Safe Control for Nonlinear Systems under Faults and Attacks via Control Barrier Functions

Hongchao Zhang, *Student Member, IEEE,* Zhouchi Li, *Student Member, IEEE,* and Andrew Clark, *Senior Member, IEEE*

***Abstract*— Safety is one of the most important properties of control systems. Sensor faults and attacks and actuator failures may cause errors in the sensor measurements and system dynamics, which leads to erroneous control inputs and hence safety violations. In this paper, we improve the robustness against sensor faults and actuator failures by proposing a class of Fault-Tolerant Control Barrier Functions (FT-CBFs) for nonlinear systems. Our approach maintains a set of state estimators according to fault patterns and incorporates CBF-based constraints to ensure safety under sensor faults. We then propose a framework for joint safety and stability by integrating FT-CBFs with Control Lyapunov Functions. By utilizing redundancy, we proposed High order CBF-based approach to ensure safety when actuator failures occur. We propose a sum-of-squares (SOS) based approach to verify the feasibility of FT-CBFs for both sensor faults and actuator failures. We evaluate our approach via two case studies, namely, a wheeled mobile robot (WMR) system in the presence of a sensor attack and a Boeing 747 lateral control system under actuator failures.**

***Index Terms*— Fault-tolerant control; higher-order control barrier functions; stochastic control barrier functions; analysis of reliability and safety; sensor faults, attacks; actuator failures.**

I. Introduction

A control system is safe if it remains within a predetermined safe region for all time [1]. In applications including medicine, transportation and energy, safety violations can cause catastrophic economic damage and loss of human life [2].

Approaches verifiable safe control systems include Hamilton–Jacobi–Isaacs (HJI) equation [3], barrier certificates [4], and Control Barrier Functions (CBFs) [5]. Among those methods, CBFs have the advantage that they can be readily integrated into existing control policies by adding linear constraints on the control input. A CBF is a function of the system state that goes to zero as the system approaches the unsafe region. Thus, ensuring safety is equivalent to ensuring that the CBF stays non-negative. CBF-based approaches show great promise in safety-critical systems. However, they rely on sensor measurements, which sensor faults, attacks, and actuator failures may compromise.

Sensor faults and malicious attacks provide inaccurate, arbitrary readings. These inaccurate measurements bias estimates of the system state, leading to erroneous control signals that drive the true system state to an unsafe operating point. When actuator failures occur, actuators lose effectiveness and hence render the control system unable to ensure safety or stability [6]. Sensor faults, attacks and actuator failures can cause arbitrary errors in the sensor measurements and system dynamics, which is challenging for existing CBF-based approaches such as [7] that assume that noises and disturbances are either bounded or come from a known probability distribution.

Countermeasures such as Fault-Tolerant Control (FTC) [8], [9] have been proposed to accommodate faults, attacks and failures. Existing FTC approaches focus on maintaining performance and do not provide provable safety guarantees. Countermeasures incorporating disturbance observer-based CBF are proposed to ensure robust safety of systems with model uncertainties [10], [11] and model-free safe reinforcement learning [12]. With the growing attention on faults and attacks, safety guarantees on systems under faulty components or adversarial environments have become an active research area.

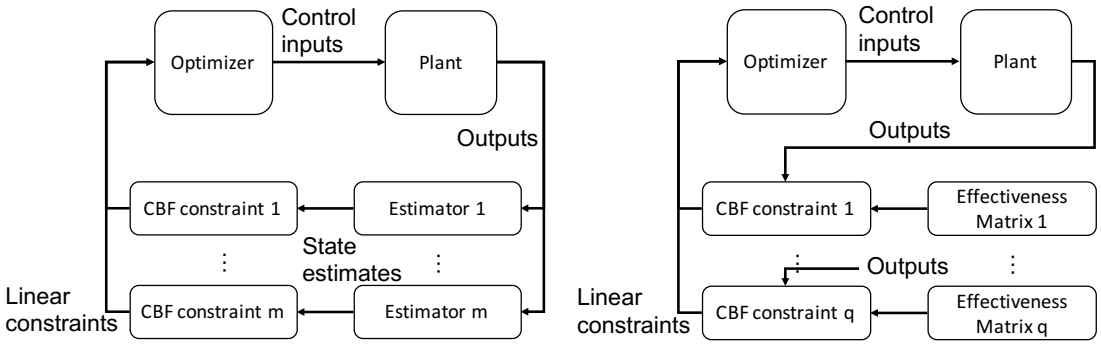
In this paper, we propose safe control algorithms for nonlinear systems under sensor and actuator faults. To ensure safety of nonlinear systems under sensor faults and attacks, we propose a class of CBFs, which is shown in Figure 1(a) and constructed as follows. We maintain a set of state estimators, each omitting a set of sensors associated with one fault pattern and then use CBF constraints to ensure that each of the estimated states remains within the safe region. The intuition of our approach is that one can simply ignore the faulty measurements by omitting the sensors of each fault pattern and construct CBFs for each estimate. However, it may be infeasible to satisfy all CBF constraints using a single control input when faults occur and the state estimates deviate due to the fault. To resolve conflicts, we propose a threshold-based method to exclude outlier estimates and relax the corresponding CBF constraints.

To ensure safety of nonlinear systems under actuator failures, we propose a class of CBFs, which is shown in Figure 1(b) and constructed as follows. We maintain a set of effectiveness matrices according to failure patterns and then use CBFs to ensure the system with each actuator failure remains within the safe region. The basic idea is to find a single control input

H. Zhang, and A. Clark are with the Electrical and Systems Engineering Department, McKelvey School of Engineering, Washington University in St. Louis, St. Louis, MO 63130 USA (e-mail:{hongchao,andrewclark}@wustl.edu)

Z. Li is with Black Sesame Technologies Inc, San Jose, CA 95131 USA (e-mail: lizhouchi@gmail.com).

This work was supported by National Science Foundation grants CNS-1941670, CMMI-2418806, and Air Force Office of Scientific Research grant FA9550-22-1-0054.



(a) Scheme for system under sensor faults and attacks (b) Scheme for system under actuator failures.

Fig. 1: Schematic illustration of our proposed approach. Both schemes ensure safety via CBFs in a passive manner.

that satisfies all CBF constraints constructed for each failure pattern. We propose a sum-of-squares (SOS) based scheme to verify the feasibility of CBFs for sensor faults/attacks and actuator failures, respectively.

We make the following specific contributions:

- We propose High-order Stochastic CBFs (HOSCBF) for the system with high relative degree and propose FT-SCBFs with high order degree to ensure finite time safety when sensor faults occur. We propose an SOS-based scheme to verify the feasibility of constraints of FT-SCBFs with high relative degree.
- We compose HOSCBFs with Control Lyapunov Functions (CLFs) to provide joint guarantees on safety and stability under sensor faults.
- We formulate an HOCBF approach to ensure safety when actuator failures occur and propose an SOS-based scheme to verify the feasibility of CBF constraints throughout the safe region.
- We evaluate our approach via two case studies. The proposed HOSCBF-CLF ensures safety and convergence of a wheeled mobile robot (WMR) system in the presence of a sensor attack. The proposed HOCBF-based method ensures safety of a Boeing 747 lateral control system under actuator failures.

The remainder of this paper is organized as follows. Section II presents the related work. Section III presents background and preliminaries. Section IV proposes a HOSCBF-based control policy for systems under sensor faults and attacks as well as a scheme to verify the feasibility of HOSCBFs. Section V proposes a framework for joint safety and stability via HOSCBF-CLFs. Section VI proposes an HOCBF-based control policy for systems under actuator failures. Section VII presents our case studies. Section VIII concludes the paper.

II. Related Work

Fault-tolerant control systems (FTCS) aim to accommodate faults and maintain stability of the system with little or acceptable degradation in performance. See [8] for an in-depth treatment. FTCS are classified into two main types, namely, active FTCS and passive FTCS [9]. In active FTCS, Fault Detection and Isolation (FDI) plays a significant role and has been studied for decades. See [13] for more details.

Active FTCS against sensor faults and attacks include statistical hypothesis testing for stochastic systems [14], and unknown input observers for deterministic systems [15]. Kalman Filter (KF) and Extended Kalman Filter (EKF) are extensively used in FDI applications such as [16] for MIMO system. More recently, data-driven approaches to fault tolerance have shown promise [17], [18]. While the approach of using Kalman filter residues to identify potential faults is related to our conflict resolution approach, safety of the system under faults and attacks is not addressed. Sliding-mode control, as one of popular PFTCS, is proposed for singularly perturbed systems [19], switched systems [20], fuzzy systems, [21], and Markov Jump Systems [22], [23]. Several of these works aim to guarantee stability in the presence of faults [24], which is related to but distinct from the safety criteria we consider.

Passive FTCS against actuator failures is proposed due to its advantage of fast response. Reliable control for Linear time-invariant (LTI) system under actuator failure is proposed in [25] and implemented for LTI aircraft model in [26]. Actuator failure compensation control (AFCC) has been employed for LTI system [6] and nonlinear systems [27], [28]. Robust adaptive FTC is presented in [29] for linear systems with time-varying parameter uncertainty, external disturbance and actuator faults. However, the aforementioned methods focus on ensuring stability but leave safety guarantees less studied. In this paper, we address the safety of nonlinear systems under sensor faults/attacks or actuator failures.

Safety verification of control systems is an area of extensive research, with popular methods including finite-state approximations [30], HJI equation [3], barrier certificates [4], [31], simulation-driven approaches [32], [33], and counterexample-guided synthesis [34]. Barrier function-based approaches, which formulate the safety constraint as inequality over the control input, have been proposed to guarantee safety [7], [35]–[40]. Among these methods, CBFs were proposed in [5]. CBFs for stochastic systems were investigated in [7]. Higher-order CBFs were presented in [37], [41], [42]. CBFs for safe reinforcement learning were introduced in [43]–[47]. Applications of CBFs to specific domains such as multi-agent systems [48], autonomous vehicles [49], and UAVs [50] have also been considered. Recent research [5], [45], [51], [52] have investigated joint objectives of safety and stability. The work

[53] investigates the resilience of adversarial agents in multi-agent systems with higher-order CBFs. None of these existing works, however, incorporated the effects of faults and attacks on actuators and sensors.

Ensuring safety under sensor faults and attacks has attracted growing research attention. Barrier certificate based fault-tolerant Linear quadratic Gaussian (LQG) tracking is investigated in [54] for LTI system under sensor fault and false data injection attack and generalized to multiple possible compromised sensor sets in [55]. Compared with barrier certificate method, CBF-based approaches have more advantages on flexibility. In the preliminary conference version of this work [56], we investigated fault-tolerant control of nonlinear systems, in which CBF constraints are constructed and imposed to ensure safety of the system under sensor faults and attacks. However, systems under actuator failures and fault-tolerant control via CBF with high relative degree have drawn less attention. Recent work [57] proposed a model-free learning framework for an output-based neural fault-detector to detect actuator faults. However, verifying the feasibility of CBFs under faults has not yet been investigated. In this paper, we propose FT-CBFs with high relative degree for nonlinear systems under sensor faults and attacks and actuator failures. We also propose a systematical approach to verify the feasibility of CBFs.

CBF constructions depend on factors including system dynamics and disturbance, which leave it an open problem to verify whether such constraints can always be satisfied. A systematic approach to verify the feasibility of CBFs is proposed in [58] to enable broader adoption of CBFs. However, feasibility verification of CBFs in faulty or adversarial environments has not yet been studied. In this work, we propose an SOS-based scheme to verify the feasibility of constraints of SCBFs with high relative degree and HOCBFs for the system under sensor faults/attacks and actuator failures, respectively.

III. Preliminaries

In this section, we present the system model and provide background on the EKF and CBFs.

A. System Model

Notations. For a set S, we denote $\text{int}(S)$ and ∂S as the interior and boundary of S, respectively. For any vector v, we let $[v]_i$ denote the i-th element of v. We let $\overline{\lambda}(A)$ denote the magnitude of the largest eigenvalue of matrix A, noting that this is equal to the largest eigenvalue when A is symmetric and positive definite. When the value of A is clear, we write $\overline{\lambda}$.

We consider a nonlinear control system with state $x_t \in \mathbb{R}^n$ and input $u_t \in \mathbb{R}^p$ at time t. The state dynamics and the system output $y_t \in \mathbb{R}^q$ are described by the stochastic differential equations

$$dx_t = (f(x_t) + g(x_t)u_t)\ dt + \sigma_t\ dW_t \tag{1}$$

$$dy_t = cx_t\ dt + \nu_t\ dV_t \tag{2}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n\times p}$ are locally Lipschitz, $\sigma_t \in \mathbb{R}^{n\times n}$, W_t is an n-dimensional Brownian motion, $c \in \mathbb{R}^{q\times n}$, $\nu_t \in \mathbb{R}^{q\times q}$, and V_t is a q-dimensional Brownian motion.

The safety conditions of a system are specified in terms of forward invariance of a pre-defined safe region. We define the safe region as follows.

Definition 1 (Safe Region): The safe region of the system is a set $\mathcal{C} \subseteq \mathbb{R}^n$ defined by

$$\mathcal{C} = \{x : h(x) \geq 0\}, \quad \partial\mathcal{C} = \{x : h(x) = 0\} \tag{3}$$

where $h : \mathbb{R}^n \to \mathbb{R}$ is twice-differentiable on $\mathcal{C}$.

We assume throughout the paper that $x_0 \in \text{int}(\mathcal{C})$, i.e., the system is initially safe. Let $\overline{f}(x,u) = f(x) + g(x)u$. The uniform detectability property is defined as follows.

Definition 2 (Uniform Detectability): The pair $[\frac{\partial \overline{f}}{\partial x}(x,u), c]$ is uniformly detectable if there exists a bounded, matrix-valued function $\Theta(x)$ and a real number $\eta > 0$ such that

$$w^T \left(\frac{\partial \overline{f}}{\partial x}(x,u) + \Theta(x)c\right) w \leq -\eta||w||^2$$

for all w, u, and x.

B. Background and Preliminary Results

Let $\hat{x}_t$ denote the EKF estimate of x_t. The EKF for the system described by (1) and (2) is defined by

$$d\hat{x}_t = (f(\hat{x}_t) + g(\hat{x}_t)u_t)dt + K_t(dy_t - c\hat{x}_t),$$

where $K_t = P_t c^T R_t^{-1}$ and $R_t = \nu_t\nu_t^T$. The matrix P_t is the positive-definite solution to

$$\frac{dP}{dt} = F_tP_t + P_tF_t^T + Q_t - P_tc^TR_t^{-1}cP_t$$

where $Q_t = \sigma_t\sigma_t^T$ and $F_t = \frac{\partial \overline{f}}{\partial x}(\hat{x}_t, u_t)$. To utilize EKF, we make the following assumptions.

Assumption 1: The SDEs (1) and (2) satisfy the conditions:

1) There exist constants β_1 and β_2 such that $\mathbf{E}(\sigma_t\sigma_t^T) \geq \beta_1 I$ and $\mathbf{E}(\nu_t\nu_t^T) \geq \beta_2 I$ for all t.
2) The pair $[\frac{\partial \overline{f}}{\partial x}(x,u), c]$ is uniformly detectable.
3) Let ϕ be defined by

$$\overline{f}(x,u) - \overline{f}(\hat{x},u) = \frac{\partial \overline{f}}{\partial x}(x - \hat{x}) + \phi(x,\hat{x},u).$$

Then there exist real numbers k_ϕ and ϵ_ϕ such that

$$||\phi(x,\hat{x},u)|| \leq k_\phi||x - \hat{x}||_2^2$$

for all x and $\hat{x}$ satisfying $||x - \hat{x}||_2 \leq \epsilon_\phi$.

One can obtain k_ϕ by considering a compact subset $\kappa \subseteq \mathbb{R}$. For instance, given a function f that is twice differentiable with respect to x in subset κ, we have $k_\phi = \max_{1\leq i\leq n} \sup_{x\in\kappa} \|\nabla^2 f_i(x,u)\|$, where $\nabla^2 f$ is the Hessian matrix. The bounds on estimation error ϵ_ϕ can be calculated via an integral formula [59, Chapter 20]. More details can be found in [60]. The following result describes the accuracy of the EKF.

Theorem 1 ([60]): Suppose that the conditions of Assumption 1 hold. Then there exists $\delta > 0$ such that $\sigma_t\sigma_t^T \leq \delta I$ and $\nu_t\nu_t^T \leq \delta I$. For any $0 < \epsilon < 1$, there exists $\gamma > 0$ such that

$$Pr\left(\sup_{t\geq 0} ||x_t - \hat{x}_t||_2 \leq \gamma\right) \geq 1 - \epsilon.$$

We next provide background and preliminary results on stochastic CBFs. The following theorem provides sufficient conditions for safety of a stochastic system.

Proposition 1 ([61, Proposition III.5]): Given a finite time T, suppose the mapping h is a continuous function with linear function kx as the class-κ function, where $k \geq 0$. Let the control input u_t be chosen to satisfy

$$\frac{\partial h}{\partial x}(f(x_t)+g(x_t)u_t)+\frac{1}{2}tr\left(\sigma_t^T\frac{\partial^2 h}{\partial x^2}(x_t)\sigma_t\right)\geq -kh(x_t). \tag{4}$$

Let $\zeta=\sup_{x\in\mathcal{C}}h(x)$ and $x_0\in int(C)$. Then we have

$$Pr(x_t\in int(\mathcal{C}),0\leq t\leq T)\geq(\frac{h(x_0)}{\zeta})e^{-\zeta T}.$$

Theorem 2: For a system (1)–(2) with safety region defined by (3), define

$$\overline{h}_\gamma=\sup_{x,x^0}\{h(x):||x-x^0||_2\leq\gamma \text{ and } h(x^0)=0\},$$

where $\hat{h}(x):=h(x)-\overline{h}_\gamma$. Let $\frac{\partial h}{\partial x}$ denote the derivative $\frac{\partial h}{\partial x}|_{x=\hat{x}}$ for simplicity. Let $z_t\in\mathbb{R}^n$ represent the estimation error, where $z_t=(x_t-\hat{x}_t)$. Suppose that u_t is chosen to satisfy

$$\frac{\partial h}{\partial x}(f(\hat{x}_t)+g(\hat{x}_t)u_t)-\left\|\frac{\partial h}{\partial x}K_tc\right\|_2\gamma + \frac{1}{2}tr\left(\nu_t^TK_t^T\frac{\partial^2 h}{\partial x^2}(\hat{x}_t)K_t\nu_t\right)\geq-\hat{h}(\hat{x}_t). \tag{5}$$

Then $Pr(x_t\in\mathcal{C},0\leq t\leq T|\ ||x_t-\hat{x}_t||_2\leq\gamma)\geq(\frac{\hat{h}(x_0)}{\zeta})e^{-\zeta T}$.

Proof: We have the estimate $\hat{x}$ yields

$$\begin{aligned}d\hat{x}_t&=\bar{f}\left(\hat{x}_t,u_t\right)dt+K_t\left(cx_tdt+\nu_tdV_t-c\hat{x}_tdt\right)\\&=\left(\bar{f}\left(\hat{x}_t,u_t\right)+K_tc\left(x_t-\hat{x}_t\right)\right)dt+K_t\nu_tdV_t\end{aligned}$$

Given $||x_t-\hat{x}_t||_2\leq\gamma$, we have

$$\frac{\partial\hat{h}}{\partial x}K_tc\left(x_t-\hat{x}_t\right)\geq-\left\|\frac{\partial\hat{h}}{\partial x}K_tc\right\|_2\|z_t\|_2\geq-\left\|\frac{\partial\hat{h}}{\partial x}K_tc\right\|_2\gamma$$

We then choose u_t to satisfy (5). Then, we have

$$\begin{aligned}&\frac{\partial\hat{h}}{\partial x}\left(f\left(\hat{x}_t\right)+g\left(\hat{x}_t\right)u_t+K_tc\left(x_t-\hat{x}_t\right)\right)+\\&\qquad\frac{1}{2}\operatorname{tr}\left(\nu_t^TK_t^T\left(\frac{\partial^2\hat{h}}{\partial x^2}\right)K_t\nu_t\right)+\hat{h}(\hat{x}_t)\geq\\&\frac{\partial h}{\partial x}(f(\hat{x}_t)+g(\hat{x}_t)u_t)-\left\|\frac{\partial h}{\partial x}K_tc\right\|_2\gamma+\\&\qquad\frac{1}{2}tr\left(\nu_t^TK_t^T\frac{\partial^2 h}{\partial x^2}(\hat{x}_t)K_t\nu_t\right)+\hat{h}(\hat{x}_t)\geq0\end{aligned}$$

Hence, by Proposition 1, we have $Pr(x_t\in\mathcal{C},0\leq t\leq T|\ ||x_t-\hat{x}_t||_2\leq\gamma)\geq(\frac{\hat{h}(x_0)}{\zeta})e^{-\zeta T}$. ∎

Intuitively, Eq. (5) implies that as the state approaches the boundary, the control input is chosen such that the rate of increase of the barrier function decreases to zero. Hence Theorem 2 implies that if there exists an SCBF for a system, then the safety condition is satisfied with probability greater or equal to $(1-\epsilon)(\frac{\hat{h}(x_0)}{\zeta})e^{-\zeta T}$ when an EKF is used as an estimator and the control input is chosen at each time t to satisfy (5). We next present a probabilistic guarantee for HOCBFs within a finite time horizon.

Definition 3 (SCBF): The function h is a *Stochastic Control Barrier Function (SCBF)* of the system, if for all $\hat{x}_t$ satisfying $h(\hat{x}_t)>0$ there exists u_t s.t. $\forall z_t$ with $\|z_t\|\leq\gamma$ (5) is satisfied.

There may not exist a value of u_t to satisfy (5) when $\frac{\partial h}{\partial x}g(x)=0$. To address this problem, CBF with high relative degree has been proposed in deterministic [41] and stochastic [7] settings.

Let $d=0,1,\ldots$ and define d^{th} order differentiable function $h^d(x)$ as

$$\begin{aligned}h^0(x)&=h(x),\\h^1(x)&=\frac{\partial h^0}{\partial x}\overline{f}(x,u)+\frac{1}{2}\operatorname{tr}\left(\sigma^T\left(\frac{\partial^2h^0}{\partial x^2}\right)\sigma\right)+h^0(x),\\&\vdots\\h^{d+1}(x)&=\frac{\partial h^d}{\partial x}\overline{f}(x,u)+\frac{1}{2}\operatorname{tr}\left(\sigma^T\left(\frac{\partial^2h^d}{\partial x^2}\right)\sigma\right)+h^d(x).\end{aligned}$$

Define $\mathcal{C}^d=\left\{x:h^d(x)\geq0\right\}$. The following theorem provides sufficient conditions for safety of a high-degree system.

Theorem 3: Let $\overline{\mathcal{C}}=\bigcap_{d=0}^{d'}\mathcal{C}^d$. Suppose that there exists d' such that $\frac{\partial h^{d'}}{\partial x}g(x)u\neq0$. If u_t is chosen to satisfy

$$\frac{\partial h^{d'}}{\partial x}(\overline{f}(x,u))+\frac{1}{2}\operatorname{tr}\left(\sigma^T\frac{\partial^2h^{d'}}{\partial x^2}\sigma\right)\geq-h^{d'}(x). \tag{6}$$

Let $\zeta^d=\sup_{x\in\mathcal{C}^d}h^d(x)$. Then $Pr(x_t\in\mathcal{C},0\leq t\leq T)\geq\prod_{d=0}^{d'}(\frac{h^d(x_0)}{\zeta^d})e^{-\zeta^dT}$ if $x_0\in\overline{\mathcal{C}}$.

Proof: Suppose that u_t satisfying the conditions of the theorem is chosen at each time t. Theorem 2 implies that $h^{d'}(x_t)\geq0$ when $0\leq t\leq T$ with probability greater or equal to $(\frac{h^{d'}(x_0)}{\zeta^{d'}})e^{-\zeta^{d'}T}$. By definition of $h^d(x)$ we have that $\frac{\partial h^{d'-1}}{\partial x}g(x)u=0$. We also have $Pr(x_t\in\mathcal{C}^{d'-1},0\leq t\leq T|x_t\in\mathcal{C}^{d'})\geq(\frac{h^{d'-1}(x_0)}{\zeta^{d'-1}})e^{-\zeta^{d'-1}T}$. Proceeding inductively, we have $Pr(x_t\in\mathcal{C},0\leq t\leq T)\geq P$ where

$$P=\prod_{i=1}^{d'}Pr(x_t\in\mathcal{C}^{i-1},0\leq t\leq T\ |x_t\in\mathcal{C}^i)\cdot Pr(x_t\in\mathcal{C}^{d'},0\leq t\leq T).$$

Finally, we have $Pr(x_t\in\mathcal{C},0\leq t\leq T)\geq\prod_{d=0}^{d'}(\frac{h^d(x_0)}{\zeta^d})e^{-\zeta^dT}$. ∎

Definition 4: The function $h^{d'}$ is a high-order CBF (HOCBF) of relative degree d' for system (1)-(2) if for all $x\in\overline{\mathcal{C}}$ there exists u satisfying (6).

The following theorem provides equivalent conditions for the existence of solutions to systems of polynomial equations and inequalities. This allows us to verify sufficient and necessary conditions for safety.

Theorem 4 (Positivstellensatz [62]): Let $(\phi_j)_{j=1,\ldots,s}$, $(\chi_k)_{k=1,\ldots,t}$, $(\psi_\ell)_{\ell=1,\ldots,u}$ be finite families of polynomials in $\mathbb{R}[x_1,\ldots,x_n]$. Denote by C the cone generated by

$(\phi_j)_{j=1,\dots,s}$, M the multiplicative monoid generated by $(\chi_k)_{k=1,\dots,t}$, and D the ideal generated by $(\psi_\ell)_{\ell=1,\dots,u}$. Then, the following properties are equivalent:

1) The set
$$\left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \phi_j(x) \geq 0, & j = 1,\dots,s \\ \chi_k(x) \neq 0, & k = 1,\dots,t \\ \psi_\ell(x) = 0, & l = 1,\dots,u \end{array} \right\}$$
is empty.
2) There exist $\phi \in C, \chi \in M, \psi \in D$ such that $\phi + \chi^2 + \psi = 0$.

Lemma 1 (Farkas' Lemma [63]): Let A be a real matrix with m rows and n columns, and let $b \in \mathbb{R}^n_u$ be a vector. One of the following conditions holds. i) The system of inequalities $Ax \leq b$ has a solution. ii) There exists y such that $y \geq \mathbf{0}$, $y^T A = \mathbf{0}^T$ and $y^T b < 0$, where $\mathbf{0}$ denote a zero vector.

IV. Safe Control Under Sensor Faults and Attacks

In this section, we consider the system under sensor faults and derive safety guarantees.

We consider a nonlinear control system whose output may be affected by one of m sensor faults. The set of possible faults is indexed as $\{r_1, \dots, r_m\}$. Each fault r_i maps to a set of affected observations $\mathcal{F}(r_i) \subseteq \{1, \dots, q\}$. We assume that $\mathcal{F}(r_i) \cap \mathcal{F}(r_j) = \emptyset$ for $i \neq j$. Let $r \in \{r_1, \dots, r_m\}$ denote the index of the fault experienced by the system. The system dynamics are described as

$$dx_t = (f(x_t) + g(x_t)u_t)\, dt + \sigma_t\, dW_t \quad (7)$$
$$dy_t = (cx_t + a_t)\, dt + \nu_t\, dV_t, \quad (8)$$

where the vector $a_t \in \mathbb{R}^q$ represents the impact of the fault and is constrained by $\text{supp}(a_t) \subseteq \mathcal{F}(r)$. Hence, if fault r_i occurs, then the outputs of any of the sensors indexed in $\mathcal{F}(r_i)$ can be arbitrarily modified by the fault. The sets $\mathcal{F}(r_1), \dots, \mathcal{F}(r_m)$ are known, but the value of r_i is unknown. In other words, the set of possible faults is known, but the exact fault that has occurred is unknown to the controller.

To illustrate, we take an autonomous vehicle system as an example. Consider an autonomous system equipped with two INS sensors, a GNSS and one LiDAR system indexed as $\{1, 2, 3, 4\}$ for localization measurements denoted as $y = \{y_1, y_2, y_3, y_4\}$. We consider three possible attacks: an attack on one of the INS sensors, an attack on another INS sensor, or a simultaneous GPS/LiDAR spoofing attack. The corresponding fault patterns are given as $\mathcal{F}(r_1) = \{1\}$, $\mathcal{F}(r_2) = \{2\}$, $\mathcal{F}(r_3) = \{3, 4\}$.

Define $\bar{c}_i$ to be the c matrix with the corresponding rows indexed in $\mathcal{F}(r_i)$ removed, $\bar{y}_{t,i}$ to be equal to the vector y_t with the entries indexed in $\mathcal{F}(r_i)$ removed, and $\bar{\nu}_{t,i}$ to be the matrix ν_t with rows and columns indexed in $\mathcal{F}(r_i)$ removed. Define $\bar{c}_{i,j}$ to be the c matrix with the corresponding rows indexed in $\mathcal{F}(r_i)$ and $\mathcal{F}(r_j)$ removed, where $i \neq j$.

We make the following assumption for the sensor fault scenario.

Assumption 2: The system (7)-(8) and the sensor fault patterns $\mathcal{F}(r_1), \dots, \mathcal{F}(r_m)$ satisfy the conditions:

1) The system is controllable.
2) For each $i, j \in \{1, \dots, m\}$, the pair $[\frac{\partial \bar{f}}{\partial x}(x, u), \bar{c}_{i,j}]$ is uniformly detectable.

Problem Statement: Given a finite time T, a safe set $\mathcal{C}$ defined in (3) and a parameter $\epsilon \in (0, 1)$, compute a control policy that, at each time t, maps the sequence $\{y_{t'} : t' \in [0, t)\}$ to an input u_t such that, for any fault $r \in \{r_1, \dots, r_m\}$, $Pr(x_t \in \mathcal{C},\ 0 < t < T) \geq (1 - \epsilon)\mathcal{T}(T)$, for some function $\mathcal{T} : \mathbb{R}^n \to (0, 1)$.

A. Sensor FTC Strategy Definition

We propose a CBF-based strategy with safety guarantees for a system satisfying Assumption 2. The strategy that accommodates sensor faults and attacks is an FTC in a passive manner. The goal of FTC is to ensure the robustness of the control system to accommodate multiple component faults without striving for optimal performance for any specific fault condition.

The intuition behind our approach is as follows. Since we do not know the fault pattern r, we construct estimators excluding faulty sensors by maintaining m EKFs. Each EKF corresponds to a different possible fault pattern in $\{r_1, \dots r_m\}$. We ensure safety with desired probability by defining m corresponding SCBFs, each of which results in a different linear constraint on the control input.

The potential drawback is that the safety guarantees of Theorem 2 rely on the existence of a control input satisfying the safety constraint at each time-step. This assumption may not hold for two reasons. Firstly, feasible control input u may not exist when $\frac{\partial h}{\partial x} g(x) = 0$, since u does not affect states x. Secondly, a feasible solution may not exist when faulty sensor measurements cause the state estimates to diverge. To address the first reason, we define higher-order SCBFs such that for d-th degree $\frac{\partial h^d}{\partial x} g(x) \neq 0$. Then, we choose control input u to satisfy constraints constructed by higher-order SCBFs. To address the second reason, we define a set of $\binom{m}{2}$ EKFs to resolve conflicts between the constraints. Each EKF estimator omits all sensors affected by either fault r_i or fault r_j for some $i, j \in \{1, \dots, m\}$, $i \neq j$. These estimators will be used to resolve any deviations between the state estimates from sensors $\{1, ..., m\} \backslash \mathcal{F}(r_i)$ and $\{1, ..., m\} \backslash \mathcal{F}(r_j)$.

Let $\mathcal{C}_\gamma := \{x : \hat{h}^d(x) \geq 0\}$ where $\hat{h}^d(x) = h^d(x) - \bar{h}^d_\gamma$ and

$$\bar{h}^d_\gamma = \sup_{x, x^{d,0}} \left\{ h^d(x) : \|x - x^{d,0}\|_2 \leq \gamma \text{ and } h^d(x^{d,0}) = 0 \right\} \quad (9)$$

Let $\overline{\mathcal{C}}_\gamma = \bigcap_{d=0}^{d'} \mathcal{C}^d_\gamma$. To ensure safety as defined in (3), we need to show that $Pr(x_t \in \mathcal{C},\ 0 \leq t \leq T) \geq \prod_{d=0}^{d'} (\frac{\hat{h}^d(x_0)}{\zeta^d}) e^{-\zeta^d T}$ if $x_0 \in \bar{\mathcal{C}}_\gamma$ given $x_0 \in \bar{\mathcal{C}}_\gamma$ and $||x_t - \hat{x}_t||_2 \leq \gamma,\ \forall t$.

Proposition 2: For a system (7)-(8) with safety region defined by (3), suppose there exists d', such that $\frac{\partial h^{d'}}{\partial x} g(x) \neq 0$. Suppose that u_t is chosen to satisfy

$$\frac{\partial h^{d'}}{\partial x} \bar{f}(\hat{x}_t, u_t) - \left\| \frac{\partial h^{d'}}{\partial x}(\hat{x}_t) K_t c \right\|_2 \gamma + \frac{1}{2} tr\left(\nu_t^T K_t^T \frac{\partial^2 h^{d'}}{\partial x}(\hat{x}_t) K_t \nu_t \right) \geq -\hat{h}^{d'}(\hat{x}_t). \quad (10)$$

Then $Pr(x_t \in \mathcal{C},\ 0 \le t \le T|\ ||x_t - \hat{x}_t||_2 \le \gamma\ \forall t) \ge \prod_{d=0}^{d'}(\frac{\hat{h}^d(x_0)}{\zeta^d})e^{-\zeta^d T}$ if $x_0 \in \bar{\mathcal{C}}_\gamma$.

Proof: Given $||x_t - \hat{x}_t||_2 \le \gamma\ \forall t$, we suppose that u_t is chosen to satisfy (10). By Theorem 2 and (10), we have $h^{d'}(x_t) \ge 0$ when $0 \le t \le T$ with probability greater or equal to $(\frac{\hat{h}^{d'}(x_0)}{\zeta^{d'}})e^{-\zeta^{d'}T}$. By definition of relative degree, we have $\frac{\partial h^d}{\partial x}g(x)u = 0$ for $d < d'$. By definition of $h^{d'}(x_t)$, we have $\frac{\partial h^d}{\partial x}\overline{f}(x,0) + \frac{1}{2}\operatorname{tr}\left(\sigma^T\left(\frac{\partial^2 h^d}{\partial x^2}\right)\sigma\right) + h^d(x) \ge 0$, where $d = d'-1$. This implies $h^{d'-1}(x_t) \ge 0$. Similar to the proof of Theorem 3, by proceeding inductively, we then have $Pr(x_t \in \mathcal{C},\ 0 \le t \le T|\ ||x_t - \hat{x}_t||_2 \le \gamma\ \forall t) \ge \prod_{d=0}^{d'}(\frac{\hat{h}^d(x_0)}{\zeta^d})e^{-\zeta^d T}$ if $x_0 \in \bar{\mathcal{C}}_\gamma$. ∎

The function $h^{d'}$ ensures safety of the system with relative degree d' by Proposition 2. Hence we define the function as follows.

Definition 5: The function $h^{d'}$ is a *higher order SCBF (HOSCBF)* of relative degree d' for system (1)-(2) if for all $\hat{x}_t \in \overline{\mathcal{C}}$ there exists u_t satisfying (10).

We next present a scheme to resolve conflicts between constraints in the case of faults and attacks. Let $\overline{R}_{t,i} = \overline{\nu}_{t,i}\overline{\nu}_{t,i}^T$ and $K_{t,i} = \overline{P}_{t,i}\bar{c}_i^T(\overline{R}_{t,i})^{-1}$. Here $\overline{P}_{t,i}$ is the solution to the Riccati differential equation

$$\frac{d\overline{P}_{t,i}}{dt} = F_{t,i}\overline{P}_{t,i} + \overline{P}_{t,i}F_{t,i}^T + Q_t - \overline{P}_{t,i}\bar{c}_i^T\overline{R}_{t,i}^{-1}\bar{c}_i\overline{P}_{t,i}$$

with $F_{t,i} = \frac{\partial \overline{f}}{\partial x}(\hat{x}_{t,i}, u_t)$. Define a set of m EKFs with estimates denoted $\hat{x}_{t,i}$ via

$$d\hat{x}_{t,i} = (f(\hat{x}_{t,i}) + g(\hat{x}_{t,i})u_t)\ dt + K_{t,i}(d\overline{y}_{t,i} - \bar{c}_i\hat{x}_{t,i}\ dt). \quad (11)$$

Each of these EKFs represents the estimate obtained by removing the sensors affected by fault r_i. Furthermore, define $\overline{y}_{t,i,j}$, $\overline{\nu}_{t,i,j}$, $\bar{c}_{i,j}$, $\overline{R}_{t,i,j}$, and $K_{t,i,j}$ in an analogous fashion with entries indexed in $\mathcal{F}(r_i) \cup \mathcal{F}(r_j)$ removed. We assume throughout that the $\overline{R}$ matrices are invertible. We then define a set of $\binom{m}{2}$ estimators $\hat{x}_{t,i,j}$ as

$$d\hat{x}_{t,i,j} = (f(\hat{x}_{t,i,j}) + g(\hat{x}_{t,i,j})u_t)\ dt + K_{t,i,j}(d\overline{y}_{t,i,j} - \bar{c}_{i,j}\hat{x}_{t,i,j}\ dt). \quad (12)$$

When $\mathcal{F}(r_i) \cup \mathcal{F}(r_j) = \{1, \dots, q\}$, the open-loop estimator is used for $\hat{x}_{t,i,j}$.

We then select parameters $\gamma_1, \dots, \gamma_m \in \mathbb{R}_+$, and $\{\theta_{ij} : i < j\} \subseteq \mathbb{R}_+$. The set of feasible control inputs is defined at each time t using the following steps:

1) Define $Z_t = \{1, \dots, m\}$. Define a collection of sets Ω_i, $i \in Z_t$, by

$$\Omega_i \triangleq \left\{u : \frac{\partial h_i^{d'}}{\partial x}\left(\bar{f}(\hat{x}_{t,i}, u_t)\right) - \left\|\frac{\partial h_i^{d'}}{\partial x}(\hat{x}_t)K_t c\right\|_2 \gamma_i + \frac{1}{2}tr\left(\nu_t^T K_t^T \frac{\partial^2 h_i^{d'}}{\partial x}(\hat{x}_{t,i})K_t\nu_t\right) \ge -\hat{h}_i^{d'}(\hat{x}_{t,i})\right\} \quad (13)$$

Select u_t satisfying $u_t \in \bigcap_{i \in Z_t}\Omega_i$. If no such u_t exists, there exists conflicts between constraints, i.e., $\exists i, j,\ i \ne j$ s.t. $\Omega_i \cap \Omega_j = \emptyset$. Then go to Step 2.

2) For each i, j with $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 > \theta_{ij}$, set $Z_t = Z_t \setminus \{i\}$ (resp. $Z_t = Z_t \setminus \{j\}$) if $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 > \theta_{ij}/2$ (resp. $||\hat{x}_{t,j} - \hat{x}_{t,i,j}||_2 > \theta_{ij}/2$). If $\bigcap_{i \in Z_t}\Omega_i \ne \emptyset$, then select $u_t \in \bigcap_{i \in Z_t}\Omega_i$. Else, go to Step 3. This step resolves conflicts between estimations by comparing the difference between estimations against thresholds θ_{ij}.
3) Remove the indices i from Z_t corresponding to the estimators with the largest residue values $\overline{y}_{t,i} - \bar{c}_i\hat{x}_{t,i}$ until there exists $u_t \in \bigcap_{i \in Z_t}\Omega_i$.

We next provide sufficient conditions for this control policy to guarantee safety.

Theorem 5: Given $x_0 \in \bar{\mathcal{C}}_\gamma$, define

$$\overline{h}_{\gamma_i}^d = \sup_{x, x^0}\left\{h^d(x) : ||x - x^{d,0}||_2 \le \gamma_i \ \text{ and } h^d(x^{d,0}) = 0\right\}$$

and $\hat{h}_i^d(x) = h^d(x) - \overline{h}_{\gamma_i}^d$. Suppose $\gamma_1, \dots, \gamma_m$, and θ_{ij} for $i < j$ are chosen such that the following conditions are satisfied:

1) Define $\Lambda_i^d(\hat{x}_{t,i}) = \frac{\partial h_i^d}{\partial x}(\hat{x}_{t,i})g(\hat{x}_{t,i})$. For all $i, j \in Z_t$ with $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 \le \theta_{ij}$, there exists u such that
$$\Lambda_i^{d'}(\hat{x}_{t,i})u > 0 \quad (14)$$
for all $i \in Z_t'$.
2) For each i, when $r = r_i$,
$$Pr(||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \le \frac{\theta_{ij}}{2}\ \forall j, ||\hat{x}_{t,i} - x_t||_2 \le \gamma_i\ \forall t) \ge 1 - \epsilon. \quad (15)$$

Then $Pr(x_t \in \mathcal{C},\ 0 \le t \le T) \ge (1-\epsilon)\prod_{d=0}^{d'}(\frac{\hat{h}^d(x_0)}{\zeta^d})e^{-\zeta^d T}$ for any fault pattern $r \in \{r_1, \dots, r_m\}$.

Proof: Suppose that the fault $f = f_i$. We will show that, if $||\hat{x}_{t,i} - x_t||_2 \le \gamma_i$ and $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \le \theta_{ij}/2$ for all t, then $u_t \in \Omega_i$ holds. Hence $x_t \in \mathcal{C}$ for $0 \le t \le T$ with probability greater or equal to $\prod_{d=0}^{d'}(\frac{\hat{h}^d(x_0)}{\zeta^d})e^{-\zeta^d T}$ by Proposition 2.

At time t, suppose that $\hat{h}_i^{d'}(\hat{x}_{t,i}) \ge 0$, and that $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \le \theta_{ij}/2$. We consider three cases, namely (i) $||\hat{x}_{t,j} - \hat{x}_{t,k}||_2 \le \theta_{jk}$ for all $j, k \in Z_t$, (ii) $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 \le \theta_{ij}$ for all $j \in Z_t$, but there exist $j, k \in Z_t \setminus \{i\}$ such that $||\hat{x}_{t,j} - \hat{x}_{t,k}||_2 > \theta_{jk}$, and (iii) $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 > \theta_{ij}$ for some $j \in Z_t$.

Case (i): We will show that there exists $u \in \cap_{j \in Z_t}\Omega_j$, and hence in particular u_t satisfies Ω_i. Each Ω_j can be written in the form

$$\Omega_j = \{u : \Lambda_j^{d'}(\hat{x}_{t,j})u_t \ge \overline{\omega}_j^{d'}\} \quad (16)$$

where $\overline{\omega}_j^{d'}$ is a real number that does not depend on u_t. Under the assumption 1) of the theorem, there exists u satisfying (14) for all $i \in Z_t'$. Choose

$$u_t = \left(\max_j\{|\overline{\omega}_j^{d'}|\}/||u||_2\right)u.$$

This choice of u_t satisfies $u_t \in \bigcap_{j \in Z_t}\Omega_j$, in particular $u_t \in \Omega_i$.

Case (ii): In this case, Step 2 of the procedure is reached and constraints Ω_j are removed until all indices in Z_t satisfy $||\hat{x}_{t,j} - \hat{x}_{t,k}||_2 \le \theta_{jk}$. Since $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 \le \theta_{ij}$ already holds for all $j \in Z_t$, i will not be removed from Z_t during this step. After Step 2 is complete, the analysis of Case (i) holds and

there exists a u which satisfies all the remaining constraints, including Ω_i.

Case (iii): Suppose j satisfies $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 > \theta_{ij}$. We have

$$\begin{aligned} \theta_{ij} &< ||\hat{x}_{t,i} - \hat{x}_{t,i,j} + \hat{x}_{t,i,j} - \hat{x}_{t,j}||_2 \\ &\leq ||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 + ||\hat{x}_{t,i,j} - \hat{x}_{t,j}||_2 \quad (17) \\ &\leq \theta_{ij}/2 + ||\hat{x}_{t,i,j} - \hat{x}_{t,j}||_2 \quad (18) \end{aligned}$$

where Eq. (17) follows from the triangle inequality and (18) follows from the assumption that $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \leq \theta_{ij}/2$. Hence $||\hat{x}_{t,j} - \hat{x}_{t,i,j}||_2 > \theta_{ij}/2$ and j is removed from Z_t. By applying this argument to all such indices j, we have that i is not removed during Step 2 of the procedure, and thus the analyses of Cases (i) and (ii) imply that $u_t \in \Omega_i$.

From these cases, we have that Ω_i holds whenever $\hat{h}_i^d(\hat{x}_{t,i}) \geq 0$. Therefore, by Proposition 2, we have

$$Pr(x_t \in \mathcal{C},\ 0 \leq t \leq T|\ ||\hat{x}_{t,i} - \hat{x}_t||_2 \leq \gamma_i, \\ ||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 \leq \theta_{ij}/2\ \forall t) \geq \prod_{d=0}^{d'} (\frac{\hat{h}^d(x_0)}{\zeta^d}) e^{-\zeta^d T}$$

and $Pr(x_t \in \mathcal{C},\ 0 \leq t \leq T) \geq (1-\epsilon)\prod_{d=0}^{d'}(\frac{\hat{h}^d(x_0)}{\zeta^d})e^{-\zeta^d T}$ by (15). ■

The bank of functions in Proposition 5 ensures the safety of the system with faulty components. Hence we define the functions as follows.

Definition 6: The bank of functions $h_1^{d'}, \ldots, h_m^{d'}$ are *Fault-Tolerant High Order Stochastic Control Barrier Functions (FT-HOSCBFs)* of relative degree d' for system (1)-(2) if conditions in Theorem 5 are satisfied.

B. Feasibility Verification

In order for Theorem 5 to guarantee system safety, the linear constraint (14) must hold for all time t. In what follows, we develop an SOS-based scheme to verify the feasibility of SCBF, FT-SCBF and FT-HOCBF constraints for both fault-free case and the case with sensor faults and attacks.

We focus on verification for a constant-gain Kalman filter. In the case where the system is LTI with constant noise, the steady-state Kalman filter gain is optimal and hence satisfies the stochastic stability criteria by Theorem 1. In this subsection, we omit the time subscript of x_t, $\hat{x}_t$ and z_t, i.e., (x, $\hat{x}$ and z) to simplify the expression. We consider an LTI system described by (1) and (2), where $f(x) = F$, $g(x) = G$, the matrices $R_t = R$ and $Q_t = Q$. For an LTI system, P_t is the covariance matrix for the estimation error and will converge to a steady-state value P. The Kalman filter has a constant gain given by $K = Pc^T R^{-1}$. We introduce an SOS-based approach to verify the feasibility for this case.

1) Verification for SCBF: We first present the verification for an SCBF in an attack-free scenario, in which one SCBF-based safety constraint must be satisfied. We have the following initial result.

Proposition 3: Suppose Assumption 1 holds. The function $h(\hat{x})$ is a SCBF if and only if there is no $\hat{x} \in \mathcal{C}_\gamma$, $z \in \mathbb{R}^n$ satisfying $\frac{\partial h}{\partial x} g(\hat{x}) = 0$, $z^T z - \gamma^2 \leq 0$ and $\xi(\hat{x}) < 0$ where

$$\xi(\hat{x}) = \frac{\partial h}{\partial x} f(\hat{x}) + \frac{1}{2}\mathbf{tr}\left(\nu^T K^T \frac{\partial^2 h}{\partial x^2}(\hat{x}) K\nu\right) - \left\|\frac{\partial h}{\partial x}(\hat{x}) Kc\right\|_2 \gamma + \hat{h}(\hat{x}). \quad (19)$$

Proof: By Theorem 2, the set $\mathcal{C}_\gamma$ is positive invariant given $\|x - \hat{x}\| \leq \gamma$ if for all time t u_t is chosen to satisfy (5) $\forall z_t$ with $\|z_t\| \leq \gamma$. By Definition 3, we have that $h(\hat{x})$ is a SCBF if and only if (5) holds for all $\hat{x} \in \mathcal{C}_\gamma := \{x : \hat{h}(x) \geq 0\}$. If $\frac{\partial h}{\partial x} g(\hat{x}) \neq 0$, we can choose u s.t.

$$\frac{\partial h}{\partial x} g(\hat{x}) u \geq \sup_{\|z\| \leq \gamma} \left\{ -\frac{\partial h}{\partial x} f(\hat{x}) - \frac{1}{2}\mathbf{tr}\left(\nu^T K^T \frac{\partial^2 h}{\partial x^2}(\hat{x}) K\nu\right) + \left\|\frac{\partial h}{\partial x}(\hat{x}) Kc\right\|_2 \gamma - \hat{h}(\hat{x}) \right\}.$$

Since $\|z\| \leq \gamma$ is a compact set, such a u always exists. Hence, (4) fails if and only if $\exists \hat{x}$ and z with $\|z\| \leq \gamma$ s.t. (i) $\frac{\partial h}{\partial x} g(\hat{x}) = 0$, and (ii) $\xi(\hat{x}) < 0$ hold simultaneously. ■

Based on the proposition, we can formulate the following conditions via the Positivstellensatz.

Lemma 2: A polynomial $h(\hat{x})$ is an SCBF for system (1)–(2) if and only if there exist polynomials $\rho(\hat{x}, z)$, sum-of-squares polynomials $q_S(\hat{x}, z)$, integers r_1 such that

$$\phi(\hat{x}, z) + \chi(\hat{x}, z) + \psi(\hat{x}) = 0, \quad (20)$$

and

$$\phi(\hat{x}, z) = \sum_{S \subseteq \{1,\ldots,3\}} q_S(\hat{x}, z) \prod_{i \in S} \phi_i(\hat{x}, z)$$
$$\chi(\hat{x}, z) = (\xi(\hat{x}))^{2r_1}$$
$$\psi(\hat{x}) = \sum_{i=1}^{m} \rho_i(\hat{x}, z) \left[\frac{\partial h}{\partial x} g(\hat{x})\right]_i,$$

where $\phi_1(\cdot) = -\xi(\hat{x})$, $\phi_2(\cdot) = -z^T z + \gamma^2$ and $\phi_3(\cdot) = \hat{h}(\hat{x})$.

Proof: By Proposition 3, we have $h(\hat{x})$ is an SCBF iff there exist no $\hat{x}$, z such that $\frac{\partial h}{\partial x} g(\hat{x}) = 0$, $z^T z - \gamma^2 \leq 0$ and $-\xi(\hat{x}) > 0$. The latter two conditions are equivalent to $-z^T z + \gamma^2 \geq 0$, and $-\xi(\hat{x}) \geq 0$, $\xi(\hat{x}) \neq 0$. These conditions are equivalent to (20) by the Positivstellensatz. ■

2) Verification for FT-SCBF: We now extend the result into the case where sensors may experience faults and attacks. Specifically, we consider a nonlinear control system whose output may be affected by one of m sensor faults described by (7) and (8).

In this case, we need to verify the feasibility of u to satisfy m SCBF constraints under m possible sensor faults. To achieve this, we extend Proposition 3 to verify the feasibility of a set of SCBF constraints via Farkas' Lemma.

Corollary 1: Define $A(x)$ and $\Xi(x)$ as follows.

$$A(x) = (A_1(x) \ldots A_m(x))^T$$
$$\Xi(x) = [\xi_1(x) \ldots \xi_m(x)]^T \quad (21)$$

Control input u that is chosen to satisfy a set of linear constraints can be written as

$$A(x)u \leq \Xi(x, z). \quad (22)$$

By Farkas's Lemma, the system $A(x)u \leq \Xi(x)$ has a solution $u \in \mathbb{R}^p$, if and only if there does not exist $y \in \mathbb{R}^m$ such that

$$A^T(\hat{x})y = 0,\ y \geq 0,\ \Xi^T(\hat{x})y < 0. \tag{23}$$

We define $A(x)$ and $\Xi(x)$ as follows

$$A(\hat{x}) = \left(-\frac{\partial h}{\partial x}g(\hat{x}_1)\ldots-\frac{\partial h}{\partial x}g(\hat{x}_m)\right)^T,$$
$$\Xi(\hat{x}) = [\xi(\hat{x}_1)\ldots\xi(\hat{x}_m)]^T.$$

Proposition 4: Suppose Assumption 1 and conditions in Theorem 5 hold. There exists a feasible solution u satisfying a set of m SCBF constraints if and only if there is no $\hat{x}_1,\ldots,\hat{x}_m \in \mathcal{C}_\gamma$, $z_1,\ldots,z_m \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ satisfying $z_j^T z_j - \gamma^2 \leq 0$, $(\hat{x}_j-\hat{x}_k)^T(\hat{x}_j-\hat{x}_k)-\gamma^2 \leq 0$, $A^T(\hat{x}_j)y = 0$, $y \geq 0$ and $\Xi^T(\hat{x}_j)y < 0$, $\forall j,k \in \{1,\ldots,m\}$.

Proof: By Definition 3, we have $\hat{h}(\hat{x}_j) = h(\hat{x}_j) - \overline{h}_\gamma \geq 0$ for all $\hat{x}_j \in \mathcal{C}_\gamma$. By Theorem 1, we have $z_j^T z_j - \gamma^2 \leq 0$ for all j. By Theorem 5, we have $(\hat{x}_j-\hat{x}_k)^T(\hat{x}_j-\hat{x}_k) \leq \gamma^2$, $\forall j,k \in \{1,\ldots,m\}$. For the case where $\hat{h}(\hat{x}) = 0$, u can be chosen to satisfy (5) for all j. By Corollary 1, we have the existence of u if and only if there does not exist $y \in \mathbb{R}^m$ such that (23) hold. Conversely, if for some $\hat{x}_0$, $\hat{x}_1$, z_0 and y_0 satisfying $\frac{\partial h}{\partial x}g(\hat{x}_0) = 0$, $(\hat{x}_0-\hat{x}_1)^T(\hat{x}_0-\hat{x}_1)-\gamma^2 \leq 0$, $z_0^T z_0 - \gamma^2 \leq 0$, $A^T(\hat{x}_0)y_0 = 0$, $y_0 \geq 0$ and $\Xi^T(\hat{x}_0)y_0 < 0$, the set $\mathcal{C}$ is not positive invariant. ■

Then, we can formulate the following conditions via the Positivstellensatz.

Lemma 3: There exists a feasible solution u satisfying a set of m SCBF constraints if and only if there exist polynomials $\rho(\hat{x}_j,y,z_j)$, sum-of-squares polynomials $q(\hat{x}_j,y,z_j)$, integers $s = 4m+m^2$, $r_1,\ldots,r_m$ such that

$$\phi(\hat{x}_j,\hat{x}_k,y,z_j)+\chi(\hat{x}_j,\hat{x}_k,y,z_j)+\psi(\hat{x}_j,\hat{x}_k,y,z_j) = 0, \tag{24}$$

and

$$\phi(\cdot) = \sum_{S\subseteq\{1,\ldots,s\}} q_S(\hat{x}_j,\hat{x}_k,y,z_j)\prod_{i\in S}\phi_i(\hat{x}_j,\hat{x}_k,y,z_j)$$
$$\chi(\cdot) = \prod_{\forall j\in\{1,\ldots,m\}}\left(-\Xi^T(\hat{x}_j)y\right)^{2r_j}$$
$$\psi(\cdot) = \sum_{j=1}^m\left(\sum_{i=1}^p \rho_i^0(\hat{x}_j,\hat{x}_k,y,z_j)\left[A^T(\hat{x}_j)y\right]_i\right),$$

where $\phi_{\{1,\ldots,m\}}(\cdot) = -\Xi^T(\hat{x}_j)y_j$, $\phi_{\{m+1,\ldots,2m\}}(\cdot) = \hat{h}(\hat{x}_j)$, $\phi_{\{2m+1,\ldots,3m\}}(\cdot) = -z_j^T z_j + \gamma^2$, $\phi_{\{3m+1,\ldots,4m\}}(\cdot) = y_j$ and $\phi_{\{4m+1\ldots4m+m^2\}}(\cdot) = -(\hat{x}_j-\hat{x}_k)^T(\hat{x}_j-\hat{x}_k)+\gamma^2$.

Proof: By Proposition 4, we have $h(\hat{x})$ is an SCBF if and only if there exist no $\hat{x}_1,\ldots,\hat{x}_m$, $z_1,\ldots,z_m$ and y satisfying $(\hat{x}_j-\hat{x}_k)^T(\hat{x}_j-\hat{x}_k)-\gamma^2 \leq 0, \forall j,k \in \{1,\ldots,m\}$, $z_j^T z_j-\gamma^2 \leq 0\ \forall j$, $A^T(\hat{x})y_j = \mathbf{0}$, $y_j \geq 0$ and $\Xi^T(\hat{x})y_j < 0$. The conditions are equivalent to $\forall j,k \in \{1,\ldots,m\}$,

$$-z_j^T z_j+\gamma^2 \geq 0,$$
$$-(\hat{x}_j-\hat{x}_k)^T(\hat{x}_j-\hat{x}_k)+\gamma^2 \geq 0,$$
$$A^T(\hat{x}_j)y = \mathbf{0},\ y \geq 0,\ -\Xi^T(\hat{x}_j)y \geq 0, \Xi^T(\hat{x}_j)y \neq 0$$

These conditions are equivalent to (24) by the Positivstellensatz. ■

3) Verification for FT-SCBF with high relative degree: We further extend the proposition 4 and Lemma 3 to verify the feasibility of a set of HOSCBF constraints.

We define $A(x)$ and $\Xi(x)$ as follows

$$A(\hat{x}) = \left(-\frac{\partial h^{d'}}{\partial x}g(\hat{x}_1),\ldots,-\frac{\partial h^{d'}}{\partial x}g(\hat{x}_m)\right)^T,$$
$$\Xi(\hat{x}) = [\xi_1(\hat{x}_1),\ldots,\xi_m(\hat{x}_m)]^T, where$$
$$\xi_i(\hat{x}) = \frac{\partial h^{d'}}{\partial x}f(\hat{x}) + \frac{1}{2}\mathbf{tr}\left(\nu^T K^T\frac{\partial^2 h_i^{d'}}{\partial x^2}(\hat{x})K\nu\right) - \left\|\frac{\partial h_i^{d'}}{\partial x}(\hat{x})Kc\right\|_2\gamma_i + \hat{h}_i^{d'}(\hat{x}).$$

Proposition 5: Suppose Assumption 1 and conditions in Theorem 5 hold. There exists a feasible solution u satisfying a set of m HOSCBF constraints with relative degree d if and only if there is no $\hat{x}_1,\ldots,\hat{x}_m \in \mathcal{C}_\gamma$, $z_1,\ldots,z_m \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ satisfying $\hat{h}^d(\hat{x}_j) \geq 0,\ \forall j,\forall d \leq d'$, $(\hat{x}_j-\hat{x}_k)^T(\hat{x}_j-\hat{x}_k)-\gamma^2 \leq 0$, $\forall j,k \in \{1,\ldots,m\}$, $z_j^T z_j-\gamma^2 \leq 0\ \forall j$, $A^T(\hat{x})y = 0$, $y \geq 0$ and $\Xi^T(\hat{x})y < 0$.

Proof: By Theorem 1, we have $z_j^T z_j - \gamma^2 \leq 0$ for all j with m EKFs. By the Definition 5, we have $\hat{h}^d(\hat{x}_j) = h^d(\hat{x}_j) - \overline{h}_\gamma \geq 0$ for all $\hat{x}_j \in \overline{\mathcal{C}}$, if and only if the following three conditions are satisfied. For all $\hat{x}_j \in \overline{\mathcal{C}}$, $\hat{h}^d(\hat{x}) \geq 0$ for all $d \leq d'$. Next, by Theorem 5, $(\hat{x}_j-\hat{x}_k)^T(\hat{x}_j-\hat{x}_k) \leq \gamma^2$, $\forall j,k \in \{1,\ldots,m\}$. Moreover, $\frac{\partial h^{d'}}{\partial x}g(\hat{x}_j) \neq 0$ and u are chosen to satisfy (10) for all j. By Corollary 1, we have a solution $u \in \mathbb{R}^p$ exists, if and only if there does not exist $y \in \mathbb{R}^m$ such that (23) holds. Conversely, if for some $\hat{x}_0$, $\hat{x}_1$, z_0 and y_0 satisfying $\hat{h}^d(\hat{x}_0) \geq 0$, $(\hat{x}_0-\hat{x}_1)^T(\hat{x}_0-\hat{x}_1)-\gamma^2 \leq 0$, $z_0^T z_0 - \gamma^2 \leq 0$, $A^T(\hat{x})y_0 = 0$, $y_0 \geq 0$ and $\Xi^T(\hat{x})y_0 < 0$, the set $\mathcal{C}$ is not positive invariant. ■

Lemma 4: There exists a feasible solution u satisfying a set of m HOSCBF constraints if and only if there exist polynomials $\rho(\hat{x}_j,\hat{x}_k,y,z_j)$, sum-of-squares polynomials $q_S(\hat{x}_j,\hat{x}_k,y,z_j)$, integers $d_1',\ldots,d_m'$, $s = 3m+m^2+\sum_{j=1}^m d_j'$, $r_1,\ldots,r_m$ such that

$$\phi(\hat{x}_j,\hat{x}_k,y,z_j)+\chi(\hat{x}_j,\hat{x}_k,y,z_j)+\psi(\hat{x}_j,\hat{x}_k,y,z_j) = 0, \tag{25}$$

and

$$\phi(\cdot) = \sum_{S\subseteq\{1,\ldots,s\}} q_S(\hat{x}_j,\hat{x}_k,y,z_j)\prod_{i\in S}\phi_i(\hat{x}_j,\hat{x}_k,y,z_j)$$
$$\chi(\cdot) = \prod_{\forall j\in\{1,\ldots,m\}}\left(-\Xi^T(\hat{x}_j)y\right)^{2r_j}$$
$$\psi(\cdot) = \sum_{j=1}^m\left(\sum_{i=1}^p \rho_i^0(\hat{x}_j,\hat{x}_k,y,z_j)\left[A^T(\hat{x}_j)y\right]_i\right)$$

where $\phi_{\{1,\ldots,m\}}(\cdot) = -\Xi^T(\hat{x}_j)y$, $\phi_{\{m+1,\ldots,2m\}}(\cdot) = y_j$, $\phi_{\{2m+1,\ldots,3m\}}(\cdot) = -z_j^T z_j + \gamma^2$, $\phi_{\{3m+1\ldots3m+m^2\}}(\cdot) = -(\hat{x}_j-\hat{x}_k)^T(\hat{x}_j-\hat{x}_k)+\gamma^2$ and for $d_j \in \{0,\ldots,d_j'\}$, $\phi_{\{3m+m^2+1,\ldots,3m+m^2+\sum_{j=1}^m d_j'\}}(\cdot) = \hat{h}^{d_j}(\hat{x}_j)$.

Proof: By proposition 5, we have $h^d(\hat{x})$ are HOSCBFs if and only if there exist no $\hat{x}_1,\ldots,\hat{x}_m \in \mathcal{C}_\gamma$, $z_1,\ldots,z_m \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ satisfying $\hat{h}^d(\hat{x}_j) \geq 0,\ \forall j,\forall d \leq d'$, $(\hat{x}_j-$

$\hat{x}_k)^T(\hat{x}_j - \hat{x}_k) - \gamma^2 \leq 0$, $\forall j,k \in \{1,\ldots,m\}$, $z_j^T z_j - \gamma^2 \leq 0$ $\forall j$, $A^T(\hat{x})y = 0$, $y \geq 0$ and $\Xi^T(\hat{x})y < 0$. The conditions are equivalent to $\forall j,k \in \{1,\ldots,m\}$,

$$\begin{aligned}&\hat{h}^{d_j}(\hat{x}_j) \geq 0, \forall d_j \leq d_j' \\ &-(\hat{x}_j - \hat{x}_k)^T(\hat{x}_j - \hat{x}_k) + \gamma^2 \geq 0, -z_j^T z_j + \gamma^2 \geq 0, \\ &A^T(\hat{x}_j)y = \mathbf{0},\ y \geq 0,\ -\Xi^T(\hat{x}_j)y \geq 0, \Xi^T(\hat{x}_j)y \neq 0\end{aligned}$$

These conditions are equivalent to (25) by the Positivstellensatz. ∎

V. Joint Safety and Stability Under Sensor Faults and Attacks

We next present a framework to ensure joint safety and stability for systems with sensor faults and attacks via CLFs and HOSCBFs. Such an approach has been widely used in fault-free scenarios.

Define the goal set $\mathcal{G} \subseteq \mathcal{C}$ by $\mathcal{G} = \{x : w(x) \geq 0\}$ for some function w for some equilibrium point $x_e \in G$, $f(x_e) = 0$ and $g(x_e) = 0$. Define $\tau(\mathcal{G})$ as the first time when x_t reaches $\mathcal{G}$.

Problem Statement: Given a goal set $\mathcal{G}$, a safe set $\mathcal{C}$ and a parameter $\epsilon \in (0,1)$, compute a control policy that, at each time t, maps the sequence $\{y_{t'} : t' \in [0,t)\}$ to an input u_t such that, given a finite stopping time T, for any fault $r \in \{r_1,\ldots,r_m\}$, $Pr(x_t \in \mathcal{C},\ 0 \leq t \leq T) \geq (1-\epsilon)\mathcal{T}(T)$, for some function $\mathcal{T} : \mathbb{R}^n \to (0,1)$ and $Pr(\tau(\mathcal{G}) < \infty) > 1 - \epsilon$.

A. HOSCBF-CLF

Our approach towards through asymptotically convergence to goal set $\mathcal{G}$ is through the use of *stochastic Control Lyapunov Functions*. A function $V : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is a stochastic CLF for the SDE (1) if, for each x_t, we have

$$\inf_u \left\{ \frac{\partial V}{\partial x}\overline{f}(x_t,u_t) + \frac{1}{2}\mathbf{tr}\left(\sigma^T \frac{\partial^2 V}{\partial x^2}\sigma\right)\right\} < -\rho V(x_t)^\eta \quad (26)$$

for some $\rho > 0$ and $0 < \eta < 1$.

The following result describes the stochastic stability of systems using CLFs with [64, Theorem 3.1] providing sufficient conditions for the following result. As a preliminary, define $\tau(z) = \inf\{t : V(x_t) \leq z\}$.

Proposition 6 ([64, Theorem 3.1]): Suppose there exists a $\overline{V}$ such that, whenever $V(x_t) \geq \overline{V}$, we choose u_t to satisfy

$$\frac{\partial V}{\partial x}f(x_t) + \frac{\partial V}{\partial x}g(x_t)u_t + \frac{1}{2}\mathbf{tr}\left(\sigma^T\frac{\partial^2 V}{\partial x^2}\sigma\right) < -\rho V(x_t)^\eta$$

for some $\rho > 0$ and $0 < \eta < 1$. For $x \in \mathbb{R}^n \backslash \{x | V(x) = 0\}$, $Pr(\tau(\overline{V}) < \infty | x_0 = x) = 1$.

In the case with sensor faults and attacks, we consider a system with dynamics (1) and an Extended Kalman Filter estimator $\hat{x}_t$. The following result is an extension of Proposition 6 to this case.

Lemma 5: Suppose that there exist constants $M > 0$ and $k \in \mathbb{N}$ such that, for any x and x', $|V(x) - V(x')| \leq M||x - x'||_2^k$. Suppose $Pr(||\hat{x}_t - x_t||_2 \leq \gamma\ \forall t) > 1 - \epsilon$ and, at each time t when $V(\hat{x}_t) > \overline{V}$, we have

$$\begin{aligned}&\frac{\partial V}{\partial x}(\hat{x}_t)(f(\hat{x}_t) + g(\hat{x}_t)u_t) + \gamma||\frac{\partial V}{\partial x}(\hat{x}_t)K_t c||_2 \\ &+ \frac{1}{2}\mathbf{tr}\left(\nu_t^T K_t^T \frac{\partial^2 V}{\partial x^2}(\hat{x}_t)K_t\nu_t\right) < -\rho(V(\hat{x}_t) + M\gamma^k)^\eta.\end{aligned} \quad (27)$$

Then

$$Pr(\tau(\overline{V} + M\gamma^k) < \infty) > 1 - \epsilon$$

for all $x_t \in \mathbb{R}^n \backslash \{x_t | V(x_t) = 0\}$.

Proof: The dynamics of $\hat{x}_t$ are given by

$$d\hat{x}_t = \overline{f}(\hat{x}_t,u_t) + K_t c(x_t - \hat{x}_t) + K_t\nu_t dV_t.$$

If $||\hat{x}_t - x_t||_2 \leq \gamma$, the differential generator $LV(\hat{x}_t)$ satisfies

$$\begin{aligned}LV(\hat{x}_t) &= \frac{\partial V}{\partial x}(\hat{x}_t)\overline{f}(\hat{x}_t,u_t) + \frac{\partial V}{\partial x}(\hat{x}_t)K_t c(x_t - \hat{x}_t) \\ &\quad + \frac{1}{2}\mathbf{tr}\left(\nu_t^T K_t^T \frac{\partial^2 V}{\partial x^2}(\hat{x}_t)K_t\nu_t\right) \\ &\leq \frac{\partial V}{\partial x}(\hat{x}_t)\overline{f}(\hat{x}_t,u_t) + \gamma||\frac{\partial V}{\partial x}(\hat{x}_t)K_t c||_2 \\ &\quad + \frac{1}{2}\mathbf{tr}\left(\nu_t^T K_t^T \frac{\partial^2 V}{\partial x^2}(\hat{x}_t)K_t\nu_t\right).\end{aligned}$$

since

$$\begin{aligned}\frac{\partial V}{\partial x}(\hat{x}_t)K_t c(x_t - \hat{x}_t) &\leq ||\frac{\partial V}{\partial x}(\hat{x}_t)K_t c||_2||x_t - \hat{x}_t||_2 \\ &\leq \gamma||\frac{\partial V}{\partial x}(\hat{x}_t)K_t c||_2.\end{aligned}$$

Since $|V(x_t) - V(\hat{x}_t)| \leq M||x_t - \hat{x}_t||_2^k$, we have $V(x_t) \leq V(\hat{x}_t) + M||x_t - \hat{x}_t||_2^k$ and $\rho(V(x_t))^\eta \leq \rho(V(\hat{x}_t) + M\gamma^k)^\eta$, for some $\rho > 0$ and $0 < \eta < 1$.

Hence, if (27) holds, then

$$Pr\left(\inf\{t : V(\hat{x}_t) \leq \overline{V}\} < \infty \mid ||x_t - \hat{x}_t||_2 \leq \gamma\ \forall t\right) = 1$$

by Proposition 6. Since $|V(x_t) - V(\hat{x}_t)| \leq M||x_t - \hat{x}_t||_2^k$, we have $V(x_t) \leq V(\hat{x}_t) + M||x_t - \hat{x}_t||_2^k$, and so

$$\begin{aligned}&Pr(V(x_t) > \overline{V} + M\gamma^k |\ ||x_t - \hat{x}_t||_2 \leq \gamma\ \forall t) \\ &\leq Pr(V(\hat{x}_t) + M\gamma^k > \overline{V} + M\gamma^k |\ ||x_t - \hat{x}_t||_2 \leq \gamma\ \forall t) \\ &= Pr(V(\hat{x}_t) > \overline{V} |\ ||x_t - \hat{x}_t||_2 \leq \gamma\ \forall t)\end{aligned}$$

Hence $Pr(\tau(\overline{V} + M\gamma^k) < \infty |\ ||x_t - \hat{x}_t||_2 \leq \gamma\ \forall t) = 1$ and thus $Pr(\tau(\overline{V} + M\gamma^k) < \infty) > 1 - \epsilon$. ∎

Motivated by this result, we next state a control policy that combines CLFs and HOSCBFs to ensure safety and stability. At each time t, the set of feasible control actions is defined as follows:

1) Define $Y_t(\overline{V}) = \{j : V(\hat{x}_{t,j}) > \overline{V}\}$, and initialize $U_t = Y_t(\overline{V})$. Define a collection of sets Υ_i, $i \in U_t$, by

$$\begin{aligned}\Upsilon_i \triangleq \Big\{u : &\frac{\partial V_i}{\partial x}\overline{f}(\hat{x}_{t,i},u) + \gamma_i||\frac{\partial V_i}{\partial x}(\hat{x}_{t,i})c||_2 \\ &+ \frac{1}{2}\mathbf{tr}\left(\overline{\nu}_{t,i}^T K_{t,i}^T \frac{\partial^2 V}{\partial x^2}(\hat{x}_{t,i})K_{t,i}\overline{\nu}_{t,i}\right) \\ &< -\rho_i(V(\hat{x}_t) + M\gamma^k)^{\eta_i}\Big\}\end{aligned} \quad (28)$$

for some ρ_i, η_i, $i = 1, \ldots, m$. Select any

$$u_t \in \left(\bigcap_{i \in Z_t} \Omega_i \right) \cap \left(\bigcap_{j \in U_t} \Upsilon_j \right),$$

where Ω_i is defined as in (13). If no such u_t exists, go to Step 2.

2) For each i, j with $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 > \overline{\theta}_{ij}$, set $Z_t = Z_t \backslash \{i\}$ and $U_t = U_t \backslash \{i\}$ (resp. $Z_t = Z_t \backslash \{j\}$ and $U_t = U_t \backslash \{j\}$) if $||\hat{x}_{t,i} - \hat{x}_{t,i,j}||_2 > \theta_{ij}/2$ (resp. $||\hat{x}_{t,j} - \hat{x}_{t,i,j}||_2 > \theta_{ij}/2$). If

$$\left(\bigcap_{i \in Z_t} \Omega_i \right) \cap \left(\bigcap_{j \in U_t} \Upsilon_j \right) \neq \emptyset,$$

then select u_t from this set. Else go to Step 3.

3) Remove the sets Ω_i and Υ_i corresponding to the estimators with the largest residue values until there exists a feasible u_t.

This policy is similar to the HOSCBF-based approach of Section IV, with additional constraints to satisfy the stability condition. This leads to another m linear inequalities. The following result gives sufficient conditions for safety and stability.

Theorem 6: Suppose that $\hat{h}_1^{d'}, \ldots, \hat{h}_m^{d'}$, $\gamma_1, \ldots, \gamma_m$, V, $\overline{V}$, and θ_{ij} satisfy the constraints of Theorem 5, as well as the following: (i) The function V satisfies $\{x : V(x) \leq \overline{V} + M\gamma_i^k\} \subseteq \mathcal{G}$ for all i. (ii) Define $\Gamma_i(\hat{x}_{t,i}) = \frac{\partial V_i}{\partial x} g(\hat{x}_{t,i})$. Let $X_t' \subseteq X_t^{d'}(\delta)$ and $Y_t' \subseteq Y_t(\overline{V})$ be sets satisfying $||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 \leq \theta_{ij}$ for all $i \in X_t'$ and $j \in Y_t'$. Then there exists u with

$$\Lambda_i^{d'}(\hat{x}_{t,i})u > 0, \quad \Gamma_j(\hat{x}_{t,j})u < 0 \tag{29}$$

for all $i \in X_t'$ and $j \in Y_t'$. If conditions (i) and (ii) hold, then $Pr(x_t \in \mathcal{C},\ 0 \leq t \leq T) \geq (1-\epsilon) \prod_{d=0}^{d'} (\frac{\hat{h}^d(x_0)}{\zeta^d}) e^{-\zeta^d T}$ and $Pr(\tau(\mathcal{G}) < \infty) > 1 - \epsilon$ for any fault pattern $r \in \{r_1, \ldots, r_m\}$, where $\tau(\mathcal{G})$ is the first time when x_t reaches $\mathcal{G}$.

Proof: Suppose there exists relative degree d'. By the argument of Theorem 5, $i \in X_t^{d'}(\delta)$ implies that Ω_i is a constraint on u_t at time t. An analogous argument yields that Υ_i is a constraint as well. By selecting u_t satisfying (29) at each time t, we have that $Pr(x_t \in \mathcal{C},\ 0 \leq t \leq T) \geq (1 - \epsilon) \prod_{d=0}^{d'} (\frac{\hat{h}^d(x_0)}{\zeta^d}) e^{-\zeta^d T}$ by Theorem 5 and $Pr(\tau(\overline{V} + M\gamma^k) < \infty) > 1 - \epsilon$ by Lemma 5. Hence $Pr(\tau(\mathcal{G}) < \infty) > 1 - \epsilon$ by assumption (i) of the theorem. ■

A controller that reaches a goal set defined by a function V while satisfying a safety constraint $\mathcal{C} = \{x : h(x) \geq 0\}$ can be obtained by solving the optimization problem

$$\begin{array}{ll} \text{minimize} & u_t^T R u_t \\ \text{s.t.} & \Lambda_i^{d'}(\hat{x}_{t,j}) u_t \geq \overline{\omega}_j^{d'} \ \forall j \in X_t^{d'}(\delta) \quad \text{(HOSCBF)} \\ & \Gamma_i(\hat{x}_{t,i}) u_t \leq \overline{\tau}_i \ \forall i \in Y_t(\overline{V}) \quad \text{(CLF)} \end{array} \tag{30}$$

at each time step, where R is a positive definite matrix representing the cost of exerting control.

B. HOSCBF-CLF Construction

As in the case of a single HOSCBF constraint, satisfaction of (29) will depend on the geometry of the safe region and goal set as well as the values of γ_i and θ_{ij}. We consider a linear system with dynamics

$$dx_t = (Fx_t + Gu_t)\, dt + \sigma dW_t. \tag{31}$$

The goal set is ellipsoidal, so that $w(x) = V(x) = (x - x'')^T \Psi (x - x'')$, and the safe region $\mathcal{C}$ is given by a hyperplane constraint $a^T x - b \geq 0$. We next construct SCBF-CLF as a special case of HOSCBF-CLF to ensure safety and stability of the cases where $\text{rank}(G) = n$ and $\text{rank}(G) < n$.

Proposition 7: Suppose that $\text{rank}(G) = n$ and the following conditions hold:

$$a^T x'' - b > 0 \tag{32}$$

$$\left\{ (x - x'')^T \Psi (x - x'') \leq \frac{\overline{\theta}^2 \overline{\lambda}}{2} \right\} \cap \{a^T x - b \leq 0\} = \emptyset \tag{33}$$

Then there exists $\delta > 0$ such that, at each time t, there exists u satisfying (29) when $\overline{V} = \frac{\overline{\theta}^2 \overline{\lambda}(\Phi)}{2}$.

Proof: Select δ such that $\delta < a^T x''$. We consider three cases. In the first case, $X_t(\delta) \neq \emptyset$ and $Y_t(\overline{V}) = \emptyset$. In the second case, $X_t(\delta) = \emptyset$ and $Y_t(\overline{V}) \neq \emptyset$. In the third case, $X_t(\delta) \neq \emptyset$ and $Y_t(\overline{V}) \neq \emptyset$.

If $X_t(\delta) \neq \emptyset$ and $Y_t(\overline{V}) = \emptyset$, then u satisfying $a^T G u > 0$ suffices to ensure safety by Lemma 1 in [56]. If $Y_t(\overline{V}) \neq \emptyset$ and $X_t(\delta) = \emptyset$, then choose u such that $Gu = -(\hat{x}_{t,i} - x'')$ for some $i \in Y_t(\overline{V})$. By Proposition 1 in [56], for any positive definite matrix Φ and $x' \in \mathbb{R}^n$ if

$$||\hat{x}_{t,i} - \hat{x}_{t,j}||_2 \leq \overline{\theta} \leq \overline{\lambda}(\Phi)^{-1/2} \sqrt{2}$$

and $\hat{x}_{t,i}$ and $\hat{x}_{t,j}$ both satisfy $(x - x')^T \Phi (x - x') > (1 - \epsilon)$ for ϵ sufficiently small, then

$$(\hat{x}_{t,i} - x')^T \Phi (\hat{x}_{t,j} - x') > 0.$$

Choosing $\Phi = \frac{1}{2\overline{\theta}^2} \Psi$ and $x' = x''$ yields $\overline{\lambda}(\Phi) = \frac{1}{2\overline{\theta}^2}$, and hence $\overline{\theta} \leq \overline{\lambda}(\Phi)^{-1/2} \sqrt{2}$ holds by construction. Hence we have

$$(\hat{x}_{t,i} - x'')^T \left(\frac{1}{2\overline{\theta}^2 \overline{\lambda}} \Psi \right) (\hat{x}_{t,j} - x'') > 0$$

when $\hat{x}_{t,i}$ and $\hat{x}_{t,j}$ satisfy

$$(x - x'')^T \left(\frac{1}{2\overline{\theta}^2 \overline{\lambda}} \Psi \right) (x - x'') \geq 1,$$

or equivalently, when they satisfy

$$(x - x'')^T \Psi (x - x'') \geq \frac{\overline{\theta}^2 \overline{\lambda}}{2}.$$

Finally, suppose that $Y_t(\overline{V}) \neq \emptyset$ and $X_t(\delta) \neq \emptyset$. Choosing u such that $Gu = -(\hat{x}_{t,i} - x'')$ for some $i \in Y_t(\overline{V})$. By the preceding discussion, (29) holds for all $j \in Y_t(\overline{V})$. By choice of δ, $i \in X_t(\delta)$ implies that $a^T \hat{x}_{t,i} < a^T x''$. Hence, we have

$$a^T G u = a^T (x'' - \hat{x}_{t,i}) > 0,$$

and therefore $\Lambda(\hat{x}_{t,i}) u_t < 0$ is satisfied for all $i \in X_t'$. ■

We next turn to the case where $\text{rank}(G) < n$. As in the case of the SCBF construction in [56], we add a hyperplane constraint $(x - x'')^T \Psi v < 0$ to ensure that the SCBF-CLF constraints are satisfied.

Proposition 8: Suppose that $v \in \text{span}(G)$ and satisfies the following conditions: (i) $a^T v > 0$, and (ii) the initial state x' satisfies $(x' - x'')^T \Psi v < 0$. Then there exists u satisfying (29) at each time t.

Proof: At each time t, choose $Gu = v$. We need to verify both SCBF constraints and the CLF constraint for each i. First, for the constraint $h(x) = a^T x - b > 0$, we must have $-a^T G u < 0$, which is equivalent to assumption (i) of the proposition. For the constraint $(x - x'')^T \Psi G u < 0$, the choice of $v = Gu$ and the hyperplane constraint $(x - x'')^T \Psi v < 0$ implies that the CLF constraint is satisfied. Finally, the hyperplane constraint $(x - x'')^T \Psi v < 0$ can be satisfied if $-v^T \Psi G u < 0$, or equivalently, if $-v^T \Psi v < 0$, which holds since Ψ is positive definite. ■

VI. Safe Control Under Actuator Failures

Motivated by FT-CBFs for sensor faults, in this section, we propose a passive fault-tolerant control based on CBFs to mitigate actuator failures. We consider a nonlinear control system affected by actuator failures in a noise-free scenario. The system dynamics can be described as

$$dx_t = (f(x_t) + g(x_t) u_t^F)\, dt \quad (34)$$

$$dy_t = c x_t\, dt, \quad (35)$$

where u_t^F represents the actuator's output with failures.

We consider loss of control effectiveness failure [65] in which a subset of actuators produce a zero output. Let $\mathcal{I}$ denote the set of failed actuators. The actuator failure model is given by $u_t^F = L u_t$, where L is a $p \times p$ diagonal matrix with $L_{ii} = 0$ if $i \in \mathcal{I}$ and $L_{ii} = 1$ otherwise.

The concept of actuator redundancy for LTI systems is proposed in [25]. We extend the definition to nonlinear systems.

Definition 7 (Actuator Redundancy): A nonlinear system (34)-(35) is said to have r actuator redundancy if the system remains controllable for all failure patterns $L_j \in \mathcal{L}_r = \{L_{\mathcal{I}} \mid \mathcal{I} \subsetneq \{1, \dots p\}, |\mathcal{I}| \leq r\}$, where $|\mathcal{I}|$ denotes the cardinality of $\mathcal{I}$.

In this section, we assume that the system described in (34) and (35) has r actuator redundancy.

Problem Statement: Given a set $\mathcal{C}$ defined in (3), construct a control policy that, at each time t, maps the sequence $\{y_{t'} : t' \in [0, t)\}$ to an input u_t and, for any failure $\mathcal{L}_j \in \mathcal{L}$, ensure $x_t \in \mathcal{C}$, $\forall t$.

The goal is to ensure the safety defined in (3) of the system when actuator failures occur. The intuition behind our approach is to choose a control input u that is safe for all possible actuator failure patterns. To achieve this, we examine the system dynamics with m possible failure patterns $\mathcal{L}_j \in \mathcal{L}$ at each time and choose u_t to satisfy all safety constraints.

Lemma 6: For a system (34)–(35) with safety region defined by (3), u_t is chosen to satisfy constraints $\bigcap_{\mathcal{L}_j \in \mathcal{L}} \Omega_j$ defined as follows.

$$\Omega_j = \{u : \frac{\partial h}{\partial x} (f(x_t) + g(x_t) L_j u_t) \geq -\alpha(h(x_t))\} \quad (36)$$

Then when the set $\bigcap_{L_j \in \mathcal{L}} \Omega_j$ is non-empty, the safety can be guaranteed when any failure pattern $L_j \in \mathcal{L}$ happens.

Proof: By Corollary 2 in [35], safe set $\mathcal{C}$ is forward invariant if u satisfies the CBF constraint. Safety of the system under actuator failure i can be ensured by choosing $u \in \bigcap_{\mathcal{L}_j \in \mathcal{L}} \Omega_j \subseteq \Omega_i$. ■

However, a feasible control input u may not exist when $\frac{\partial h}{\partial x} g(x) L_j = 0$, since u does not affect states x due to dynamics or actuator failures. To address this problem, we define higher order HOCBFs for actuator failure such that for d-th degree $\frac{\partial h^d}{\partial x} g(x) L_j \neq 0$. Then, we choose control input u to satisfy constraints constructed by higher order CBFs.

Lemma 7: For a system (34)–(35) with safety region defined by (3), suppose there exist m relative degrees d_j', for each L_j, $j \in \{1, \dots, m\}$ such that $\frac{\partial h^{d_j'}}{\partial x} g(x_t) L_j \neq 0$. For all $x_0 \in \overline{\mathcal{C}} := \cap_{j=1}^m \cap_{d=0}^{d_j'} \mathcal{C}_j^d$, u_t is chosen to satisfy constraints $\bigcap_{L_j \in \mathcal{L}} \Omega_j$ defined as follows.

$$\Omega_j = \{u : \frac{\partial h^{d_j'}}{\partial x} (f(x_t) + g(x_t) L_j u) \geq -\alpha(h^{d_j'}(x_t))\} \quad (37)$$

Then when the set $\bigcap_{L_j \in \mathcal{L}} \Omega_j$ is non-empty, the safety can be guaranteed when any failure pattern $L_j \in \mathcal{L}$ happens.

Proof: For a given $j \in \{1, \dots, m\}$, the safe set $\cap_{d=0}^{d_j'} \mathcal{C}_j^d$ remains forward invariant by Theorem 3, if u_t is chosen to satisfy the corresponding HOCBF constraint (37). For an unknown actuator failure i, the forward invariance of the set $\overline{\mathcal{C}} = \cap_{j=1}^m \cap_{d=0}^{d_j'} \mathcal{C}_j^d$ is ensured if $u_t \in \bigcap_{\mathcal{L}_j \in \mathcal{L}} \Omega_j \subseteq \Omega_i$ for $x_0 \in \overline{\mathcal{C}}$. ■

Although the approach in Lemma 6 and Lemma 7 can ensure the safety of the system, excessive constraints make the existence of a feasible solution problematic. In what follows, we present feasibility verification for safe control under actuator failures.

A. Feasibility Verification

We provide feasibility verification for both CBF and HOCBF of the system with actuator failures.

1) Verification for CBF Under Actuator Failures: We first show the verification for CBF of the system with actuator failures. We denote $A(x)$ and $\Xi(x)$ as follows

$$A(x) = \left(-\frac{\partial h}{\partial x} g(x) L_1, \dots, -\frac{\partial h}{\partial x} g(x) L_p \right)^T,$$

$$\Xi(x) = [\xi(x), \dots, \xi(x)]^T,$$

where

$$\xi(x) = \frac{\partial h}{\partial x} f(x_t) + h(x_t)$$

Proposition 9: There exists a feasible solution u satisfying a set of m CBF constraints if and only if there is no x and y satisfying $A^T(x) y = 0$, $y \geq 0$ and $\Xi^T(x) y < 0$.

Proof: By Corollary 1, we have a solution $u \in \mathbb{R}^p$, if and only if there does not exist $y \in \mathbb{R}^m$ such that (23) hold. Conversely, if for some x_0 and y_0 satisfying $\frac{\partial h}{\partial x} g(x_0) L_j = 0$ for some j, $A^T(x) y_0 = 0$, $y_0 \geq 0$ and $\Xi^T(x) y_0 < 0$, the set $\mathcal{C}$ is not positive invariant. ■

Based on this proposition, we can formulate the following conditions via the Positivstellensatz.

Lemma 8: There exists a feasible solution u satisfying a set of m CBF constraints if and only if there exist polynomials $\rho_i^0(x,y)$, $\rho_i^1(x,y)$, sum-of-squares polynomials $q_S(x,y)$, and integers r such that

$$\phi(x,y)+\chi(x,y)+\psi(x,y)=0, \tag{38}$$

and

$$\begin{aligned}\phi(x,y)&=\sum_{S\subseteq\{1,\dots,3\}} q_S(x,y)\prod_{i\in S}\phi_i(x,y)\\ \chi(x,y)&=\left(\Xi^T y\right)^{2r}\\ \psi(x,y)&=\sum_{j=1}^{m}\left(\sum_{i=1}^{m}\rho_i^0(x,y)\left[A^T y\right]_i\right),\end{aligned}$$

where $\phi_1(\cdot)=y$, $\phi_2(\cdot)=-\Xi^T(x)y$ and $\phi_3(\cdot)=h(x)$.

Proof: By Proposition 9, we have $h(x)$ is an CBF if and only if there exist no x and y satisfying $y_j\geq 0$, $\frac{\partial h}{\partial x}g(x)L_j=\mathbf{0}\ \forall j\in\{1,\dots,m\}$, $A^Ty=\mathbf{0}$ and $\Xi^Ty<0$. The conditions are equivalent to $\forall j\in\{1,\dots,m\}$,

$$h(x)\geq 0,\ A^Ty=\mathbf{0},\ y_j\geq 0,\ -\Xi^T(x)y\geq 0,\Xi^T(x)y\neq 0$$

These conditions are equivalent to (38) by the Positivstellensatz. ∎

2) Verification for HOCBF of Actuator Failures: We next verify the feasibility of a set of HOCBF constraints. We denote A and Ξ as follows

$$\begin{aligned}A(x)&=\left(-\frac{\partial h^{d_1'}}{\partial x}g(x)L_1,\dots,-\frac{\partial h^{d_p'}_p}{\partial x}g(x)L_p\right)^T,\\ \Xi(x)&=[\xi_1(x),\dots,\xi_p(x)],\end{aligned}$$

where

$$\xi_j(x)=\frac{\partial h^{d_j'}}{\partial x}f(x_t)+h(x_t)$$

Proposition 10: There exists a feasible solution u satisfying a set of m CBF constraints if and only if there is no x and y satisfying $A^Ty=0$, $y_j\geq 0$ and $\Xi^T(x)y<0$, $\forall j\in\{1,\dots,m\}$.

Proof: By Definition 4, we have $h^d(x_j)\geq 0$ for all $x\in\overline{\mathcal{C}}$, if and only if the following two conditions are satisfied. For all $x\in\overline{\mathcal{C}}$, $h^d(x)\geq 0$ for all $d\leq d'$. Moreover, $\frac{\partial h^{d'}}{\partial x}g(x)L_j\neq 0$ and u are chosen to satisfy (6) for all j at the boundary. By Corollary 1, we have a solution $u\in\mathbb{R}^p$, if and only if there does not exist $y\in\mathbb{R}^m$ such that (23) holds. Conversely, if for some x_0 and y_0 satisfying $A^Ty_0=0$, $y_0\geq 0$ and $\Xi^T(x)y_0<0$, the set $\mathcal{C}$ is not positive invariant. ∎

Based on this proposition, we can formulate the following conditions via the Positivstellensatz.

Lemma 9: There exists a feasible solution u satisfying a set of m HOCBF constraints if and only if there exist polynomials $\rho_i^0(x,y)$, $\rho_i^1(x,y)$, sum-of-squares polynomials $q_S(x,y)$, integers $s=2+\sum_{j=1}^m d_j'$, and $r_1,\dots,r_m$ such that

$$\phi(x,y)+\chi(x,y)+\psi(x,y)=0, \tag{39}$$

and

$$\begin{aligned}\phi(x,y)&=\sum_{S\subseteq\{1,\dots,s\}} q_S(x,y)\prod_{i\in S}\phi_i(x,y)\\ \chi(x,y)&=\prod_{\forall j\in\{1,\dots,m\}}\left(-\Xi^T(\hat{x}_j)y\right)^{2r_j}\\ \psi(x,y)&=\sum_{j=1}^{m}\left(\sum_{i=1}^{m}\rho_i^0(x,y)\left[A^T y\right]_i\right),\end{aligned}$$

where $\phi_1(\cdot)=y$, $\phi_2(\cdot)=-\Xi^T(x)y$ and for $d_j\in\{0,\dots,d_j'\}$ $\phi_{\{3,2+\sum_{j=1}^m d_j'\}}(\cdot)=\hat{h}^{d_j}(\hat{x}_j)$.

Proof: By Proposition 4, we have $h(x)$ is an HOCBF if and only if there exist no x and y such that $h^{d_j'}(x)\geq\mathbf{0}$, for all $x\in\overline{\mathcal{C}}$, $y_j\geq 0$, $A^Ty=\mathbf{0}$, and $\Xi^T(x)y<0$. The conditions are equivalent to $\forall j\in\{1,\dots,m\}$,

$$\begin{aligned}&h^{d_j}(x)\geq 0,\forall d_j\in\{1,\dots,d_j'\}\\ &A^T(x)y=\mathbf{0},\ y\geq 0,\ -\Xi^T(x)y\geq 0,\Xi^T(x)y\neq 0\end{aligned}$$

These conditions equal to (39) by the Positivstellensatz. ∎

VII. Case Study

In this section, we present a case study of a wheeled mobile robot under sensor faults and a case study of a Boeing 747 under actuator Failure. We first describe the system models then then present the results.

A. Sensor Fault and Attack

We consider a wheeled mobile robot (WMR) with dynamics

$$\begin{pmatrix}[\dot{x}_t]_1\\ [\dot{x}_t]_2\\ \dot{\theta}_t\end{pmatrix}=\begin{pmatrix}\cos\theta_t & 0\\ \sin\theta_t & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}[\omega_t]_1\\ [\omega_t]_2\end{pmatrix}+\mathbf{w}_t \tag{40}$$

where $([x_t]_1,[x_t]_2,\theta_t)^T$ is the vector of the horizontal, vertical, and orientation coordinates for the wheeled mobile robot, $([\omega_t]_1,[\omega_t]_2)^T$ (the linear velocity of the robot and the angular velocity around the vertical axis) is taken as the control input, and $\mathbf{w}_t$ is the process noise. The feedback linearization [66] is utilized to transform the original state vector and the WMR model into the new state variable $x_t=([x_t]_1,[x_t]_2,[\dot{x}_t]_1,[\dot{x}_t]_2)^T$ with control input $u_t=([u_t]_1,[u_t]_2)^T$ and the controllable linearized model defined as follow.

$$\dot{x}_t=Fx_t+Gu_t+\mathbf{w}_t' \tag{41}$$

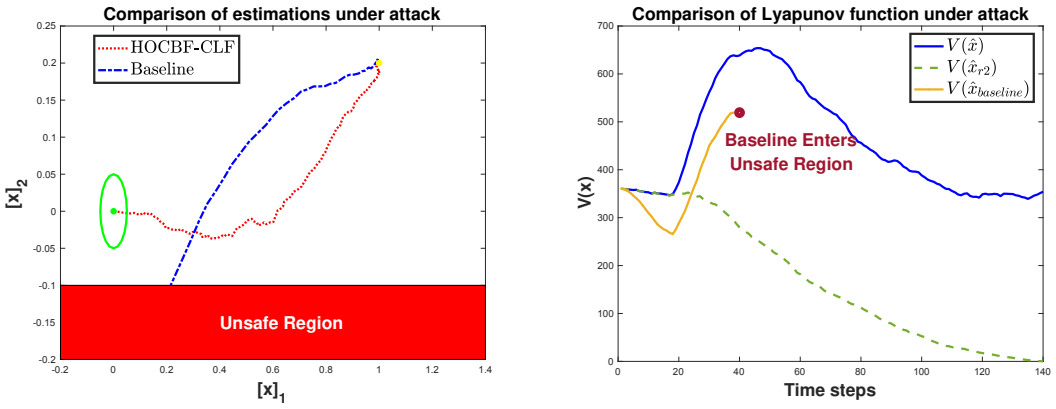
where the process noise $\mathbf{w}_t'\in\mathbb{R}^4$ has distribution $\mathcal{N}(0,\sigma_w I)$, where $\sigma_w=0.05$. The matrices F and G are defined as

$$F=\begin{pmatrix}0&0&1&0\\0&0&0&1\\0&0&0&0\\0&0&0&0\end{pmatrix},\ G=\begin{pmatrix}0&0\\0&0\\1&0\\0&1\end{pmatrix}.$$

The following compensator is used to calculate the input $[\omega_t]_1$ and $[\omega_t]_2$ into (40)

$$[\omega_t]_1 = \int_{t^-}^{t^+}[u_t]_1\cos\theta_t+[u_t]_2\sin\theta_t\ dt \tag{42}$$

$$[\omega_t]_2 = ([u_t]_2\cos\theta_t-[u_t]_1\sin\theta_t)/[\omega_t]_1. \tag{43}$$



(a) (b)

Fig. 2: Comparison of actual trajectory and Lyapunov function between HOSCBF-CLF and baseline on WMR system under sensor false data injection attacks. In (a), the baseline entered unsafe region while proposed method remains safe and converge to goal region. In (b), the Lyapunov function of real states decreases and converges to zero.

Here we assume that the observation for the orientation coordinate θ_t is attack-free and noise-free, which enables feedback linearization based on the variable θ_t.

In the linearized model, we use the observation equation

$$\begin{pmatrix}[y_t]_1\\[y_t]_2\\[y_t]_3\\[y_t]_4\\[y_t]_5\\[y_t]_6\end{pmatrix} = \begin{pmatrix}1&0&0&0\\1&0&0&0\\0&1&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}\begin{pmatrix}[x_t]_1\\[x_t]_2\\[\dot{x}_t]_1\\[\dot{x}_t]_2\end{pmatrix} + \mathbf{a}_t + \mathbf{v}_t \tag{44}$$

where the measurement noise $\mathbf{v}_t \in \mathbb{R}^6$ has distribution $\mathcal{N}(0, \sigma_v I)$, where $\sigma_v = 0.05$. The impact of the attack is denoted as $\mathbf{a}_t$. The attack signal satisfies that

$$\mathbf{a}_t = \begin{cases}\mathbf{0}, & t < 1\\ [0,0,0,2,0,0]^T, & t \geq 1\end{cases}$$

Note that there is one redundant sensor for the horizontal coordinate and one for the vertical coordinate.

Here we let the safe region $\mathcal{C} = \{x_t : h(x_t) = [x_t]_2 + 0.1 \geq 0, t \geq 0\}$ and the goal region $\mathcal{G} = \{x_t : \omega(x_t) = d - ||x_t - x_g||_2 \geq 0\}$, where x_g is $(0,0)$ and $d = 0.05$ is the radius of the goal region. The baseline utilizes a fault detection scheme [13, Chapter 7.3] to detect and identify sensor faults by comparing EKF residuals against the threshold 0.1 and recomputes control input with an LQR controller. We then compare with our proposed HOCBF-based and CLF-based method. To keep the system remaining within safe region, we systematically construct the FT-SCBF with relative degree 1 by using the following class of sets

$$C^{(0)} = \{x \mid h^0(x_t) = a^T x_t + b \geq 0, \forall t \geq 0\}$$
$$C^{(1)} = \{x \mid h^1(x_t) = a^T F x_t + a^T x_t + b \geq 0, \forall t \geq 0\},$$

where $a^T = [0,1,0,0]$ and $b^T = [0,0.1,0,0]$. This differs from our previous work [56] which solves the problem by manually tuning the parameters and constructing the CBFs for high relative degree. In order to reach the goal region without violating the safety constraint, we choose the CLF

$$V(x) = (x_t - x_g)^T P_d (x_t - x_g) \tag{45}$$

where $P_d = \begin{pmatrix}\frac{1}{d}I & 0\\ 0 & I\end{pmatrix} P_L \begin{pmatrix}\frac{1}{d}I & 0\\ 0 & I\end{pmatrix}$, P_L is the solution of the Lyapunov equation $F^T P_L + P_L F = -I$, and I is the identity matrix [42], [67]. We set $\rho = 0.2$, $\eta = 0.8$ and $M = 2$ in the CLF constraint. The control input u_t is computed at each time step by solving (30) with $R = I$.

Simulation Result: The results are shown in Fig. 2. In Fig. 2(a), we plot the first two dimensions of the state, which describe the horizontal and vertical coordinates. Note that the robot stays in the safe region and eventually reaches the goal region, and hence satisfies safety and stability. As a comparison, the baseline can identify sensor faults but still resulted in a safety violation due to the slow response time of residual-based diagnosis.

B. Actuator Failure

In lateral control of an aircraft, lower yaw rate can renders smoother flight performance to avoid package damage or harsh passenger experience. We consider the lateral dynamics of Boeing 747 with state $x(t) = [[x]_1, [x]_2, [x]_3, [x]_4]^T$, where $[x]_1$ is the side-slip angle, $[x]_2$ is the yaw rate, $[x]_3$ is the roll rate, $[x]_4$ is the roll angle. In this case study we study yaw rate control with preset upper and lower boundary on yaw rate $\mathcal{C} = \{x_t : -0.025 \leq [x_t]_2 \leq 0.025, t \geq 0\}$ and reference point $x_t^T = [0,0,0,0]$. The dynamic system can be linearized and described as follows.

$$\begin{aligned}\dot{x}_t &= F x_t + G u_t^F + \mathbf{w}_t\\ y_t &= \begin{bmatrix}0 & 1 & 0 & 0\end{bmatrix} x_t + \mathbf{v}_t,\end{aligned} \tag{46}$$

where the process noise $\mathbf{w}_t \in \mathbb{R}^4$ has distribution $\mathcal{N}(0, \sigma_w I)$, where $\sigma_w = 0.001$ and the measurement noise $\mathbf{v}_t \in \mathbb{R}$ has distribution $\mathcal{N}(0, \sigma_v I)$, where $\sigma_v = 0.001$. The matrices F,

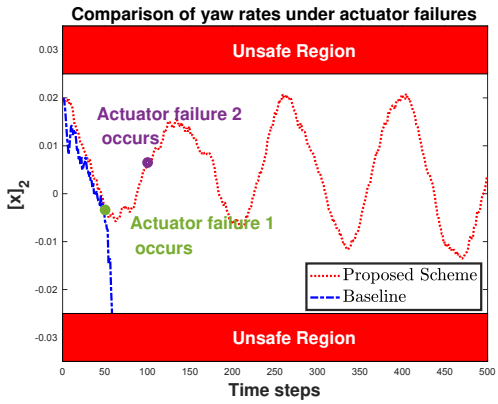


Fig. 3: Yaw rate comparison between FT-HOCBF and baseline on Boeing 747 lateral control system with rudder servo failures. The baseline enters unsafe region when actuator failure 1 happens, while proposed method remains safe under either actuator failure 1 and 2.

G are defined as

$$F = \begin{bmatrix} -0.0558 & -0.9968 & 0.0802 & 0.0415 \\ 0.598 & -0.115 & -0.0318 & 0 \\ -3.05 & 0.388 & -0.465 & 0 \\ 0 & 0.0805 & 1 & 0 \end{bmatrix},$$

$$G = \begin{bmatrix} 0.00729 & 0.01 & 0.005 \\ -0.475 & -0.5 & -0.3 \\ 0.153 & 0.2 & 0.1 \\ 0 & 0 & 0 \end{bmatrix}.$$

We assume that the system has two redundant actuators. The control input u_t^F contains three control signals representing three rudder servos, which may fail and output zero when failure happens. We simulate the actuator failure by denoting $u_t^F = Lu_t$ with two potential failure patterns $L_1 = diag([1, 0, 1])$ and $L_2 = diag([0, 1, 1])$. Failure L_1 starts from $2.5s$ to $5s$ and failure L_2 occurs since $5s$. We set $\rho = 0.2$, $\eta = 0.8$ and $M = 2$ in the CLF constraint.

The baseline utilizes a fault detection scheme [13, Chapter 7.3] to detect and identify actuator failures by comparing the EKF residuals against the threshold 0.02. Once identifying actuator failure L_i, the baseline recomputes control input based on reconfigured $G = GL_i$ by solving a CBF-CLF-based quadratic program. In order to keep the system remaining within safe region, we further define the following class of set for the upper and lower bound of yaw rate as

$$C_0 = \{x \mid h_0(x_t) = a_0^T x_t + b_0 \geq 0, \forall t \geq 0\}$$
$$C_1 = \{x \mid h_1(x_t) = a_1^T x_t + b_1 \geq 0, \forall t \geq 0\},$$

where $a_0^T = [0, 1, 0, 0]$, $a_1^T = [0, -1, 0, 0]$ and $b_0^T = b_1^T = [0, 0.025, 0, 0]$. We impose CBF with relative degree 0 as constraints to ensure safety.

Simulation Result: As is shown in Fig. 3, we compare the yaw rate trajectory between the baseline and the proposed safe control scheme. The baseline identifies actuator failures but still results in safety violations. The trajectory of the proposed scheme stays in the safe region and converges to 0, and hence satisfies safety and stability.

VIII. Conclusion

This paper proposed a new class of SCBFs with high relative degree for safety and stability of control systems under sensor faults and attacks. Our approach maintains a set of state estimators, excludes outlier estimates and ensures safety with a CBF-based approach. We then constructed an SCBF with high order degree for each state estimator, which guaranteed safety provided that a linear constraint on the control input was satisfied at each time step. We proposed a scheme for using additional state estimators to resolve conflicts between these constraints, and derived a scheme to verify the feasibility of SCBFs. We then showed how to compose our proposed HOSCBFs with CLFs to provide joint guarantees on safety and stability of a desired goal set under sensor faults and attacks. We proposed HOCBF-based approach to ensure safety of systems under all possible actuator failures and proposed an SOS-based scheme to verify the existence of control inputs satisfying HOCBF constraints. The proposed approach against sensor faults was validated on a wheeled mobile robot and our approach against actuator failures was validated on a Boeing 747 lateral control system. Future work in this area will include attacks that jointly affect sensors and actuators.

References

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

[2] Department of Homeland Security, *Department of Homeland Security Cyber-Physical Systems Page*, Jan 2022. [Online] Available: `https://www.dhs.gov/science-and-technology/cpssec` [Accessed: Jan 2022].

[3] C. Tomlin, G. J. Pappas, and S. Sastry, "Conflict resolution for air traffic management: A study in multiagent hybrid systems," *IEEE Transactions on Automatic Control*, vol. 43, no. 4, pp. 509–521, 1998.

[4] S. Prajna, A. Jadbabaie, and G. J. Pappas, "A framework for worst-case and stochastic safety verification using barrier certificates," *IEEE Transactions on Automatic Control*, vol. 52, no. 8, pp. 1415–1428, 2007.

[5] A. D. Ames, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs with application to adaptive cruise control," in *53rd IEEE Conference on Decision and Control*, pp. 6271–6278, IEEE, 2014.

[6] G. Tao, S. Chen, and S. M. Joshi, "An adaptive actuator failure compensation controller using output feedback," *IEEE Transactions on Automatic Control*, vol. 47, no. 3, pp. 506–511, 2002.

[7] A. Clark, "Control barrier functions for stochastic systems," *Automatica*, vol. 130, p. 109688, 2021.

[8] A. A. Amin and K. M. Hasan, "A review of fault tolerant control systems: advancements and applications," *Measurement*, vol. 143, pp. 58–68, 2019.

[9] J. Jiang and X. Yu, "Fault-tolerant control systems: A comparative study between active and passive approaches," *Annual Reviews in control*, vol. 36, no. 1, pp. 60–72, 2012.

[10] Y. Wang and X. Xu, "Observer-based control barrier functions for safety critical systems," in *2022 American Control Conference (ACC)*, pp. 709–714, IEEE, 2022.

[11] E. Daş and R. M. Murray, "Robust safe control synthesis with disturbance observer-based control barrier functions," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 5566–5573, IEEE, 2022.

[12] Y. Cheng, P. Zhao, and N. Hovakimyan, "Safe and efficient reinforcement learning using disturbance-observer-based control barrier functions," in *Learning for Dynamics and Control Conference*, pp. 104–115, PMLR, 2023.

[13] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, and J. Schröder, *Diagnosis and Fault-Tolerant Control*, vol. 2. Springer, 2006.

[14] Z. Chen, Y. Cao, S. X. Ding, K. Zhang, T. Koenings, T. Peng, C. Yang, and W. Gui, "A distributed canonical correlation analysis-based fault detection method for plant-wide process monitoring," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2710–2720, 2019.

[15] L. Li, H. Luo, S. X. Ding, Y. Yang, and K. Peng, "Performance-based fault detection and fault-tolerant control for automatic control systems," *Automatica*, vol. 99, pp. 308–316, 2019.

[16] A. M. Bardawily, M. Abdel-Geliel, M. Tamazin, and A. Nasser, "Sensors fault estimation, isolation and detection using MIMO extended Kalman filter for industrial applications," in *2017 10th international conference on electrical and electronics engineering (ELECO)*, pp. 944–948, IEEE, 2017.

[17] J.-S. Wang and G.-H. Yang, "Data-driven compensation method for sensor drift faults in digital PID systems with unknown dynamics," *Journal of Process Control*, vol. 65, pp. 15–33, 2018.

[18] D. Jung and E. Frisk, "Residual selection for fault detection and isolation using convex optimization," *Automatica*, vol. 97, pp. 143–149, 2018.

[19] J. Wang, C. Yang, H. Shen, J. Cao, and L. Rutkowski, "Sliding-mode control for slow-sampling singularly perturbed systems subject to markov jump parameters," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 12, pp. 7579–7586, 2020.

[20] Y. Yang, F. Chen, J. Lang, X. Chen, and J. Wang, "Sliding mode control of persistent dwell-time switched systems with random data dropouts," *Applied Mathematics and Computation*, vol. 400, p. 126087, 2021.

[21] Y.-A. Liu, S. Tang, Y. Liu, Q. Kong, and J. Wang, "Extended dissipative sliding mode control for nonlinear networked control systems via event-triggered mechanism with random uncertain measurement," *Applied Mathematics and Computation*, vol. 396, p. 125901, 2021.

[22] H. Yang, Y. Jiang, and S. Yin, "Fault-tolerant control of time-delay Markov jump systems with Ito stochastic process and output disturbance based on sliding mode observer," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 12, pp. 5299–5307, 2018.

[23] H. Li, H. Gao, P. Shi, and X. Zhao, "Fault-tolerant control of Markovian jump stochastic systems via the augmented sliding mode observer approach," *Automatica*, vol. 50, no. 7, pp. 1825–1834, 2014.

[24] H. Yang, C. Huang, B. Jiang, and M. M. Polycarpou, "Fault estimation and accommodation of interconnected systems: a separation principle," *IEEE Transactions on Cybernetics*, vol. 49, no. 12, pp. 4103–4116, 2018.

[25] Q. Zhao and J. Jiang, "Reliable state feedback control system design against actuator failures," *Automatica*, vol. 34, no. 10, pp. 1267–1272, 1998.

[26] X. Yu and Y. Zhang, "Design of passive fault-tolerant flight controller against actuator failures," *Chinese Journal of Aeronautics*, vol. 28, no. 1, pp. 180–190, 2015.

[27] W. Wang and C. Wen, "Adaptive actuator failure compensation control of uncertain nonlinear systems with guaranteed transient performance," *Automatica*, vol. 46, no. 12, pp. 2082–2091, 2010.

[28] S. Tong, T. Wang, and Y. Li, "Fuzzy adaptive actuator failure compensation control of uncertain stochastic nonlinear systems with unmodeled dynamics," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 3, pp. 563–574, 2013.

[29] X. J. Li and G. H. Yang, "Robust adaptive fault-tolerant control for uncertain linear systems with actuator failures," *IET control theory & applications*, vol. 6, no. 10, pp. 1544–1551, 2012.

[30] A. Girard, "Controller synthesis for safety and reachability via approximate bisimulation," *Automatica*, vol. 48, no. 5, pp. 947–953, 2012.

[31] P. Jagtap, S. Soudjani, and M. Zamani, "Formal synthesis of stochastic systems via control barrier certificates," *IEEE Transactions on Automatic Control*, vol. 66, no. 7, pp. 3097–3110, 2020.

[32] S. Kikukawa, "Consequence analysis and safety verification of hydrogen fueling stations using CFD simulation," *International Journal of Hydrogen Energy*, vol. 33, no. 4, pp. 1425–1434, 2008.

[33] H. A. Blom, J. Krystul, and G. Bakker, "A particle system for safety verification of free flight in air traffic," in *Proceedings of the 45th IEEE Conference on Decision and Control*, pp. 1574–1579, IEEE, 2006.

[34] G. Frehse, S. K. Jha, and B. H. Krogh, "A counterexample-guided approach to parameter synthesis for linear hybrid automata," in *International Workshop on Hybrid Systems: Computation and Control*, pp. 187–200, Springer, 2008.

[35] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3861–3876, 2016.

[36] J. Usevitch, K. Garg, and D. Panagou, "Strong invariance using control barrier functions: A Clarke tangent cone approach," in *59th IEEE Conference on Decision and Control (CDC)*, pp. 2044–2049, IEEE, 2020.

[37] W. Xiao, C. G. Cassandras, C. A. Belta, and D. Rus, "Control barrier functions for systems with multiple control inputs," in *2022 American Control Conference (ACC)*, pp. 2221–2226, IEEE, 2022.

[38] C. Santoyo, M. Dutreix, and S. Coogan, "Verification and control for finite-time safety of stochastic systems via barrier functions," in *2019 IEEE conference on control technology and applications (CCTA)*, pp. 712–717, IEEE, 2019.

[39] C. Santoyo, M. Dutreix, and S. Coogan, "A barrier function approach to finite-time stochastic system verification and control," *Automatica*, vol. 125, p. 109439, 2021.

[40] J. J. Choi, D. Lee, K. Sreenath, C. J. Tomlin, and S. L. Herbert, "Robust control barrier–value functions for safety-critical control," in *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 6814–6821, IEEE, 2021.

[41] W. Xiao and C. Belta, "Control barrier functions for systems with high relative degree," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 474–479, IEEE, 2019.

[42] Q. Nguyen and K. Sreenath, "Exponential control barrier functions for enforcing high relative-degree safety-critical constraints," in *2016 American Control Conference (ACC)*, pp. 322–328, IEEE, 2016.

[43] 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 3387–3395, 2019.

[44] 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.

[45] J. Choi, F. Castañeda, C. J. Tomlin, and K. Sreenath, "Reinforcement learning for safety-critical control under model uncertainty, using control lyapunov functions and control barrier functions," in *Robotics: Science and Systems (RSS)*, 2020.

[46] H. Ma, J. Chen, S. Eben, Z. Lin, Y. Guan, Y. Ren, and S. Zheng, "Model-based constrained reinforcement learning using generalized control barrier function," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4552–4559, IEEE, 2021.

[47] H. Zhang, Z. Li, and A. Clark, "Model-based reinforcement learning with provable safety guarantees via control barrier functions," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 792–798, IEEE, 2021.

[48] L. Wang, A. D. Ames, and M. Egerstedt, "Safety barrier certificates for collision-free multirobot systems," *IEEE Transactions on Robotics*, vol. 33, no. 3, pp. 661–674, 2017.

[49] X. Xu, J. W. Grizzle, P. Tabuada, and A. D. Ames, "Correctness guarantees for the composition of lane keeping and adaptive cruise control," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 3, pp. 1216–1229, 2017.

[50] E. Squires, P. Pierpaoli, and M. Egerstedt, "Constructive barrier certificates with applications to fixed-wing aircraft collision avoidance," in *2018 IEEE Conference on Control Technology and Applications (CCTA)*, pp. 1656–1661, IEEE, 2018.

[51] B. Xue, N. Zhan, and M. Fränzle, "Reach-avoid analysis for polynomial stochastic differential equations," *IEEE Transactions on Automatic Control*, vol. 69, no. 3, pp. 1882–1889, 2024.

[52] B. Xue, "Reach-avoid controllers synthesis for safety critical systems," *arXiv preprint arXiv:2302.14565*, 2023.

[53] J. Usevitch and D. Panagou, "Adversarial resilience for sampled-data systems under high-relative-degree safety constraints," *IEEE Transactions on Automatic Control*, vol. 68, no. 3, pp. 1537–1552, 2022.

[54] L. Niu, Z. Li, and A. Clark, "LQG reference tracking with safety and reachability guarantees under false data injection attacks," in *2019 American Control Conference (ACC)*, pp. 2950–2957, IEEE, 2019.

[55] Z. Li, L. Niu, and A. Clark, "LQG reference tracking with safety and reachability guarantees under unknown false data injection attacks," *IEEE Transactions on Automatic Control*, pp. 1–1, 2022.

[56] A. Clark, Z. Li, and H. Zhang, "Control barrier functions for safe CPS under sensor faults and attacks," in *2020 59th IEEE Conference on Decision and Control (CDC)*, pp. 796–803, IEEE, 2020.

[57] K. Garg, C. Dawson, K. Xu, M. Ornik, and C. Fan, "Model-free neural fault detection and isolation for safe control," *IEEE Control Systems Letters*, 2023.

[58] A. Clark, "Verification and synthesis of control barrier functions," in *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 6105–6112, IEEE, 2021.

[59] H. Heuser, *Lehrbuch der Analysis*. Springer-Verlag, 2013.

[60] K. Reif, S. Gunther, E. Yaz, and R. Unbehauen, "Stochastic stability of the continuous-time extended Kalman filter," *IEE Proceedings-Control Theory and Applications*, vol. 147, no. 1, pp. 45–52, 2000.

[61] C. Wang, Y. Meng, S. L. Smith, and J. Liu, "Safety-critical control of stochastic systems using stochastic control barrier functions," in *2021*

60th IEEE Conference on Decision and Control (CDC), pp. 5924–5931, IEEE, 2021.

[62] P. A. Parrilo, "Semidefinite programming relaxations for semialgebraic problems," *Mathematical programming*, vol. 96, no. 2, pp. 293–320, 2003.

[63] J. Matousek and B. Gärtner, *Understanding and Using Linear Programming*. Springer Science & Business Media, 2006.

[64] J. Yin, S. Khoo, Z. Man, and X. Yu, "Finite-time stability and instability of stochastic nonlinear systems," *Automatica*, vol. 47, no. 12, pp. 2671–2677, 2011.

[65] D. Ye and G.-H. Yang, "Adaptive fault-tolerant tracking control against actuator faults with application to flight control," *IEEE Transactions on control systems technology*, vol. 14, no. 6, pp. 1088–1096, 2006.

[66] Z. Chen, L. Li, and X. Huang, "Building an autonomous lane keeping simulator using real-world data and end-to-end learning," *IEEE Intelligent Transportation Systems Magazine*, 2018.

[67] A. D. Ames, K. Galloway, K. Sreenath, and J. W. Grizzle, "Rapidly exponentially stabilizing control Lyapunov functions and hybrid zero dynamics," *IEEE Transactions on Automatic Control*, vol. 59, no. 4, pp. 876–891, 2014.

Hongchao Zhang is a Ph.D. candidate in the Department of Electrical and Systems Engineering at Washington University in St. Louis (WashU). He received the B.Eng. degree from the Department of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2018 and the M.Sc. degree from the Department of Electrical and Computer Engineering, Worcester Polytechnic Institute (WPI) in 2020. He received the 2023 General Motors AutoDriving Security Award at the inaugural ISOC Symposium on Vehicle Security and Privacy at the Network and Distributed System Security Symposium (NDSS). His current research interests include control and security of cyber-physical systems and safe learning-based control.

Zhouchi Li received the B.Eng. degree from the Department of Electronic Science and Technology, Huazhong University of Science and Technology, Wuhan, China, in 2013 and the M.Sc. degree from the Department of Electrical and Computer Engineering, Worcester Polytechnic Institute in 2016. He received the Ph.D. degree in Electrical Engineering from the Secure Cyber-Physical Systems Lab, Department of Electrical and Computer Engineering, at Worcester Polytechnic Institute in 2023. He received the 2023 General Motors AutoDriving Security Award at the inaugural ISOC Symposium on Vehicle Security and Privacy at the Network and Distributed System Security Symposium (NDSS). His research interests include control and security of cyber-physical systems.

Andrew Clark is an Associate Professor in the Department of Electrical and Systems Engineering at Washington University in St. Louis. He received the B.S.E. degree in Electrical Engineering and the M.S. degree in Mathematics from the University of Michigan Ann Arbor in 2007 and 2008, respectively. He received the Ph.D. degree in Electrical Engineering from the Network Security Lab (NSL), Department of Electrical Engineering, at the University of Washington-Seattle in 2014. He is author or co-author of the IEEE/IFIP William C. Carter award-winning paper (2010), the WiOpt Best Paper (2012), the WiOpt Student Best Paper (2014), and the GameSec Outstanding Paper (2018), and was a finalist for the IEEE CDC 2012 Best Student Paper Award and the ACM ICCPS Best Paper Award (2016, 2018, 2020). He received an NSF CAREER award in 2020 and an AFOSR YIP award in 2022. His research interests include control and security of complex networks, safety of autonomous systems, submodular optimization, and control theoretic modeling of network security threats.