2018 European Control Conference (ECC)
June 12-15, 2018. Limassol, Cyprus

Scalable synthesis of safety certificates from data with application to learning-based control

Kim P. Wabersich and Melanie N. Zeilinger

Abstract— **The control of complex systems faces a trade-off between high performance and safety guarantees, which in particular restricts the application of learning-based methods to safety-critical systems. A recently proposed framework to address this issue is the use of a safety controller, which guarantees to keep the system within a safe region of the state space. This paper introduces efficient techniques for the synthesis of a safe set and control law, which offer improved scalability properties by relying on approximations based on convex optimization problems. The first proposed method requires only an approximate linear system model and Lipschitz continuity of the unknown nonlinear dynamics. The second method extends the results by showing how a Gaussian process prior on the unknown system dynamics can be used in order to reduce conservatism of the resulting safe set. We demonstrate the results with numerical examples, including an autonomous convoy of vehicles.**

I. Introduction

Digitalization opens new perspectives for control engineering and automation by making large amounts of data from experiments and numerical models available. Learning-based control exploits this cumulated knowledge and potentially also performs autonomous exploration of unseen system behavior in order to find an optimal control policy. An example is deep reinforcement learning (RL), providing prominent results, one of which is the application to Atari Arcade video-games [1].

Compared with traditional control techniques, learning-based methods offer the potential to reduce modeling and controller design effort. However, many industrial applications are *safety-critical* systems, i.e. systems with physical constraints that have to be satisfied. This essentially limits the application of most available learning-based control algorithms, which do not provide safety certificates. In order to address this limitation, we present efficient and scalable methods for the synthesis of a safety strategy consisting of a safe set and corresponding safe control law, which are cheap to implement and can be applied together with existing controllers or modern learning techniques to enhance them with safety guarantees.

Contributions: We consider dynamical systems with *unknown* Lipschitzian nonlinearity and formulate the safe set and safe controller synthesis as convex optimization problems, which directly employ available data. Our analysis considers an approximate linear model of the system and uses data to incorporate the unknown nonlinear effects. The computations are based on Lyapunov's method and result in two optimization problems: The first optimization problem defines a quadratic approximation of the nonlinearity in the Lyapunov conditions and the second one describes the computation of the safe set and controller. Similar to [2], [3] the framework can be used to augment any desired controller, which is lacking safety guarantees, in particular one which is based on learning.

We extend the technique to reduce conservatism of the safe set by putting a prior on the unknown dynamics in the form of a Gaussian process model, which is beneficial, especially in case of high dimensional systems and sparse data. Due to its less conservative nature, this extension favors safe exploration beyond the system behavior seen so far and is well suited for iteratively learning in closed-loop.

We illustrate the approach using examples, including a convoy of partly non-cooperative autonomous cars.

Related work: Given its relevance in industrial applications, there has been a growing interest in safe learning methods in the past years. Extensions of existing RL methods have been developed to enable safe RL with respect to different notions of safety, see [4] for a survey. A detailed literature review regarding RL, focusing on safety with respect to state and input constraints as also considered in this work, can be found in [3]. There are few results for efficient controller tuning from data with respect to best worst-case performance (also worst-case stability under physical constraints) by Bayesian min-max optimization, see e.g. [5], or by safety constrained Bayesian optimization as e.g. in [6], [7]. In [8] a method was developed that allows to analyze a given closed-loop system (under an arbitrary RL algorithm) with respect to safety.

Recent developments include the concept of a supervisory framework, which consists of a safe set in the state space and a safety controller. As long as the system state is in the interior of the safe set, any control law (e.g. unsafe learning-based control) can be applied. The safety controller only interferes if necessary, in case that the system reaches the boundary of the safe set, see e.g. [2], [3]. Such a framework allows for certifying an arbitrary learning-based control algorithm with safety guarantees. Previously proposed techniques are based on a differential game formulation, which results in solving a min-max optimal control problem. An active field of research aims at extending these techniques to larger scale systems, mostly by considering special cases as described, e.g., in [9], [10], [11]. For some relevant cases, the existence of analytic solutions [12] has been shown. The results presented in this paper are based on the concept

Kim Wabersich (wabersich@kimpeter.de) and Melanie N. Zeilinger (mzeilinger@ethz.ch) are with the Institute for Dynamic Systems and Control, ETH Zurich, Switzerland. This work was supported by the Swiss National Science Foundation under grant no. PP00P2 157601/1.

978-3-9524-2699-9 ©2018 EUCA

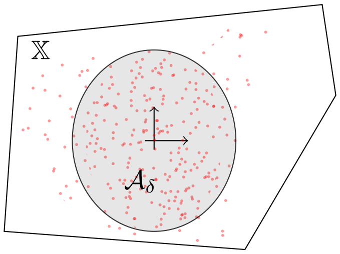

Fig. 1: Illustration of Assumption II.1. Red dots display observations $(x_i, d(x_i))$.

of a safety framework, but compared to previous work we focus on approximation techniques to improve scalability with respect to the system dimension.

Structure of the paper: In Section II we state the problem and in Section III we present our main result for safe set and controller computation. We then show an extension using a stronger assumption on the unknown system dynamics by considering Gaussian processes in Section IV in order to reduce conservatism of the safe set. The results are demonstrated on numerical examples within the respective sections. We conclude the paper in Section V.

Notation: The set of symmetric matrices of dimension n is S^n, the set of positive (semi-) definite matrices is (S^n_+) S^n_{++}, the set of integers in the interval $[a,b] \subset \mathbb{R}$ is $\mathcal{I}_{[a,b]}$, the set of integers in the interval $[a,\infty) \subset \mathbb{R}$ is $\mathcal{I}_{\geq a}$, and for $\epsilon > 0$ let $\mathcal{B}_\epsilon(\bar{x}) = \{x \in \mathbb{R} | \, \|x - \bar{x}\|_2 \leq \epsilon\}$. The boundary of an arbitrary compact set $\mathcal{C} \subset \mathbb{R}^n$ is $\partial\mathcal{C}$. Given a set $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$, let $\mathcal{D}_x = \{x_i\}_{i=1}^N$ and $\mathcal{D}_y = \{y_i\}_{i=1}^N$. Define $\Delta_{\mathcal{D}_x} : \mathbb{R}^n \to \mathcal{D}_x$ as $\Delta_{\mathcal{D}_x}(x) = \operatorname{argmin}_{\bar{x}\in\mathcal{D}_x} \|\bar{x} - x\|_2$, which picks the closest element in $\mathcal{D}_x$ with respect to $x \in \mathbb{R}^n$ under the 2-norm. Given a set $\mathcal{A} \subset \mathbb{R}^n$ and a locally Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$, the local Lipschitz constant $L \leq |f(x) - f(y)| / \|x - y\|_2$ for all $x, y \in \mathcal{A}$ is denoted by $L_{f(x)}(\mathcal{A})$. The Minkowski sum of two sets $\mathcal{A}_1, \mathcal{A}_2 \subset \mathbb{R}$ is denoted by $\mathcal{A}_1 \oplus \mathcal{A}_2$.

II. Problem description

We consider deterministic nonlinear systems of the form

$$\dot{x}(t) = Ax(t) + Bu(t) + d(x(t)) \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $d : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous. The system is subject to polytopic state constraints $x(t) \in \mathbb{X} := \{x \in \mathbb{R}^n | A_x x \leq b_x\}$, $A_x \in \mathbb{R}^{n_x \times n}$, $b_x \in \mathbb{R}^{n_x}$ and polytopic input constraints $u(t) \in \mathbb{U} := \{u \in \mathbb{R}^m | A_u u \leq b_u\}$, $A_u \in \mathbb{R}^{n_u \times m}$, $b_u \in \mathbb{R}^{n_u}$. The origin is contained in $\mathbb{X}$, (A, B) is controllable, and the system state is fully observable.

The explicit form of the nonlinearity d is unknown, hence d is assumed to be a memoryless black-box function, which will be identified from system measurements. However, we assume that B, i.e. the influences of the control input, are known. The matrix A incorporates system knowledge in form of a linear model, which will be used in the design procedure in Section III-D. The linear model can, e.g., be selected as an approximate system model. For identification of $d(x)$, we have access to finitely many observations $\mathcal{D} = \{(x_i, d(x_i))\}_{i=1}^N$ that fulfill the following property (Figure 1).

Assumption II.1. *Given a set $\mathcal{D} = \{(x_i, d(x_i))\}_{i=1}^N$ with N data tuples, where $x_i \in \mathbb{X}$, there exists a non-trivial subset $\mathcal{A}_\delta \subseteq \mathbb{X}$ such that for any $x \in \mathcal{A}_\delta$ there exists an $x_i \in \mathcal{D}_x$ such that $||x - x_i||_2 \leq \delta$.*

Remark II.2. *Assumption II.1 implies that there exists a region, in which the collected data samples are dense. Intuitively, one can think of δ together with the Lipschitz constant L, as a 'measure of knowledge' that we have about $d(x)$ inside $\mathcal{A}_\delta$. The 'knowledge' increases as δ gets smaller.*

Remark II.3. *For simplicity, we assume noise-free data $\mathcal{D}$. It would, however, be possible to incorporate bounded or stochastic noise with only minor changes.*

We consider the problem of providing a safety certificate for an arbitrary control law by means of a safe set and controller, as proposed in [2], [3]. Consider a potentially unsafe control strategy $\bar{u}(t)$, obtained for example by application of RL (which often cannot guarantee constraint satisfaction). In order to achieve minimal interference with the desired control $\bar{u}(t)$, the goal is to compute a set of states $\mathcal{S}$, for which we know a control strategy $u_\mathcal{S}(t)$ such that input and state constraints will be satisfied for all future times, in particular considering that $d(x)$ is unknown. The control $\bar{u}(t)$ can then safely be applied in the interior of S, until it becomes necessary to take a safety-ensuring action $u_\mathcal{S}$ on the boundary of $\mathcal{S}$ which guarantees that we stay in $\mathcal{S}$, i.e. that we can still provide a safe control strategy in the future. More formally:

Definition II.4. A set $\mathcal{S} \subseteq \mathbb{X}$ is called a *safe set* for system (1) if there exists a *safe control law* $u_\mathcal{S} : \mathcal{S} \to \mathbb{U}$ such that for an arbitrary (learning-based) policy $\bar{u}(t)$, the safe controller

$$u(t) = \begin{cases} u_S(x(t)), & x(t) \in \partial\mathcal{S} \vee \bar{u}(t) \notin \mathbb{U} \\ \bar{u}(t), & \text{otherwise} \end{cases} \tag{2}$$

guarantees that the state $x(t)$ is contained in $\mathcal{S}$ for all $t > 0$ if $x(0) \in \mathcal{S}$.

In particular, we aim at finding an algorithm that scales well in computational complexity with respect to the dimensionality of system (1) as well as the number of measurements.

III. Safe-sets for nonlinear systems from data

We first introduce the class of safe-sets considered. Afterwards, we motivate the proposed method and highlight its basic idea using an example, in order to then introduce the algorithm for safe set and controller computation in the remainder of the section.

A. Ellipsoidal safe set

In order to provide a scalable optimization-based approach, we restrict the form of the safe set to an ellipsoidal set of the form

$$\mathcal{S}^P(\gamma) = \{x \in \mathbb{X} | x^\top Px \leq \gamma\} \tag{3}$$

with $P \in S^n_{++}, \gamma \in \mathbb{R}^n, \gamma > 0$, and the safe controller to the class of linear state feedback control laws $u_\mathcal{S} = Kx$ with $K \in \mathbb{R}^{m\times n}$. To construct $\mathcal{S}^P(\gamma)$, we leverage Lyapunov's direct method, with a quadratic Lyapunov function $V(x) = x^\top(\gamma^{-1}P)x$. By standard Lyapunov arguments, the following sufficient conditions ensure, analogously to [2, Lemma 1], that $\mathcal{S}^P(\gamma)$ fulfills Definition II.4:

$$\mathcal{S}^P(\gamma) \subseteq \mathbb{X} \tag{4a}$$

$$Kx \in \mathbb{U} \tag{4b}$$

$$\dot V(x) \le 0 \tag{4c}$$

for all $x \in \partial\mathcal{S}^P(\gamma)$.

B. Motivating Example

Consider the system $\dot x = x + d(x) + u$ with $d(x) = -x^3$ subject to input constraints $|u| \le 2$ and state constraints $|x| \le 2$. The task is to find a *safe* interval $\mathcal{S} = [a, b]$ and a corresponding *safe control law* $u_\mathcal{S} : [-2, 2] \to [-2, 2]$ according to Definition II.4. The nonlinearity $d(x)$ is unknown, but we are given a set of noise-free observations $\mathcal{D} = \{x_i, d(x_i)\}_{i=1}^N$ such that the convex hull $\mathrm{conv}(\{x_i\}_{i=1}^N)$ equals the state space $[-2, 2]$ (this will not be a necessary assumption later, see also Assumption II.1). We consider a linear state feedback $u_\mathcal{S} = kx$, $k \in \mathbb{R}$.

Robust approach: Without any knowledge of $d(x)$, a robust approach is to consider the dynamics $\dot x = x + w + u$, with $|w| \le 8$, where the bound on w is estimated from $\mathcal{D}$. In this case, there does not exist a feasible controller gain k w.r.t. input constraints for any state and $\forall |w| \le 8$, such that $\dot x \le 0$, i.e. there does not exist a safe set.

Proposed approach: Let $V(x) = px^2$, $p > 0$ be our Lyapunov candidate function. We analyze $\dot V(x) = 2p(x^2 + kx^2 + xd(x)) \le 0$. At the boundary of the state space we have the measurements $2d(2) = -2d(-2) = -16$, providing that $\dot V(x) \le 0$. By standard Lyapunov arguments, this implies that for $k = 0$ we have that for all $x(0) \in \partial\mathcal{S} = \{-2, 2\}$, $x(t) \in \mathcal{S}$ for all $t > 0$. We conclude that $u_\mathcal{S}(t) = 0$ is a safe controller for which the state constraint set constitutes a safe set.

This example highlights that rather than taking uniform bounds on the unknown dynamics, we can provide less conservative safe sets by quantifying the effect of the unknown dynamics in the form of *state-dependent* disturbances, which can be inferred from the available data $\mathcal{D}$. In the following, we will exploit this concept for safe set and controller computations and conclude the section with two examples.

C. Computation of the safe set and controller

Given a matrix P (see Section III-D) that determines the shape of the safe set, we write the problem of finding the size of the safe set $\mathcal{S}^P(\gamma)$ and a corresponding safe controller $u_\mathcal{S}$ as two consecutive convex optimization problems.

Remark III.1. *Given the shape P of the safe set* (3)*, there exists a $\bar\gamma > 0$ small enough such that Assumption II.1 is satisfied on the sub-level set $\mathcal{S}^P(\bar\gamma)$, i.e.*

$$\mathcal{S}^P(\bar\gamma) \subseteq \mathcal{A}_\delta. \tag{5}$$

The proposed procedure first uses data to bound the effects of the nonlinearity on the Lyapunov decrease by a quadratic form in the largest possible safe set, i.e. over $\gamma \in (0, \bar\gamma]$. This quadratic bound is then used as input to the second optimization problem, which computes the controller and set size in order to take into account the nonlinearities in addition to the linear system dynamics. The restriction to a quadratic bound of the nonlinearity is motivated by the fact that it can be treated efficiently by means of a convex problem.

In order to reduce conservatism, we bound the nonlinearity on sub-regions of the safe set, described by intervals

$$\gamma \in \Gamma_i = [\gamma_1^i, \gamma_2^i],\ \gamma_1^i < \gamma_2^i,\ \gamma_2^i \le \bar\gamma,\ i = 1, 2, .., n_\Gamma, \tag{6}$$

which are defined such that $\mathcal{S}^P(\gamma) \subseteq \mathcal{A}_\delta$ for any $\gamma \in \Gamma_i$. Note that the selection of sub-intervals is possible as we will use the quadratic bound for upper bounding (4c), which is only required to hold on $\partial\mathcal{S}^P(\gamma)$. For every interval Γ_i, we then formulate two convex optimization problems in order to determine the volume of the safe set and the safe controller. In case no solution exists, the interval can be reduced. In general, the smaller the intervals are chosen, the less conservative the bound will be.

Bounding the nonlinear effects: Given an interval Γ_i, consider the neighborhood $\mathcal{R}(\Gamma_i) = \{\mathcal{S}^P(\gamma_i^2)\backslash\mathcal{S}^P(\gamma_i^1)\}\oplus\mathcal{B}_\delta(0)$. The indices of data samples inside the set $\mathcal{R}(\Gamma_i)$ are given by $\mathbb{I}_\mathcal{R}(\Gamma_i) = \{k \in \mathcal{I}_{\ge 1} | x_k \in \mathcal{D}_x,\ x_k \in \mathcal{R}(\Gamma_i)\}$. We seek to find a quadratic bound on the nonlinearity arising in the Lyapunov decrease (4c) for all $\bar x \in \mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)$, i.e. to find a $Q(\Gamma_i)$ such that

$$\begin{aligned}\dot V(\bar x) =& 2\gamma^{-1}\bar x^\top P(A + BK)\bar x + 2\gamma^{-1}\bar x^\top Pd(\bar x)\\ \le& 2\gamma^{-1}\bar x^\top P(A + BK)\bar x + 2\gamma^{-1}\bar x^\top Q(\Gamma_i)\bar x.\end{aligned} \tag{7}$$

The first optimization problem for bounding the nonlinearity over each interval is given by

$$Q(\Gamma_i) = \underset{\tilde Q \in S^n}{\mathrm{argmin}} \sum_{k \in \mathbb{I}_\mathcal{R}(\Gamma)} \left(x_k^\top \tilde Q x_k - p_k\right)^2 \tag{8a}$$

s.t. for all $k \in \mathbb{I}_\mathcal{R}(\Gamma_i)$:

$$\lambda_k \ge 0 \tag{8b}$$

$$\begin{pmatrix} -\tilde Q - \lambda_k I_n & \lambda_k x_k \\ \lambda_k x_k^\top & -\lambda_k\left(x_k^\top x_k - \delta^2\right) + p_k \end{pmatrix} \preceq 0 \tag{8c}$$

with $p_k = x_k^\top Pd(x_k) + \delta L_{x^\top Pd(x)}(\mathcal{R}(\Gamma_i))$ and δ as defined in Assumption II.1.

Lemma III.2. *Let Assumption II.1 hold and let $\bar\gamma$ satisfy* (5)*. Consider an interval Γ_i according to* (6)*. If* (8) *attains a solution, then for all $\bar x \in \mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)$ it holds that*

$$\bar x^\top Pd(\bar x) \le \bar x^\top Q(\Gamma_i)\bar x. \tag{9}$$

Proof. We prove that for all $\bar x \in \mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)$ the implication (8b), (8c) $\Rightarrow$ (9) holds. First note that for any $\bar x \in \mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)$ there exists a $k \in \mathbb{I}_\mathcal{R}(\Gamma_i)$ such that $||\bar x - x_k|| \le \delta$ by the definition of the intervals, i.e. $\gamma_i^2 \le \bar\gamma$, see also Remark III.1. For notational ease let $f(\bar x) =$

$d(\bar{x})^\top P\bar{x}$. Equation (9) reads $f(\bar{x}) - \bar{x}^\top Q(\Gamma_i)\bar{x} \leq 0$. For all $k \in \mathbb{I}_{\mathcal{R}}(\Gamma_i)$ and for all $\bar{x}_k \in \mathcal{B}_\delta(x_k)$ we have therefore by Lipschitz continuity $f(\bar{x}_k) - f(\Delta_{\mathcal{D}_x}(\bar{x}_k)) + f(\Delta_{\mathcal{D}_x}(\bar{x}_k)) - \bar{x}_k^\top Q(\Gamma_i)\bar{x}_k \leq p_k - \bar{x}_k^\top Q(\Gamma_i)\bar{x}_k$. Note that by the definition of the intervals and Remark III.1 the relation $\{\mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)\} \subset \bigcup_{i\in\mathbb{I}_{\mathcal{R}}(\Gamma_i)} \mathcal{B}_\delta(x_k)$ holds. As a consequence, if for all $k \in \mathbb{I}_{\mathcal{R}}(\Gamma_i)$ and for all $\bar{x}_k \in \mathcal{B}_\delta(x_k)$ we have that $p_k - \bar{x}_k^\top Q(\Gamma_i)\bar{x}_k \leq 0$, then the quadratic bound (9) holds for all $\bar{x} \in \mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)$. Finally, using the S-Lemma (see [13]) the condition $\bar{x} \in \mathcal{B}_\delta(x_k) \Rightarrow p_i - \bar{x}^\top Q(\gamma)\bar{x} \leq 0$ is equal to (8b),(8c) which completes the proof. □

Remark III.3. *The optimization problem in* (8) *is a convex semidefinite programming problem. In case that there are more observations* ($N \gg 0$) *than the optimization algorithm can handle in* (8)*, one can iteratively calculate* $Q(\Gamma_i)$*: Solve* (8) *using a subset of* $\mathcal{D}$ *in order to obtain* $Q^1(\Gamma_i)$*, in the next iteration choose another disjoint subset of* $\mathcal{D}$ *and add the constraint* $\tilde{Q} \succeq Q^1(\Gamma_i)$ *to* (8) *in order to obtain* $Q^2(\Gamma_i)$*. Repeat until all subsets of* $\mathcal{D}$ *are processed which yields a feasible, possibly suboptimal solution to* (8).

Problem (8) provides a bound on the nonlinear effect in the Lyapunov decrease by means of the Lipschitz constant of $x^\top Pd(x)$. In practice, a local Lipschitz constant can e.g. be obtained from data as $\hat{L}_{x^\top Pd(x)}(\mathcal{R}(\Gamma_i)) = 2\max_{k_1,k_2\in\mathbb{I}_{\mathcal{R}}(\Gamma_i)} ||x_{k_1}^\top Pd(x_{k_1}) - x_{k_2}^\top Pd(x_{k_2})||/||x_{k_1} - x_{k_2}||$.

Since $Q(\Gamma_1)$ does not have to be positive definite, stabilizing effects of the nonlinearity can be considered in (7), i.e. $\bar{x}^\top Q(\Gamma_i)\bar{x}$ can be negative and therefore contribute to rendering $\dot{V}(x)$ negative on the boundary of the safe set. This is demonstrated in Section III-E, see also Figure 2 (right).

Calculation of safe level set and safe controller: By using the quadratic bound from the first optimization problem, we are able to state the second optimization problem for safe set and controller design satisfying the conditions (4a)-(4c) as follows. Let $\gamma \in \mathbb{R}$, $E = P^{-1}$, $Y \in \mathbb{R}^{m\times n}$, Γ_i be chosen according to (6). The optimization problem is given by

$$\min_{\gamma,Y} - \gamma \tag{10a}$$

$$\text{s.t. } \gamma \in \Gamma_i \tag{10b}$$

$$AE\gamma + EA^T\gamma + BY + Y^\top B^\top + 2EQ(\Gamma_i)E\gamma \preceq 0 \tag{10c}$$

$$\forall j \in \mathcal{I}_{[0,n_x]} : \begin{pmatrix} b_{x,j}^2 & A_{x,j}E \\ EA_{x,j}^\top & \gamma E \end{pmatrix} \succeq 0 \tag{10d}$$

$$\forall k \in \mathcal{I}_{[0,n_u]} : \begin{pmatrix} b_{u,k}^2 & A_{u,k}Y \\ Y^\top A_{u,k}^\top & \gamma E \end{pmatrix} \succeq 0. \tag{10e}$$

Theorem III.4. *Let Assumption II.1 hold and let* $\bar{\gamma}$ *satisfy* (5)*. Consider an interval* Γ_i *according to* (6)*. If* (10) *attains a solution* $\{\gamma^*, Y^*\}$*, then* $\mathcal{S}^P(\gamma^*)$ *is a safe set for system* (1) *according to Definition II.4 with* $u_S(x) = Kx$, $K = {\gamma^*}^{-1}Y^*E^{-1}$.

Proof. We prove the result in two steps: 1) Conditions (10d)-(10e) imply (4a) and (4b). By [14, Section 5.2.2] we can rewrite (4b) as $A_{u,i}K(\gamma^{-1}P)^{-1}K^\top A_{u,i}^\top \leq b_{u,i}^2$ which equals (10e). The matrix inequality for the states can be derived similarly. 2) Conditions (10b)-(10c) ensure that $\mathcal{S}^P(\gamma^*)$ fulfills (4c). For all $x \in \partial\mathcal{S}^P(\gamma)$ we have to fulfill (4c) which is implied by (7). A sufficient condition for (4c) is therefore $\gamma^{-1}P(A+BK) + \gamma^{-1}(A+BK)^\top P + 2\gamma^{-1}Q(\Gamma_i) \preceq 0$, i.e. that $\dot{V}(x) \leq 0$ for all $x \in \mathbb{R}^n$. Multiplying from left and right by γP^{-1} yields (10c). We have shown that $\mathcal{S}^P(\gamma^*)$ is a safe set according to Definition II.4. The objective in (10) yields the largest safe set under these sufficient conditions. □

Note that optimizing over P and K in (10) is not possible, because the bound obtained in the first optimization step depends on P.

Problem (8) and (10) provide (semidefinite) convex optimization problems for computing a safe set and controller by directly employing data points. Such problems can be solved efficiently even for higher dimensions, see e.g. [15], [16]. While this offers a general approach with favorable scalability properties, the limitation is that the resulting safe set cannot be larger than the convex hull of the data points plus a δ-neighborhood. Exploration is therefore limited. Nevertheless, initially collected data can be iteratively extended inside $\mathcal{A}_\delta$ such that δ from Assumption II.1 gets smaller over time. Recomputation of the safe set can then reduce conservatism. We present an extension in Section IV which further reduces conservatism and improves exploration.

D. Shape of the safe set

Using the linear model of the system dynamics in (1), we define an approximate initial shape matrix P of the safe set by neglecting the unknown nonlinearity. Assume that $\mathcal{A}_\delta$ is given by $\{x \in \mathbb{R}^n | x^\top A_\delta x \leq 1\}$ with $A_\delta \in S_{++}^n$, which can be e.g. calculated as the minimum volume covering ellipse of the data points $\mathcal{D}$, see [17, p. 222]. We can find a safe set for system (1) by setting $d(x) = 0$, resulting in the following optimization problem with $E \in S_{++}^n$ and $Y \in \mathbb{R}^{m\times n}$

$$\min_{E,Y} - \text{logdet}(E) \tag{11a}$$

$$\text{s.t. } A_\delta^{-1} \succeq E \tag{11b}$$

$$AE + EA^T + BY + Y^\top B^\top \preceq 0 \tag{11c}$$

$$(10d), (10e). \tag{11d}$$

If (11) attains a solution, then analogously to Theorem III.4, we obtain a safe set for system (1) according to Definition II.4 with $P = E^{-1}$, $u_S(x) = Kx$, and $K = YE^{-1}$, but for $d(x) = 0$. Constraint (11b) ensures that the safe set is a subset of $\mathcal{A}_\delta$, i.e. the set in which Assumption II.1 is satisfied and therefore the data is dense. This implies that $(0, 1]$, i.e. $\bar{\gamma} = 1$, is the maximum set size such that Assumption II.1 as well as all constraints are fulfilled. The optimization problem (10) then aims at improving the initial approximation obtained via (11) with respect to the nonlinearity by designing a different control law and safe set size.

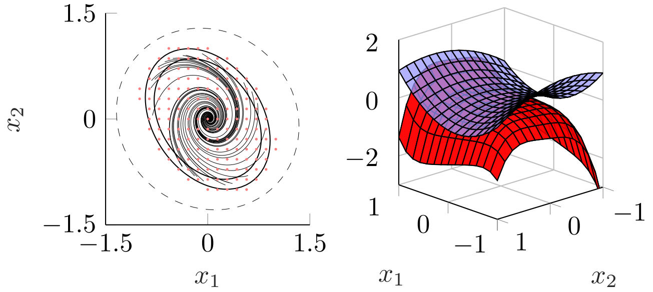

Fig. 2: **Left:** Sample trajectories under the safe control law, when starting on the boundary of the safe set $\mathcal{S}^P$. Red dots: Data points, black lines: Closed-loop system trajectories, elliptic ring $\mathcal{R}(\Gamma_1)$, which contains $\partial\mathcal{S}^P$. Dashed set: Safe set using Q_{GP} from Section IV. **Right:** Quadratic bound $x^\top Q(\Gamma_i)x$ according to Lemma III.2 (blue) of non-linear term $x^\top Pd(x)$ (red).

E. Illustrative numerical example

Consider a nonlinear system of the form (1), where $A = \begin{pmatrix} -1 & 2 \\ -3 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 0.5 \\ -2 \end{pmatrix}$, $d(x) = \begin{pmatrix} 0.5x_1^4 \\ 0.35-1.5x_2^3 \end{pmatrix}$ with input constraints $|u| \leq 3$ and state constraints $|x| \leq 2$. The origin is not a stable equilibrium, and is neither an equilibrium point. For simplicity we consider a grid of data points as illustrated in Figure 2, however any data set fulfilling Assumption II.1 could be used. The data was taken inside the set $\{x \in \mathbb{R}^n | x^\top Px \leq 1\}$ with $P = \begin{pmatrix} 0.7651 & 0.2162 \\ 0.2162 & 0.6481 \end{pmatrix}$ obtained by solving (11), which also corresponds to $\mathcal{A}_\delta$ with $\delta = 0.15$. We apply Theorem III.4 by solving (8) and (10) for $\Gamma_1 = [0.9, 1]$, $L_{x^\top Pd(x)}(\mathcal{R}(\Gamma_1)) \approx 6.02$ and obtain $K^* = (0.5261,\ 2.2953)$, $\gamma^* = 1$. The results are illustrated in Figure 2. Note that in general the quadratic bound must only hold on $\partial\mathcal{S}(\gamma)$ and could therefore be violated around the origin.

F. Simulation: Safety for autonomous convoys

Consider a convoy of cars or trucks as depicted in Figure 3. Given a target velocity v_{tar} and a possibly small target distance x_{tar}, the goal is to drive closely behind each other, in order to leverage slipstream effects for efficiency. We assume that it is possible to overwrite the local controllers, i.e. the acceleration of car 1, car 3 and car 4 in a centralized way if necessary to ensure safety. During a supervised observation phase, initial data about the system is collected. We consider the problem of finding a safe, centralized control law, and a safe set such that the cars will not crash, even if we cannot determine the acceleration of car 2 and car 5.

Let $z_{i+1\to i} = x_{\text{tar}} - x_{i+1\to i}$ be the difference between the target distance x_{tar} and the actual distance $x_{i+1\to i}$ of car $i+1$ and car i. Let v_i be the difference between the target velocity of the convoy v_{tar} and the velocity $\bar{v}_i$ of car i. The dynamics of all cars $i = 1, ..., 5$ are given as $\dot{z}_{i+1\to i} = v_{i+1} - v_i$, $\dot{v}_i = u_i$ where u_i is the applied acceleration. The control law of car 1 is given by $u_1(v_1) = -v_1$, of cars $3, 4$ by $u_i(z_{i\to i-1}, v_i) = 0.1z_{i\to i-1} - 0.3v_i$ and of cars $2, 5$ (which we cannot overwrite) by $u_2(z_{2\to 1}, v_2) = \max\{\min\{z_{2\to 1} - v_2, 0.9\}, -0.9\}$, $u_5(z_{5\to 4}, v_5) = \max\{\min\{z_{5\to 4} - v_5, 0.9\}, -0.9\}$, i.e. they apply a saturated, stabilizing state feedback law and are therefore nonlinear. The target distance between the cars is 1 meter. In order to avoid a crash the state constraints are given by $x_{i+1\to i} \geq 0$. The maximum acceleration of cars 1, 3, and 4 is 3 $[m/s^2]$, i.e. $|u_i| \leq 3,\ i = 1, 3, 4$. We are given observations of $\dot{z}_{2\to 1}, \dot{v}_2$ and $\dot{z}_{5\to 4}, \dot{v}_5$ in the interval $[-0.8, 0.8]$ with $\delta = 0.013$.

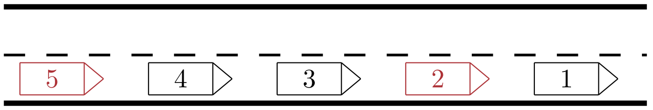

Fig. 3: Illustration of the autonomous car convoy. The acceleration of red cars cannot be controlled.

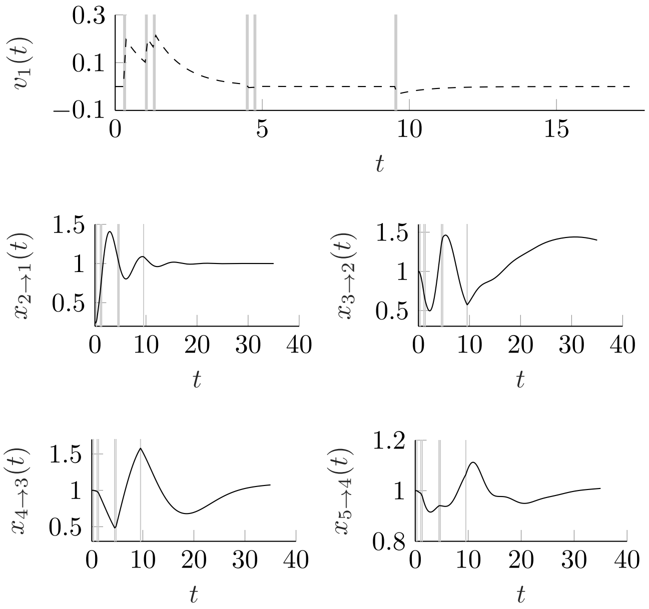

Fig. 4: Sample trajectory of the car convoy under the safety framework, when starting on the boundary of $\mathcal{S}^P(\gamma)$. Grey lines indicate times when the safe control law $u_\mathcal{S}$ is applied.

In Figure 4, a numerical simulation under the resulting safe control law (2) is shown, starting from the boundary of the safe set with $v_2(0) = 0.02\ [m/s], x_{2\to 1}(0) = -0.76\ [m]$ and the remaining states equal to zero, which represents the situation that the second car is too close to the first one and its velocity is slightly higher than the reference velocity. As we can see in Figure 4, the first car has to accelerate quickly several times during the first two seconds for safety reasons, since the second car (which cannot be controlled) would decelerate more as car 3 would be able to compensate. The same situation occurs at 4.2 seconds between car 3 and car 4 and at around 10 seconds between car 1 and car 2 again. After six safety interventions in total, the local controllers of the cars are able to stabilize the overall system.

Algorithm 1 Calculation of Q_{GP}

```
1:  procedure
2:      define $f(x) \leftarrow \max_{\hat{\sigma} \in [-c\sigma_N(x), c\sigma_N(x)]} xP(\mu_N(x) + \hat{\sigma})$
3:      initialize set $X^0 \leftarrow \{x_1^0, ..., x_N^0\}$, $x_k^0 \in \{\mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)\}$
4:      initialize set $Y^0 \leftarrow \{f(x_1^0), ..., f(x_N^0)\}$
5:      $n \leftarrow 1$, $Q^0 = 0, Q^1 = G(X^0, Y^0)$
6:      while $Q^n \neq Q^{n-1}$ do
7:          $n \leftarrow n + 1$
8:          $x = \mathrm{argmin}_{\hat{x} \in \mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)} \hat{x}^\top Q^n \hat{x} - f(\hat{x})$
9:          $X^n \leftarrow \{x\} \cup X^{n-1}$
10:         $Y^n \leftarrow \{f(x)\} \cup Y^{n-1}$
11:         $Q^n \leftarrow G(X^n, Y^n)$
12:     end while
13:     return $Q_{GP}(\Gamma_i) = Q^n$
```

IV. Safe-sets using Gaussian Processes

The previous sections are based on Lipschitz continuity of the unknown nonlinearity $d(x)$, which was incorporated into the quadratic upper bound in (10), see Lemma III.2. A shortcoming of using only Lipschitz continuity as 'prior' knowledge is the requirement of relatively dense and structured data, see Assumption II.1. This implies that almost no 'exploration' can be made beyond the data, observed so far. By putting a stronger prior on the class of functions for modeling the nonlinearity $d(x)$, we develop a less conservative quadratic bound, which can then be used in (10) and is generally expected to yield a larger safe set. In addition, it improves exploration beyond the data points allowing to iteratively improve the safe set during closed-loop operation, where initially few data points are generally available.

A. Gaussian Processes

We use a Gaussian process model (GP) in order to perform Bayesian inference on the unknown nonlinearity (see. e.g. [18]) for each element $d_i(x)$ of $d(x)$. The GP is defined by a mean function $\mu^i(x)$, together with a covariance (kernel) function $k^i(x, x')$, denoted with $\mathcal{GP}(c^i_\mu, k^i)$ for the GP prior on $d_i(x)$ in short. We set the mean prior function to be constant, i.e. $\mu^i = c^i_\mu$. Given observations $\boldsymbol{y}^i_N = [y^i_1, .., y^i_N]^T$ at locations $X_N = [x_1, .., x_N]^T$ where $y^i_j = d_i(x_j)$, the posterior distribution of $d_i(x)$ is given by

$$\mu^i_N(x) = c^i_\mu + k^i_N(x)^T {K^i_N}^{-1}(\boldsymbol{y}^i_N - \boldsymbol{c}^i_\mu) \tag{12}$$

$$k^i_N(x, x') = k^i(x, x') - k^i_N(x)^T {K^i_N}^{-1} k^i_N(x') \tag{13}$$

$$ {\sigma^i_N}^2(x) = k^i_N(x, x), \tag{14}$$

where $k^i_N(x) = [k^i(x_1, x), .., k^i(x_N, x)]^T$, $k^i_N(x, x')$ is the posterior covariance, ${\sigma^i_N}^2(x)$ is the variance, $K^i_N = [k^i(x, x')]_{x,x' \in X_N}$ is the positive definite covariance matrix matrix and $\boldsymbol{c}^i_\mu$ is a vector of N elements, each equal to c^i_μ. The posterior mean for the vector $d(x)$ is $\mu_N(x) = [\mu^0_N(x), .., \mu^n_N(x)]^\top$ and the variance $\sigma^2_N(x) = [{\sigma^0_N}^2(x), .., {\sigma^n_N}^2(x)]^\top$.

B. GP-based bounding of nonlinearity

The GP model provides a measure for the posterior variance of $d(x)$, which is used to improve the bound on the effect of the nonlinearity on the Lyapunov decrease. Instead of using Lipschitz-based arguments, as in Section III, we calculate a strict quadratic bound on highly probable, worst-case realizations of the nonlinear term $x^\top Pd(x)$ for all $x \in \mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)$, given an interval Γ_i. The intervals Γ_i in this section are not limited to a dense data region $\mathcal{A}_\delta$. Therefore one can drop the constraint (11b) in the computation of P in (11) and choose $\bar{\gamma} = 1$ for the construction of the intervals (6). By relying on the GP, the largest interval $\bar{\gamma}$ can be chosen independent of Assumption II.1 and Remark III.1.

Algorithm 1 summarizes the calculation of the quadratic bound, implementing the following idea. Let $f(x)$ be a function that has to be quadratically upper bounded by $x^\top Qx$ for all $x \in \{\mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)\}$. In order to enforce the infinite dimensional constraint $\forall x \in \{\mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)\} : f(x) \leq x^\top Qx$, we proceed iteratively by starting with a finite approximation, which will be improved until $f(x) \leq x^\top Qx$ holds for all $x \in \{\mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)\}$.

In line 2 of Algorithm 1, the function f is defined, which returns the maximum value of the nonlinear term $x^\top Pd(x)$ with a chosen probability, e.g. with 99.73% by letting $c = 3$. Starting with a finite number of samples of $f(x)$ for $x \in \mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)$ (lines 3,4) we compute an initial guess for a quadratic bound on $x^\top Pd(x)$ given by $x^\top Q^1 x$ in line 5, where $G(X, Y) = \mathrm{argmin}_{\tilde{Q} \in S^n} \sum_{(x_i, y_i) \in (X,Y)} \left(x_i^\top \tilde{Q} x_i - y_i\right)^2$ s.t. for all $(x_i, y_i) \in \mathcal{P} : y_i \leq x_i^\top \tilde{Q} x_i$, which yields a quadratic upper bound on $\{y_i\}_{i=1}^N$.

We search for potential violations of the current bound in line 8 and add it to the set of data points in lines 9 and 10. After that we update the quadratic bound in line 11. The algorithm iterates until there is no violation. This way for all $x \in \mathcal{S}^P(\gamma_i^2) \setminus \mathcal{S}^P(\gamma_i^1)$, the quadratic form $x^\top Q_{GP}(\Gamma_i)x$ will be a bound on $x^\top Pd(x)$ with high probability.

Remark IV.1. *The optimization problem in line 8 is continuous (compare for example [19, Chapter 2G]), but nonconvex. An alternative approach is to build a discretization of $\mathcal{R}(\Gamma_i)$, which we denote by $\mathcal{D}_x$, with grid size δ, $|\mathcal{D}_x| = N$, and evaluate the posterior mean and covariance $\mu_N(x), \sigma_N(x)$ for each $x \in \mathcal{D}_x$. By selecting $p_k = x_k^\top P\mu_N(x_k) + \beta_N \sum_{i=1}^n \sigma^i_N(x) + \delta L_{x^\top Pd(x)}(\mathcal{R}(\Gamma_i))$ with β_N as defined in [20, Lemma 3], the bound Q_{GP} can be approximated using* (8) *as a convex optimization problem.*

For the second step of calculating a safe set size and controller, we can simply use the bound $Q_{GP}(\Gamma_i)$ instead of $Q(\Gamma_i)$ in (10) in order to obtain a set, which is safe with the selected probability. By construction, $Q_{GP}(\Gamma_i)$ will be less conservative, or equal than $Q(\Gamma_i)$. This is due to the fact that we put a prior on the unknown nonlinearity $d(x)$, which allows for Bayesian inference and therefore improved extra- and interpolation based on the data, as opposed to using estimates based on Lipschitz continuity. In Figure 2 (left), the benefit of using Q_{GP} over the Lipschitz based bound Q is illustrated. Moreover, the main advantage is that we do *not* require particular assumptions on the data and the

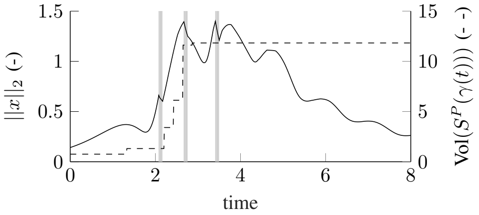

Fig. 5: Safe RL using the proposed safety framework together with the PGSD [21] algorithm. The distance of the state to the origin, as well as the volume of the safe set is shown over time. Grey-shaded time points depict application of the safe control law $u_{\mathcal{S}}$.

safe set is not limited to a subset of $\mathcal{A}_\delta$, as it is the case in Lemma III.2.

C. Numerical example: Exploration

Consider a nonlinear system of the form (1) with $A = \left(\begin{smallmatrix} -1 & 2 \\ -3 & 4 \end{smallmatrix}\right)$, $B = \left(\begin{smallmatrix} 0.5 \\ -2 \end{smallmatrix}\right)$, $d(x) = \begin{pmatrix} 0.5x_1^2\sin(6x_1) \\ -0.8x_2^3 \end{pmatrix}$, input constraints $|u| \leq 4$, and state constraints $|x| \leq 4$. We use a squared exponential kernel as defined in [18] with $\sigma_f^1 = \sigma_f^2 = 0.05$ and $l^1 = l^2 = 0.2$. Given initial data in $[-0.2, 0.2]^2$ with $\delta = 0.05$, we solve (10) using the high probability ($c = 3$, 99.73%) bound $Q_{GP}(\Gamma_i)$ with an initial P obtained via (11) and $\Gamma_i = [1 - i0.1 - 0.1, 1 - i0.1]$ for $i = 1, 2, .., 8$, i.e. we start with $i = 1$ and iterate $i = 2, 3, ..$ until we find a feasible solution. We assume that the desired control input $\bar{u}(t)$ is given by the policy gradient with signed derivative (PGSD) algorithm (see [21]), which is a policy search RL method without any safety guarantees. During closed-loop operation under the safe control law (2) we collect data $\mathcal{D}$. Every 0.2 seconds we recompute the safe level set, i.e. $\gamma(t)$, where the safe set size converges after 2.5 seconds. In Figure 5 the evolution of the safe set size (volume of the ellipse) is shown as well as the distance of the system state to the origin, which has to be minimized by the PGSD learning based control law. The unsafe RL input is 'overwritten' three times indicated by the grey lines to ensure safety, until it begins to converge.

V. Conclusion

This paper presents a safety framework that allows to enhance arbitrary learning-based and unsafe control strategies, for nonlinear and potentially larger scale systems with safety certificates. The nonlinearity is assumed to be unknown and we only require a possibly inaccurate linear system model and observations of the system. A key feature is that the proposed method directly exploits the available data, without the need of an additional learning mechanism. By relying on convex optimization problems, the proposed method is scalable with respect to the system dimension and number of data points. In order to reduce conservatism of the safe set calculations, the approach was extended using a Gaussian process model as prior on the nonlinearity. This modification enables safe exploration and thereby iterative computation of the safe set during closed-loop operation. The results were demonstrated using several numerical example problems, showing that the safety framework can be used to certify arbitrary and in particular learning-based controllers.

References

[1] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[2] A. K. Akametalu, J. F. Fisac, J. H. Gillula, S. Kaynama, M. N. Zeilinger, and C. J. Tomlin, "Reachability-based safe learning with gaussian processes," in *53rd IEEE Conference on Decision and Control (CDC)*. IEEE, 2014, pp. 1424–1431.

[3] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. Gillula, and C. J. Tomlin, "A general safety framework for learning-based control in uncertain robotic systems," *arXiv preprint arXiv:1705.01292*, 2017.

[4] J. Garcıa and F. Fernández, "A comprehensive survey on safe reinforcement learning," *Journal of Machine Learning Research*, vol. 16, pp. 1437–1480, 2015.

[5] K. P. Wabersich and M. Toussaint, "Automatic testing and minimax optimization of system parameters for best worst-case performance," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2015, pp. 5533–5539.

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

[7] F. Berkenkamp, A. P. Schoellig, and A. Krause, "Safe controller optimization for quadrotors with gaussian processes," in *Proc. of the International Conference on Robotics and Automation (ICRA)*, 2016, pp. 491–496.

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

[9] M. Chen, J. F. Fisac, S. Sastry, and C. J. Tomlin, "Safe sequential path planning of multi-vehicle systems via double-obstacle hamilton-jacobi-isaacs variational inequality," in *European Control Conference (ECC)*, July 2015, pp. 3304–3309.

[10] S. Kaynama, I. M. Mitchell, M. Oishi, and G. A. Dumont, "Scalable safety-preserving robust control synthesis for continuous-time linear systems," *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 3065–3070, 2015.

[11] J. F. Fisac and S. S. Sastry, "The pursuit-evasion-defense differential game in dynamic constrained environments," in *54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 4549–4556.

[12] 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 the Mathematical Sciences*, vol. 3, no. 1, p. 19, 2016.

[13] I. Pólik and T. Terlaky, "A survey of the S-lemma," *SIAM review*, vol. 49, no. 3, pp. 371–418, 2007.

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

[15] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3a matlab software package for semidefinite programming, version 1.3," *Optimization methods and software*, vol. 11, no. 1-4, pp. 545–581, 1999.

[16] M. ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28).*, 2016. [Online]. Available: http://docs.mosek.com/7.1/toolbox/index.html

[17] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[18] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. MIT press Cambridge, 2006, vol. 1.

[19] A. L. Dontchev and R. T. Rockafellar, "Implicit functions and solution mappings," *Springer Monogr. Math.*, 2009.

[20] F. Berkenkamp, M. Turchetta, A. P. Schoellig, and A. Krause, "Safe model-based reinforcement learning with stability guarantees," in *Proc. of the Conference on Neural Information Processing Systems (NIPS)*, 2017.

[21] J. Z. Kolter and A. Y. Ng, "Policy search via the signed derivative." in *Robotics: science and systems*, 2009, p. 34.