline that can be adopted online to over come the computation of large SoS problems. The library can be used for stability and obstacle avoidance by verifying the geometric overlap of the funnels with trajectory obstacles in real time. Branicky et al [80], developed a hybrid controller using LQR funnels at each level of the controller to drive the trajectories towards a goal state within a RoA. Branicky et al. term the LQR, Time-Varying-LQR (TVLQR). A hybrid controller switches between control policies via an indexing parameter that changes the controller at fixed regions within the state space or via threshold constraints. In this manner, highly nonlinear systems or robotic systems with discrete behaviour can be controlled more easily by adopting localised controller behaviour.

D. Nonconvex Techniques

Nonconvex techniques are needed for problems with multiple optimisation parameters problem or when the system is highly nonlinear. In the first case Lyapunov techniques using SoS optimisation to find a stabilising controller where the objective includes maximising the region of attraction are bilinear [74]. Where V_{max} is the maximum level set region of attraction and $\dot{V} = \frac{\partial V}{\partial x}(F(x) + B(x)u)$. Equation 5c is linear in $L(x), U(x)$ and linear in $V(x)$ and can solved via 2 convex subproblems, alternating between the two [74].

$$\max_{V_{max}, L, V, u} \quad V_{max} \tag{5a}$$

$$\text{s.t.} \quad V(x) \quad > 0, \tag{5b}$$

$$-\dot{V}(x) + L(x)(V(x) - V^*) \quad > 0, \tag{5c}$$

$$L(x) \quad > 0 \tag{5d}$$

Several authors have adopted bilinear optimisation approaches to solve these types of problems, [1], [64], [81]. The downside is the loss in efficieny in finding a solution due to the increased complexity and that no formal guarantees can be made on the solution. The techniques also require feasible initialisations to promote convergence to local optima typically via LQR approximations [74]. Nonconvex problems can adopt nonlinear solvers that use gradient descent methods which are commonly adopted in neural networks (NNs). For safety critical applications however, the use of nonlinear solvers is generally avoided due to both latency and reliability in finding a solution, [50].

[82] uses SoS and H_∞ control techniques to approximate the HJI equation. The technique is conservative but can be solved as a convex optimisation problem for both input and disturbance max min optimisation by using dual methods and the S-procedure to enforce positivity of the approximating polynomial SoS. The technique has yet to be compared to standard techniques, however.

E. Neural Network Approaches

Neural networks (NN) are becoming more popular due to their speed in handling high dimensional function approximations. Richards et al [62], present a NN method to learn accurate safety certificates for nonlinear, closed-loop dynamical systems. The NN technique offers the least conservatism over the true safe set when compared to LQR and SoS techniques. [83], [84] use NNs with 2 steps to both calculate the maxmium RoA for a CLF and then to verify that the RoA satisfies the Lyapunov conditions. The solution sets are checked in a falsifier or verifier step with counter examples to verify the RoA. The NN approaches produce larger regions of attraction compared to SoS and LQR techniques. The disadvantage of using neural networks is that they typically require large datasets from which to learn from and the solutions are not explainable in terms of tractable algebra. Also the bias-variance tradeoff is prone to overfitting without system expert knowledge.

F. Lyapunov Based Sampling Techniques

In many systems, the online computation of a region of attraction is too computationally time-consuming or intractable. Research has been made into sampling-based techniques where the exploration of a region of attraction is

made by pointwise analysis of the system behaviour to approximate the RoA for real-time applications [42]. Bobiti et al [43], propose an automated sampling-based Lyapunov technique to define a bounded set of states using multi-resolution sampling and hyper-rectangles as basic sampling blocks. They demonstrate the performance via benchmark examples. [42] proposed a fast sampling technique for the RoA for real-time applications. The authors exploit the local continuity of sampled points to generalise the bounds and use patching theorems to bind validated Lyapunov functions to a local function. Simulation and experimental results showed that the sampling technique quickly estimated RoAs within seconds for a second-order dynamical system. The limitations of the sampling based techniques are the lack of formal guarantees on the solution and that knowledge of the function approximation must be known to choose hyperparameters suitable for efficient sampling.

G. Nonlinear Model Predictive Control Techniques

Nonlinear model predictive control (NMPC) has recently been employed for both trajectory optimisation in learning algorithms [56], and for safety certification in combination with stochastic uncertainty, [7], [59]. A recent review of MPC techniques is detailed in [60]. The NMPC scheme is based on solving an optimal control problem (OCP) at every time step and then implementing the first control input iteratively over a time horizon, T. The cost function also serves as the quadratic Lyapunov function $V(x,t)$ to ensure stability asymptotically for each trajectory solution. By nature, the NMPC problem becomes a zero order hold discrete time problem, where x and u are indexed iteratively by subscripts k so that signals x_0, x_k, u_k are shifted in time at each iteration. A detailed description can be found in [85].

$$\min_{u \in U_T(x_k)} \quad J_T(x_k,u) = \int_{t_k}^{t_k+T} (x^T Q x + u^T R u) \quad \text{(6a)}$$

$$\text{s.t.} \quad \dot{x}(t) = F(x(t)) + B(x(t))u(t)), \quad \text{(6b)}$$

$$x(t) \in X, \quad u(t) \in U, \quad \forall t \in [t_k, t_k + T]$$

The advantage of optimisation techniques is that constraints can be included as linear or nonlinear equality or inequality equations. Stochastic techniques are a significant challenge for control theory proofs however, due to the nonlinear propagation of Gaussian distributions, [60]. Recently Grandia et al. [56] have integrated CLFs with MPC for a 4D Segway robot by either enforcing exponential convergence of the Lyapunov cost function at every MPC iteration or by restricting each iteration to a Lyapunov level set. The algorithm uses SQP with a modified Hessian. They show that compared to NMPC no tuning parameters are required and the algorithm performs smoothly to set point changes, whereas a myopic CLF-QP solver causes abrupt changes in control for changes to set point commands. Mehrez et al. [57] implement a learning model predictive controller (LMPC) that linearises the dynamics locally and uses quadratic programming to minimise the lap time of a racing car. A finite time OCP is solved in 10ms with a sampling time of 10Hz and time horizon of 12s. [7] use MPC as a safety filter. [85] show how a relaxed Lyapunov asymptotic stability criteria can be used with NMPC for a holonomic 2 wheeled robot to ensure trajectory tracking to a set point. The multiple shooting algorithm is able to sythesize a controller for a prediction horizon of 2 or 3 seconds and 10 Hz sampling time for a reasonable accuracy. [86] implement a NMPC technique via the use of recurrent neural networks (NMPC/RNN) to synthesize a controller at up to 30Hz for a model helicopter autorotation maneuver. Although there are no stability certifcates the technique uses quadratic cost functions amenable to Lyapunov analysis. The technique updates the controller after a suitable number of epochs have evolved from the RNN. The possibly nonconvex optimisation is augmented with random sampling to increase convergence guarantees. Speed of optimisation is also achieved via parallel computation of algorithm equations.

H. Contraction Theory Based Techniques

Contraction theory has been developed as a differential dynamics extension of the more established control Lyapunov function (CLF) techniques, which exploit linear eigenvalue analysis [87]. A detailed description of a contraction metric formulation using sum-of-squares techniques is given in [88] and its extension to control contraction metrics (CCM) is described in [89]. Contraction metrics refer to the vector field differential, $\delta_{x(t)} = (x(t) - x^*(t))$ of the system trajectories under the control action $u(x,t)$. A CCM forms a safety certificate for stability in the sense of a differential Lyapunov function, $V(x, \delta x)$, to ensure convergence of system trajectories to an equilibrium state via online controller synthesis. A metric is computed offline or incrementally online from sample data by learning the system dynamics.

$M(x,t)$ is a Riemannian metric commonly used in robotics and high dimensional control problems. The metric defines a vector field as a manifold in the state space where the tangent space describes the flow of the system trajectories. The length $\sqrt{\delta x^T M(x,t) \delta x}$ provides a local distance measurement and notion of orthogonality.

The solution to the CCM is found from a Linear Matrix Inequality (LMI).

$$-\dot{W} + WA^T + AW - \rho BB^T < -2\lambda W \quad (7)$$

where A is the system Jacobian and B is the control matrix. The inequality is a convex optimisation in $W(x,t)$ and $\rho(x,t)$, where the contraction rate, λ, is either fixed or optimised using bisection. W and ρ are approximated by choosing suitable basis functions and then enforcing the constraints by gridding over a region of the state space [45]. The solution is also found when it is possible to structure the dynamics function as a sum-of-squares to certify Lyapunov stability under the control constraints [90]. The resulting $W(x)$ and $\rho(x)$ form the stabilising differential

control equation:

$$u(t) - u^*(t) = -K(x(t) - x^*(t)) \tag{8}$$

$$K = -\frac{1}{2}\rho(x)B^T W(x)^{-1} \tag{9}$$

In [12], the authors extend the use of CCMs to a direct adaptive control approach for stabilisable nonlinear systems with parametric matched or extended matched uncertainty. The authors in [11], extend the use of CCMs to Robust control (RCCM) using a Schur complement and Lagarange multipliers. An LMI is solved for a 2 dimensional Moore-Greitzer system in 1.3s using Yalmip and Mosek. Singh et al. [1] use a CCM approach to learn stabilisable dynamics by regularising a dynamics optimisation problem online from demonstration data. The dynamics and metric subproblems are solved simultaneously via alternation over the biconvex problem formulation using semidefinite programming, line search, and localised gridding over a number of training samples. In [45] an online solution to a 6-state planar quadrotor trajectory optimisation is found using a robust CCM with gridding over 4 state space constraints taking 60s offline to solve. The online controller optimisation is performed via pseudospectral collocation using the SNOPT solver in around 4ms, adequate for a 1s resolve time online. [12] use a control contraction metric technique using SOS optimisation in Yalmip for the Moore-Greitzer jet engine system. The metric was computed in under 0.45s for up to degree 6 polynomials. They highlight the global solution properties as being favourable over LQR solutions.

In [36] it was shown that the pseudospectral technique using Chebyshev polynomials and Clenshaw-Curtis quadrature is optimal compared to Legendre polynomials and standard Gaussian quadrature but online predetermination of polynomial degree and number of nodes is required to ensure sufficient accuracy of the solution. Recently Finsler manifolds have been combined with the CCM framework to generate controller paths not requiring minimising geodesics [91], [92]. The Finsler framework offers a larger class of metrics to be used and may be an alternative for complex systems where minimising geodesics are too expensive to be compute [91]. Tsukamoto and Chung [61] use a Neural Contraction Metric (NCM) approach for robust estimation and control that uses deep LSTM-RNN network for global approximation of an optimal contraction metric. The controller is robust to bounded disturbances and demonstrated on a Lorenz oscillator state estimation and spacecraft optimal motion planning problem.

I. Hamilton-Jacobi-Isaacs (HJI) Reachability Techniques

Reachability analysis seeks to prove that system trajectories from an initial set subject to all possible inputs and disturbances remain within safe bounds. An overview of Reachability techniques is provided in [26]. For nonlinear systems reachability techniques rely on the solutions of the Hamilton-Jacobi-Isaacs (HJI) equations to compute the safe set of states and actions. The safe, reachable set $\mathcal{S}$ is affected by the system input values $\mathcal{U}$ and a disturbance set $\mathcal{D}$. The goal is to limit the control inputs to remain within the safe set over a given time horizon [44]. The reachability equations for dynamic systems stem from optimal control theory and dynamic programming, where the dynamics function is augmented with a disturbance, $d(s)$, (10).

$$\dot{x} = F(x) + B(x)u + B_d(d(x)) \tag{10}$$

The optimal control is found from the Hamiltonian equation using differential game theory, where one player, d, is trying to maximise the function and the other, u, is trying to minimise the function.

$$H(x, \tfrac{\partial V(x,t)}{\partial x}) = \min_u \max_d \left(\tfrac{\partial V(x,t)}{\partial x} . f(x,u,d) + c(x,u,d) \right) \tag{11}$$

Reachability using HJI computations is typically done offline over a grid of the state space and is subject to the *curse of dimensionality*. The discretisation of the grid also needs to meet the accuracy of the dynamics problem at hand [52].

Gillula et al. [93], adopt machine learning techniques to increase computational efficiency. The technique adopts optimal control actions only when approaching the border of the unsafe set. At all other times, the freedom to explore actions meant a better dynamics model could be learned. [65] use a NN-Reach tubes technique for a 2D system to create safe sets in 0.383s. Further progress has been made in decomposing higher dimensional systems into subsystems using the Jordan form highlighted in [8], [9]. By solving projections of the dynamics in lower dimensional space and reconstructing higher dimensional envelopes bounded by the projections, a faster computation of reachable sets is achieved, reducing the complexity in the decomposed state dimension vector. Sensitivity to initialisation of the HJI solutions is explored in [9] and a novel technique, FaSTrack (Fast and Safe Tracking) speeds up learning by adopting a coarse model approximation for early planning and tracking model uncertainty to verify that the system is within safe bounds [94]. FaSTrack exploits the convergence properties of model approximations to the true dynamics of the system by updating the planner with the best available knowledge at the earliest possible time.

J. Uncertainty Aware Bayesian Optimisation

Uncertainty aware (UA) based techniques promote intelligent exploration via BO and active learning techniques. For a detailed description of BO, Brochu et al. provide an excellent overview [38]. An overview of active learning can be found in [95]. Model uncertainty can be used to speed up training by prioritising exploration within regions of high uncertainty or high value and can be used to distinguish safe bounds within state constraints. Safe active learning (SAL) and safe Bayesian optimisation (SBO) are extensions of these techniques to safety critical domains [96]. The techniques share the aim of exploration, determined by an acquisition function which trades off uncertainty for optimisation of an

objective function, whilst also keeping the system within safe bounds. Uncertainty provides additional information about a function approximation in the form of a variance parameter about a mean distribution as a function of the state, this is typically achieved via Gaussian process regression. Gaussian processes (GPs) provide a powerful nonparametric framework to learn approximations of functions from sample data. The interpolation technique is also known to be the best linear unbiased prediction of the intermediate function values [18], [97]. The GP is first given a prior probability distribution over function values over an input value range and upon collecting data, a posterior over function values is updated over the range of input data. A GP is a random field of values over a vector space $\mathcal{X}$.

$$f(\boldsymbol{x}) \sim \mathcal{GP}(\mu(\boldsymbol{x}), \kappa(\boldsymbol{x}, \boldsymbol{x}')) \tag{12}$$

which is described by a mean function and a positive semidefinite covariance kernel function. The most common kernel is the Gaussian, also termed squared exponential or radial basis function (RBF) given by:

$$\kappa(x, x') = \sigma_f^2 \exp(-\frac{1}{2l^2}(x - x')^2) \tag{13}$$

where l determines the horizontal length scale over which the function varies, and σ_f^2 controls the vertical variation. For a full description of the formula for noisy observations and further analysis, see [97].

Gaussian processes only become problematic if the state space is high dimensional due to the processing required to generate a mean and variance of the function via matrix computations. [98] state that the computation is exponential in the number of dimensions, meaning that practically only 10s of dimensions can be used and there is no way to reduce the exponential computation by reducing dimensional dependencies after declaring them. Although their use has become popular for model free and hybrid use [1], [59], [99], a number of techniques have been used to reduce the computational load by introducing sparsity of basis features, [1] representer points, [59] and state range intervals, [99].

The goal is to optimise a nonlinear objective function $f(x)$, via exploration of the state space, subject to system safety constraints. For example, in [25] the objective was to learn the optimum gains for motor controllers during a quadcopter ascent and descent constrained by a safe distance to the ground subject to disturbances. BO relies on a priori estimates of the model probability distribution [38]. An acquisition function trades off between exploration and exploitation.

A program called SafeOpt employs SBO, developed by Berkenkamp and Sui [99], [100].

SBO also relies on the use of GP techniques to determine system uncertainties [2], [25]. In Fisac et al. [25], a probability limit is used to switch the control policy to a safe controller when the unsafe reachable set can be reached statistically beyond the safe limits. Once the safety function and safety policy have been computed, they are stored as look-up tables. Uncertainties arise from both the system model and from the uncertainty in environmental disturbances. Kahn et al [63], use an uncertainty-aware collision prediction model. To overcome the risk of exploring collision states, the speed of the robot is reduced when there is high uncertainty. In practice, a balance between safety and optimisation is difficult to find. Recent attempts to reduce computational loads for GPs include combinations of GPs with a nominal controller to only learn nonlinearities [57], [59]. The aim is to reduce sampling times, to be in the millisecond range. Hewing et al, use sparse GPs by selecting just 10 inducing inputs per time step heuristically along the approximate state input trajectory to maintain sampling times to 20ms. Further efficiencies could be made by selectively updating the most important parameters while using less critical parameters from the previous time step to update the dynamics model. Wabersich [101] uses a linear quadratic control Lyapunov function from sampled data points and convex optimisation within a safety framework. The sampled data is further enhanced by adopting the use of a GP model to provide a priori uncertainty. Experiments show safe set computation in 0.2s.

IV. COMPARISON OF TECHNIQUES

In this section we summarise the differences in approaches and describe some of the benefits and limitations for each of the techniques discussed.

HJI techniques are really offline techniques for nonlinear problems especially above 4 or 5 dimensions where processing times are in the 10^4 seconds. They provide one of the most accurate and tractable solutions through the use of barrier functions and provide a useful benchmark for the research of other online techniques.

Recently QP solvers have been optimised for embedded applications [70] with sub-millisecond solver times. For nonlinear systems however, multivariable semi-infinite optimisation tasks are a significant challenge.

NMPC offers attractive performance and safety verification properties by validating trajectories online but at the expense of slower response times unsuitable for highly dynamic, high dimensional or stiff systems. They also require tuning typically of the hyperparameters to determine the most accurate controllers online. In practice optimising hyperparameters on the fly is not be possible and response times in the 10s of seconds not practical for real-time dynamic systems.

By adopting a linear differential analysis of control systems dynamics in the form of CCMs, a broader class of systems can be characterised in reasonable online learning time frames via a pointwise LMI over a grid or via bi-convex optimisation and line searching, [1]. Time frames in the order of $10^0 s$ to $10^2 s$ with controller synthesis response times in the $10^{-3} s$ to $10^{-1} s$ time range. In terms of scaleability CCM techniques are a middle ground between linear techniques and nonlinear MPC solutions. A few drawbacks to CCM techniques are that SoS techniques create an exponential increase in coefficients for increasing dimensionality and

polynomial degree and in general research has shown that some dynamical systems do not admit polynomial Lyapunov functions of any degree [83]. CCM control synthesis is also a non-trivial optimisation and requires a higher level of abstraction and mathematical understanding of the problem formulation using topological vector spaces.

Uncertainty aware based techniques provide probability guarantees for safety but limit the scaleability of algorithms using GP techniques and also rely on an understanding of the model priors. SAL is an active area of research for the trade off of exploration and explotation of the policy function. Uncertainty will always increase the computation of any technique by introducing additional parameters and hyperparameters and so expert judgement must be used to maximise performance.

Hybrid approaches using a nominal model and a sparse learning function have been the most successsful in recent research for both online optimisation and uncertainty handling. Sparse Gaussian process techniques are proving to be competitive when combined with fast linear MPC or SQP implementations of nominal model behaviour.

In safety critical online learning, it is not only necessary to produce solutions in fast time frames but also to have guarantees on the quality of the solutions. For this reason model-based control theory is preferable over sampling based, model-free or neural network approaches that do not lend themselves to analytical proofs or reliable convergence properties arising from numerous localised minima. That said the speed of NNs and sampling based techniques makes them attractive for research and practical implementations.

Hybrid techniques using model-based and model-free learning are a promising future research direction. Parallelisation and decomposition of systems offer engineers opportunities to overcome the challenges of state space explosion.

V. POTENTIAL FUTURE WORK DIRECTONS

Based on our survey of approaches we have identified two of the most promising techniques currently for extension to online learning problems. Namely the CCM technique and hybrid MPC techniques with residual dynamics learning. These two approaches have gained the most interest and success in real world testing and warrant further investigation. Further the limitations in scaleability owing to the choice of optimisation technique and function classes are of interest and will be investigated further. Model reduction, parallelisation or other decomposition techniques will be investigated for the development of efficient learning algorithms and function approximations investigated for robustness to both process noise and system uncertainty.

VI. CONCLUSION

In this review we have compared state-of-the-art safe online learning algorithms for online nonlinear dynamic systems. We highlight some of the challenges in terms of performance and scaleability. The most promising developments and future directions have been discussed. This paper is intended to provide a useful introduction to the subject, an overview of the current avenues of research and the potential challenges posed by each. The research is aimed to help progress fully autonomous real-time learning for nonlinear dynamic control systems.

ACKNOWLEDGMENT

This work has been jointly funded by the EPSRC and BAE Systems under an Industrial CASE studentship. The authors would also like to thank the following researchers for their kind assistance. Sumeet Singh, Ian Manchester and Johan Löfberg.

REFERENCES

[1] S. Singh, S. M. Richards, V. Sindhwani, J.-J. E. Slotine, and M. Pavone, "Learning stabilizable nonlinear dynamics with contraction-based regularization," *The International Journal of Robotics Research*, 2020. [Online]. Available: https://doi.org/10.1177/0278364920949931

[2] F. Berkenkamp, M. Turchetta, A. P. Schoellig, and A. Krause, "Safe model-based reinforcement learning with stability guarantees," *Advances in Neural Information Processing Systems*, vol. 2017-Decem, no. Nips, pp. 909–919, 2017.

[3] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete problems in ai safety," *arXiv e-prints*, pp. 1–29, 2016. [Online]. Available: http://arxiv.org/abs/1606.06565

[4] U. Eren, A. Prach, B. B. Koçer, S. V. Rakovic, E. Kayacan, and B. Açikmese, "Model predictive control in aerospace systems: Current state and opportunities," *Journal of Guidance, Control, and Dynamics*, vol. 40, no. 7, pp. 1541–1566, 2017.

[5] H. Mo and G. Farid, "Nonlinear and adaptive intelligent control techniques for quadrotor uav – a survey," *Asian Journal of Control*, vol. 21, no. 2, pp. 989–1008, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/asjc.1758

[6] Y. Zhou, E.-J. Van Kampen, and Q. Chu, "Incremental model based heuristic dynamic programming for nonlinear adaptive flight control," *Proceedings of the international micro air vehicles conference and competition (IMAV)*, 2016.

[7] K. P. Wabersich and M. N. Zeilinger, "Safe exploration of nonlinear dynamical systems: A predictive safety filter for reinforcement learning," *arXiv*, 2018. [Online]. Available: http://arxiv.org/abs/1812.05506

[8] S. Bansal, R. Calandra, T. Xiao, S. Levine, and C. J. Tomiin, "Goal-driven dynamics learning via Bayesian optimization," *2017 IEEE 56th Annu. Conf. Decis. Control. CDC 2017*, vol. 2018-Janua, no. 1545126, pp. 5168–5173, 2018.

[9] S. L. Herbert, S. Ghosh, S. Bansal, and C. J. Tomlin, "Reachability-based safety guarantees using efficient initializations," *arXiv*, 2019. [Online]. Available: http://arxiv.org/abs/1903.07715

[10] N. Fulton and A. Platzer, "Safe reinforcement learning via formal methods: Toward safe control through proof and learning," *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, pp. 6485–6492, 2018.

[11] I. R. Manchester and J. J. E. Slotine, "Robust control contraction metrics: A convex approach to nonlinear state-feedback control," *IEEE Control Systems Letters*, vol. 2, no. 3, pp. 333–338, 2018.

[12] B. T. Lopez and J.-J. E. Slotine, "Contraction Metrics in Adaptive Nonlinear Control," *arXiv*, no. December, 2019. [Online]. Available: http://arxiv.org/abs/1912.13138

[13] H. van Hasselt, Y. Doron, F. Strub, M. Hessel, N. Sonnerat, and J. Modayil, "Deep reinforcement learning and the deadly triad," *arXiv e-prints*, 2018. [Online]. Available: http://arxiv.org/abs/1812.02648

[14] W. F. . B. U. Koch, "Flight controller synthesis via deep reinforcement learning," Ph.D. dissertation, Boston University, 2019.

[15] A. Ray, J. Achiam, and D. Amodei, "Benchmarking safe exploration in deep reinforcement learning," *SemanticScholar*, 2019. [Online]. Available: https://d4mucfpksywv.cloudfront.net/safexp-short.pdf

[16] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, and F. L. Lewis, "Optimal and autonomous control using reinforcement learning: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2042–2062, 2018.

[17] R. Kamalapurkar, P. Walters, J. Rosenfeld, and W. Dixon, *Reinforcement Learning for Optimal Feedback Control*. Springer, 2018.

[18] M. Ghavamzadeh, S. Mannor, J. Pineau, and A. Tamar, *Bayesian reinforcement learning: A survey*. Now Publishers, 2015, vol. 27, no. 5-6.

[19] W. Gu, K. P. Valavanis, M. J. Rutherford, and A. Rizzo, "A Survey of Artificial Neural Networks with Model-based Control Techniques for Flight Control of Unmanned Aerial Vehicles," in *2019 Int. Conf. Unmanned Aircr. Syst.* IEEE, 2019, pp. 362–371.

[20] A. S. Polydoros and L. Nalpantidis, "Survey of model-based reinforcement learning: Applications on robotics," *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 86, no. 2, pp. 153–173, 2017.

[21] U. Rosolia and F. Borrelli, "Learning model predictive control for iterative tasks. A data-driven control framework," *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 1883–1896, 2018.

[22] H. Chen and F. Allgöwer, "A quasi-infinite horizon nonlinear model predictive control scheme with guaranteed stability," *ECC 1997 - European Control Conference*, vol. 34, no. 10, pp. 1421–1426, 1997.

[23] D. Wang, H. He, and D. Liu, "Adaptive critic nonlinear robust control: A survey," *IEEE Transactions on Cybernetics*, vol. 47, no. 10, pp. 3429–3451, 2017.

[24] C. Finn, S. Levine, and P. Abbeel, "Guided cost learning: Deep inverse optimal control via policy optimization," *33rd International Conference on Machine Learning, ICML 2016*, vol. 1, pp. 95–107, 2016.

[25] J. Fisac, A. Akametalu, M. Zeilinger, S. Kaynama, J. Gillula, and C. Tomlin, "A general safety framework for learning-based control in uncertain robotic systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2737–2752, 2019.

[26] S. Bansal, M. Chen, S. Herbert, and C. J. Tomlin, "Hamilton-jacobi reachability: A brief overview and recent advances," *2017 IEEE 56th Annual Conference on Decision and Control, CDC 2017*, vol. 2018-January, pp. 2242–2253, 2018.

[27] J. Darbon and S. Osher, "Algorithms for overcoming the curse of dimensionality for certain hamilton–jacobi equations arising in control theory and elsewhere," *Research in Mathematical Sciences*, vol. 3, no. 1, 2016.

[28] R. Tedrake, I. R. Manchester, M. Tobenkin, and J. W. Roberts, "LQR-trees: Feedback motion planning via sums-of-squares verification," *Int. J. Rob. Res.*, vol. 29, no. 8, pp. 1038–1052, 2010.

[29] V. Sindhwani, S. Tu, and M. Khansari, "Learning contracting vector fields for stable imitation learning," *arXiv e-prints*, 2018. [Online]. Available: http://arxiv.org/abs/1804.04878

[30] M. J. Zeestraten, I. Havoutis, J. Silverio, S. Calinon, and D. G. Caldwell, "An approach for imitation learning on riemannian manifolds," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1240–1247, 2017.

[31] T. Hester, T. Schaul, A. Sendonaris, M. Vecerik, B. Piot, I. Osband, O. Pietquin, D. Horgan, G. Dulac-Arnold, M. Lanctot, J. Quan, J. Agapiou, J. Z. Leibo, and A. Gruslys, "Deep q-learning from demonstrations," *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, pp. 3223–3230, 2018.

[32] P. Abbeel and A. Y. Ng, "Exploration and apprenticeship learning in reinforcement learning," *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*, 2005.

[33] I. R. Manchester and J. J. E. Slotine, "Control contraction metrics: Convex and intrinsic criteria for nonlinear feedback design," *IEEE Transactions on Automatic Control*, vol. 62, no. 6, pp. 3046–3053, 2017.

[34] I. Michael Ross, Q. Gong, F. Fahroo, and W. Kang, "Practical stabilization through real-time optimal control," *Proceedings of the American Control Conference*, vol. 2006, pp. 304–309, 2006.

[35] S. Heyer, D. Kroezen, and E.-J. V. Kampen, *Online Adaptive Incremental Reinforcement Learning Flight Control for a CS-25 Class Aircraft*. AIAA Scitech 2020 Forum, 2020. [Online]. Available: https://arc.aiaa.org/doi/abs/10.2514/6.2020-1844

[36] K. Leung and I. R. Manchester, "Nonlinear stabilization via control contraction metrics: A pseudospectral approach for computing geodesics," *Proc. Am. Control Conf.*, pp. 1284–1289, 2017.

[37] S. Singh, V. Sindhwani, J.-J. E. Slotine, and M. Pavone, "Learning Stabilizable Dynamical Systems via Control Contraction Metrics," *arXiv e-prints*, p. arXiv:1808.00113, Jul. 2018.

[38] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv*, 2010. [Online]. Available: http://arxiv.org/abs/1012.2599

[39] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2016.

[40] Y. Sui, vincent Zhuang, J. Burdick, and Y. Yue, "Stagewise safe Bayesian optimization with Gaussian processes," *Proceedings of the 35th International Conference on Machine Learning*, vol. 80, pp. 4781–4789, 10–15 Jul 2018. [Online]. Available: http://proceedings.mlr.press/v80/sui18a.html

[41] B. Lütjens, M. Everett, and J. P. How, "Safe reinforcement learning with model uncertainty estimates," *2019 International Conference on Robotics and Automation (ICRA)*, pp. 8662–8668, 2019.

[42] E. Najafi, R. Babuška, and G. A. Lopes, "A fast sampling method for estimating the domain of attraction," *Nonlinear Dynamics*, vol. 86, no. 2, pp. 823–834, 2016.

[43] R. Bobiti and M. Lazar, "Automated-sampling-based stability verification and doa estimation for nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 63, no. 11, pp. 3659–3674, 2018.

[44] A. K. Akametalu, S. Kaynama, J. F. Fisac, M. N. Zeilinger, J. H. Gillula, and C. J. Tomlin, "Reachability-based safe learning with gaussian processes," *Proceedings of the IEEE Conference on Decision and Control*, vol. 2015-Febru, no. February, pp. 1424–1431, 2014.

[45] S. Singh, A. Majumdar, J. J. Slotine, and M. Pavone, "Robust online motion planning via contraction theory and convex optimization," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 5883–5890, 2017.

[46] A. Bajcsy, S. Bansal, E. Bronstein, V. Tolani, and C. J. Tomlin, "An Efficient Reachability-Based Framework for Provably Safe Autonomous Navigation in Unknown Environments," *Proc. IEEE Conf. Decis. Control*, vol. 2019-December, pp. 1758–1765, 2019.

[47] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, and P. Tabuada, "Control barrier functions: Theory and applications," *2019 18th European Control Conference, ECC 2019*, pp. 3420–3431, 2019.

[48] F. Berkenkamp and A. P. Schoellig, "Safe and robust learning control with gaussian processes," *2015 European Control Conference, ECC 2015*, pp. 2496–2501, 2015.

[49] H. K. Khalil and J. W. Grizzle, *Nonlinear systems*. Prentice hall Upper Saddle River, NJ, 2002, vol. 3.

[50] R. Tedrake, "Underactuated robotics: Algorithms for walking, running, swimming, flying, and manipulation (course notes for mit 6.832)." [Online]. Available: Downloaded on 02/12/2020 from http://underactuated.mit.edu/

[51] B. Recht, "A tour of reinforcement learning: The view from continuous control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 2, no. 1, pp. 253–279, 2018.

[52] M. Chen, S. L. Herbert, M. S. Vashishtha, S. Bansal, and C. J. Tomlin, "Decomposition of Reachable Sets and Tubes for a Class of Nonlinear Systems," *IEEE Trans. Automat. Contr.*, vol. 63, no. 11, pp. 3675–3688, 2018.

[53] W. Xiang, P. Musau, A. A. Wild, D. M. Lopez, N. Hamilton, X. Yang, J. Rosenfeld, and T. T. Johnson, "Verification for machine learning, autonomy, and neural networks survey," *arXiv*, pp. 1–51, 2018.

[54] A. Majumdar, R. Vasudevan, M. M. Tobenkin, and R. Tedrake, "Convex optimization of nonlinear feedback controllers via occupation measures," *The International Journal of Robotics Research*, vol. 33, no. 9, pp. 1209–1230, 2014. [Online]. Available: https://doi.org/10.1177/0278364914528059

[55] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos, "The explicit linear quadratic regulator for constrained systems," *Automatica*, vol. 38, no. 1, pp. 3–20, 2002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109801001741

[56] R. Grandia, A. Taylor, A. Singletary, M. Hutter, and A. Ames, "Nonlinear model predictive control of robotic systems with control lyapunov functions," *Robotics: Science and Systems XVI*, Jul 2020. [Online]. Available: http://dx.doi.org/10.15607/RSS.2020.XVI.098

[57] U. Rosolia and F. Borrelli, "Learning how to autonomously race a car: A predictive control approach," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 6, pp. 2713–2719, 2020.

[58] H. Tsukamoto and S. J. Chung, "Robust controller design for stochastic nonlinear systems via convex optimization," *IEEE Transactions on Automatic Control*, pp. 1–1, 2020.

[59] L. Hewing, J. Kabzan, and M. N. Zeilinger, "Cautious model predictive control using gaussian process regression," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 6, pp. 2736–2743, 2020.

[60] L. Hewing, K. P. Wabersich, M. Menner, and M. N. Zeilinger, "Learning-based model predictive control: Toward safe learning in control," *Annu. Rev. Control. Robot. Auton. Syst.*, vol. 3, no. 1, pp. 269–296, 2020.

[61] H. Tsukamoto and S. Chung, "Neural contraction metrics for robust estimation and control: A convex optimization approach," *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 211–216, 2021.

[62] S. M. Richards, F. Berkenkamp, and A. Krause, "The lyapunov neural network: Adaptive stability certification for safe learning of dynamical systems," *arXiv*, no. CoRL, 2018. [Online]. Available: http://arxiv.org/abs/1808.00924

[63] G. Kahn, A. Villaflor, V. Pong, P. Abbeel, and S. Levine, "Uncertainty-aware reinforcement learning for collision avoidance," *arXiv e-prints*, 2017. [Online]. Available: http://arxiv.org/abs/1702.01182

[64] A. Majumdar and R. Tedrake, "Funnel libraries for real-time robust feedback motion planning," *Int. J. Rob. Res.*, vol. 36, no. 8, pp. 947–982, 2017.

[65] W. Xiang and T. T. Johnson, "Reachability analysis and safety verification for neural network control systems," *arXiv*, pp. 1–21, 2018. [Online]. Available: http://arxiv.org/abs/1805.09944

[66] M. Althoff, O. Stursberg, and M. Buss, "Reachability analysis of nonlinear systems with uncertain parameters using conservative linearization," *Proceedings of the IEEE Conference on Decision and Control*, pp. 4042–4048, 2008.

[67] F. Berkenkamp, A. Krause, and A. P. Schoellig, "Bayesian optimization with safety constraints: safe and automatic parameter tuning in robotics," *arXiv preprint arXiv:1602.04450*, 2016.

[68] A. A. Ahmadi and A. Majumdar, "Some applications of polynomial optimization in operations research and real-time decision making," *Optim. Lett.*, vol. 10, no. 4, pp. 709–729, 2016.

[69] P. Patrinos and A. Bemporad, "An accelerated dual gradient-projection algorithm for embedded linear model predictive control," *IEEE Trans. Automat. Contr.*, vol. 59, no. 1, pp. 18–33, 2014.

[70] G. Cimini, D. Bernardini, S. Levijoki, and A. Bemporad, "Embedded model predictive control with certified real-time optimization for synchronous motors," *IEEE Trans. Control Syst. Technol.*, pp. 1–8, 2020.

[71] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. SIAM, 1994, vol. 15.

[72] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[73] J. Löfberg, "Pre- and post-processing sum-of-squares programs in practice," *IEEE Transactions on Automatic Control*, vol. 54, no. 5, pp. 1007–1011, 2009.

[74] A. Majumdar, A. A. Ahmadi, and R. Tedrake, "Control design along trajectories with sums of squares programming," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 4054–4061, 2013.

[75] A. Papachristodoulou, J. Anderson, G. Valmorbida, S. Prajna, P. Seiler, and P. Parrilo, "Sostools version 3.00 sum of squares optimization toolbox for matlab," *arXiv*, 2013. [Online]. Available: http://arxiv.org/abs/1310.4716

[76] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," *Proc. IEEE Int. Symp. Comput. Control Syst. Des.*, pp. 284–289, 2004.

[77] M. Aps, "Mosek modeling cookbook," 2020. [Online]. Available: https://docs.mosek.com/modeling-cookbook/index.html#%0Ahttps://www.mosek.com/

[78] A. Megretski, "Spot (systems polynomial optimization tools) manual," 2010.

[79] Y. Zheng, G. Fantuzzi, and A. Papachristodoulou, "Sparse sum-of-squares (SOS) optimization : A bridge between DSOS/SDSOS and SOS optimization for sparse polynomials," in *2019 Am. Control Conf.*, 2019, pp. 5513–5518.

[80] M. Branicky and M. Curtiss, "Nonlinear and hybrid control via RRTs," *Proc. Intl. Symp. Math. Theory Networks ...*, pp. 1–10, 2002.

[81] A. A. Ahmadi, A. Chaudhry, V. Sindhwani, and S. Tu, "Safely learning dynamical systems from short trajectories," *arXiv*, no. 1, pp. 1–19, 2020.

[82] A. P. Pang, Z. He, M. H. Zhao, G. X. Wang, Q. M. Wu, and Z. T. Li, "Sum of squares approach for nonlinear h control," *Complexity*, vol. 2018, 2018.

[83] Y. C. Chang, N. Roohi, and S. Gao, "Neural lyapunov control," *arXiv*, no. NeurIPS, 2020.

[84] R. Cheng, G. Orosz, R. M. Murray, and J. W. Burdick, "End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 3387–3395, Jul. 2019. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/4213

[85] M. W. Mehrez, K. Worthmann, J. P. Cenerini, M. Osman, W. W. Melek, and S. Jeon, "Model Predictive Control without terminal constraints or costs for holonomic mobile robots," *Rob. Auton. Syst.*, vol. 127, no. February, 2020.

[86] K. Dalamagkidis, K. P. Valavanis, and L. A. Piegl, "Nonlinear model predictive control with neural network optimization for autonomous autorotation of small unmanned helicopters," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 4, pp. 818–831, 2011.

[87] W. Lohmiller and J.-J. E. Slotine, "On contraction analysis for non-linear systems," *Automatica*, vol. 34, no. 6, pp. 683–696, 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109898000193

[88] E. M. Aylward, P. A. Parrilo, and J.-J. E. Slotine, "Stability and robustness analysis of nonlinear systems via contraction metrics and sos," *Automatica*, vol. 44, no. 8, pp. 2163–2170, 2008.

[89] I. R. Manchester and J.-J. E. Slotine, "Control contraction metrics, robust control and observer duality," *arXiv*, 2014. [Online]. Available: http://arxiv.org/abs/1403.5364

[90] I. R. Manchester, J. Z. Tang, and J.-J. E. Slotine, "Unifying robot trajectory tracking with control contraction metrics," in *Robotics Research*. Springer, 2018, pp. 403–418.

[91] T. L. Chaffey and I. R. Manchester, "Control Contraction Metrics on Finsler Manifolds," *Proc. Am. Control Conf.*, vol. 2018-June, pp. 3626–3633, 2018.

[92] F. Forni and R. Sepulchre, "A differential lyapunov framework for contraction analysis," *IEEE Trans. Automat. Contr.*, vol. 59, no. 3, pp. 614–628, 2014.

[93] J. H. Gillula, G. M. Hoffmann, H. Huang, M. P. Vitus, and C. J. Tomlin, "Applications of hybrid reachability analysis to robotic aerial vehicles," *The International Journal of Robotics Research*, vol. 30, no. 3, pp. 335–354, 2011. [Online]. Available: https://doi.org/10.1177/0278364910387173

[94] S. L. Herbert, M. Chen, S. Han, S. Bansal, J. F. Fisac, and C. J. Tomlin, "FaSTrack: A modular framework for fast and guaranteed safe motion planning," in *2017 IEEE 56th Annu. Conf. Decis. Control. CDC 2017*, vol. 2018-Janua, 2018, pp. 1517–1522.

[95] B. Settles, "Active Learning Literature Survey," University of Wisconsin-Madison, Tech. Rep. 26 January, 2010.

[96] M. Schillinger, B. Hartmann, P. Skalecki, M. Meister, D. Nguyen-Tuong, and O. Nelles, "Safe active learning and safe bayesian optimization for tuning a pi-controller," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 5967–5972, 2017. [Online]. Available: https://doi.org/10.1016/j.ifacol.2017.08.1258

[97] K. P. Murphy, *Machine Learning A Probabilistic Perspective*. London, England: MIT Press, 2012.

[98] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[99] F. Berkenkamp, R. Moriconi, A. P. Schoellig, and A. Krause, "Safe learning of regions of attraction for uncertain, nonlinear systems with gaussian processes," *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 4661–4666, 2016.

[100] Y. Sui, A. Gotovos, J. W. Burdick, and A. Krause, "Safe exploration for optimization with gaussian processes," *32nd International Conference of Machine Learning, ICML 2015*, vol. 2, pp. 997–1005, 2015.

[101] K. P. Wabersich and M. N. Zeilinger, "Scalable synthesis of safety certificates from data with application to learning-based control," *2018 Eur. Control Conf. ECC 2018*, pp. 1691–1697, 2018.

A review of safe online learning for nonlinear control systems

Osborne, Matthew

2021-07-19

Osborne M, Shin H-S, Tsourdos A. (2021) A review of safe online learning for nonlinear control systems. In: 2021 International Conference on Unmanned Aircraft Systems (ICUAS), 15-18 June 2021, Athens

https://doi.org/10.1109/ICUAS51884.2021.9476765