# Safe exploration of nonlinear dynamical systems: A predictive safety filter for reinforcement learning

# Kim P. Wabersich <sup>1</sup> Melanie N. Zeilinger <sup>1</sup>

#### **Abstract**

The transfer of reinforcement learning (RL) techniques into real-world applications is challenged by safety requirements in the presence of physical limitations. Most RL methods, in particular the most popular algorithms, do not support explicit consideration of state and input constraints. In this paper, we address this problem for nonlinear systems with continuous state and input spaces by introducing a predictive safety filter, which is able to turn a constrained dynamical system into an unconstrained safe system, to which any RL algorithm can be applied 'out-ofthe-box'. The predictive safety filter receives the proposed learning input and decides, based on the current system state, if it can be safely applied to the real system, or if it has to be modified otherwise. Safety is thereby established by a continuously updated safety policy, which is based on a model predictive control formulation using a data-driven system model and considering state and input dependent uncertainties.

#### 1. Introduction

Reinforcement learning (RL) has demonstrated its success in solving complex and high-dimensional control tasks, see for example Lillicrap et al. (2015); Mnih et al. (2015); Schulman et al. (2015); Levine et al. (2016). These results motivate a more wide-spread transfer to real-world applications in order to enable automated design of high performance controllers with little need for expert knowledge. In physical systems, such as mechanical, thermal, biological, or chemical systems, physical limitations naturally arise as constraints, such as limited torque in case of a robot arm or limited power supply in building control. In addition to physical constraints, many relevant applications in industry require satisfaction of safety specifications, preventing, e.g., an autonomous car or aircraft from crashing, which

can typically be formulated in terms of constraints on the system state. The simultaneous satisfaction of safety constraints under physical limitations during RL constitutes one of the main open problems in AI safety as discussed e.g. in Amodei et al. (2016, Section 3).

Significant progress in the safe operation of constrained systems has been made through model predictive control techniques, which provide rigorous constraint satisfaction, see e.g. Mayne (2014). While model-based RL techniques (Kamthe & Deisenroth, 2017; Clavera et al., 2018) are conceptually tightly related to model predictive control, comparably few methods consider safety guarantees so far.

Learning-based model predictive control aims at combining the benefits of both fields, see for example Aswani et al. (2013). In addition to the fact that designing such algorithms to ensure safety is rather challenging, often times conservative, and requires a considerable amount of expert knowledge, the approach is inherently restricted to a model-based control policy. More precisely, at each time step, a finite horizon optimal control problem is solved in a receding horizon fashion in order to approximate a potentially infinite horizon optimal control policy.

**Concept:** This paper presents a general framework, called predictive safety filter, which is able to turn highly nonlinear and safety-critical dynamical systems into inherently *safe systems*, to which any RL algorithm without safety certificates can be applied 'out-of-the-box', see also Figure 1. The *predictive safety filter* is realized based on an available state and input dependent statistical state transition model (e.g. Gaussian Process or Neural Network). If the input proposed by the RL algorithm would potentially be unsafe, the safety filter is entitled to modify the input as little as necessary in order to maintain safe operation.

Different from recently proposed related concepts presented in Gillula & Tomlin (2011); Fisac et al. (2017); Wabersich & Zeilinger (2018a;b), we use the notion of a *safe system* in Figure 1, as similarly introduced in Wieland & Allgöwer (2007) within the context of safety barrier functions. The concept emphasizes the possibility that *any* RL algorithm that would have been used to control the original system (1) can be applied to the *safe system* instead,

<sup>&</sup>lt;sup>1</sup>Institute for Dynamic Systems and Control, ETH Zurich, Zurich, Switzerland. Correspondence to: Kim P. Wabersich <a href="mailto:kwabersich@kimpeter.de">kimpeter.de</a>.

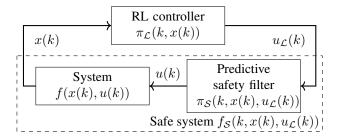


Figure 1. Concept of predictive safety filter: Based on the current state x(k), a learning-based algorithm provides a control input  $u_{\mathcal{L}}(k) = \pi_{\mathcal{L}}(k, x(k)) \in \mathbb{R}^m$ , which is processed by the safety filter  $u(k) = \pi_{\mathcal{S}}(k, x(k), u_{\mathcal{L}}(k))$  and applied to the real system.

yielding a certified safe RL application.

The predictive safety filter provides *safety* at a desired level of probability, *modularity* in terms of the employed RL controller, and *minimal intervention* by filtering RL input signals only if we cannot guarantee safety at the specified probability level, similar to Gillula & Tomlin (2011); Fisac et al. (2017). On a conceptual level, the proposed safety filter decides based on the current system state x(k) whether it is safe to apply a learning-based control input  $u_{\mathcal{L}}(k)$ , or if it is necessary to modify  $u_{\mathcal{L}}(k)$  such that safety can be guaranteed for all future times.

**Contributions:** Based on a probabilistic model of the system dynamics, this paper presents a predictive safety filter for general nonlinear systems, by generalizing the safety certification method for linear systems proposed by Wabersich & Zeilinger (2018b). Safety of an RL input is thereby estimated in real-time by searching for a safe back-up trajectory for the next time step towards a known set of safe states. The proposed technique enables this approach for nonlinear and potentially complex system descriptions, obtained, e.g., from deep learning, by relying on systemtheoretic properties rather than, e.g., on Lipschitz-based uncertainty estimates of system predictions, which can be prohibitively large. This approach leads to a theoretical analysis that rigorously relates parameters of the predictive safety filter and accuracy of its system model to safety in probability. We illustrate the approach using a simulated pendulum swing-up task, in which only little initial data around the downward position is available and overshoots of the upward position are prohibited, imposing challenging safety constraints on the system.

**Discussion:** While the focus of this paper is the certification of RL algorithms, the concept can also be used together with, e.g., human inputs. For example, in case of autonomous driving, the safety filter could be used to ensure safety of either an RL-based controller or a human driver, and can be viewed as a driver assistance system that

is able to overrule the student driver (or RL algorithm), if necessary for safety. Note, that the safety filter only has the task of keeping the system safe, but is not necessarily able to control the system 'well' with respect to a certain objective (e.g. comfort criteria). The problem of finding a safety filter is therefore in general less complex than finding a desired optimal policy with respect to some objective and subject to constraints motivating the combination of a safety filter with an RL algorithm.

#### 2. Related Work

Safe model-free reinforcement learning: There is a growing awareness of safety questions in the domain of artificial intelligence (Amodei et al., 2016), and several safe reinforcement learning techniques have been proposed, see e.g. Garcia & Fernández (2015) for an overview. Achiam et al. (2017), e.g., provide safety in expectation based on a trust-region approach with respect to the policy gradient. Other approaches are based on Bayesian optimization in order to carefully tune parametric policies (Berkenkamp & Schoellig, 2015), also with respect to worst case scenarios (Wabersich & Toussaint, 2015; Marzat et al., 2016; Ghosh et al., 2018).

Most notions of safety considered in this line of research, e.g. constraint satisfaction in expectation, tend to be less strict compared with the probabilistic safety requirements at all time steps considered in this paper. More importantly, since most techniques are policy-based, safety is coupled to a specific policy and therefore potentially also to a specific task, limiting generalization of the safety certificates.

Learning-based model predictive control: Originating from concepts in robust model predictive control (MPC), extensions of MPC schemes to safe learning-based methods have been proposed, e.g. in Aswani et al. (2013). In addition, various results have investigated combinations of MPC with learning-based online model identification techniques (Ostafew et al., 2016; Limon et al., 2017; Hewing & Zeilinger, 2017; Kamthe & Deisenroth, 2017; Koller et al., 2018; Clavera et al., 2018; Soloperto et al., 2018), also in an adaptive manner (Tanaskovic et al., 2013; Lorenzen et al., 2017). In the context of robotics, similar concepts exist, which are often referred to as funneling, see e.g. Majumdar & Tedrake (2017) and references therein, as well as so called LQR-trees (Tedrake et al., 2010).

While some of these techniques have been demonstrated to work well in practice (Bouffard et al., 2012; Ostafew et al., 2016; Hewing et al., 2017), they typically either lack rigorous theoretical safety guarantees, tend to be overly conservative by relying on Lipschitz-based estimates in the prediction of the uncertain system evolution, or are restricted to a very specific class of systems.

Model-based policy certification and safety frameworks: Using Bayesian model estimates from data, certification techniques were proposed that validate the resulting closed-loop system (Berkenkamp et al., 2016; 2017). The techniques share similar limitations with safe modelfree RL methods, namely that they are tailored to a specific task. In order to decouple safety from a specific task, the concept of a safety framework has been introduced (Gillula & Tomlin, 2011), which consists of a model-based computation of a safe set of system states and a corresponding safe control policy, which is entitled to override a potentially unsafe RL algorithm in order to ensure invariance with respect to the safe set of system states i.e. containment within the safe set at all times. This concept was further developed in several papers, providing methods to compute the safe set as well as the corresponding safe policy (Fisac et al., 2017; Wabersich & Zeilinger, 2018a; Larsen et al., 2017; Wabersich & Zeilinger, 2018b), which build the foundation of the safety filter presented in this paper.

The aforementioned techniques related to a safety framework either suffer from limited scalability to higher dimensional and complex systems, or only support simple model classes, such as linear parametric models. While also building on the same high-level concept, this paper addresses these limitations by 1) considering a general system model belief, which is well suited for exploration of highly nonlinear and unstable system dynamics, and 2) a unified optimization-based formulation for the safety policy (predictive safety filter), which avoids the explicit certification via a safe set.

## 3. Problem Statement

**Notation:** The set of integers in the interval  $[a,b] \subset \mathbb{R}$  is  $\mathcal{I}_{[a,b]}$ , and the set of integers in the interval  $[a,\infty) \subset \mathbb{R}$  is  $\mathcal{I}_{\geq a}$ . The i-th row and i-th column of a matrix  $A \in \mathbb{R}^{n \times m}$  is denoted by  $\mathrm{row}_i(A)$  and  $\mathrm{col}_i(A)$ . By  $\mathbb{I}_n$  we denote the vector of ones with length n.

Consider deterministic discrete-time systems of the form

$$x(k+1) = f(x(k), u(k); \theta_{\mathcal{R}}), \ \forall k \in \mathcal{I}_{>0},$$
 (1)

with dynamics  $f: \mathbb{X} \times \mathbb{U} \to \mathbb{R}^n$  parametrized by  $\theta_{\mathcal{R}} \in \mathbb{R}^p$  and deterministic initial condition  $x(0) = x_{\text{init}} \in \mathbb{X}$ . The system is subject to polyhedral state and input constraints  $\mathbb{X} := \{x \in \mathbb{R}^n | A_x x \leq \mathbb{1}_{n_x}\}$  and  $\mathbb{U} := \{u \in \mathbb{R}^m | A_u u \leq \mathbb{1}_{n_u}\}$ , originating from physical limitations and safety requirements. We consider the case of unknown 'real' parameters  $\theta_{\mathcal{R}}$ , but assume availability of a distribution

$$\theta \sim p(\theta)$$
 with mean  $\mathbb{E}[\theta]$ , (2)

which can be estimated from data. The proposed approach can be conceptually also extended to non-parametric model classes, which have bounded norm in a reproducing kernel Hilbert space.

The overall objective is to *safely* find a policy  $\pi_{\mathcal{L}}: \mathcal{I}_{\geq 0} \times \mathbb{X} \to \mathbb{U}$  that minimizes an objective, commonly chosen as the infinite horizon cost

$$J_{\infty}(x(k)) = \mathbb{E}\left[\sum_{i=k}^{\infty} \gamma^{i} \ell(x(i), \pi_{\mathcal{L}}(i, x(i)))\right]$$

with stage cost  $\ell: \mathbb{X} \times \mathbb{U} \to \mathbb{R}$  and discount factor  $\gamma \in (0,1)$ . In order to prescribe a desired level of cautiousness, we consider *safety* in terms of constraint satisfaction in probability as follows.

**Definition 3.1.** System (1) is operated safely at probability level  $p_S > 0$  if

$$\Pr\left(\forall k \in \mathcal{I}_{>0} : x(k) \in \mathbb{X}, u(k) \in \mathbb{U}\right) \ge p_{\mathcal{S}}. \tag{3}$$

This paper addresses the problem of implementing a safety filter as shown in Figure 1, which ensures closed-loop safety according to Definition 3.1. The filter enables application of any RL algorithm to the virtual input of the safe system, i.e.  $u_{\mathcal{L}}(k) = \pi_{\mathcal{L}}(k,x(k))$ , with the goal of minimizing the objective, while ensuring safety by selecting the input to the real system as  $u(k) = \pi_{\mathcal{S}}(k,x(k),u_{\mathcal{L}}(k))$ . In other words, the approach turns a safety-critical task into an unconstrained task with respect to the safe system dynamics  $f_{\mathcal{S}}(k,x(k),u_{\mathcal{L}}(k)) := f(x(k),\pi_{\mathcal{S}}(k,x(k),u_{\mathcal{L}}(k)))$  such that any RL algorithm can be safely applied. To further specify the desired properties of  $\pi_{\mathcal{S}}$ , consider the following definition of a safety certified learning-based control input.

**Definition 3.2.** An input  $u_{\mathcal{L}}(\bar{k})$  is *certified as safe* for system (1) at time step  $\bar{k}$  and state  $x(\bar{k})$  with respect to a *safety filter*  $\pi_{\mathcal{S}}: \mathcal{I}_{\geq 0} \times \mathbb{X} \times \mathbb{U} \to \mathbb{U}$ , if  $\pi_{\mathcal{S}}(\bar{k}, x(\bar{k}), u_{\mathcal{L}}(\bar{k})) = u_{\mathcal{L}}(\bar{k})$  and application of  $u(k) = \pi_{\mathcal{S}}(k, x(k), u_{\mathcal{L}}(k))$  for  $k \geq \bar{k}$  implies safety for all times according to (3).

Following this definition, the goal is to provide a safety filter  $\pi_{\mathcal{S}}$  that restricts learning as little as possible by certifying a possibly large set of learning inputs  $u_{\mathcal{L}}$  for a given state x(k). If the learning input cannot be certified as safe, the safety filter provides an alternative safe input, i.e.  $u(k) = \pi_{\mathcal{S}}(k, x(k), u_{\mathcal{L}}(k)) \neq u_{\mathcal{L}}(k)$ , where the filter aims at the smallest possible modification by, e.g., minimizing  $\|\pi_{\mathcal{S}}(k, x(k), u_{\mathcal{L}}(k)) - u_{\mathcal{L}}(k)\|_2$ . The following section introduces the mechanisms of the proposed predictive safety filter, which builds on a predictive constrained control formulation planning safe trajectories based on a probabilistic model belief to ensure safe system operation at all times according to Definition 3.1.

## **Nominal Predictive Safety Filter**

#### Nominal online problem:

$$\min_{\{u_{i|k}\}} \|u_{\mathcal{L}} - u_{0|k}\|_2 \tag{4a}$$

s.t.  $\forall i \in \mathcal{I}_{[0,N-1]}$ :

$$x_{i+1|k} = f(x_{i|k}, u_{i|k}; \bar{\theta}),$$
 (4b)

$$x_{i|k} \in \mathbb{X},$$
 (4c)

$$u_{i|k} \in \mathbb{U},$$
 (4d)

$$(x_{i|k}, u_{i|k}) \in \mathbb{Z}_c, \tag{4e}$$

$$x_{N|k} \in \mathcal{S}^t,$$
 (4f)

$$x_{0|k} = x(k). (4g)$$

## Algorithm 1 (Nominal predictive safety filter):

1: **function**  $\pi_{\mathcal{NS}}(k, x(k), u_{\mathcal{L}}(k))$ 

2: **if** (4) is feasible for horizon N **then** 

3: Define  $\bar{k} := k$ 

4: **return**  $u_{0|k,N}^*$ 

5: else if  $k < N + \bar{k}$  then

6: Solve (4) for horizon  $N - (k - \bar{k})$ 

7. wetrum a.\*

7: **return**  $u_{0|k,N-(k-\bar{k})}^*$ 

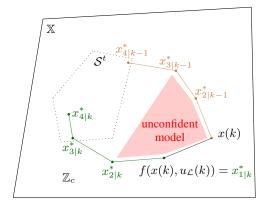
8: else

9: **return**  $\pi_{\mathcal{S}}^t(x(k))$ 

10: **end if** 

11: end function

#### Illustration of nominal safety filter:



## **Predictive Safety Filter**

#### Online problem:

$$\min_{\{v_{i|k}\}} \|u_{\mathcal{L}} - v_{0|k}\|_2 \tag{5a}$$

s.t.  $\forall i \in \mathcal{I}_{[0,N-1]}$ :

$$\mu_{i+1|k} = f(\mu_{i|k}, v_{i|k}; \bar{\theta}),$$
 (5b)

$$\mu_{i|k} \in \bar{\mathbb{X}}_i,$$
 (5c)

$$v_{i|k} \in \bar{\mathbb{U}}_i,$$
 (5d)

$$\mathcal{E}_{p_{\mathcal{S}}}(\mu_{i|k}, v_{i|k}) \subseteq \bar{\mathcal{E}}_{i},$$
 (5e)

$$\mu_{N|k} \in \bar{\mathcal{S}}_N^t, \tag{5f}$$

$$\mu_{0|k} = x(k). \tag{5g}$$

## Algorithm 2 (Predictive safety filter):

- 1: **function**  $\pi_{\mathcal{S}}(k, x(k), u_{\mathcal{L}}(k))$
- 2: **if** (5) is feasible for horizon N **then**
- 3: Define  $\bar{k} := k$
- 4: return  $v_{0|k,N}^*$
- 5: else if  $k < N + \bar{k}$  then
- 6: Solve (5) for horizon  $N (k \bar{k})$
- 7: return  $v_{0|N-(k-\bar{k})}^*$
- 8: else
- 9: **return**  $\pi_{\mathcal{S}}^t(x(k))$
- 10: **end if**
- 11: end function

#### Illustration of safety filter under uncertainty:

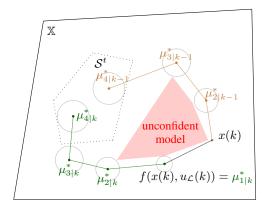


Figure 2. The basic idea of the predictive safety filter explained using a nominal, simplified version in the left column and the final method on the right. The illustrations show the system state at time k with safe backup plan for a shorter horizon obtained from the solution at time k-1, depicted in brown, and areas with poor model quality in red. An arbitrary learning input  $u_{\mathcal{L}}$  is certified if a feasible solution towards the terminal safe set  $\mathcal{S}^t$  can be found, as shown in green. If this new backup solution cannot be found and the planning problem (4)/(5) is infeasible, the system can be driven to the safe set  $\mathcal{S}^t$  along the brown previously computed trajectory. **Left (NPSF):** By assuming perfect system knowledge, the computed backup plans correspond exactly to the true state dynamics and constraints are guaranteed to be satisfied using the nominal backup trajectory. **Right (PSF):** Backup plans are computed w.r.t. the nominal expected state  $\mu$ . The true state trajectory lies within a growing tube around the nominal state with probability  $p_{\mathcal{S}}$ , which needs to be considered using tightened constraints according to (8).

## 4. Predictive safety filter

We first develop an intuitive understanding of the predictive safety filter by considering a simplified setting and assuming perfect model knowledge in Section 4.1, which is then extended in Section 4.2 to an uncertain model (1), (2) inferred from data, for which rigorous proofs are provided. As it will be shown, the presented method establishes safety by relying on controllability of (1) along system trajectories, in combination with an efficient mechanism enforcing the system to carefully enter uncertain areas within the state and input space.

#### 4.1. Nominal (simplified) predictive safety filter

Consider the simplified situation where the real system dynamics (1) are perfectly known for some subset of the state and input space, as specified in the following.

**Assumption 4.1.** There exists a set  $\mathbb{Z}_c \subseteq \mathbb{X} \times \mathbb{U}$ , such that for all  $(x, u) \in \mathbb{Z}_c$  and some  $\bar{\theta} \in \mathbb{R}^q$  it holds  $f(x, u; \bar{\theta}) = f(x, u; \theta_R)$ .

Similar to Wabersich & Zeilinger (2018b), we propose a predictive safety filter that is not pre-computed, but defined via an optimization problem and computed on-the-fly. The main working mechanism is the construction of safe backup plans that, if applied, would keep the system provably safe, see Figure 2 (left) for an illustration.

The backup plans are defined via (4), where  $\{x_{i|k,N}\}$  denote the planned states computed at the current time step k and predicted i time steps into the future with planning horizon N using the corresponding input sequence  $\{u_{i|k,N}\}$ . One of the key challenges in computing the backup plans is to deal with the fact that a good model is not known in unexplored regions of the state-space, i.e.  $\mathbb{X} \setminus \mathbb{Z}_c$ , shown as red (unconfident model) sets in Figure 2. In the nominal setting, we simply address this problem by enforcing the system to strictly stay within the confident model subset  $\mathbb{Z}_c$  via (4e). One of the main problems addressed in the next section will be to relax this constraint in order to enable cautious exploration of such unconfident subsets. The purpose of the remaining constraints in (4) is to construct backup plans that lead the system within state and input constraints  $\mathbb{X}$  and  $\mathbb{U}$  (4c), (4d) into a safe terminal set  $S^t$  in N steps (4f).

The objective for constructing the backup plans in (4) is to minimize the deviation between the first element of the input sequence  $u_{0|k}$  and the requested input  $u_{\mathcal{L}}(k)$  from the RL algorithm, such that  $u_{0|k} = u_{\mathcal{L}}(k)$  if  $u_{\mathcal{L}}(k)$  is safe. The resulting nominal predictive safety filter is then given by  $\pi_{\mathcal{NS}}(k,x(k),u_{\mathcal{L}}(k)) = u_{0|k}^*$ , with  $u_{0|k}^*$  being the optimal first control input obtained from (4).

In order to ensure constraint satisfaction beyond the plan-

ning horizon, (4) utilizes a mechanism common in predictive control (see e.g. Chen & Allgöwer (1998)), by requiring the last state of the sequence  $\{x_{N|k}\}$  to lie in a safe terminal set of system states  $\mathcal{S}^t$ , for which a locally valid safety filter  $\pi^t_{\mathcal{S}}$  is known.

**Assumption 4.2.** There exists a terminal safe set  $S^t := \{x \in \mathbb{R}^n | a_{\mathcal{S}}(x) \leq \mathbb{1}_{n_{\mathcal{S}}}\} \subseteq \mathbb{X}$ , with  $a_{\mathcal{S}}$  Lipschitz continuous with Lipschitz constant  $L_{\mathcal{S}}$ , and a corresponding terminal safety filter  $\pi_{\mathcal{S}}^t : \mathcal{I}_{\geq 0} \times \mathbb{X} \times \mathbb{U} \to \mathbb{U}$ , such that if  $x(\bar{k}) \in S^t$ , then application of  $u(k) = \pi_{\mathcal{S}}^t(k, x(k), u_{\mathcal{L}}(k))$  implies that  $x(k) \in \mathbb{X}$  and  $u(k) \in \mathbb{U}$  for all  $k > \bar{k}$ .

A terminal safe set  $\mathcal{S}^t$  and the corresponding controller  $\pi^t_{\mathcal{S}}$  can be chosen, e.g., as a classical terminal set for nonlinear (robust) MPC (Chen & Allgöwer, 1998; Yu et al., 2013), regions around stable steady-states of system (1), or using expert system knowledge as it is demonstrated in Section 5.

Based on problem (4), the predictive safety filter  $\pi_{\mathcal{S}}$  is defined by Algorithm 1 (Figure 2, left). At every time step, we attempt to solve optimization problem (4). If problem (4) is feasible at time k, safety, i.e.,  $x(k) \in \mathbb{X}$ ,  $u(k) \in \mathbb{U}$ , directly follows from (4c), (4d). Due to the generality of the terminal safe set, however, problem (4) may become infeasible for some state x(k), even after being feasible at the previous time step x(k-1). Algorithm 1 implements an additional mechanism to provide a safe trajectory and input sequence towards the terminal safe set also for this case.

Assume that (4) was feasible at time k-1 with corresponding optimal input sequence  $\{u_{i|k-1,N}^*\}$ . Application of  $u(k-1)=u_{0|k-1,N}^*$  results in a safe state x(k) as depicted in Figure 2 (left), because  $(x(k-1),u(k-1))\in\mathbb{Z}_c$  by (4e) and therefore  $x(k)=f(x(k-1),u(k-1);\bar{\theta})\in\mathbb{X}$  by (4c). At the next time step k, if (4) is not feasible, we can still solve (4) with a reduced planning horizon N-1. This can be easily verified by noting that  $u_{i|k}=u_{i+1|k-1}^*$  for  $i\in\mathcal{I}_{[0,N-2]}$ , i.e. the tail of the previously computed feasible trajectory from time step k-1, is a feasible solution as depicted by the brown trajectory in Figure 2 (left). Feasibility of (4) for a reduced horizon again directly provides  $x(k)\in\mathbb{X}, u(k)\in\mathbb{U}$ .

The same holds true in case that j < N steps were consecutively infeasible for planning horizon N, i.e. (4) will then be feasible with horizon N-j until we reach the safe terminal set. This shortening of the horizon is implemented in lines 6-7 of Algorithm 1. If the horizon length reaches 0, the state is in the terminal set and  $\pi_{\mathcal{S}}^t$  can be applied to ensure  $x(k) \in \mathbb{X}$ ,  $u(k) \in \mathbb{U}$  (line 9). Note again that if (4) is feasible at time k (line 3-4),  $u_{0|k}^*$  can be applied, which ideally results in  $u_{\mathcal{L}}(k)$  (i.e. objective (4a) is zero) as shown in Figure 2 (left) together with the optimal backup plan in green. Algorithm 1 therefore ensures constraint satisfac-

tion at all time steps, realizing a predictive safety filter in a receding horizon fashion with varying prediction length.

The next section will extend the previously introduced basic concept of the predictive safety filter to consider a datadriven approximate system belief, represented by (1), (2), subject to probabilistic constraint satisfaction (3).

## 4.2. Predictive safety filter

A key goal of the safety filter is to support exploration beyond available data via the learning policy  $\pi_{\mathcal{L}}(k)$ , in which case Assumption 4.1 does not necessarily hold. While fast approximate computation of the backup trajectories can still be performed online using the mean estimate  $\bar{\theta}$  of the parameter  $\theta_{\mathcal{R}}$ , we need to safely handle the resulting non-vanishing model error

$$e(k,\bar{\theta}) := f(x(k),u(k);\theta_{\mathcal{R}}) - f(x(k),u(k);\bar{\theta}).$$
 (6)

In the following, we first treat uncertainty via a uniform error bound to introduce the safety filter for uncertain systems, which is then extended to consider a less conservative bound and impose it as a constraint in the filter planning problem, in order to reduce conservatism.

**Uniformly bounded model error.** Assume that the model error with respect to the point estimate  $\bar{\theta}$  can be reasonably bounded as

$$\Pr(e(k, \bar{\theta}) \in \mathcal{E} \text{ for all } k \in \mathcal{I}_{>0}) \ge p_{\mathcal{S}},$$
 (7)

with  $\mathcal{E} \subseteq \mathbb{R}^n$  compact. In this case, the filter can still compute backup plans using the point estimate  $\bar{\theta}$ , however, in contrast to the nominal case in Section 4.1, the constraints in (4) are modified such that prediction errors induced by (6) are compensated to ensure constraint satisfaction.

We denote the nominal (expected) system states as  $\{\mu_{i|k}\}$ , corresponding to the nominal input sequence  $\{v_{i|k}\}$  according to  $\mu_{i+1|k} = f(\mu_{i|k}, v_{i|k}; \bar{\theta})$ . Due to the model error (6), we need to address the fact that potentially  $x(k+1) \notin$  $\mathbb{X}$ , i.e.  $A_x x(k+1) > \mathbb{1}_{n_x}$ , when applying the nominal input  $v_{0|k}$ , even though the corresponding nominal predicted state satisfies  $\mu_{1|k}^* \in \mathbb{X}$ . A common strategy for achieving robustness in predictive control is to tighten the constraints by leveraging controllability along any possible predicted state sequence  $\{\mu_{i|k}\}$  (Mayne, 2014). Intuitively speaking, controllability enables efficient compensation of deviations  $x(i) - \mu_{i|k}$  via feedback control. More precisely, the possible deviations can be bounded by a decay constant, expressed by parameter  $\rho$ , at which a controller can compensate disturbances of a certain magnitude, captured by parameter  $\epsilon$ . Using these two measures, deviations from the planned nominal trajectory are compensated via an iterative tightening of the constraints, which allows to flexibly respond to upcoming disturbances at consecutive time steps via replanning, thereby enabling overall constraint satisfaction. Following Köhler et al. (2018b), we tighten the constraints (4c), (4d), and (4f) in the computation of the backup plans as

$$\bar{\mathbb{X}}_i := \{ x \in \mathbb{R}^n | A_x x \le (1 - \epsilon_i) \mathbb{1}_{n_x} \}, \tag{8a}$$

$$\bar{\mathbb{U}}_i := \{ u \in \mathbb{R}^m | A_u u \le (1 - \epsilon_i) \mathbb{1}_{n_u} \}, \tag{8b}$$

$$\bar{\mathcal{S}}_N^f := \{ x \in \mathbb{R}^n | a_{\mathcal{S}}(x) \le (1 - \epsilon_N) \mathbb{1}_{n_{\mathcal{S}}} \}, \qquad (8c)$$

implementing a trade-off between compensation and magnitude of disturbances via the converging recursion

$$\begin{cases}
\epsilon_0 = 0 \\
\epsilon_{i+1} \coloneqq \epsilon_i + \sqrt{\rho^i} \epsilon
\end{cases} \Rightarrow \epsilon_i = \epsilon \frac{1 - \sqrt{\rho^i}}{1 - \sqrt{\rho}}, \quad (9)$$

with design parameter  $\epsilon > 0$  and parameter  $\rho \in (0,1)$  that depends on system (1) as follows.

**Assumption 4.3.** There exists a control policy  $\pi: \mathbb{X} \times \mathbb{X} \times \mathbb{U} \to \mathbb{R}^m$ , a function  $V: \mathbb{X} \times \mathbb{X} \to \mathbb{R}_{>0}$ , which is continuous in its first argument and satisfies V(x,x)=0 for all  $x \in \mathbb{X}$ , and parameters  $c_l, c_u, \delta, \pi_{\max} \in \mathbb{R}_{>0}$ ,  $\rho \in (0,1)$ , such that for a given  $\bar{\theta} \in \mathbb{R}^q$  the following properties hold for all  $x, z \in \mathbb{X}$  with  $V(x,z) < \delta$  and  $v \in \mathbb{U}$ :

$$c_l \|x - z\|_2^2 \le V(x, z) \le c_u \|x - z\|_2^2$$
 (10a)

$$\|\pi(x, z, v) - v\|_2 \le \pi_{\max} \|x - z\|_2$$
 (10b)

$$V\left(f(x,\pi(x,z,v),\bar{\theta}),f(z,v,\bar{\theta})\right)\leq \rho V(x,z). \tag{10c}$$

Informally, Assumption 4.3 defines how well the uncertain system can be controlled in a neighborhood of predicted nominal backup plans  $\{\mu_{i|k}^*\}$ . Intuitively speaking, considering the task of tracking a reference trajectory as an optimal control problem with value function V (using for example a linear quadratic regulator in the linear dynamics setting), parameter  $\rho$  defines 'how fast' a reference can be reached, measured in terms of the contraction rate of the optimal tracking cost V. Interestingly, this translates into a system-theoretic requirement on system (1), or more precisely to local incremental stabilizability, which can be formally verified based on a system linearization as discussed in (Köhler et al., 2018a, Prop. 1). The condition is provided in the Appendix. It is important to note that the final algorithm only requires existence of the policy  $\pi$  and the corresponding function V rather than their explicit form.

These concepts lead to a robustified version of the nominal predictive safety filter defined in (5) and Algorithm 2 (Figure 2, right), where we omit (5e) in the case of uniformly bounded errors (7). Assumption 4.3 ties the model uncertainty (7) to the constraint tightening (8) in order to ensure existence of a safe backup plan at all times and allows for extending the arguments for the nominal case to a probabilistic model belief. If (5) is feasible at time k-1 and the

error bound  $\mathcal{E}$ , i.e.  $\max_{e \in \mathcal{E}} \|e\|_2$ , is sufficiently small with probability  $p_{\mathcal{S}}$ , then at time k, the input sequence based on the plan computed at time step k-1

$$v_{i|k} \coloneqq \pi(\mu_{i|k}, \mu_{i+1|k-1,N}^*, v_{i+1|k-1,N}^*)$$
 (11)

for  $i \in \mathcal{I}_{[0,N-2]}$  with  $\mu_{0|k} = x(k)$ ,  $\pi$  according to Assumption 4.3, and  $\mu_{i|k}$  according to (5b), always provides a feasible solution to (5) with planning horizon N-1 (Algorithm 2, line 6) with probability  $p_{\mathcal{S}}$ . Again, the tracking policy  $\pi$  is only used in order to show that there exists a solution to (5), but it is not needed for implementation of the approach. The same argument holds true for all  $\tilde{k} \in \mathcal{I}_{[k+1,k+N-1]}$  until the terminal set is reached (line 10), which allows us to establish safety at all times in a similar fashion as in the nominal case. A formal proof will be given in the following for the more general case including a constraint on model confidence.

Planning in confident subspaces: In order to reduce conservatism introduced by uniformly overbounding the uncertainty in (7), a central novelty in the proposed safety filter is the ability to restrict planning to regions in the state and input space  $\mathbb{Z}_c$  where we are sufficiently confident about the system dynamics and the assumed error bound holds, i.e. ' $\mathbb{Z}_c := \mathbb{X} \setminus \text{unconfident model'}$ , compare also with Figure 2. A simple approach would be to compute such a region offline and add it as an additional state and input constraint, as it was similarly done for the case of linear dynamics with state dependent uncertainties by Soloperto et al. (2018). However, it is in general difficult to compute  $\mathbb{Z}_c$  analytically and in addition, the set needs to be recomputed once the model belief (1), (2) is updated based on observed data. We therefore reformulate the requirement to stay inside  $\mathbb{Z}_c$  as an implicit constraint, avoiding the explicit computation of  $\mathbb{Z}_c$ , and include it in the computation of the safety filter in (5) using the following definition.

**Definition 4.4.** A set-valued map  $\mathcal{E}_{p_{\mathcal{S}}}: \mathbb{X} \times \mathbb{U} \to 2^{\mathcal{E}}$  with  $\mathcal{E} \subset \mathbb{R}^n$  is said to be a *set-valued model confidence map* associated with (1), (2), for a given  $\bar{\theta} \in \mathbb{R}^q$  at probability level  $p_{\mathcal{S}} > 0$  if, at probability greater or equal to  $p_{\mathcal{S}}$ , it holds for all  $k \in \mathcal{I}_{>0}$ ,  $x(k) \in \mathbb{X}$ ,  $u(k) \in \mathbb{U}$  that

$$e(k,\bar{\theta}) \in \mathcal{E}_{p_{\mathcal{S}}}(x(k),u(k))$$
 (12)

with  $e(k, \bar{\theta})$  as defined in (6).

Note that according to Definition 4.4 it is not sufficient to guarantee that (12) holds for some k, but it has to hold for all  $k \ge 0$  in order to ensure safety for all times, including also the case  $k \to \infty$ . An example for how to design (12) from data is given in the following.

**Examples.** Consider a Bayesian description of (1) with prior distribution  $p(\theta)$  and posterior estimate  $\theta \sim$ 

 $p(\theta|\mathcal{D})$ , which is based on available system data  $\mathcal{D} := \{(x(k),u(k)),f(x(k),u(k);\theta_{\mathcal{R}})\}_{k=1}^{N_{\mathcal{D}}}$ . Define a confidence region  $\mathcal{C}_{p_{\mathcal{S}}}(p(\theta|\mathcal{D}))$  at probability level  $p_{\mathcal{S}}>0$  of the random parameters  $\theta$  as  $\Pr(\theta \in \mathcal{C}_{p_{\mathcal{S}}}(p(\theta|\mathcal{D}))) \geq p_{\mathcal{S}}$ . Using the above notation, a set-valued model confidence map according to Definition 4.4 is given by

$$\mathcal{E}_{p_{\mathcal{S}}}(x, u) = \{ e \in \mathbb{R}^n | e = f(x, u, \theta) - f(x, u; \bar{\theta}), \theta \in \mathcal{C}_{p_{\mathcal{S}}}(p(\theta|\mathcal{D})) \}.$$

Note that similar set-valued model confidence maps can be obtained when using non-parametric Gaussian process regression, by assuming that the system dynamics (1) has bounded norm in a reproducing kernel Hilbert space (Chowdhury & Gopalan, 2017, Theorem 2).

In case of large amounts of available data on the whole state and input space, e.g., for models from deep learning, one can simply choose a uniform confidence map, i.e.,  $\forall x \in \mathbb{X}, \ u \in \mathbb{U}: \ \mathcal{E}_{\mathcal{PS}}(x,u) = \bar{\mathcal{E}} \subset \mathbb{R}^n$ , reducing to the special case (7).

As discussed for the case of uniformly bounded errors, the tightened constraints (8) ensure safety if (7) holds for  $\max_{e \in \bar{\mathcal{E}}} \|e\|_2$  small enough. Since  $e(k, \bar{\theta})$  is unknown and we cannot simply impose  $e(k, \bar{\theta}) \in \bar{\mathcal{E}}$  in (5), we make use of the model confidence map in Definition 4.4 to enforce

$$\mathcal{E}_{p_{\mathcal{S}}}(x(k), u(k)) \subseteq \bar{\mathcal{E}},$$
 (13)

implying  $e(k,\bar{\theta}) \in \bar{\mathcal{E}}$  with probability  $p_{\mathcal{S}}$ . In the following, we consider error bounds  $\bar{\mathcal{E}}$  of the form  $\bar{\mathcal{E}} \coloneqq \{e \in \mathbb{R}^n | a_{\mathcal{E}}(e) \leq \mathbb{1}_{n_{\mathcal{E}}}\}$ , allowing us to enforce condition (13) by imposing (5e) on the nominal plan  $\{\mu_{i|k}\}, \{v_{i|k}\}$ , where constraint (13) is tightened similarly to (8) using

$$\bar{\mathcal{E}}_i := \{ e \in \mathbb{R}^n | a_{\mathcal{E}}(e) \le (1 - \epsilon_i) \mathbb{1}_{n_{\mathcal{E}}} \}. \tag{14}$$

The tightening again ensures existence of a feasible solution when replanning with a shorter horizon (Algorithm 2, line 6). In order for the filter to ensure safety in probability using (5e), we additionally require that small changes of the nominal predicted trajectory must not lead to arbitrary large changes in the model confidence by assuming that the set-valued model confidence map is Lipschitz continuous in terms of the Hausdorff metric (see Definition A.2 in the appendix).

**Assumption 4.5.** There exists a set-valued confidence map  $\mathcal{E}_{p_{\mathcal{S}}}$  associated with (1), (2), which is Lipschitz continuous with Lipschitz constant  $L_{\mathcal{E}_{p_{\mathcal{S}}}}$  under the Hausdorff metric with respect to  $d_{\mathbb{R}^m}(a,b) \coloneqq \|a-b\|_2$ .

Note that for many common models, Assumption 4.5 is generally fulfilled, compare also with Fisac et al. (2017).

The above assumptions allow for extending the ideas from the uniform error bound to make use of a potentially reduced error bound that is ensured by imposing (5e) on the backup plan, and thereby again characterize the relation between the tightening  $\epsilon$  in (8),(14) and the tolerated model error  $\bar{\mathcal{E}}$ . This leads us to the main result of the paper, showing that the proposed predictive safety filter guarantees safety in probability at all times according to Definition 3.1.

**Theorem 4.6.** Let Assumptions 4.2, 4.3 and 4.5 hold. For every  $\epsilon > 0$  there exists a corresponding  $\hat{e} > 0$ , such that  $\max_{e \in \bar{\mathcal{E}}} \|e\|_2 \leq \hat{e}$  and initial feasibility of (5) for x(0) implies that  $u(k) = \pi_{\mathcal{S}}(k, x(k), u_{\mathcal{L}}(k))$  as defined in Algorithm 2 ensures safe system operation according to Definition 3.1.

The proof is provided in the appendix. Theorem 4.6 implies that one can always specify a constraint tightening and impose a corresponding sufficiently small admissible error set  $\bar{\mathcal{E}}$ , such that if (5) is initially feasible for x(0), application of Algorithm 2 will keep the system safe in probability. Note that a particular  $\bar{\mathcal{E}}$  can lead to initial infeasibility of (5) due to (5e), in which case either the constraint tightening would have to be adjusted or the model needs to be improved.

Remark 4.7. Given a constraint tightening (9), a simple design procedure in order to determine the corresponding model error set can be obtained as follows. Let  $\mathcal{E}(\hat{e})$  be a parametrized set such that  $\max_{e \in \bar{\mathcal{E}}} \|e\|_2 \leq \hat{e}$  holds. Define a desired number of samples (e.g. related to Hoeffdings inequality)  $\mathbb{S} := (\tilde{x}_0, \tilde{\theta}) \sim (\mathcal{X}_N, p(\theta))$ , where  $\mathcal{X}_N$  is the feasible set of (5) with planning horizon N. Reduce  $\hat{e}$  until all closed-loop simulations using  $\mathbb{S}$  satisfy the tightened state and input constraints.

Remark 4.8. Similar results can be obtained using a different constraint tightening usually employed in tube-based MPC, see for example Limon et al. (2009); Mayne et al. (2011); Yu et al. (2013). However, such a tightening tends to be either more difficult to compute or to be more conservative compared to the proposed approach.

#### 5. Numerical example

We investigate the classical control problem of swinging up a pendulum from the downward position (angle  $\alpha=0$  deg) to the upward position ( $\alpha=180$  deg) with limited input authority and under challenging safety constraints of the form -60 deg  $\leq \alpha \leq 185$  deg, such that the pendulum is not allowed to overshoot. Constraints of this type in combination with unknown system dynamics have not been addressed with another control approach according to our knowledge. Further details on the system can be found in Appendix A.3.

The transition model (1), (2) for states  $x_1 = \alpha$ ,  $x_2 = \dot{\alpha}$  is obtained via linear Bayesian regression, i.e.,  $f(x, u) = \theta^{\top} \phi(x)$ , with  $\theta \in \mathbb{R}^{2 \times 9}$  and polynomial features  $\phi(x) = \frac{1}{2} (x_1 + y_2)^{-1}$ 

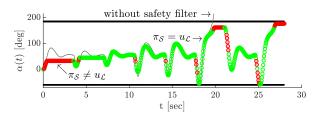


Figure 3. Learning to safely swing-up a pendulum under challenging position constraints. The markers show the closed-loop trajectory under the safety filter  $\pi_{\mathcal{S}}$ . Green dots indicate unmodified application of the learning policy, while red dots represent the magnitude of safety ensuring modifications. The gray line shows the closed-loop behavior without safety filter.

 $[\alpha, \alpha^3, \alpha^5, \dot{\alpha}, \dot{\alpha}^3, \dot{\alpha}^5, u, u^3, u^5]^{\top}$ . The set-valued model confidence map according to Definition 4.4 is defined as a box, the size of which is defined by the standard deviation of the predicted state transitions  $\sigma_{f_i}(x)$  for i =1,2. The tightening was experimentally chosen to  $\rho =$  $0.99, \epsilon = 0.02$  using Monte Carlo sampling as described in Remark 4.7. The corresponding admissible error set  $\bar{\mathcal{E}}$  is defined as the 1-norm Ball with radius 0.02 and therefore constraint (5e) results in  $\sum_{i=1}^2 \sigma_{f_i}([\mu_{i|k}^\top, v_{i|k}]) \leq 0.02$ , which can be efficiently implemented. As the terminal safe set we select  $S^t := \{\alpha, \dot{\alpha} | -30 \deg \leq \alpha \leq 30 \deg, |\dot{\alpha}| \leq$  $30 \ \mathrm{deg/sec}$  with  $\pi_{\mathcal{S}}^t = 0$ . The resulting problem (5) with planning horizon N=20 was solved in real-time using IPOPT (Wächter & Biegler, 2006) together with the CasADi framework (Andersson et al., 2018). For learning the swing-up task, we make use of simple stochastic policy search with a bang-bang policy, see appendix, where one episode consists of 70 time steps. After each episode, we update the model belief using the acquired data.

As shown in Figure 3, after 8 episodes (30 sec), the pendulum was successfully controlled into the upward position without constraint violations. The number of interventions of the safety filter decreases as the model gets more accurate. In contrast, direct application of the learning policy without the safety filter leads to constraint violations.

## 6. Conclusion

This paper has addressed the problem of safe RL by introducing a predictive safety filter, which enables modularity in terms of safety and the employed RL algorithm. An optimization-based formulation was proposed that provides rigorous safety guarantees using a possibly data-driven approximate system model. By its capability to consider nonlinear and complex system descriptions without being overly conservative, we believe that the proposed approach is an important step towards safe RL for realistic applications.

#### References

- Achiam, J., Held, D., Tamar, A., and Abbeel, P. Constrained policy optimization. In *International Conference on Machine Learning*, pp. 22–31, 2017.
- Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané, D. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*, 2016.
- Andersson, J. A. E., Gillis, J., Horn, G., Rawlings, J. B., and Diehl, M. CasADi A software framework for non-linear optimization and optimal control. *Mathematical Programming Computation*, 2018.
- Aswani, A., Gonzalez, H., Sastry, S. S., and Tomlin, C. Provably safe and robust learning-based model predictive control. *Automatica*, 49(5):1216 – 1226, 2013.
- Berkenkamp, F. and Schoellig, A. P. Safe and robust learning control with Gaussian Processes. In *European Control Conference (ECC)*, pp. 2496–2501, July 2015.
- Berkenkamp, F., Moriconi, R., Schoellig, A. P., and Krause, A. Safe learning of regions of attraction for uncertain, nonlinear systems with Gaussian Processes. In *55th IEEE Conference on Decision and Control (CDC)*, pp. 4661–4666, Dec 2016.
- Berkenkamp, F., Turchetta, M., Schoellig, A., and Krause, A. Safe model-based reinforcement learning with stability guarantees. In *Advances in Neural Information Processing Systems 30*, pp. 908–918. 2017.
- Bouffard, P., Aswani, A., and Tomlin, C. Learning-based model predictive control on a quadrotor: Onboard implementation and experimental results. In *2012 IEEE International Conference on Robotics and Automation*, pp. 279–284, May 2012.
- Chen, H. and Allgöwer, F. A quasi-infinite horizon non-linear model predictive control scheme with guaranteed stability. *Automatica*, 34(10):1205 1217, 1998.
- Chowdhury, S. R. and Gopalan, A. On kernelized multiarmed bandits. In *International Conference on Machine Learning (ICML)*, pp. 844–853, 2017.
- Clavera, I., Nagabandi, A., Fearing, R. S., Abbeel, P., Levine, S., and Finn, C. Learning to adapt: Metalearning for model-based control. *arXiv preprint arXiv:1803.11347*, 2018.
- Fisac, J. F., Akametalu, A. K., Zeilinger, M. N., Kaynama, S., Gillula, J., and Tomlin, C. J. A general safety framework for learning-based control in uncertain robotic systems. arXiv preprint arXiv:1705.01292, 2017.

- Garcia, J. and Fernández, F. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learn*ing Research, 16:1437–1480, 2015.
- Ghosh, S., Berkenkamp, F., Ranade, G., Qadeer, S., and Kapoor, A. Verifying controllers against adversarial examples with Bayesian optimization. In *Proc. of the International Conference on Robotics and Automation (ICRA)*, 2018.
- Gillula, J. H. and Tomlin, C. J. Guaranteed safe online learning of a bounded system. In *Intelligent Robots and Systems (IROS)*, 2011 IEEE/RSJ International Conference on, pp. 2979–2984. IEEE, 2011.
- Hewing, L. and Zeilinger, M. N. Cautious model predictive control using Gaussian Process regression. *arXiv* preprint arXiv:1705.10702, 2017.
- Hewing, L., Liniger, A., and Zeilinger, M. N. Cautious NMPC with Gaussian Process dynamics for miniature race cars. *arXiv* preprint arXiv:1711.06586, 2017.
- Kamthe, S. and Deisenroth, M. P. Data-efficient reinforcement learning with probabilistic model predictive control. *arXiv* preprint arXiv:1706.06491, 2017.
- Köhler, J., Müller, M. A., and Allgöwer, F. Nonlinear reference tracking: An economic model predictive control perspective. *IEEE Transactions on Automatic Control*, pp. 1–1, 2018a.
- Köhler, J., Müller, M. A., and Allgöwer, F. A novel constraint tightening approach for nonlinear robust model predictive control. In *2018 Annual American Control Conference (ACC)*, pp. 728–734, June 2018b.
- Koller, T., Berkenkamp, F., Turchetta, M., and Krause, A. Learning-based model predictive control for safe exploration and reinforcement learning. arXiv preprint arXiv:1803.08287, 2018.
- Larsen, R. B., Carron, A., and Zeilinger, M. N. Safe learning for distributed systems with bounded uncertainties. *20th IFAC World Congress*, 50(1):2536 2542, 2017.
- Levine, S., Finn, C., Darrell, T., and Abbeel, P. End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, 17(1):1334–1373, 2016.
- Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971, 2015.
- Limon, D., Alamo, T., Raimondo, D., De La Peña, D. M., Bravo, J., Ferramosca, A., and Camacho, E. Input-tostate stability: a unifying framework for robust model

- predictive control. In *Nonlinear model predictive control*, pp. 1–26. Springer, 2009.
- Limon, D., Calliess, J., and Maciejowski, J. Learning-based nonlinear model predictive control. *IFAC-PapersOnLine*, 50(1):7769 7776, 2017.
- Lorenzen, M., Allgöwer, F., and Cannon, M. Adaptive model predictive control with robust constraint satisfaction. *IFAC-PapersOnLine*, 50(1):3313 3318, 2017. 20th IFAC World Congress.
- Majumdar, A. and Tedrake, R. Funnel libraries for realtime robust feedback motion planning. *The International Journal of Robotics Research*, 36(8):947–982, 2017.
- Marzat, J., Walter, E., and Piet-Lahanier, H. A new expected-improvement algorithm for continuous minimax optimization. *Journal of Global Optimization*, 64 (4):785–802, aug 2016.
- Mayne, D. Q. Model predictive control: Recent developments and future promise. *Automatica*, 50(12):2967–2986, 2014.
- Mayne, D. Q., Kerrigan, E. C., Van Wyk, E., and Falugi, P. Tube-based robust nonlinear model predictive control. *International Journal of Robust and Nonlinear Control*, 21(11):1341–1353, 2011.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. Human-level control through deep reinforcement learning. *Nature*, 518 (7540):529–533, 2015.
- Ostafew, C. J., Schoellig, A. P., and Barfoot, T. D. Robust constrained learning-based NMPC enabling reliable mobile robot path tracking. *The International Journal of Robotics Research*, 35(13):1547–1563, 2016.
- Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P. Trust region policy optimization. In *International Conference on Machine Learning*, pp. 1889–1897, 2015.
- Soloperto, R., Müller, M. A., Trimpe, S., and Allgöwer, F. Learning-based robust model predictive control with state-dependent uncertainty. *IFAC-PapersOnLine*, 51 (20):442–447, 2018.
- Tanaskovic, M., Fagiano, L., Smith, R., Goulart, P., and Morari, M. Adaptive model predictive control for constrained linear systems. In *Control Conference (ECC)*, 2013 European, pp. 382–387. IEEE, 2013.
- Tedrake, R., Manchester, I. R., Tobenkin, M., and Roberts, J. W. Lqr-trees: Feedback motion planning via sumsof-squares verification. *The International Journal of Robotics Research*, 29(8):1038–1052, 2010.

- Wabersich, K. P. and Toussaint, M. Automatic testing and minimax optimization of system parameters for best worst-case performance. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5533–5539, 2015.
- Wabersich, K. P. and Zeilinger, M. N. Scalable synthesis of safety certificates from data with application to learning-based control. In *2018 European Control Conference* (*ECC*), pp. 1691–1697, 2018a.
- Wabersich, K. P. and Zeilinger, M. N. Linear model predictive safety certification for learning-based control. *arXiv* preprint arXiv:1803.08552, 2018b.
- Wächter, A. and Biegler, L. T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57, 2006.
- Wieland, P. and Allgöwer, F. Constructive safety using control barrier functions. *IFAC Proceedings Volumes*, 40 (12):462–467, 2007.
- Yu, S., Maier, C., Chen, H., and Allgöwer, F. Tube MPC scheme based on robust control invariant set with application to Lipschitz nonlinear systems. *Systems & Control Letters*, 62(2):194–200, 2013.

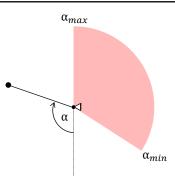


Figure 4. Learning to safely swing-up a pendulum under challenging position constraints.

# A. Appendix

## A.1. Sufficient condition for Assumption 4.3

According to Köhler et al. (2018a, Prop. 1), the following verifiable assumption implies Assumption 4.3 and relates it to controllability of system (1).

**Assumption A.1.** Let  $r:=(z,v)\in\mathbb{X}\times\mathbb{U}$  and define the linearization  $A_r:=\frac{\partial f}{\partial x}(r,\bar{\theta}),\ B_r:=\frac{\partial f}{\partial u}(r,\bar{\theta}).$  For any  $r\in\mathbb{X}\times\mathbb{U}$ , the pair  $(A_r,B_r)$  is stabilizable, i.e. there exist  $K_r\in\mathbb{R}^{m\times n},\ P_r,Q_r\in\mathbb{R}^{n\times n}$  positive definite and continuous in r, such that

$$P_r - (A_r + B_r K_r)^{\top} P_r (A_r + B_r K_r) = Q_r.$$

Furthermore, there exists a constant  $c \in \mathbb{R}_{>0}$ , such that for any  $r^+ = (z^+, v^+) \in \mathbb{X} \times \mathbb{U}$  with  $z^+ = f(z, v, \bar{\theta})$ , the corresponding matrix  $P_{r^+}$  satisfies:

$$\lambda_{\max}(P_r^{-1}P_{r^+})Q_r \ge (\lambda_{\max}(P_r^{-1}P_{r^+}) - 1)P_r + cI_n.$$
(15)

Given Assumption A.1, we can choose  $V(x,z)=(x-z)^{\top}P_r(x-z)$  in Assumption 4.3, with (15) bounding the rate at which V(x,z) can possibly change in any time step when applying  $u=\pi(x,z,v)=v+K_r(x-z)$ .

#### A.2. Hausdorff metric

**Definition A.2.** The *Hausdorff metric* between two sets  $\mathcal{A}$  and  $\mathcal{B}$  in a metric space  $(M, d_{\mathrm{M}})$  is defined as

$$d_{\mathrm{H}}(\mathcal{A}, \mathcal{B}) \coloneqq \max \left\{ \sup_{a \in \mathcal{A}} \inf_{b \in \mathcal{B}} d_{\mathrm{M}}(a, b), \inf_{a \in \mathcal{A}} \sup_{b \in \mathcal{B}} d_{\mathrm{M}}(a, b) \right\}.$$

## A.3. Details of numerical example

The discretized dynamics x(k+1) = f(x(k), u(k)) of the nonlinear pendulum in Figure 4 are described as

$$x(k) + h \left( \frac{x_2(k)}{-\frac{g}{l}\sin(x_1(k)) - \frac{\eta}{ml^2}x_2(k) + \frac{1}{ml^2}u(k)} \right)$$

where  $x_1(k)=\alpha(k)$  is the angle,  $x_2(k)=\dot{\alpha}(k)$  is the angular velocity at time step  $k,\ h=0.05\ \mathrm{s}$  is the discretization interval,  $g=9.81\ \mathrm{m/s^2}$  is the gravity constant,  $l=0.5\ \mathrm{m}$  is the length,  $m=0.15\ \mathrm{kg}$  is the mass,  $\eta=0.1\ \mathrm{Nms/rad}$  is the friction and the input torque u is restricted to  $|u|\leq 0.6\ \mathrm{Nms/rad}$ . As an increased challenge to the classical setting, we impose safety constraints of the form  $-60\ \mathrm{deg} \leq \alpha \leq 185\ \mathrm{deg}$  such that the pendulum is not allowed to overshoot, which makes the swing-up task challenging for unknown system dynamics.

For learning the swing-up task, we make use of simple stochastic policy search with a bang-bang policy

$$\pi_{\mathcal{L}}(k, x; k_s) = \begin{cases} -0.6, & k \le k_s \\ 0.6, & \text{else}, \end{cases}$$
 (16)

parametrized by the switching time  $k_s$ . The objective is given by  $\|\alpha_{70} - \pi\|_2$ , i.e. the deviation between the last angle  $\alpha_{70}$  of an episode and the upward position  $\pi$ .

#### A.4. Proof of Theorem 4.6

We begin by deriving a bound on the amount at which small changes in the planned nominal trajectory  $\{\mu_{i|k}^*, v_{i|k}^*\}$  affect the set membership constraint (5e) in Lemma A.3. Based on Assumption 4.3 together with Lipschitz continuity of the state, input, and terminal constraints, as well as the aforementioned bound on the set membership constraint, we then show that feasibility of (5) for planning horizon N at time k together with  $e(k, \bar{\theta}) \in \mathcal{E}_{p_{\mathcal{S}}}(x(k), u(k))$  implies the existence of a feasible solution at time N-1 in Lemma A.4. Finally, we iteratively apply Lemma A.4, to prove Theorem 4.6.

In the following, we consider a model error set  $\bar{\mathcal{E}}:=\{e\in\mathbb{R}^n|a_{\mathcal{E}}(e)\leq\mathbb{1}_{n_{\mathcal{E}}}\}$ , a safe terminal set  $\mathcal{S}^t:=\{x\in\mathbb{R}^n|a_{\mathcal{S}}(x)\leq\mathbb{1}_{n_{\mathcal{E}}}\}$  a cacording to Assumption 4.2 with  $a_{\mathcal{E},i}:\mathbb{R}^n\to\mathbb{R}$  and  $a_{\mathcal{S},i}:\mathbb{R}^n\to\mathbb{R}$ , both Lipschitz continuous functions with constants  $L_{\mathcal{E}},L_{\mathcal{S}}$ . In the predictive safety filter optimization problem (5) the constraints are defined according to the tightening (8) and (14). We denote an optimal solution of (5) at time k with planning horizon N as nominal input sequence  $v_{0|k,N}^*,..,v_{N-1|k,N}^*$  with corresponding nominal state sequence  $\mu_{0|k,N}^*,..,\mu_{N|k,N}^*$  (5b).

**Lemma A.3.** Let Assumption 4.5 hold. If  $x, x^+ \in \mathbb{X}$ ,  $u, u^+ \in \mathbb{U}$  and  $\mathcal{E}_{p_{\mathcal{S}}}(x, u) \subseteq \bar{\mathcal{E}}$ , then  $\mathcal{E}_{p_{\mathcal{S}}}(x^+, u^+) \subseteq \bar{\mathcal{E}}^+(\|\Delta z\|_2)$ , where  $\Delta z \coloneqq \|[x^\top, u^\top] - [x^{+\top}, u^{+\top}]\|_2$  and  $\bar{\mathcal{E}}^+(\|\Delta z\|_2) \coloneqq \{e \in \mathbb{R}^n | a_{\mathcal{E}}(e) \leq \mathbb{1}_{n_{\mathcal{E}}} + L_{a_{\mathcal{E}}} L_{\mathcal{E}} \|\Delta z\|_2 \mathbb{1}_{n_{\mathcal{E}}} \}.$ 

*Proof.* The essential observation is that all  $e^+ \in \mathcal{E}_{p_S}(x^+, u^+)$  can be written as  $e^+ = e^* + \Delta e$  with  $e^* \coloneqq \operatorname{arginf}_{e \in \mathcal{E}_{p_S}(x,u)} \|e^+ - e\|_2$ ,  $\Delta e \in \mathbb{R}^n$  for which we have

by Assumption 4.5 that

$$\begin{split} \left\|\Delta e\right\|_{2} &= \left\|e^{+} - e^{*}\right\|_{2} = \inf_{e \in \mathcal{E}_{\mathcal{P}_{\mathcal{S}}}(x, u)} \left\|e^{+} - e\right\|_{2} \\ &\leq \sup_{e^{+} \in \mathcal{E}_{\mathcal{P}_{\mathcal{S}}}(x^{+}, u^{+})} \inf_{e \in \mathcal{E}_{\mathcal{P}_{\mathcal{S}}}(x, u)} \left\|e^{+} - e\right\|_{2} \\ &\leq L_{\mathcal{E}} \left\|\Delta z\right\|_{2}. \end{split}$$

This allows us to derive

$$a_{\mathcal{E}}(e^{+}) = a_{\mathcal{E}}(e^{*}) + a_{\mathcal{E}}(e^{+}) - a_{\mathcal{E}}(e^{*})$$

$$\leq a_{\mathcal{E}}(e^{*}) + L_{a_{\mathcal{E}}} \|\Delta e\|_{2} \mathbb{1}_{n_{\mathcal{E}}}$$

$$\leq a_{\mathcal{E}}(e^{*}) + L_{a_{\mathcal{E}}} L_{\mathcal{E}_{p_{\mathcal{S}}}} \|\Delta z\|_{2} \mathbb{1}_{n_{\mathcal{E}}}$$

$$\leq \mathbb{1}_{n_{\mathcal{E}}} + L_{a_{\mathcal{E}}} L_{\mathcal{E}_{p_{\mathcal{S}}}} \|\Delta z\|_{2} \mathbb{1}_{n_{\mathcal{E}}}$$

$$(17)$$

since by definition  $e^* \in \mathcal{E}_{p_{\mathcal{S}}}(x,u) \subseteq \bar{\mathcal{E}}$ . Therefore, for all  $e^+ \in \mathcal{E}_{p_{\mathcal{S}}}(x^+,u^+)$ , (17) holds, which implies  $\mathcal{E}_{p_{\mathcal{S}}}(x^+,u^+) \subseteq \bar{\mathcal{E}}^+(\|\Delta z\|_2)$ , completing the proof.  $\square$ 

**Lemma A.4.** Let Assumptions 4.3 and 4.5 hold. For every  $\epsilon > 0$ , there exists a corresponding  $\hat{e} > 0$  such that if 1)  $\max_{e \in \bar{\mathcal{E}}} \|e\|_2 \leq \hat{e} \leq \epsilon$ , 2) Problem (5) is feasible at time k with prediction horizon N > 0, 3)  $u(k) = \pi_{\mathcal{S}}(k, x(k), u_{\mathcal{L}}(k)) = v_{0|k}^*$  is applied to (1), and 4)  $e(k, \bar{\theta}) \in \mathcal{E}_{p_{\mathcal{S}}}(x(k), u(k))$  with probability 1, then the input sequence

$$\tilde{v}_{i|k+1} \coloneqq \pi(\tilde{\mu}_{i|k+1}, \mu^*_{i+1|k,N}, v^*_{i+1|k,N}) \text{ for } i \in \mathcal{I}_{[0,N-2]}$$

with  $\pi$  according to Assumption 4.3, corresponding nominal state sequence  $\tilde{\mu}_{i+1|k+1} = f(\tilde{\mu}_{i|k+1}, \tilde{v}_{i|k+1}, \bar{\theta})$ ,  $\tilde{\mu}_{0|k+1} = x(k+1)$ , is a feasible solution to (5) at time k+1 with prediction horizon N-1.

*Proof.* The following proof is an extension of (Köhler et al., 2018b, Proposition 5), which considers nonlinear systems with additive disturbances of the form x(k+1) = f(x(k), u(k)) + w(k), to address model (1), (2) in combination with the set-valued confidence and terminal safe set constraints (5e) and (5f). For completeness, we provide the entire proof here, although the first half is a straightforward adoption from Köhler et al. (2018b).

The proof makes use of the conditions in Assumption 4.3 to derive bounds on the difference between the optimal plan at time k, and the constructed plan  $\{\tilde{\mu}_{i|k+1}\}$ ,  $\{\tilde{v}_{i|k+1}\}$  at time k+1, which is in turn used to show that the constraint tightening implies that the constructed plan is a feasible solution for (5) with planning horizon N-1. We start by deriving a bound on  $\hat{e}$  such that  $V(x(k+1), \mu^*_{1|k,N}) \leq \delta$  holds. Select

$$\hat{e} \leq \hat{e}_1 \coloneqq \sqrt{\frac{\delta}{c_n}}$$

and note that by assumption and constraint (5e) we have  $x(k+1) - \mu_{1|k,N}^* \in \bar{\mathcal{E}},$  i.e.

$$\left\| x(k+1) - \mu_{1|k,N}^* \right\|_2 \le \hat{e}$$

and therefore by Assumption 4.3

$$V(x(k+1), \mu_{1|k,N}^*) \le c_u \left\| x(k+1) - \mu_{1|k,N}^* \right\|_2^2$$
  
$$\le c_u \hat{e}^2 \le c_u \hat{e}_1^2 \le \delta. \tag{18}$$

By selecting  $\hat{e} \leq \hat{e}_1$ , we are now able to show that  $\tilde{v}_{i|k+1}$  is a a feasible candidate input sequence to (5) with planning horizon N-1. We first bound the deviation between  $\tilde{\mu}_{i|k+1}$  and  $\mu^*_{i+1|k}$  using Assumption 4.3, which allows us in a second step to construct sufficient bounds on  $\hat{e}$  implying feasibility via the tightening sequence (9). To this end, due to Assumption 4.3,  $\tilde{\mu}_{0|k+1} = x(k+1)$  and definition of  $\tilde{\mu}_{i|k+1}$ , we have

$$c_{u}\hat{e}^{2} \geq V(\tilde{\mu}_{0|k+1}, \mu_{1|k,N}^{*})$$

$$\geq \rho^{-1}V(\tilde{\mu}_{1|k+1}, \mu_{2|k,N}^{*})$$

$$\geq \rho^{-2}V(\tilde{\mu}_{2|k+1}, \mu_{3|k,N}^{*})$$

$$\geq \dots$$

$$\geq \rho^{1-N}V(\tilde{\mu}_{N-1|k+1}, \mu_{N|k,N}^{*})$$

and consequently

$$V(\tilde{\mu}_{i|k+1}, \mu^*_{i+1|k,N}) \le \rho^i c_u \hat{e}^2 \le \delta \text{ for all } i \in \mathcal{I}_{[0,N-1]}$$

since  $\rho \in (0,1)$ . By Assumption (4.3) it holds  $c_l \left\| \tilde{\mu}_{i|k+1} - \mu_{i+1|k,N}^* \right\|_2^2 \leq V(\tilde{\mu}_{i|k+1}, \mu_{i+1|k,N}^*)$  yielding

$$\left\| \tilde{\mu}_{i|k+1} - \mu_{i+1|k,N}^* \right\|_2^2 \le \rho^i \frac{c_u}{c_i} \hat{e}^2.$$

Based on the above, we can select  $\hat{e}$  according to the state constraints (8) to satisfy

$$\hat{e} \le \hat{e}_2 \coloneqq \sqrt{\frac{c_l}{c_u}} \frac{\epsilon}{\|A_x\|_{\infty}}$$

yielding, together with the row sum norm  $\|A_x\|_{\infty}$  and the fact that  $\|a\|_2 \|b\|_2 \le \|a\|_1 \|b\|_2$  for all  $a,b \in \mathbb{R}^{n_x}$ , that

$$\begin{split} A_x \tilde{\mu}_{i|k+1} &\leq A_x \mu_{i+1|k}^* + \|A_x\|_{\infty} \sqrt{\rho^i \frac{c_u}{c_l}} \hat{e} \mathbb{1}_{n_x} \\ &\leq (1 - \epsilon_{i+1}) \mathbb{1}_{n_x} + \sqrt{\rho^i} \epsilon \mathbb{1}_{n_x} \leq (1 - \epsilon_i) \mathbb{1}_{n_x} \end{split}$$

for all  $i\in\mathcal{I}_{[0,N-2]}$ , which proves constraint satisfaction of the candidate state sequence  $\tilde{\mu}_{i|k+1}$  with respect to state constraints. Let in addition

$$\hat{e} \le \hat{e}_3 := \sqrt{\frac{c_l}{c_u}} \frac{\epsilon}{L_S},$$

implying in a similar fashion that

$$\begin{aligned} a_{\mathcal{S}}(\tilde{\mu}_{N-1|k+1}) &\leq a_{\mathcal{S}}(\mu_{N|k}^*) + L_{\mathcal{S}} \sqrt{\rho^i \frac{c_u}{c_l}} \hat{e} \mathbb{1}_{n_s} \\ &\leq (1 - \epsilon_{N-1}) \mathbb{1}_{n_{\mathcal{S}}}, \end{aligned}$$

showing terminal constraint satisfaction of (5f). Furthermore, let

$$\hat{e} \leq \hat{e}_4 \coloneqq \sqrt{\frac{c_l}{c_u}} \frac{\epsilon}{\|A_u\|_{\infty}} \pi_{\max},$$

yielding together with Assumption 4.3

$$\left\| \tilde{v}_{i|k+1} - v_{i+1|k,N}^* \right\|_2^2 \le \pi_{\max}^2 \left\| \tilde{\mu}_{i|k+1} - \mu_{i|k,N}^* \right\|_2^2$$

$$\le \pi_{\max}^2 \rho^i \frac{c_u}{c_l} \hat{e}^2,$$

providing that

$$A_{u}\tilde{v}_{i|k+1} \leq A_{u}\mu_{i+1|k}^{*} + \|A_{u}\|_{\infty} \pi_{\max} \sqrt{\rho^{i} \frac{c_{u}}{c_{l}}} \hat{e} \mathbb{1}_{n_{u}}$$

$$\leq (1 - \epsilon_{i}) \mathbb{1}_{n_{u}} \text{ for all } i \in \mathcal{I}_{[0, N-2]},$$

showing input constraint satisfaction (5d) of the candidate input sequence. Finally, for the uncertainty constraint (5e) we have by Lemma A.3

$$\mathcal{E}_{p_{\mathcal{S}}}(\tilde{\mu}_{i|k+1}, \tilde{v}_{i|k+1}) \subseteq \bar{\mathcal{E}}_{i+1}^+(\|\Delta z\|_2)$$

with  $\bar{\mathcal{E}}_{i+1}^+ \coloneqq \{e \in \mathbb{R}^n | a_{\mathcal{E}}(e) \leq (1 - \epsilon_{i+1}) \mathbb{1}_{n_{\mathcal{E}}} + L_{\mathcal{E}} \|\Delta z_i\|_2 \mathbb{1}_{n_{\mathcal{E}}} \}$  and  $\Delta z \coloneqq [\tilde{\mu}_{i|k+1}^\top, \tilde{v}_{i|k+1}^\top]^\top - [\mu_{i+1|k}^*, v_{i+1|k}^*]^\top$ . For the latter it holds

$$\|\Delta z_i\|_2^2 \le \|\tilde{\mu}_{i|k+1} - \mu_{i+1|k}^*\|_2^2 + \|\tilde{v}_{i|k+1} - v_{i+1|k}^*\|_2^2$$
  
$$\le (1 + \pi_{\max})^2 \rho^i \frac{c_u}{c_l} \hat{e}^2$$

for all  $i \in \mathcal{I}_{[0,N-2]}$ . Enforcing

$$\hat{e} \le \hat{e}_5 := \sqrt{\frac{c_l}{c_u}} \frac{\epsilon}{L_{\mathcal{E}}(1 + \pi_{\max})}$$

gives the desired relation

$$\begin{split} \mathcal{E}_{p_{\mathcal{S}}}(\tilde{\mu}_{i|k+1}, \tilde{v}_{i|k+1}) &\subseteq \bar{\mathcal{E}}_{i+1}^{+}(\|\Delta z\|_{2}) \\ &\subseteq \{e \in \mathbb{R}^{n} | a_{\mathcal{E}}(e) \leq (1 - \epsilon_{i+1}) \mathbb{1}_{n_{\mathcal{E}}} \\ &\quad + L_{\mathcal{E}}(1 + \pi_{\max}) \sqrt{\rho^{i} \frac{c_{u}}{c_{l}}} \hat{e} \mathbb{1}_{n_{\mathcal{E}}} \} \\ &\subseteq \{e \in \mathbb{R}^{n} | a_{\mathcal{E}}(e) \leq (1 - \epsilon_{i+1}) \mathbb{1}_{n_{\mathcal{E}}} + \sqrt{\rho^{i}} \epsilon \mathbb{1}_{n_{\mathcal{E}}} \} \\ &\subseteq \{e \in \mathbb{R}^{n} | a_{\mathcal{E}}(e) \leq (1 - \epsilon_{i}) \mathbb{1}_{n_{\mathcal{E}}} \} \\ &\subseteq \bar{\mathcal{E}}_{i} \end{split}$$

which shows feasibility of (5e).

Therefore, for any  $\epsilon$ , there exists a  $\hat{e} := \min\{\hat{e}_1,\hat{e}_2,\hat{e}_3,\hat{e}_4,\hat{e}_5\} > 0$ , such that  $\{\tilde{v}_{i|k+1}\}$  is a feasible solution to (5) at time k+1 with planning horizon N-1 and error bound  $\max_{e \in \bar{\mathcal{E}}} \|e\|_2 \le \hat{e}$ , proving the desired statement.  $\square$ 

We can now utilize Lemma A.4 to prove Theorem 4.6, i.e. that application of Algorithm 2 implies safe system operation according to Definition 3.1.

*Proof of Theorem 4.6.* Select  $\hat{e}$  according to Lemma A.4. By considering the different cases in Algorithm 2, we show safety for all k according to (3) by utilizing the relation

$$\Pr(\forall k : x(k) \in \mathbb{X}, u(k) \in \mathbb{U})$$

$$\geq \Pr(\forall k : x(k) \in \mathbb{X}, u(k) \in \mathbb{U}, e(k, \bar{\theta}) \in \mathcal{E}_{p_{\mathcal{S}}}(k))$$

$$= \Pr(\forall k : x(k) \in \mathbb{X}, u(k) \in \mathbb{U} | \forall k : e(k, \bar{\theta}) \in \mathcal{E}_{p_{\mathcal{S}}}(k))$$

$$\cdot \Pr(\forall k : e(k, \bar{\theta}) \in \mathcal{E}_{p_{\mathcal{S}}}(k)). \tag{19}$$

Since  $\Pr(\forall k : e(k, \bar{\theta}) \in \mathcal{E}_{p_{\mathcal{S}}}(k)) \ge p_{\mathcal{S}}$  by Assumption 4.5, relation (19) allows us to prove (3) by establishing

$$\Pr(\forall k : x(k) \in \mathbb{X}, u(k) \in \mathbb{U} | \forall k : e(k, \bar{\theta}) \in \mathcal{E}_{p_{\mathcal{S}}}(k)) = 1.$$
(20)

The proof therefore reduces to the deterministic case showing that  $x(k) \in \mathbb{X}$ ,  $u(k) \in \mathbb{U}$ , given  $e(k, \bar{\theta}) \in \mathcal{E}_{ps}(k)$ ), at any time step k, which implies directly (20) for all times k and therefore via (19) chance constraint satisfaction according to (3), i.e., safe system operation with respect to Definition 3.1.

In order to show  $x(k) \in \mathbb{X}$  and  $u(k) \in \mathbb{U}$ , note that if (5) is feasible at time k for any planning horizon  $\tilde{N}>0$  it follows due to the state and input constraints (5c), (5d) that  $x(k) \in \mathbb{X}$  and  $u(k) \in \mathbb{U}$ . This implies directly that  $x(k) \in \mathbb{X}$  and  $u(k) \in \mathbb{U}$  for any time step k for which (5) is feasible for horizon N, as well as for all  $\tilde{k} \in \mathcal{I}_{[k+1,N+k-1]}$ , for which (5) is infeasible for horizon N, since feasibility of (5) with horizon  $N-(\tilde{k}-k)$  is obtained from iteratively applying Lemma A.4 due to the condition that  $\forall k: e(k,\bar{\theta}) \in \mathcal{E}_{p_S}(k)$ .

For all  $\tilde{k} \geq N+k$ , it follows from containment in the terminal safe set via (5f) and Assumption 4.2 that  $x(\tilde{k}) \in \mathbb{X}$  and  $u(\tilde{k}) = \pi_S^t(x(k)) \in \mathbb{U}$ .

Overall this shows (20) and therefore via (19) that  $\Pr(\forall k : x(k) \in \mathbb{X}, u(k) \in \mathbb{U}) \geq p_{\mathcal{S}}$ , completing the proof.