IEEE Control Systems Letters paper presented at
2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Learning Safety Filters for Unknown Discrete-Time Linear Systems

Farhad Farokhi, Alex S. Leong, Mohammad Zamani, and Iman Shames

***Abstract*— A learning-based safety filter is developed for discrete-time linear time-invariant systems with unknown models subject to Gaussian noises with unknown covariance. Safety is characterized using polytopic constraints on the states and control inputs. The empirically learned model and process noise covariance with their confidence bounds are used to construct a robust optimization problem for minimally modifying nominal control actions to ensure safety with high probability. The optimization problem relies on tightening the original safety constraints. The magnitude of the tightening is larger at the beginning since there is little information to construct reliable models, but shrinks with time as more data becomes available.**

I. Introduction

It is often desired to ensure *safety* of a controlled system. Safety can be defined as maintaining the system's states and control inputs inside a well-defined set, referred to as *safety set*. For instance, robots must be maneuvered in complicated previously-unseen environments without collisions, and phases and voltages in power system must be maintained within pre-defined bands to avoid blackouts. Controllers often ensure safety using reliable models of the system and environment. Models are required to extrapolate the behaviour of the system given the current state and the designed input sequences. Models are however subject to unknown uncertainties or might even be entirely unknown. Irrespective of the accuracy of the model in laboratory conditions, unknown or varying environmental factors, such as slippage and wind, can render the model uncertain. When facing uncertainties, we can consider their worst-case magnitude to ensure safety robustly. However, robust safety can result in conservative controllers. Alternatively, we can utilize real-time data to "learn" representations or models of the uncertainty. Thus we must ensure safety based on inaccurate time-varying models that fit the data on the fly. This is the topic of the current paper.

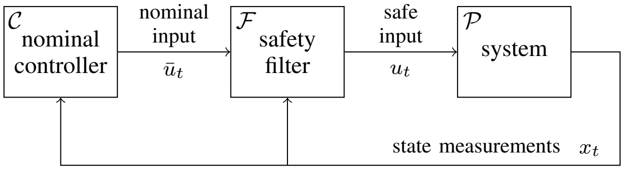

Fig. 1. Safety filter $\mathcal{F}$ ensures satisfaction of state and control constraints for linear discrete-time system $\mathcal{P}$ with known model subject to zero-mean Gaussian uncertainty with unknown covariance. The safety filter minimally modifies the control input from nominal controller $\mathcal{C}$ via an optimization program to guarantee that the states in the next time step remain in the safe set with high probability.

In this paper, a learning-based *safety filter* is developed for systems with unknown discrete-time linear time-invariant dynamics subject to a zero-mean Gaussian process noise with unknown covariance. The safety set is characterized using polytopic constraints on the states and control inputs. A block diagram of the closed-loop system using the safety filter is illustrated in Figure 1. The safety filter relies on two learning-based components: regression for learning the system model and empirical covariance of the process noise. The learned model and empirical process noise covariance, along with their confidence bounds, are used to construct a robust optimization problem for minimally modifying (in terms of a distance metric) nominal control actions to ensure safety with high probability. The nominal control can be generated by standard controllers, such as proportional-integral-derivative controllers, or by learning-based controllers, such as reinforcement learning. Finally, we propose directly optimizing the closed-loop performance by solving a model predictive control problem with tightened constraints instead of projecting nominal control inputs into the set of control signals to ensure safety. Similar to the projection-based safety filter, the magnitude of constraint tightening, which is dictated by the confidence in the learned models, is larger at the beginning since there is little information to construct reliable models, but shrinks with time as more data becomes available. Note that individual elements of this paper, such as regularized least-squares learning of the model, empirical estimation of the covariance matrices of the process noise, and robust optimization for modifying control inputs, are traditionally investigated separately in the literature. The *main* contributions of this paper are to combine these methods to develop a rigorous analysis of learning-based safety filters for unknown systems, and to develop computationally-efficient safety filters that are missing from the literature.

Safe learning-based control, where the system model and uncertainties are unknown and must be estimated, has gained much attention recently [1]–[5]. A popular approach is to learn models using Gaussian processes [2]–[4]. Also, there are many definitions for safety in reinforcement learning [6], but the approach of this paper relates more to reinforcement learning with constraints [2], [7]–[9]. However, constraints

An early version of this paper was presented at NeurIPS 2021 Workshop on Safe and Robust Control of Uncertain Systems, which does not have a proceedings and is only intended for soliciting feedback from the research community.

F. Farokhi and M. Zamani are with the University of Melbourne ({farhad.farokhi,mohammad.zamani}@unimelb.edu.au). A. S. Leong is with the Defence Science and Technology (DST) Group, Australia (alex.leong@defence.gov.au). I. Shames is with the CIICADA Lab, Australian National University (iman.shames@anu.edu.au). At the time of working on this paper, M. Zamani was also affiliated with the DST Group.

The work of F. Farokhi is supported by a research contract (ID10298) from the Defence Science and Technology (DST) Group, Australia. The work of I. Shames is partially supported by the Australian Government, via grant AUSMURIB000001 associated with ONR MURI N00014-19-1-2571.

in this paper are stage-wise as opposed to constraints on accumulated penalties over the planning horizon in constrained reinforcement learning. The above-mentioned studies share a common assumption that the learning of the models and uncertainties are done prior to control or that we can alternate between learning and control with batch learning [1]–[5]. In contrast, our main interest is to perform learning and control simultaneously while new state measurements arrive, and to maintain safety based on time-varying inaccurate models.

There are few studies that consider safety in simultaneous learning and control. The work of [10] uses confidence of learned Gaussian processes when modelling non-linear systems and environments to make decisions regarding safety based on the number of measurements used for learning. Although powerful, that work does not provide computationally efficient methods for ensuring safety, as their framework relies on Lyapunov functions, which can be difficult to find or compute for general systems. Another relevant study is [11], which proposes computationally-efficient methods for projecting control signals into safe sets by computing the confidence of learning additive Gaussian models. However, that work only considers learning stochastic disturbances caused by the environment and assumes that the underlying model of the system is known. Most recently, a single-trajectory learning-based feedback scheme that ensures safety was proposed in [12]. In contrast to that study, the current paper does not focus on computing feedback functions but rather emphasizes constraint tightening for computing control actions using optimization problems. The approach of this paper results in a less complex for optimization problem; however, it requires sequentially solving optimization problems, which can be only done if only there is dedicated on-board computational capability. Similarly, learning-based optimal control over an infinite horizon was considered in [13]. The emphasize in that paper was also on computing linear feedback policies using semi-definite programming. Safe learning for stochastic jump linear systems using semi-definite programming was considered in [14]. A myopic safety-constrained optimization was presented for water distribution networks in [15]. However, in that paper, constraint shrinking in response to learned model uncertainty was not considered. Learning-based model predictive control with safety constraints have been proposed in [16]–[18]. These studies prove recursive feasibility and stability. However, computational issues, such as relying on polytopic sets for learning the model (with increasing numbers of polytopes or vertices with time), development of robust positively invariant sets, and requiring potentially high-dimensional parametric feedback functions, can stifle their implementation in practice.

The rest of the paper is organized as follows. First, the mathematical problem formulation is presented in Section II. Section III overviews confidence bounds for regularized least-squares learning of the system model. The data-driven safety filter is presented and analyzed in Section IV. In Section V, we reduce the conservatism of the safety filter by using the empirical covariance of the process noise in addition to learning the model parameters. Finally, Section VI concludes the paper and presents directions for future research.

Notation: Sets are denoted by calligraphic letters, such as $\mathcal{A}$. Matrices are denoted by capital Roman letters, such as A. The i-th row of A is denoted by A_i. The entry in the i-th row and the j-th column of matrix A is a_{ij}. Scalars and vectors are denoted by lowercase Roman and Greek letters, such as x and θ. Similarly, the i-th entry of vector x is denoted by x_i. Let $\mathcal{S}^n_{++}$ and $\mathcal{S}^n_{+}$ refer to the sets of symmetric positive definite and positive semi-definite matrices in $\mathbb{R}^{n\times n}$. In what follows, $A \succ B$ and $A \succeq B$, respectively, signify that $A - B \in \mathcal{S}^n_{++}$ and $A - B \in \mathcal{S}^n_{+}$. The smallest and the largest singular values of matrix Y are, respectively, denoted by $\sigma_{\min}(Y)$ and $\sigma_{\max}(Y)$. Vector e_i denotes the column-vector with all entries zero except the i-th entry, which is equal to one. For any $x \in \mathbb{R}^n$, $\|x\|$ denotes its Euclidean norm, i.e., $\|x\| = (\sum_{i=1}^n x_i^2)^{1/2}$. For any $A \in \mathbb{R}^{n\times m}$, $\|A\|$ denotes the induced matrix norm $\|A\| = \sup_{\|x\|=1}\|Ax\|$ and $\|A\|_F$ denotes the Frobenius norm $\|A\|_F = (\sum_{i=1}^n\sum_{j=1}^m a_{ij}^2)^{1/2}$. For any set $\mathcal{A} \subseteq \mathbb{R}^n$, $\mathrm{rad}(\mathcal{A})$ is its radius, i.e., $\mathrm{rad}(\mathcal{A}) = \sup_{a\in\mathcal{A}}\|a\|$. For any signal $x[\cdot]$, $x[k_0 : k_1]$ with $k_1 \geq k_0$ denotes the sequence $(x[k_0], x[k_0+1], \ldots, x[k_1])$. For $a, b \in \mathbb{R}^n$, $a \leq b$ signifies that the inequality holds entry-wise.

II. Problem Formulation

Consider a linear time-invariant discrete-time system:

$$x[k+1] = Ax[k] + Bu[k] + w[k], \tag{1}$$

where $x[k] \in \mathbb{R}^n$ is the state, $u[k] \in \mathbb{R}^m$ is the control input, and $w[k] \in \mathbb{R}^n$ is the process noise. The process noise is composed of a sequence of independently and identically distributed (i.i.d.) zero-mean Gaussian random variables with covariance $W \in \mathcal{S}^n_{+}$. *Model parameters A, B, and W are unknown and must be learned.* Safety is encoded by time-varying polytopic constraints:

$$x[k] \in \mathcal{X}_k := \{x \,|\, H[k]x \leq h[k]\}. \tag{2}$$

The control action is also constrained by

$$u[k] \in \mathcal{U} := \{u \,|\, Eu \leq f\}. \tag{3}$$

We make the following standing assumptions on covariance of the process noise, magnitude of the model parameters, and radii of the control and state constraint sets.

Assumption 1: There exists known constants:

a: $r > 0$ such that $W \preceq rI$.
b: $s > 0$ such that $\|[A \quad B]\|_F \leq s$.
c: $d > 0$ such that $\mathrm{rad}(\mathcal{X}_k) + \mathrm{rad}(\mathcal{U}) \leq d$.

When controlling a system with unknown model, the uncertainty of the learned model gets multiplied by the states and control inputs at the current time to determine the uncertainty of the state in the next time step; see (5) and (7) below. Therefore, if the state and the control input are unbounded, the uncertainty of the state after making a decision can become large, which can complicate ensuring safety. Assumption 1.c ensures that the state and the control input are bounded so that we can avoid this problem. In practice, this assumption can be relaxed. At the beginning when the uncertainty of the learned model is high, we can keep the states and the control actions restricted to small sets but, as our confidence in the learned model improves, we relax this assumption by gradually increasing the radii of the sets. Subsection IV-B presents another approach that partially relaxes Assumption 1.c and removes the need for requiring that the states remain within a bounded set with *a priori* known radius for all times.

Problem 1: At $k \in \mathbb{N} \cup \{0\}$, given state measurements $x[0], \dots, x[k]$, find a procedure to compute a modified control input $u[k] \in \mathcal{U}$ based on a nominal control input $\bar{u}[k]$ by minimizing $\mathbf{d}(u[k], \bar{u}[k])$, where $\mathbf{d}(.,.)$ is a distance metric[1], subject to potentially tightened state and control constraints to ensure the state in the next time step remains safe, i.e., $x[k+1] \in \mathcal{X}_{k+1}$, with high probability.

III. Preliminary Results

We use (regularized) least-squares to learn the model:

$$(\hat{A}[k], \hat{B}[k]) \in \underset{(\bar{A},\bar{B})}{\arg\min} \left[\sum_{t=0}^{k-1} \|x[t+1] - (\bar{A}x[t] + \bar{B}u[t])\|^2 \right. \\ \left. + \lambda(\|\bar{A}\|_F^2 + \|\bar{B}\|_F^2) \right], \tag{4}$$

where $\lambda > 0$ is the regularization weight. Before we gather enough measurements, i.e., if $k < n(n+m)$, the least-squares problem (4) admits infinitely-many solutions without regularization, i.e., if $\lambda = 0$. Regularization ensures that the least-squares problem (4) is strictly convex with a unique solution even in the absence of enough measurements. This also enables computing the confidence bounds for the learned model at all times.

To analyze the safety filter, we need to better understand the moments of the random variable:

$$v[k', k] := (A - \hat{A}[k])x[k'] + (B - \hat{B}[k])u[k'], \\ \forall k' \geq k \geq 0. \tag{5}$$

Note that we can rewrite the system dynamics in (1) as

$$x[k'+1] = \hat{A}[k]x[k'] + \hat{B}[k]u[k'] + v[k', k] + w[k']. \tag{6}$$

Therefore, the random variable $v[k', k]$ captures the error of forecasting the state at time $k'+1$, i.e., $x[k'+1]$, by using the learned model based on the measurements up to time k, i.e., $(\hat{A}[k], \hat{B}[k])$. When $k' = k$, with slight abuse of notation, we write

$$v[k] := v[k, k] = (A - \hat{A}[k])x[k] + (B - \hat{B}[k])u[k]. \tag{7}$$

Proposition 1: If $x[k'] \in \mathcal{X}_{k'}$ and $u[k'] \in \mathcal{U}$, then

$$\mathbb{P}\left\{\|v[k', k]\| \leq \zeta n \beta_k(\delta/n)\right\} \geq 1 - \delta, \quad \forall k' \geq k \geq 0,$$

where $\zeta := d/\sqrt{\sigma_{\min}(V[k])}$ and

$$\beta_k(\delta) := r\sqrt{2\log\left(\det(V[k])^{1/2}/(\lambda^{n/2}\delta)\right)} + \lambda^{1/2}s \tag{8}$$

with $V[k] := \lambda I + \hat{V}[k]$ and

$$\hat{V}[k] := \begin{bmatrix} x[0]^\top & u[0]^\top \\ x[1]^\top & u[1]^\top \\ \vdots & \vdots \\ x[k-1]^\top & u[k-1]^\top \end{bmatrix}^\top \begin{bmatrix} x[0]^\top & u[0]^\top \\ x[1]^\top & u[1]^\top \\ \vdots & \vdots \\ x[k-1]^\top & u[k-1]^\top \end{bmatrix}.$$

Proof: See [19, Appendix II]. ■

Before presenting the following result, we need to define persistence of excitation, which is a common assumption in system identification and adaptive control [20].

Definition 1 (Persistence of Excitation): The system in (1) is persistently excited if there exists constants $\gamma \geq \alpha > 0$ and an integer $T_0 > 0$ such that, $\forall k \in \mathbb{N} \cup \{0\}$,

$$\alpha I \preceq \begin{bmatrix} \displaystyle\sum_{t=k}^{k+T_0-1} x[t]x[t]^\top & \displaystyle\sum_{t=k}^{k+T_0-1} x[t]u[t]^\top \\ \displaystyle\sum_{t=k}^{k+T_0-1} u[t]x[t]^\top & \displaystyle\sum_{t=k}^{k+T_0-1} u[t]u[t]^\top \end{bmatrix} \preceq \gamma I.$$

Proposition 2: If $x[k'], x[k''] \in \mathcal{X}_{k'}$, $u[k'], u[k''] \in \mathcal{U}$, and the persistence of excitation holds, then

$$\mathbb{P}\left\{\sqrt{\|v[k', k]\| \, \|v[k'', k]\|} \leq \zeta'_k n \beta_k(\delta/n)\right\} \geq 1 - \delta, \\ \forall k' \geq k \geq 0,$$

where $\zeta'_k := d/\sqrt{\lfloor k/T_0 \rfloor \alpha + \lambda}$.

Proof: See [19, Appendix III]. ■

Proposition 3: Assume that $n \geq 2$. If $x[k'], x[k''] \in \mathcal{X}_{k'}$, $u[k'], u[k''] \in \mathcal{U}$, and the persistence of excitation holds, then

$$\mathbb{E}\{\|v[k', k]\|^2\} \leq L_2(k), \quad \forall k', k'' \geq k \geq 0$$
$$\mathbb{E}\{\|v[k', k]\|^2 \|v[k'', k]\|^2\} \leq L_4(k), \quad \forall k', k'' \geq k \geq 0$$

where

$$L_2(k) := \frac{\lambda s^2 d^2 n^2}{\lfloor k/T_0 \rfloor \alpha + \lambda} \\ + \frac{((\lfloor k/T_0 \rfloor + 1)\gamma + \lambda)}{\lambda} \frac{r d^2 n^{\frac{7}{2}} (r n^{\frac{1}{2}} + \sqrt{\pi \lambda} s)}{(\lfloor k/T_0 \rfloor \alpha + \lambda)},$$

$$L_4(k) := \frac{\lambda^2 s^4 d^4 n^4}{(\lfloor k/T_0 \rfloor \alpha + \lambda)^2} \\ + \frac{((\lfloor k/T_0 \rfloor + 1)\gamma + \lambda)}{\lambda} \frac{8 r d^4 n^{\frac{11}{2}} (r^3 n^{\frac{3}{2}} + \sqrt{\pi} \lambda^{\frac{3}{2}} s^3)}{(\lfloor k/T_0 \rfloor \alpha + \lambda)^2}.$$

Evidently, $L_2(k) = \mathcal{O}(1)$ and $L_4(k) = \mathcal{O}(1/k)$.

Proof: See [19, Appendix IV]. ■

With these preliminary results in hand, we are ready to investigate the effectiveness of the safety filter.

IV. Data-Driven Safety Filter

In this paper, we modify a nominal control input $\bar{u}[k]$ at each iteration to ensure safety. Projection of the control action $\bar{u}[k]$ to a safe set can be done by solving:

$$u[k] \in \underset{u \in \mathcal{U}}{\arg\min}\, \mathbf{d}(u, \bar{u}[k]), \tag{9a}$$
$$\text{s.t.} \quad H[k+1](\hat{A}[k]x[k] + \hat{B}[k]u) \\ \leq h[k+1] - \bar{e}[k+1], \tag{9b}$$

[1] An example of the distance metric is $\mathbf{d}(x, x') = \|x - x'\|$.

where

$$\bar{e}_i[k+1] = \left(\frac{dn\beta_k\left(\dfrac{\delta}{2n}\right)}{\sqrt{\sigma_{\min}(V[k])}} + \sqrt{\frac{2rn}{\delta}}\right)\|H_i[k+1]^\top\|, \quad (10)$$

and $\delta \in (0,1)$ is a design parameter determining the probability of violating the safety constraints, $w \in \mathbb{R}^n$ is an uncertainty term linked with the process noise, and $v \in \mathbb{R}^n$ is an uncertainty term linked with the (in)accuracy of the learned model.

Theorem 1: Assume that problem (9) is feasible. Then, by implementing the control action $u[k]$ extracted from (9), $\mathbb{P}\{x[k+1] \in \mathcal{X}_{k+1}\} \geq 1-\delta$.

Proof: See Appendix II. ■

The constraint-tightening term in (9) is composed of two independent terms: one is caused by the uncertainty of the learned model and the other stems from the process noise. We can show that the constraint-tightening term due to the uncertainty of the learned model goes to zero under persistence of excitation.

A. Persistence of Excitation for Safety Filter

Persistence of excitation is a common assumption in system identification and adaptive control, which ensures that the error of learning the model converges to zero almost surely as more samples are gathered. This is done by exciting the system along all directions.

Proposition 4: Assume that $\|H_i[k+1]^\top\|$ is uniformly bounded and system (1) is persistently excited. Then,

$$\lim_{k\to\infty}(dn/\sqrt{\sigma_{\min}(V[k])})\beta_k\,(\delta/(2n))\,\|H_i[k+1]^\top\| = 0.$$

Proof: See [19, Appendix VI]. ■

Proposition 4 shows that, assuming persistence of excitation, the effect of the uncertainty caused by learning the model in the constraint tightening of (10) tends to zero as more measurements are gathered. Therefore, in the large k regime, we can solve (9) with $\bar{e}_i[k+1] = \sqrt{2rn/\delta}\|H_i[k+1]^\top\|$. The remaining constraint tightening term in this optimization problem is caused by the process noise. Note that, because we have not attempted at learning the statistics of the process noise, we consider the worst-case scenario in light of Assumption 1.a. After recovering the model parameters, the techniques of [11] can be used to learn the statistics of the noise and also shrink this term. This is formalized in Section V.

B. Conservatism in Constraint Tightening

In (9), the worst-case magnitude of the uncertainty term v, linked to the inaccuracy of the learned model, scales quadratically with d, which is an upper bound on the radii of $\mathcal{X}_k$ and $\mathcal{U}$. This is because the model uncertainty gets multiplied by the state and the control input, and can result in conservative behaviour when $\mathcal{X}_k$ and $\mathcal{U}$ are large sets. Furthermore, according to Assumption 1.c, we need to assume existence of a bounded set to which $x[k]$ belongs for all k. These factors can combine to increase the conservatism of the projection-based approach. By examining the steps of the proof of Proposition 1, which is used to prove Theorem 1, we can show that $\mathbb{P}\{\|v[k]\|^2 \leq n^2(\|x[k]\|^2 + \mathrm{rad}(\mathcal{U})^2)\beta_k^2(\delta/(2n))/\sigma_{\min}(V[k])\} \geq 1-\delta/2$. Therefore, we can relax (9) to

$$u[k] \in \operatorname*{arg\,min}_{u\in\mathcal{U}} \mathbf{d}(u, \bar{u}[k]), \quad (11a)$$

$$\text{s.t.} \quad H[k+1](\hat{A}[k]x[k] + \hat{B}[k]u) \leq h[k+1] - \hat{e}[k+1], \quad (11b)$$

where

$$\hat{e}_i[k+1] = \left(\frac{n\sqrt{\|x[k]\|^2 + \mathrm{rad}(\mathcal{U})^2}}{\sqrt{\sigma_{\min}(V[k])}}\beta_k\left(\frac{\delta}{2n}\right) + \sqrt{\frac{2rn}{\delta}}\right)\|H_i[k+1]^\top\|. \quad (12)$$

Similarly, it can be proved that, by implementing the control action $u[k]$ extracted from the optimization problem (11), if feasible, $x[k+1]$ is safe with probability of at least $1-\delta$. This clearly yields an improved performance because $\|x[k]\|^2 + \mathrm{rad}(\mathcal{U})^2 \leq (\|x[k]\| + \mathrm{rad}(\mathcal{U}))^2 \leq d^2$ for all k due to Assumption 1.c. Furthermore, we do not need to assume *a priori* knowledge of $\sup_{k\geq 0}\|x[k]\|$.

C. Combining Controller and Safety Filter

Instead of projecting nominal control inputs into the set of control signals that ensure the safety of the system, we can directly optimize the closed-loop performance by solving:

$$\operatorname*{arg\,min}_{\substack{\bar{u}[k:k+T-1]\\ \bar{x}[k+1:k+T]}} \sum_{t=k}^{k+T-1}\bar{u}[t]^\top R_t\bar{u}[t] + \sum_{t=k+1}^{k+T}(\bar{x}[t]^\top Q_t\bar{x}[t] + q_t^\top\bar{x}[t]), \quad (13a)$$

$$\text{s.t.} \quad \bar{u}[k:k+T-1] \in \mathcal{U}^T \quad (13b)$$

$$\bar{x}[t+1] = \hat{A}[t]\bar{x}[t] + \hat{B}[t]\bar{u}[t], \quad \forall t \in \{k, \ldots, k+T-1\}, \quad (13c)$$

$$\bar{x}[k] = x[k], \quad (13d)$$

$$H[k+1]\bar{x}[k+1] \leq h[k+1] - \bar{e}[k+1], \quad (13e)$$

$$\bar{x}[t] \in \mathcal{X}_t, \forall t \in \{k+2, \ldots, k+T\}, \quad (13f)$$

where $T \in \mathbb{N}$ denotes the decision making horizon, $\mathcal{U}^T$ denotes the T-fold Cartesian product of the set $\mathcal{U}$, $\bar{e}[k+1]$ is defined in (10), and $R_t \in \mathcal{S}^m_{++}$, $Q_t \in \mathcal{S}^n_{+}$, and $q_t \in \mathbb{R}^n$ are the parameters of the cost function. This optimization problem is similar to the one solved in model predictive control [21], with the exception that the safety constraints on the state for the next time step, i.e., $x[k+1]$, is tightened to ensure safety despite modelling uncertainty and process noise. Note that other safety constraints can be tightened following a similar line of reasoning; however, the conservatism increases for them as new measurements are not available or taken into consideration for shrinking the magnitude of the constraint tightening. Assuming that problem (13) is feasible and, by implementing the control action $u[k]$ from the solution $u[k:k+H-1]$ of (13), $x[k+1]$ is safe with probability of at least $1-\delta$.

One positive aspect of the model predictive control formulation, as opposed to instantaneous or myopic projection of nominal control actions to ensure safety, is that the optimization problem is more likely to remain feasible. For instance, in obstacle avoidance, model predictive control looks ahead to avoid future states that can cause infeasibility down the track. However, this comes at the cost of an increased computational burden because of the longer horizon and increased dimension. An important direction for future research is to establish recursive feasibility of the proposed learning-based model predictive control, i.e., establishing conditions under which, if (13) is feasible at time k, it is also feasible at time $k+1$. To be able to establish recursive feasibility, we need to prove that the uncertainty sets for the model matrices and the covariance matrix are recursively contained, i.e., access to more measurements does not increase uncertainty in some directions. Furthermore, we must search over the set of feedback policies rather than control inputs. Given these properties in addition to a robust positively invariant safe set, we can use standard recursive feasibility arguments from robust model predictive control. These requirements however can limit the computationally-friendly nature of the constraint-tightening projection-based approach in this paper.

V. Learning of Process Noise Covariance

In this section, the covariance of the process noise is estimated empirically to reduce the conservatism of working with only the upper bound in Assumption 1.a. In particular, we use the empirical covariance of the process noise:

$$\widehat{W}[k,k_0] = \frac{1}{k-k_0}\sum_{t=k_0+1}^{k}\hat{w}[t|k_0]\hat{w}[t|k_0]^\top,$$

where $\hat{w}[k,k_0] := x[k+1] - (\hat{A}[k_0]x[k] + \hat{B}[k_0]u[k])$. For all $k > k_0 \geq 0$, we ensure safety by projecting the control action $\bar{u}[k]$ using

$$u[k] \in \operatorname*{arg\,min}_{u\in\mathcal{U}} \mathbf{d}(u,\bar{u}[k]), \tag{14a}$$

$$\text{s.t.} \quad H[k+1](\hat{A}[k_0]x[k] + \hat{B}[k_0]u) \leq h[k+1] - \tilde{e}[k+1], \tag{14b}$$

where

$$\tilde{e}_i[k+1] = \frac{dn}{\sqrt{\sigma_{\min}(V[k])}}\beta_k\left(\frac{\delta}{3n}\right)\|H_i[k+1]^\top\| + \sqrt{\frac{3n}{\delta}}\left\|\Pi_{k,k_0}^{1/2}H_i[k+1]^\top\right\|, \tag{15}$$

and

$$\Pi_{k,k_0}^{-1} := \widehat{W}[k-1,k_0] + \sqrt{\frac{3}{\delta}\left(2L_4(k_0)^2 + \frac{8rL_2(k_0)}{k-k_0} + \frac{2r^2n(n+1)}{k-k_0}\right)}I.$$

Theorem 2: Assume that problem (14) is feasible and $n \geq 2$. Then, by implementing the control action $u[k]$ extracted from (14), $\mathbb{P}\{x[k+1] \in \mathcal{X}_{k+1}\} \geq 1-\delta$.

Proof: See [19, Appendix VII]. ■

Remark 1: The need for the assumption $n \geq 2$ in Theorem 2 arises from an inequality (i.e., $\delta^{n/2} \leq \delta$ for $\delta \in (0,1)$ and $n \geq 2$) used to prove Proposition 3. Although this assumption seems to be technical, we have not been able to relax it.

Remark 2: By increasing k_0, $L_4(k_0)$ decreases, which can potentially reduce the constraint tightening term. This is because, by increasing k_0, the accuracy of the learned model improves. However, by increasing k_0, $k-k_0$ gets smaller, which can potentially increase the constraint tightening. This trade-off stems from the fact that only k_0 measurements are used to learn the model parameters $(\hat{A}[k_0], \hat{B}[k_0])$ (so by increasing k_0 the learned model becomes more accurate) while the remaining $k-k_0$ measurements are used to empirically estimate the covariance of the process noise (so by increasing k_0 the empirical covariance becomes less reliable). This fundamental trade-off cannot be avoided unless the entire set of measurements are used to simultaneously learn the model parameters and estimate the covariance of the process noise. However, this approach complicates the proofs significantly and worsens the tightness of the bounds by generating extra cross-correlation terms. This is a trade-off that must be considered when choosing k_0.

VI. Conclusions

We considered safe learning-based control for discrete-time linear time-invariant dynamical systems when the system model and the process noise covariance are unknown but bounded. We used regularized least-squares estimation to learn the model online and used the empirical covariance of the noise. We relied on the confidence bounds of the learned system model and the empirical process noise covariance to modify the control inputs via a robust optimization problem with time-varying safety constraints. We reformulated the problem in a computationally-friendly optimization problem for ensuring safety based on constraint tightening. Future work can focus on noisy output measurements and learning nonlinear systems using Gaussian processes.

References

[1] A. J. Taylor, A. Singletary, Y. Yue, and A. D. Ames, "Learning for safety-critical control with control barrier functions," in *Proc. Conf. Learning for Dynamics and Control*, 2020.

[2] R. Cheng, G. Orosz, R. M. Murray, and J. W. Burdick, "End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks," in *Proc. AAAI-19*, (Honolulu, USA), Jan. 2019.

[3] R. Cheng, M. J. Khojasteh, A. D. Ames, and J. W. Burdick, "Safe multi-agent interaction through robust control barrier functions with learned uncertainties," in *Proceedings of the 59th IEEE Conference on Decision and Control (CDC)*, pp. 777–783, 2020.

[4] P. Jagtap, G. J. Pappas, and M. Zamani, "Control barrier functions for unknown nonlinear systems using gaussian processes," in *Proceedings of the 59th IEEE Conference on Decision and Control (CDC)*, pp. 3699–3704, 2020.

[5] J. Choi, F. Castaneda, C. J. Tomlin, and K. Sreenath, "Reinforcement learning for safety-critical control under model uncertainty, using control Lyapunov functions and control barrier functions." arXiv preprint arXiv:2004.07584, 2020.

[6] J. Garcıa and F. Fernández, "A comprehensive survey on safe reinforcement learning," *Journal of Machine Learning Research*, vol. 16, no. 1, pp. 1437–1480, 2015.

[7] Z. Marvi and B. Kiumarsi, “Safe reinforcement learning: A control barrier function optimization approach,” *International Journal of Robust and Nonlinear Control*, vol. 31, no. 6, pp. 1923–1940, 2021.

[8] 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,” *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2737–2752, 2018.

[9] N. Fulton and A. Platzer, “Safe reinforcement learning via formal methods: Toward safe control through proof and learning,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

[10] A. Devonport, H. Yin, and M. Arcak, “Bayesian safe learning and control with sum-of-squares analysis and polynomial kernels,” in *Proc. IEEE Conf. Decision and Control*, (Jeju Island, South Korea), pp. 3159–3165, Dec. 2020.

[11] F. Farokhi, A. S. Leong, I. Shames, and M. Zamani, “Safe learning of uncertain environments,” 2021. arXiv preprint arXiv:2103.01413v2 [cs.LG] https://arxiv.org/abs/2103.01413.

[12] Y. Li, S. Das, J. Shamma, and N. Li, “Safe adaptive learning-based control for constrained linear quadratic regulators with regret guarantees,” *arXiv preprint arXiv:2111.00411*, 2021.

[13] S. Dean, S. Tu, N. Matni, and B. Recht, “Safely learning to control the constrained linear quadratic regulator,” in *2019 American Control Conference (ACC)*, pp. 5582–5588, IEEE, 2019.

[14] M. Schuurmans, P. Sopasakis, and P. Patrinos, “Safe learning-based control of stochastic jump linear systems: a distributionally robust approach,” in *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 6498–6503, IEEE, 2019.

[15] J. Val, R. Wisniewski, and C. S. Kallesoe, “Safe reinforcement learning control for water distribution networks,” in *2021 IEEE Conference on Control Technology and Applications (CCTA)*, pp. 1148–1153, 2021.

[16] A. Didier, K. P. Wabersich, and M. N. Zeilinger, “Adaptive model predictive safety certification for learning-based control,” in *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 809–815, IEEE, 2021.

[17] M. Lorenzen, M. Cannon, and F. Allgöwer, “Robust MPC with recursive model update,” *Automatica*, vol. 103, pp. 461–471, 2019.

[18] K. P. Wabersich and M. N. Zeilinger, “Linear model predictive safety certification for learning-based control,” in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 7130–7135, IEEE, 2018.

[19] F. Farokhi, A. S. Leong, M. Zamani, and I. Shames, “Learning safety filters for unknown discrete-time linear systems,” 2023. arXiv preprint arXiv:2111.00631 [cs.LG] https://arxiv.org/abs/2111.00631.

[20] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence and Robustness*. Dover Books on Electrical Engineering Series, Dover Publications, 2011.

[21] J. B. Rawlings and D. Q. Mayne, *Model Predictive Control: Theory and Design*. Nob Hill Pub., 2009.

[22] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton Series in Applied Mathematics, Princeton University Press, 2009.

[23] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

Appendix I
Useful Lemma

Lemma 1: For $W \succeq 0$ and $d \geq 0$, $\{u \,|\, a^\top u + b^\top w \leq c, \forall w : w^\top W w \leq d\} = \{u \,|\, a^\top u \leq c - \sqrt{d}\|W^{-1/2}b\|\}$.

Proof: With the change of variables $\bar{w} = W^{1/2}w$ and $\bar{b} = W^{-1/2}b$, we have $\{u \,|\, a^\top u + b^\top w \leq c, \forall w : w^\top W w \leq d\} = \{u \,|\, a^\top u + \bar{b}^\top \bar{w} \leq c, \forall \bar{w} : \bar{w}^\top \bar{w} \leq d\}$. Then, following the approach of [22, Example 1.3.3], we can obtain $\{u \,|\, a^\top u + \bar{b}^\top \bar{w} \leq c, \forall \bar{w} : \bar{w}^\top \bar{w} \leq d\} = \{u \,|\, \sqrt{d}\|\bar{b}\| \leq c - a^\top u\}$. ■

Appendix II
Proof of Theorem 1

We first show that the projection of the control action $\bar{u}[k]$ to a safe set can be done by solving:

$$u[k] \in \operatorname*{arg\,min}_{u \in \mathcal{U}} \mathbf{d}(u, \bar{u}[k]), \tag{16a}$$

$$\begin{aligned} \text{s.t.}\quad & H[k+1](\hat{A}[k]x[k] + \hat{B}[k]u + v + w) \\ & \qquad\qquad \leq h[k+1], \\ & \forall w : w^\top w \leq \frac{2rn}{\delta}, \\ & \forall v : v^\top v \leq \frac{n^2 d^2}{\sigma_{\min}(V[k])}\beta_k^2\left(\frac{\delta}{2n}\right), \end{aligned} \tag{16b}$$

Note that $x[k+1] = \hat{A}x[k] + \hat{B}u[k] + v[k] + w[k]$, where $v[k] = (A - \hat{A}[k])x[k] + (B - \hat{B}[k])u[k]$. Therefore, proving the safety of the projected control action in (16) follows from bounding the noise and perturbation terms $v[k]$ and $w[k]$ with high probability. Proposition 1 implies that

$$\begin{aligned} \mathbb{P}\left\{\|v[k]\|^2 \leq \zeta^2 n^2 \beta_k^2\left(\frac{\delta}{2n}\right)\right\} =& \mathbb{P}\left\{\|v[k]\| \leq \zeta n \beta_k\left(\frac{\delta}{2n}\right)\right\} \\ \geq& 1 - \frac{\delta}{2}, \end{aligned}$$

where $\zeta = d/\sqrt{\sigma_{\min}(V[k])}$. For the process noise, we have

$$\begin{aligned} \mathbb{P}\{w[k]^\top (rI)^{-1} w[k] \leq \varepsilon\} \geq& \mathbb{P}\{w[k]^\top W^{-1} w[k] \leq \varepsilon\} \\ \geq& 1 - \frac{\mathbb{E}\{w[k]^\top W^{-1} w[k]\}}{\varepsilon} \\ =& 1 - \frac{n}{\varepsilon}, \end{aligned}$$

where the first inequality follows from Assumption 1.a and the second inequality follows from an application of Markov's inequality for scalar random variables [23, § 2.1]. Selecting $\varepsilon = (2n)/\delta$ gives $\mathbb{P}\{w[k]^\top (rI)^{-1} w[k] \leq (2n)/\delta\} \geq 1 - \delta/2$. Finally, we note that

$$\begin{aligned} & \mathbb{P}\left\{w[k]^\top w[k] \leq \frac{2rn}{\delta} \bigwedge \|v[k]\| \leq \zeta n \beta_k\left(\frac{\delta}{2n}\right)\right\} \\ & = 1 - \mathbb{P}\left\{w[k]^\top w[k] > \frac{2rn}{\delta} \bigvee \|v[k]\| > \zeta n \beta_k\left(\frac{\delta}{2n}\right)\right\} \\ & \geq 1 - \mathbb{P}\left\{w[k]^\top w[k] > \frac{2rn}{\delta}\right\} \\ & \qquad - \mathbb{P}\left\{\|v[k]\| > \zeta n \beta_k\left(\frac{\delta}{2n}\right)\right\} \\ & = 1 - \delta, \end{aligned}$$

where the inequality follows from the union bound.

Finally, Lemma 1 can be used to eliminate v in (16) to obtain

$$\begin{aligned} u[k] \in & \operatorname*{arg\,min}_{u \in \mathcal{U}} \mathbf{d}(u, \bar{u}[k]), \\ & \text{s.t.}\quad H[k+1](\hat{A}[k]x[k] + \hat{B}[k]u + w) \leq h[k+1] \\ & \qquad\qquad - e[k+1], \quad \forall w : w^\top w \leq \frac{2rn}{\delta}, \end{aligned}$$

where $e_i[k+1] = (dn/\sqrt{\sigma_{\min}(V[k])})\beta_k(\delta/(2n))\|H_i[k+1]^\top\|$. An additional application of Lemma 1 to eliminate w concludes the proof.