Output-Feedback Safety-Preserving Control

Mahdi Yousefi, Klaske van Heusden, Ian M. Mitchell and Guy A. Dumont

***Abstract*— Safety verification of control systems is mostly discussed for full-state measurable systems and there are scant results on safety analysis of output-feedback control systems in which the states are partially measurable. This paper proposes a safety-preserving control scheme for output-feedback control systems. We specify a viable tube for the states of an observer (estimated states) based on constraints on the actual states. Furthermore, we discuss the existence of a control input which maintains trajectories of the observer within the viable tube. We prove that a control action which keeps the estimated states inside the specified tube also maintains the actual states within the original set of constraints. Finally, we compare the proposed approach in this paper with a recent technique proposed by Lesser and Abate [1] and show that our approach reduces conservatism.**

I. INTRODUCTION

Verification of safety-critical control systems, such as aviation [2], process control [3] and automated drug delivery [4], is required for reliable operation. In these applications, safety is guaranteed if we can show that the states of a controlled system can be maintained within a set of constraints (safe region) [5]. Formal model checking techniques provide us with powerful tools to analyse the behaviour of such systems [6]. There are two common approaches to study safety in safety-critical control systems: 1) to investigate the existence of a control input which preserves safety, 2) to investigate if a given feedback controller maintains safety.

Safety-preserving control techniques (approach 1) formulate a control policy which keeps the system states within the safe region. The first step in safety-preserving control techniques is to approximate the viability kernel [5]. The viability kernel is the set of all initial states for which there exists a control action that keeps trajectories of a system starting from those states within the safe region (for viability kernel approximation, see [7], [8]). Gao et al. [9] extend the definition of the viability kernel with the discriminating kernel for the case where the evolution of a system is perturbed by disturbances. We call a control input safety-preserving if it generates a viable trajectory. If the viability kernel (discriminating kernel) is empty, it means no controller can provide safety. However, if the set is not empty, one needs to synthesize a controller to preserve safety. Lygeros et al. in a series of papers [10], [11], [12] as well as Mitchell et al. in [13] used optimal control formulations based on Hamilton-Jacobi equations to synthesize a controller which satisfies safety specifications. Girard [14], [15] uses approximate bisimulation to design a safety-preserving controller. He shows that a controller which preserves safety of an approximately bisimilar abstraction of a system also maintains safety of the original system. Kaynama et al. in [16] combine a safety-preserving control law with an arbitrary controller (performance controller). This hybrid scheme is capable of satisfy performance criteria while preserving system safety [16].

To address safety using approach 2, one needs to approximate the feedback invariant for a given controller. The feedback invariant is the set of initial conditions for which the controller maintains trajectories starting from those states within the safe region. Raković et al. [17], [18], [19] use this approach to address safety in model predictive control (Tube MPC). Artstein et al. [20] employ invariant sets to verify safety in feedback systems with a given controller.

In the techniques discussed above there is an implicit assumption that the states are fully measurable. The main contribution of this paper is to introduce a safety-preserving control scheme for output-feedback control systems. Given a system with a stable state observer and a bound on the difference between initial conditions of the system and the observer, we quantify a viable tube for the observer based on constraints on the actual states. We prove that a control input which keeps trajectories of the observer within the specified viable tube also keeps trajectories of the actual system within the safe region. This approach is not limited to a specific type of controller and any safety-preserving control techniques can be employed to design a safety-preserving controller for the observer.

There are a few other papers discussing safety verification of output-feedback control systems. For instance, Lesser and Abate in [1] recently proposed a general framework to approximate the feedback invariant for output-feedback control systems (see also [21] and [22]). For a given observer, they suggest to calculate an upper bound on the state estimation error and reduce the size of the safe region accordingly. They show that for a given controller which only sees an estimate of the states, if we choose initial conditions of an output-feedback system from the feedback invariant calculated based on the contracted safe region, the controller keeps the states of the system within the actual safe region. There are two major concerns regarding this method. First, in output-feedback systems we do not have access to true initial conditions. However, if we quantify the difference between initial conditions of the system and initial conditions of the

*This work was supported by NSERC/CIHR Collaborative Health Research Projects (CHRP) Program #446511-13 (Dumont) and NSERC Discovery Grant #298211 (Mitchell).

Mahdi Yousefi, Klaske van Heusden and Guy A. Dumont are with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver BC. {mahdiyou,klaskeh,guyd }@ece.ubc.ca.

Ian M. Mitchell is with the Department of Computer Science, The University of British Columbia, Vancouver BC. mitchell@cs.ubc.ca.

observer, we can address this concern by further reducing the size of the safe region. Second, the feedback invariant is controller-specific, which means it can only be used for the controller used to approximate the feedback invariant. It does not provide us with any extra information about other feedback control laws that may result in a bigger set of viable trajectories. In this paper, we compare our approach with this technique and show that our technique results in less conservative solutions.

This paper is outlined as follows. In section II, we describe the framework of Lesser and Abate [1] for safety-verification of output-feedback controllers. Later in section IV, we show the improvement we achieve by using the approach proposed in this paper in terms of performance and conservatism. In section III, we introduce output-feedback safety-preserving control and show how maintaining the states of an observer within a viable set results in preserving safety of the actual system. Finally, section V concludes the paper and discusses further research on this topic.

A. Notations

The Minkowski sum of any two non-empty convex sets $\mathcal{P} \subset \mathcal{R}^n$ and $\mathcal{Q} \subset \mathcal{R}^n$ is $\mathcal{P} \oplus \mathcal{Q} := \{p+q| \ p \in \mathcal{P}, \ q \in \mathcal{Q}\}$; their Pontryagin difference (the erosion of $\mathcal{P}$ by $\mathcal{Q}$) is $\mathcal{P} \ominus \mathcal{Q} := \{p| \ \forall q \in \mathcal{Q}, \ p+q \in \mathcal{P}\}$. The set $\mathcal{B}(\kappa) := \{x| \ \|x\|_2 \leq \kappa\}$ denotes the closed 2-norm ball of radius $\kappa > 0$. For the set $\mathcal{P}$, $\breve{\mathcal{P}}$ refers to the interior of $\mathcal{P}$. For vectors $q \in \mathcal{Q}$ and $p \in \mathcal{P}$, $< q, p >$ denotes the inner product of the vectors.

Lemma 1: For non-empty convex sets $\mathcal{P}, \mathcal{Q} \subset \mathcal{R}^n$ the following relation holds:

$$(\mathcal{P} \ominus \mathcal{Q}) \oplus \mathcal{Q} \subseteq \mathcal{P}. \tag{1}$$

Proof: See [23]. ■

II. Background: Safety Verification of Output-Feedback Control Systems

In this section, we describe the general framework Lesser and Abate proposed for safety verification of output-feedback control systems.

Consider $\mathcal{U} \subset \mathbb{R}^m$ and $\mathcal{K} \subset \mathbb{R}^n$ as convex sets specifying constraints on inputs and states of a system, respectively. To verify that a closed-loop system is safe using the approach proposed in [1], one needs to show that the closed-loop controller provides a constrained control input which maintains states of a system inside $\mathcal{K}$ for all $t \in \mathbb{T}$, where $\mathbb{T} = \{t| \ t \in [0, \tau]\}$ and τ specifies the final time.

Consider the following state-space equation:

$$\begin{aligned} X: \quad \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t), \end{aligned} \tag{2}$$

with $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$. In (2), $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$ and $y(t) \in \mathbb{R}^p$ denote the vectors of states, inputs and outputs, respectively.

Definition 1: The finite-time feedback invariant set $\mathcal{F}$ is the set of all initial states from which a given feedback controller ($u(t) = g(x(t))$) maintains the closed-loop trajectories within the safe region without violating the input constraint $\mathcal{U}$.

Using the feedback invariant set to verify safety requires full knowledge of the states. However, for state-space model (2) in which only $y(t)$ is measurable, an observer can be implemented to estimate the states. Then, the question is if the controller which only sees the estimated states ($u(t) = g(\hat{x}(t))$) still satisfies the feedback invariance property. Relying on fast convergence of high-gain observers, Lesser and Abate in [1] use a high-gain observer of the following form to estimate the states of (2):

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + H(\epsilon)(y(t) - C\hat{x}(t)). \tag{3}$$

The observer gain $H(\epsilon)$ is characterized by a positive parameter ϵ:

$$H(\epsilon) = \begin{bmatrix} \frac{\alpha_1}{\epsilon} & \frac{\alpha_2}{\epsilon^2} & \dots & \frac{\alpha_n}{\epsilon^n} \end{bmatrix}^T. \tag{4}$$

The α_is are selected such that the following polynomial is Hurwitz:

$$s^n + \alpha_1 s^{n-1} + \alpha_2 s^{n-2} + \cdots + \alpha_{n-1}s + \alpha_n. \tag{5}$$

The selection of α_is guarantees that the error dynamics defined over the error signal.

$$e(t) = x(t) - \hat{x}(t), \tag{6}$$

is stable. By choosing ϵ arbitrarily small, the error signal converges to zero arbitrarily fast. Lesser and Abate quantified an upper bound on the estimation error ($e(t)$) as a function of ϵ:

$$\|e(t)\|_2 = \|x(t) - \hat{x}(t)\|_2 \leq \delta(\epsilon). \tag{7}$$

Consider a closed-loop system X_C which is formed by combining state-space equation (2), the high-gain observer (3) and the feedback control policy $u(t) = g(\hat{x}(t))$. The following procedure is proposed in [1] to calculate the feedback invariant set for the output-feedback control system X_C:

1) Compute $\delta(\epsilon)$ for the high-gain observer,
2) Compute $\tilde{\mathcal{F}}$ using formal techniques for fully observable systems with the state constraint $\bar{\mathcal{K}} = \mathcal{K} \ominus \mathcal{B}(\delta(\epsilon))$ and the input constraint $\mathcal{U}$,

They prove that closed-loop trajectories of X_C starting from $x(0) \in \tilde{\mathcal{F}}$ and under the control policy $g(\hat{x})$, stay within the safe set $\mathcal{K}$. One of the drawbacks of this approach is that $x(0)$ is unknown and we cannot show that $x(0) \in \tilde{\mathcal{F}}$. However, if the uncertainty on $x(0)$ can be specified as $\|x(0) - \hat{x}(0)\|_2 \leq \lambda$, by choosing $\hat{x}(0)$ from $\bar{\mathcal{F}} = \tilde{\mathcal{F}} \ominus \mathcal{B}(\lambda)$ it can be shown that $x(0) \in \tilde{\mathcal{F}}$.

III. Output-Feedback Safety-Preserving Control

One of the shortcomings of the approach described in section II is that the feedback invariance is limited to a specific controller and cannot be used for a different controller. Moreover, there might be a bigger set of initial conditions which result in viable trajectories with a different choice of controller. In this section, we propose a more general approach to preserve safety of systems in which the states are not fully measurable. In this design, we follow the scheme proposed by Kaynama et al. [16] to preserve safety while meeting performance specifications.

A. Notations And Definitions

Consider the safe tube $\mathcal{K}_{\mathbb{T}}$ which characterizes constraints on trajectories of X:

$$\mathcal{K}_{\mathbb{T}} = \{x(\cdot)|\ \forall t \in \mathbb{T},\ x(t) \in \mathcal{K}_t\}. \tag{8}$$

Since the constraint on the states of X does not change over time, $\mathcal{K}_t = \mathcal{K}$ in (8).

Definition 2: Output-feedback safety-preserving control guarantees that there is a constrained input that keeps closed-loop trajectories of X within $\mathcal{K}_{\mathbb{T}}$ despite the fact that the states of X are not fully measurable and only their estimates are available.

We use a state observer of the following form to estimate the states of X:

$$\begin{aligned} O:\ \dot{\hat{x}}(t) =&(A - LC)\hat{x}(t) + Bu(t) + Ly(t) \\ =&(A - LC)\hat{x}(t) + Bu(t) + LCx(t). \end{aligned} \tag{9}$$

In the above equation L must be designed such that $(A-LC)$ is stable. Since $H(\epsilon)$ stabilizes $(A - H(\epsilon)C)$, we can also use $H(\epsilon)$ defined in (4) instead of L in (9). Consider the estimation error defined in (6) as the difference between the states of X and their estimates. We can formulate the error dynamics illustrating the evolution of the estimation error, as follows:

$$E:\ \dot{e}(t) = (A - LC)e(t). \tag{10}$$

Consider $\mathcal{E}_0 \subset \mathbb{R}^n$ as a convex set of all possible initial errors $e(0)$ and $\mathcal{E}_t \subset \mathbb{R}^n$ as a set of all reachable errors at time t calculated based on the evolution of the error dynamics starting from $\mathcal{E}_0$. Now we can define the error tube which is the set of all error trajectories $(e(\cdot))$ starting from $\mathcal{E}_0$:

$$\mathcal{T}^E_{\mathbb{T}} = \{e(\cdot)|\ \forall t \in \mathbb{T},\ e(t) \in \mathcal{E}_t\}, \tag{11}$$

Due to the stability of the error dynamics, $\mathcal{E}_t$ becomes smaller as time goes forward. According to the observer dynamics defined in (9), the states of X have a direct influence on the states of the observer. However, since we can specify the set of all estimation errors, the observer dynamics can be reformulated independently of the states of X:

$$O:\ \dot{\hat{x}}(t) = A\hat{x} + Bu(t) + LCe(t), \quad e(t) \in \mathcal{E}_t. \tag{12}$$

By looking at (2) and (12), we can see that not only X and O share the same input, but also have the same dynamics. The only difference is that the states of O are perturbed by the external variable $e(t) \in \mathcal{E}_t$.

Definition 3: The finite-horizon discriminating kernel of $\mathcal{K}_{\mathbb{T}}$ for O defined in (12) is a subset of $\mathcal{K}_0$ ($\mathcal{K}_t$ for $t = 0$) which specifies all initial conditions such that for all $t \in \mathbb{T}$ and for all $e(t) \in \mathcal{E}_t$, there exists an admissible input $u(\cdot) \in \mathcal{U}_{\mathbb{T}}$ that keeps trajectories $(\hat{x}(\cdot))$ of O starting from those initial states within $\mathcal{K}_{\mathbb{T}}$:

$$\begin{aligned} Disc_0(\mathcal{K}_{\mathbb{T}},\mathcal{U}_{\mathbb{T}},\mathcal{T}^E_{\mathbb{T}},O) = \{&\hat{x}(0) \in \mathcal{K}_0|\ \forall e(\cdot) \in \mathcal{T}^E_{\mathbb{T}}, \\ &\exists u(\cdot) \in \mathcal{U}_{\mathbb{T}}\ s.t.\ \hat{x}(\cdot) \in \mathcal{K}_{\mathbb{T}}\}. \end{aligned} \tag{13}$$

In (13), $\mathcal{U}_{\mathbb{T}} = \{u(\cdot)|\ \forall t \in \mathbb{T},\ u(t) \in \mathcal{U}\}$ where $\mathcal{U}_t = \mathcal{U}$, is a set of all admissible inputs over $\mathbb{T}$. Without loss of generality, we can reformulate *Definition 2* as follows:

Definition 4: Output-feedback safety-preserving control guarantees that there is a constrained input which keeps the state of O within a given set of constraints (a set which we can under-approximate), and that input will also keep the states of X within $\mathcal{K}$.

B. Main Result

The next proposition formulates output-feedback safety-preserving control assuming $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}},\mathcal{U}_{\mathbb{T}},\mathcal{T}^E_{\mathbb{T}},O)$ is non-empty, where

$$\hat{\mathcal{K}}_{\mathbb{T}} = \mathcal{K}_{\mathbb{T}} \ominus \mathcal{T}^E_{\mathbb{T}}. \tag{14}$$

Proposition 1: For the state-space model X defined in (2), the state observer O formulated in (12), the initial error set $\mathcal{E}_0$ and the safe tube $\hat{\mathcal{K}}_{\mathbb{T}}$ defined in (14), assuming $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}},\mathcal{U}_{\mathbb{T}},\mathcal{T}^E_{\mathbb{T}},O)$ is not empty, a safety-preserving control input that keeps trajectories of O (estimated states) within $\hat{\mathcal{K}}_{\mathbb{T}}$, also maintains trajectories of X within $\mathcal{K}_{\mathbb{T}}$.

Proof: Since $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}},\mathcal{U}_{\mathbb{T}},\mathcal{T}^E_{\mathbb{T}},O)$ is not empty, we can define the following viable tube as a set of all viable trajectories of the observer inside $\hat{\mathcal{K}}_{\mathbb{T}}$:

$$\mathcal{T}^O_{\mathbb{T}} = \{\hat{x}(\cdot)|\forall t \in \mathbb{T},\ \hat{x}(t) \in Disc_t(\hat{\mathcal{K}}_{[t,\tau]},\mathcal{U}_{[t,\tau]},\mathcal{T}^E_{[t,\tau]},O)\}, \tag{15}$$

In (15), $Disc_t(\hat{\mathcal{K}}_{[t,\tau]},\mathcal{U}_{[t,\tau]},\mathcal{T}^E_{[t,\tau]},O)$ is a set of states at time t for which there exists $u(\cdot) \in \mathcal{U}_{[t,\tau]}$ that keeps trajectories of O starting from t within $\hat{\mathcal{K}}_{[t,\tau]}$. Due to the fact that the defined trajectories start from $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}},\mathcal{U}_{\mathbb{T}},\mathcal{T}^E_{\mathbb{T}},O)$, $\mathcal{T}^O_{\mathbb{T}}$ is non-empty. This means that there exists $u(\cdot) \in \mathcal{U}_{\mathbb{T}}$ to keep trajectories of O viable. Moreover, since $\mathcal{T}^O_{\mathbb{T}}$ includes all trajectories of O inside $\hat{\mathcal{K}}_{\mathbb{T}}$, we have $\mathcal{T}^O_{\mathbb{T}} \subseteq \hat{\mathcal{K}}_{\mathbb{T}}$. As we discussed previously, the set $\mathcal{T}^E_{\mathbb{T}}$ defined in (11) is a set of all trajectories of the error between the actual states and the estimated states starting from $\mathcal{E}_0$. Moreover, based on the definition of the estimation error (6), we can calculate the states of X as $x(t) = \hat{x}(t) + e(t)$. Thus, using the same input which generates $\mathcal{T}^O_{\mathbb{T}}$, we can characterize all resulting trajectories of X as follows:

$$\mathcal{T}^X_{\mathbb{T}} = \mathcal{T}^O_{\mathbb{T}} \oplus \mathcal{T}^E_{\mathbb{T}}. \tag{16}$$

Since $\mathcal{T}^O_{\mathbb{T}} \subseteq \hat{\mathcal{K}}_{\mathbb{T}}$ and therefore $\mathcal{T}^O_{\mathbb{T}} \subseteq \mathcal{K}_{\mathbb{T}} \ominus \mathcal{T}^E_{\mathbb{T}}$, the viable trajectories of the actual system $\mathcal{T}^X_{\mathbb{T}}$ is inside the safe tube $\mathcal{K}_{\mathbb{T}}$:

$$\mathcal{T}^X_{\mathbb{T}} = \mathcal{T}^O_{\mathbb{T}} \oplus \mathcal{T}^E_{\mathbb{T}} \subseteq (\mathcal{K}_{\mathbb{T}} \ominus \mathcal{T}^E_{\mathbb{T}}) \oplus \mathcal{T}^E_{\mathbb{T}}, \tag{17}$$

and according to *Lemma 1*, $\mathcal{T}^X_{\mathbb{T}} \subseteq \mathcal{K}_{\mathbb{T}}$. ■

According to *Proposition 1*, a safety-preserving input that keeps the states of an observer designed to estimate the state of an output-feedback system within the contracted version of the safe region also maintains the states of the actual system within the safe region. In this proposition, we assume that $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}},\mathcal{U}_{\mathbb{T}},\mathcal{T}^E_{\mathbb{T}},O)$ is not empty. However, if the set is empty, it means there exists no controller providing safety for a given output-feedback system with a given set of constraints. The main difference between this approach and the one described in [1] is that the proposed approach is not controller-specific. Once $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}},\mathcal{U}_{\mathbb{T}},\mathcal{T}^E_{\mathbb{T}},O,)$ is approximated, we can formulate any sort of safety-preserving controller to maintain trajectories of the observer within $\hat{\mathcal{K}}_{\mathbb{T}}$ (see [10], [16], [24]). Moreover, to approximate feedback invariance for output-feedback systems, the safe region needs to be eroded by an upper bound on the estimation error. The proposed approach in this paper suggests to calculate the evolution of the error dynamics and erode the safe region by a set of all possible errors at each time. Accordingly, due to the stability of the error dynamics, the error set gets smaller as time goes forward which results in less erosion of the safe region and consequently, less conservative results.

C. Discriminating Kernel Approximation

Kaynama et al. [16] show that the discriminating kernel (viability kernel) can be under-approximated by recursive approximation of maximal reachable sets. Due to the fact that $\hat{\mathcal{K}}_t$ changes over time, we cannot employ this method to approximate $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}},\mathcal{U}_{\mathbb{T}},\mathcal{T}^E_{\mathbb{T}},O)$. Here, we extend the algorithm proposed in [8] to the case where the safe region (14) is time varying.

Definition 5: The maximal reachable set of $\mathcal{K}$ for O at time t is a set of initial states for which there exists an input such that trajectories of O starting from those states reach $\mathcal{K}$ exactly at t:

$$\begin{aligned} \mathit{Reach}^{\#}_t(\mathcal{K},\mathcal{U}_{[0,t]},\mathcal{T}^E_{[0,t]},O) = \{\hat{x}(0)|\ \forall e(\cdot) \in \mathcal{T}^E_{[0,t]},\\ \exists u(\cdot) \in \mathcal{U}_{[0,t]}\ s.t.\ \hat{x}(t) \in \mathcal{K}\}. \end{aligned} \tag{18}$$

Consider $\mathbb{P}^l_{\mathbb{T}} = \{t_0, t_1, \ldots, t_l\}$ as a set of points where $t_0, t_1, \ldots, t_l \in \mathbb{T}$, $t_0 < t_1 < \cdots < t_l$ and $t_0 = \min_{t\in\mathbb{T}} t$ and $t_l = \max_{t\in\mathbb{T}} t$. Moreover, we define $|\mathbb{P}^l_{\mathbb{T}}|$ as $|\mathbb{P}^l_{\mathbb{T}}| := \max\{t_{k+1} - t_k|\ k = 0, 1 \ldots, l-1,\ t_k \in \mathbb{P}^l_{\mathbb{T}}\}$. Assume O is bounded by M on $\hat{\mathcal{K}}_{\mathbb{T}}$, $\mathcal{U}_{\mathbb{T}}$ and $\mathcal{T}^E_{\mathbb{T}}$, which means for all $t \in \mathbb{T}$ and for all $\hat{x}(t) \in \hat{\mathcal{K}}_t$, $u(t) \in \mathcal{U}_t$, $e(t) \in \mathcal{E}(t)$, $\|\dot{\hat{x}}(t)\|_2 \leq M$. We define an under-approximation of $\hat{\mathcal{K}}_{\mathbb{T}}$ as:

$$\hat{\mathcal{K}}^{\downarrow}_{\mathbb{T}} = \{\hat{x}(\cdot)|\ \forall t \in \mathbb{T},\ x(t) \in \hat{\mathcal{K}}^{\downarrow}_t\}. \tag{19}$$

In (19), $\hat{\mathcal{K}}^{\downarrow}_t = \hat{\mathcal{K}}_t \ominus \mathcal{D} \ominus \mathcal{B}(M|\mathbb{P}^l_{\mathbb{T}}|)$, and $\mathcal{D} = \arg\max\{vol(\mathcal{Z})|\ \mathcal{Z} = \hat{\mathcal{K}}_{t_{i-1}} \ominus \hat{\mathcal{K}}_t,\ t_i \in \mathbb{P}^l_{\mathbb{T}}, t \in [t_{i-1}, t_i)\}$. In the above equation, $vol(\mathcal{Z})$ denotes the volume of the set $\mathcal{Z}$. $\mathcal{D}$ shows the maximum contraction of $\hat{\mathcal{K}}_{\mathbb{T}}$ in each interval $[t_{i-1}, t_i)$. In the next proposition, we show that if we keep the states of O for all $t_i \in \mathbb{P}^l_{\mathbb{T}}$ within $\hat{\mathcal{K}}^{\downarrow}_{t_i}$, they will stay within $\hat{\mathcal{K}}_t$ for all $t \in [t_i, t_{i+1})$.

Consider the following l-step recursion:

$$\begin{aligned} &\mathcal{R}_{t_l} = \hat{\mathcal{K}}^{\downarrow}_{t_l},\\ &\mathcal{R}_{t_i-1} = \hat{\mathcal{K}}^{\downarrow}_{t_i-1} \bigcap \mathit{Reach}^{\#}_{t_i-t_{i-1}}\left(\mathcal{R}_{t_i}, \mathcal{U}_{[t_{i-1},t_i]}, \mathcal{T}^E_{[t_{i-1},t_i]}, O\right),\\ &\text{for } i = l, l-1, \ldots, 1. \end{aligned} \tag{20}$$

Proposition 2: For the state-space equation defined in (12) which is bounded by M on $\hat{\mathcal{K}}_{\mathbb{T}}$, $\mathcal{U}_{\mathbb{T}}$ and $\mathcal{T}^E_{\mathbb{T}}$, the final set $\mathcal{R}_0$ defined by the recursive relation (20) satisfies:

$$\mathcal{R}_0 \subseteq Disc_0(\hat{\mathcal{K}}_{\mathbb{T}},\mathcal{U}_{\mathbb{T}},\mathcal{T}^E_{\mathbb{T}},O). \tag{21}$$

Proof: Calculating $\mathcal{R}_0$ using the recursive relation (20) indicates that starting from any points in $\mathcal{R}_{t_i-1}$ there exists $u_{t_i}(\cdot) \in \mathcal{U}_{[t_{i-1},t_i]}$ such that the states of O can reach $\mathcal{R}_{t_i}$ at $t = t_i - t_{i-1}$. By taking the concatenation of the inputs $u_{t_i}(\cdot)$ for all $t_i \in \mathbb{P}^l_{\mathbb{T}}$ [8], we can define an input $u_{\mathbb{T}}(\cdot) \in \mathcal{U}_{\mathbb{T}}$ which, starting from any points in $\mathcal{R}_0$, the states of X at $t_i \in \mathbb{P}^l_{\mathbb{T}}$ stay within $\hat{\mathcal{K}}_{t_i}$:

$$x(t_i) \in \mathcal{R}_{t_i} \subseteq \hat{\mathcal{K}}^{\downarrow}_{t_i} \subseteq \hat{\mathcal{K}}_{t_i}. \tag{22}$$

To show that (21) holds, we need to show for all $t \in \mathbb{T}$, $u_{\mathbb{T}}(\cdot)$ maintains $\hat{x}(t) \in \hat{\mathcal{K}}_t$. Since any $t \in \mathbb{T}$ lies in some interval ($t \in [t_i, t_{i+1}]$) and due to the fact that O is bounded by M, we have:

$$\begin{aligned} \|\hat{x}(t) - \hat{x}(t_i)\|_2 \leq \| \int_{t_i}^{t} \dot{\hat{x}}(\tau) d\tau \|_2 \leq M(t - t_i)\\ < M(t_{i+1} - t_i) \leq M|\mathbb{P}^l_{\mathbb{T}}|. \end{aligned} \tag{23}$$

According to (23), $\nu = \hat{x}(t) - \hat{x}(t_i) \in \mathcal{B}(M|\mathbb{P}^l_{\mathbb{T}}|)$. Since, $\hat{x}(t_i) \in \hat{\mathcal{K}}^{\downarrow}_{t_i}$ and according to the definition of ν as well as *Lemma 1*, we have:

$$\begin{aligned} \hat{x}(t) &\in \hat{\mathcal{K}}^{\downarrow}_{t_i} \oplus \mathcal{B}(M|\mathbb{P}^l_{\mathbb{T}}|)\\ &\subseteq \left(\hat{\mathcal{K}}_{t_i} \ominus \mathcal{D} \ominus \mathcal{B}(M|\mathbb{P}^l_{\mathbb{T}}|)\right) \oplus \mathcal{B}(M|\mathbb{P}^l_{\mathbb{T}}|)\\ &\subseteq \hat{\mathcal{K}}_{t_i} \ominus \mathcal{D} \subseteq \hat{\mathcal{K}}_{t_i} \ominus (\hat{\mathcal{K}}_{t_i} \ominus \hat{\mathcal{K}}_t) \subseteq \hat{\mathcal{K}}_t. \end{aligned} \tag{24}$$

This result implies that starting from any point in $\mathcal{R}_0$ there is an input to keep the states of O inside $\hat{\mathcal{K}}_t$ at time t. Thus, (21) holds. ■

D. Safety-Preserving Control Synthesis

In section III-B, we proved that a safety-preserving controller that keeps the states of the observer inside $\hat{\mathcal{K}}_{\mathbb{T}}$ maintains the states of the actual system within $\mathcal{K}_{\mathbb{T}}$ as well. We employ the method described in [16] to design a safety-preserving controller to keep trajectories of the observer viable. Kaynama et al. in [16] combine a safety-preserving control law with an arbitrary controller (performance controller) to satisfy performance criteria while preserving system safety. In this scheme, to maintain trajectories of O within $\hat{\mathcal{K}}_{\mathbb{T}}$, the input of the system is selected as follows:

$$u(t) = \begin{cases} u_{pr}(t), & \hat{x}(t) \in D\breve{i}sc_t(\hat{\mathcal{K}}_{[t,\tau]},\mathcal{U}_{[t,\tau]},\mathcal{T}^E_{[t,\tau]},O);\\ u_{sp}(t), & \hat{x}(t) \notin D\breve{i}sc_t(\hat{\mathcal{K}}_{[t,\tau]},\mathcal{U}_{[t,\tau]},\mathcal{T}^E_{[t,\tau]},O). \end{cases} \tag{25}$$

In the above equation $u_{pr}(t)$ is a control input provided by the performance controller. Moreover, the safety-preserving control input $u_{sp}(t)$ is obtained based on the safety-preserving control policy formulated in [25]. To avoid high-frequency switching between the two control modes in (25), Kaynama et al. [16] use a convex combination of them:

$$u(t) = (1 - \zeta) u_{pr}(t) + \zeta u_{sp}(t). \tag{26}$$

In the above equation, ζ is calculated based on the difference between $\hat{x}(t)$ and the boundaries of $Disc_t(\hat{\mathcal{K}}_{[t,\tau]}, \mathcal{U}_{[t,\tau]}, \mathcal{T}^E_{[t,\tau]}, O)$ [16].

IV. Computational Example

In this section, we apply the proposed output-feedback safety-preserving control to the example Lesser and Abate discussed in [1]. We regenerate the results presented in [1] and compare them with the results of the technique proposed in this paper. We show that there exists a safety-preserving controller for the closed-loop system defined below that generates viable trajectories starting from some initial conditions from outside of $\bar{\mathcal{F}}$ but inside $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}}, \mathcal{U}_{\mathbb{T}}, \mathcal{T}^E_{\mathbb{T}}(\cdot), O)$.

A. Model Definition

Consider dynamics of the double integrator:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t),$$

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}. \qquad (27)$$

A controller of the following form is used in [1] to stabilize (27):

$$u(t) = \begin{bmatrix} -\beta & -\beta \end{bmatrix} x(t). \qquad (28)$$

Since $x_2(t)$ cannot be measured, the following high-gain observer is designed in [1] to estimates the states of (27):

$$\dot{\hat{x}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \hat{x}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) + \begin{bmatrix} \frac{\alpha_1}{\epsilon} \\ \frac{\alpha_2}{\epsilon^2} \end{bmatrix} (x_1(t) - \hat{x}_1(t)). \qquad (29)$$

Thus, (28) needs to be reformulated as

$$u(t) = \begin{bmatrix} -\beta & -\beta \end{bmatrix} \hat{x}(t). \qquad (30)$$

In this example, the states are required not to leave the safe region $\mathcal{K} = \{x(t) | \; |x_1(t)| \leq 4 \;\; |x_2(t)| \leq 3\}$ over $\mathbb{T} = \{t | \; t \in [0, 10]\}$. We assume the uncertainty on $x(0)$ to be $\|x(0) - \hat{x}(0)\|_2 = 0.5$. The absolute value of the input is restricted to be less than 1, $\mathcal{U} = \{u(t) | \; |u(t)| \leq 1\}$. The parameters of the observer are set to $\alpha_1 = \alpha_2 = 4$ and $\epsilon = 0.01$. For this observer the upper bound on the 2-norm of the estimation error is $\delta(\epsilon = 0.01) = 0.1768$ [1]. Lesser and Abate in [1] showed that for the system and controller defined in (27) and (28), setting $\beta = 0.2$ results in the largest feedback-invariant set.

To approximate the discriminating kernel for the observer defined in (29), we calculate the evolution of the error dynamics defined below with the initial error set $\mathcal{E}_0 = \{e(0) \in \mathbb{R}^2 | \; \|e(0)\|_2 \leq 0.5\}$:

$$\dot{e}(t) = \left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \frac{\alpha_1}{\epsilon} \\ \frac{\alpha_2}{\epsilon^2} \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \right) e(t). \qquad (31)$$

We use the same set of initial errors to calculate $\bar{\mathcal{F}}$ as described in section II.

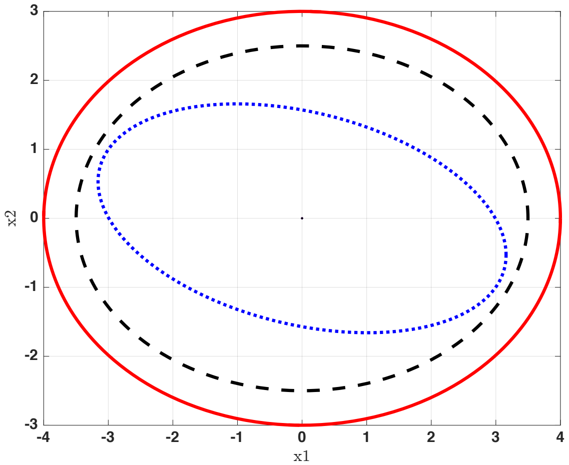

Fig. 1. Solid-line red ellipse: Ellipsoidal under-approximation of safe region $\mathcal{K}$, dashed-line black ellipse: $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}}, \mathcal{U}_{\mathbb{T}}, \mathcal{T}^E_{\mathbb{T}}, O)$, dotted-line blue ellipse: $\bar{\mathcal{F}}$.

B. Results

In this paper, we employ the ellipsoidal techniques [25] implemented in [26] to represent sets and conduct operations on them. Moreover, we utilize the level set toolbox developed by Mitchell [27] for reachability analyses. Following the steps we described in sections III and IV, we calculate the feedback invariant ($\bar{\mathcal{F}}$) of the system described in (27), (28) and (29), as well as the discriminating kernel for the observer $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}}, \mathcal{U}_{\mathbb{T}}, \mathcal{T}^E_{\mathbb{T}}, O)$. The results are illustrated in Fig. (1). Comparing $\bar{\mathcal{F}}$ with $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}}, \mathcal{U}_{\mathbb{T}}, \mathcal{T}^E_{\mathbb{T}}, O)$ depicts that we can even have a bigger viable set if we use a different controller rather than the one defined in (28). In other words, Fig. 1 shows that there is a bigger set of initial conditions that can result in viable trajectories.

In the next step, we employ the safety-preserving controller described in section III to maintain the state of the observer defined in (29) within the viable tube $Disc_{\mathbb{T}}(\hat{\mathcal{K}}_{\mathbb{T}}, \mathcal{U}_{\mathbb{T}}, \mathcal{T}^E_{\mathbb{T}}, O)$. We employ the same controller defined in (28) with $\beta = 0.2$ as the performance controller in the proposed output-feedback safety-preserving control scheme. Fig. 2 shows an observed trajectory of (27) under safety-preserving control with $\eta = 0.8$ starting from $\hat{x}(0) = (1.5, 2)$ which is inside $Disc_0(\hat{\mathcal{K}}_{\mathbb{T}}, \mathcal{U}_{\mathbb{T}}, \mathcal{T}^E_{\mathbb{T}}, O)$ (dotted green line). Fig. 2 also depicts the actual trajectory of (27) generated with the same input sequence but a different initial condition ($x(0) = (2, 2)$) due to the uncertainty (solid black line). Due to the fast convergence of the high-gain observer, the estimated states converge to the actual states right after the starting time. In Fig. 2, we can see that the safety-preserving control input (solid blue line in Fig. 3) that keeps the state of the observer viable preserves safety for (27). Fig. 2 illustrates a trajectory of (27) under the feedback control law (27) with the same choice of $\beta = 0.2$ (dashed blue line). As we expected, since $(1.5, 2)$ is not inside $\bar{\mathcal{F}}$, the controller cannot keep the trajectory inside the safe region.

V. Conclusions and Future Work

In this paper, we proposed a novel approach to preserve safety of output-feedback control systems. Given an observer

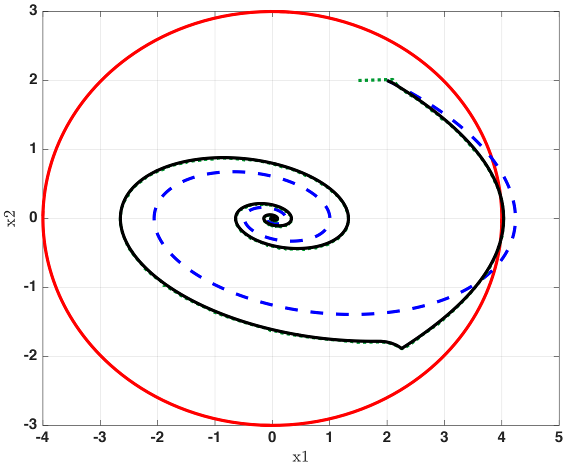

Fig. 2. Closed-loop trajectories. Red ellipse: the safe region $\mathcal{K}$; solid black line: the trajectory of X controlled by output-feedback safety-preserving control; dotted green line: the observed trajectory of X; dashed blue line: the trajectory of X controlled by the state-feedback controller (28).

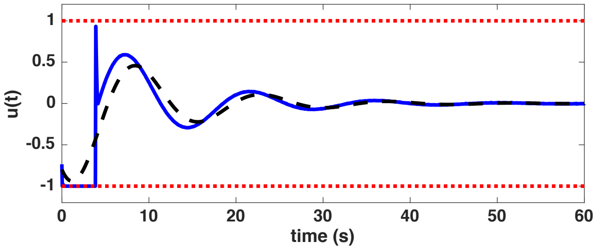

Fig. 3. Control input. Solid blue line: a control input sequence generated by the output-feedback safety-preserving control; dashed black line: a control input sequence generated by the feedback controller (28); dotted red lines: upper and lower bounds.

designed to estimate the states of an output-feedback system, we showed that a safety-preserving controller that keeps the state of the observer (estimated states) within a contracted safe region is capable of maintaining the state of the actual system within the original safe region. In contrast to other approaches suggested in the literature, the proposed technique is not limited to a specific type of controller and any safety-preserving controller can be employed. We compared the proposed technique with a recent approach for safety verification of output control systems [1] and showed that our method provides less conservative solutions.

In the proposed scheme, we only consider the case where a model of the system is known. The proposed technique can be extended to the case where we have an uncertain model of an output-feedback system. As a part of future work on this topic, we will extend the developed technique to the case of uncertain systems in which the model used in observer design does not necessarily represent the true behaviour of the system.

References

[1] K. Lesser and A. Abate. Safety verification of output feedback controllers for nonlinear systems. In *Control Conference (ECC), 2016 European*, pages 413–418. IEEE, 2016.

[2] K. Margellos and J. Lygeros. Air traffic management with target windows: An approach using reachability. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 145–150, Dec 2009.

[3] J. Hasenauer, P. Rumschinski, S. Waldherr, S. Borchers, F. Allgöwer, and R. Findeisen. Guaranteed steady-state bounds for uncertain chemical processes. In *Proceedings of International Symposium on Advanced Control of Chemical Processes, ADCHEM'09, Istanbul, Turkey*, pages 674–679, 2009.

[4] M. Yousefi, K. van Heusden, G.A. Dumont, and J.M. Ansermino. Safety-preserving closed-loop control of anesthesia. In *Engineering in Medicine and Biology Society (EMBC), 37th Annual International Conference of the IEEE*, pages 454–457. IEEE, 2015.

[5] J. P. Aubin, A. M Bayen, and P. Saint-Pierre. *Viability theory: new directions*. Springer Science & Business Media, 2011.

[6] E.M. Clarke. The birth of model checking. In *25 Years of Model Checking*, pages 1–26. Springer, 2008.

[7] K. Margellos and J. Lygeros. Viable set computation for hybrid systems. *Nonlinear Analysis: Hybrid Systems*, 10:45–62, 2013.

[8] J. N. Maidens, S. Kaynama, I. M. Mitchell, M. K. Oishi, and G. A. Dumont. Lagrangian methods for approximating the viability kernel in high-dimensional systems. *Automatica*, 49(7):2017–2029, 2013.

[9] Y. Gao, J. Lygeros, and M. Quincampoix. On the reachability problem for uncertain hybrid systems. *Automatic Control, IEEE Transactions on*, 52(9):1572–1586, 2007.

[10] J. Lygeros, C. Tomlin, and S. Sastry. Controllers for reachability specifications for hybrid systems. *Automatica*, 35(3):349–370, 1999.

[11] J. Lygeros. On reachability and minimum cost optimal control. *Automatica*, 40(6):917–927, 2004.

[12] K. Margellos and J. Lygeros. Hamilton–Jacobi formulation for reach–avoid differential games. *IEEE Transactions on Automatic Control*, 56(8):1849–1861, 2011.

[13] I.M. Mitchell, A.M. Bayen, and C.J. Tomlin. A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games. *IEEE Transactions on automatic control*, 50(7):947–957, 2005.

[14] A. Girard. Synthesis using approximately bisimilar abstractions: state-feedback controllers for safety specifications. In *Proceedings of the 13th ACM international conference on Hybrid systems: computation and control*, pages 111–120. ACM, 2010.

[15] A. Girard. Controller synthesis for safety and reachability via approximate bisimulation. *Automatica*, 48(5):947–953, 2012.

[16] S. Kaynama, I.M. Mitchell, M. Oishi, and G.A. Dumont. Scalable safety-preserving robust control synthesis for continuous-time linear systems. *IEEE Transactions on Automatic Control*, 60(11):3065–3070, 2015.

[17] S.V. Raković, E.C. Kerrigan, Konstantinos I. Kouramas, and D.Q. Mayne. Invariant approximations of the minimal robust positively invariant set. *Automatic Control, IEEE Transactions on*, 50(3):406–410, 2005.

[18] S.V. Raković. Set theoretic methods in model predictive control. In *Nonlinear Model Predictive Control*, pages 41–54. Springer, 2009.

[19] S.V. Raković, B. Kouvaritakis, R. Findeisen, and M. Cannon. Homothetic tube model predictive control. *Automatica*, 48(8):1631–1638, 2012.

[20] Z. Artstein and S.V. Raković. Set invariance under output feedback: a set-dynamics approach. *International Journal of Systems Science*, 42(4):539–555, 2011.

[21] S. Haesaert, A. Abate, and P.M.J. Van den Hof. Correct-by-design output feedback of LTI systems. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 6159–6164. IEEE, 2015.

[22] D.Q. Mayne, S.V. Raković, R. Findeisen, and F. Allgöwer. Robust output feedback model predictive control of constrained linear systems. *Automatica*, 42(7):1217–1222, 2006.

[23] P.K. Ghosh and R.M. Haralick. Mathematical morphological operations of boundary-represented geometric objects. *Journal of Mathematical Imaging and Vision*, 6(2-3):199–222, 1996.

[24] M. Yousefi, K. van Heusden, G.A. Dumont, I.M. Mitchell, and J.M. Ansermino. Model-invariant safety-preserving control. In *American Control Conference (ACC), 2016*, pages 6689–6694. IEEE, 2016.

[25] A.B. Kurzhanski and I. Vályi. *Ellipsoidal calculus for estimation and control*. Nelson Thornes, 1997.

[26] A.A. Kurzhanski and P.N. Varaiya. Ellipsoidal toolbox. *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2006-46*, 2006.

[27] I. M. Mitchell. A toolbox of level set methods. *Departmental of Computer Science, University of British Columbia, Vancouver, BC, Canada, http://www.cs.ubc.ca/~mitchell/ToolboxLS/toolboxLS.pdf, Technical Report TR-2004-09*, 2004.