

Safe Output-Feedback Adaptive Optimal Control of Input-Constrained Control-Affine Nonlinear Systems

Tochukwu E. Ogri ^a, Muzaffar Qureshi ^a, Zachary I. Bell ^b,
Rushikesh Kamalapurkar ^a

^aDepartment of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida, USA

^bAir Force Research Laboratory, Florida, USA

Abstract

In this paper, a novel online, safe output-feedback, critic-only, adaptive optimal control framework is developed for safety-critical control of partially observable systems. The developed framework ensures system stability and safety, regardless of the lack of full-state measurements, while learning and implementing a near-optimal controller. The approach leverages linear matrix inequality-based observer design methods to efficiently search for observer gains for effective state estimation. Then, approximate dynamic programming is used to develop an approximate controller that uses simulated experiences to guarantee the safety and stability of the closed-loop system. Safety is enforced by adding a recentered robust Lyapunov-like barrier function to the cost function that effectively enforces safety constraints, even in the presence of state estimation errors. Lyapunov-based stability analysis is used to guarantee uniform ultimate boundedness of the trajectories of the closed-loop system and ensure safety. Simulation studies are performed to demonstrate the effectiveness of the developed method through two real-world safety-critical scenarios, specifically one ensuring that the state trajectories of a given system remain within a given set, and the other ensuring that the system avoids an obstacle.

Key words: Optimal control; adaptive control; state estimation; linear matrix inequality; control barrier function.

1 Introduction

In recent times, safe, learning-based optimal control (SLOC) has gained traction due to the need for robust control of safety-critical systems. While there are several SLOC strategies for nonlinear systems [6, 10, 12, 14, 15, 19, 22, 34, 35, 59], most of these methods rely on full-state feedback, which is not always available in real-world systems. Without full state measurement, SLOC techniques like those in [2, 13–15, 19, 22, 34, 35, 38, 54, 57, 63] could be implemented using a naïve replacement of the actual

state with state estimates, but the accompanying stability guarantees are no longer valid. Consequently, there is a need for adaptive output-feedback optimal control techniques for nonlinear systems that guarantee safety and stability regardless of limited state information from sensors.

Related work on SLOC

To achieve the safety of a dynamical system, a widely employed technique in the field of safe adaptive optimal control is to perform a nonlinear coordinate transformation, initially introduced in [21] and further extended by [22, 34, 59] for nonlinear control-affine systems. Barrier functions are employed to transform the constrained optimal control problem into an unconstrained one; however, the safety constraints they can handle are typically limited to box constraints.

Approaches such as [14, 35] utilize control barrier functions (CBFs), introduced in [7, 23, 55] to solve the SLOC problem. A drawback of these techniques is that there is a chance that the trajectories of the nonlinear system

* This research was supported in part by the Air Force Research Laboratory under contract numbers FA8651-24-1-0019 and FA8651-23-1-0006 and the Office of Naval Research under contract number N00014-21-1-2481. Any opinions, findings, or recommendations in this article are those of the author(s), and do not necessarily reflect the views of the sponsoring agencies.

Email addresses: tochukwu.ogri@ufl.edu (Tochukwu E. Ogri), muzaffar.qureshi@ufl.edu (Muzaffar Qureshi), zachary.bell.10@us.af.mil (Zachary I. Bell), rkamalapurkar@ufl.edu (Rushikesh Kamalapurkar).

converge to an undesired equilibrium point, as stated in [47] due to the added safety constraint. Results such as [15, 16] avoid such behavior by decoupling the safety and learning objectives. In particular, safe control commands that are closest to a desired stabilizing controller are implemented to guarantee safety in a minimally invasive fashion. The technique developed in this paper differs from [15] due to its use of a recentered robust Lyapunov-like control barrier function (RLCBF) to promote safety by incorporating the barrier function into the cost function. Re-centering of the barrier function, first developed in [56], guarantees system convergence to a desired equilibrium point regardless of safety objectives. The SLOC technique employed in this paper also differs from the strategies in [19] and [14], where a user-defined smooth scheduling function is employed to selectively penalize trajectories near the boundary of the safe set, whereas our formulation imposes safety as a hard constraint rather than a soft penalty, ensuring safety without relying on trajectory-dependent penalization.

More recently, [6, 10] proposed a novel approach to SLOC by minimizing a Lagrangian that includes both the cost function and a user-defined barrier Lyapunov function (BLF) to enforce state constraints. Specifically, [6] minimizes the constrained generalized Hamilton-Jacobi-Bellman (HJB) equation using a modified Galerkin successive approximation approach to approximate the value function and synthesize the safe controller. Conversely, [10] introduced an Actor-Critic-Identifier-Lagrangian algorithm designed to learn optimal control policies online while ensuring safety.

Recent advances in multi-objective safety-critical control seek to balance safety with other performance objectives. The authors in [4, 5] achieve this through the use of barrier states (BaS) to embed safety conditions directly into control design, creating safety-embedded systems where stabilizing the augmented system ensures safe operation of the original system. The safety-embedded system framework in [5] allows the use of existing control techniques to synthesize safe controllers that enforce safety, handle input constraints, and guarantee input-to-state safe stability. In [4], the BaS framework was combined with the State-Dependent Riccati Equation (SDRE) approach, yielding the BaS-SDRE method, which synthesizes safe, nearly optimal feedback laws in real time.

While effective, the existing SLOC methods described above utilize full-state feedback and cannot be trivially extended to systems where full state information is unavailable. For example, as demonstrated in Section 4, the methods in [15, 16] cannot ensure safety if naïvely implemented using state estimates.

SLOC using output-feedback

Early output-feedback optimal control methods developed in results such as [57, 58] assume that the pseudoinverse of the output matrix exists, which implicitly restricts the technique to systems where the number of outputs is larger than the number of states. More recent output-feedback SLOC techniques such as [34] also rely on structural restrictions imposed on the output matrix so that each output is a direct measurement of a state variable.

To avoid assumptions on the form or rank of the output matrix, our preliminary work in [40, 41] developed an observer for control-affine nonlinear systems using the multiplier matrix approach, a technique introduced in works such as [39, 43, 45, 46, 51, 52, 60–62]. For certain nonlinear systems, this approach guarantees asymptotic convergence of estimation errors under Lyapunov-derived conditions. For example, [45] established sufficient stability conditions using eigenvalues and eigenvectors of the linear stability matrix for Lipschitz systems. Building on this, [61] reformulated the error dynamics as an LPV system via the differential mean value theorem (DMVT) and applied Lyapunov analysis to derive linear matrix inequalities (LMIs) for observer gains. Later works [43, 52] developed LMI-based designs for bounded Jacobian systems. In particular, [52] extended the Luenberger-type observer [8] by adding a third gain to cancel nonconvex terms in the LMI stability condition. Although LMI-based methods [39, 51, 52] ensure uniform asymptotic convergence, they remain restricted to specific classes of nonlinear systems. More recently, [24] exploited the combined benefits of matrix multipliers-based LMIs and NNs to develop an exponentially stable neuro-adaptive observer for state and nonlinear function estimation in Lipschitz systems with partially modeled process dynamics.

In our previous work, we developed novel observers for control-affine systems whose dynamics have bounded Jacobians. The observer developed in [41] was integrated with an online critic-only adaptive control architecture to learn a controller that keeps an input-constrained nonlinear system stable during the learning phase. The architecture in [41] differs from the existing neural network (NN)-based observers in literature like [1, 20, 30, 57, 58] whose convergence analysis relies solely on negative terms that result from a σ -modification-like term added to the weight update laws. In these results, similar to adaptive control, the convergence of the observer weights to their true values cannot be expected, and the convergence of state estimates to the true state is not robust to disturbances and approximation errors.

While our previous works [41] and [40] extend the LMI-based observer design to a broader class of nonlinear systems in the control-affine form, they do not address sys-

tem safety. Incorporating state estimators into SLOC architectures is challenging, as estimation errors can propagate through value function updates, potentially destabilizing the learning process or resulting in suboptimal policies. The fact that safety constraints must be satisfied even in the presence of estimation errors necessitates robust formulations. While robustification of CBFs to ensure safety in the presence of estimation errors has been studied in results such as [3, 17], to the best of our knowledge, there are no existing output-feedback SLOC methods that incorporate observer-based state estimation with formal safety guarantees under input constraints. The development of a unified framework that couples observer design with robust CBF-based safety filters for safe learning of optimal controllers under partial observations and input constraints is thus well-motivated.

Contribution

This paper focuses on extending our preliminary work in [41] and [40] to address the problem of safety guarantees in online nonlinear output-feedback optimal control. To estimate the unknown states of the nonlinear control-affine system required to compute the controller, we design a nonlinear observer and use Lyapunov-based analysis to develop LMIs that characterize the observer gain matrices. Due to the use of state estimates in the control law, CBFs fail to guarantee safety of the closed-loop system, as demonstrated in Figure 1 of Section 4. In this paper, a robust control barrier function (RCBF) is developed to guarantee safety in the presence of state estimation errors using a time-varying robustifying term that relies on a shrinking bound on the state estimation error.

The time-varying term avoids an overly conservative controller behavior, which is a common pitfall of many robust CBF-based controllers [17, 18]. The key technical challenge in the use of RCBFs to solve SLOC problems is the fact that the incorporation of the RCBF constraint results in a time-varying optimization problem. Due to non-stationarity, such problems are difficult to solve online. In this paper, we address this technical challenge by utilizing the fact that the optimal control problem can be made time-invariant by using an augmented system state. We then develop a critic-only SLOC architecture to learn the optimal value function as a function of the augmented state to yield a safe and approximately optimal controller. Unlike actor-critic methods popular in the literature [26, 27], the framework in this paper utilizes a critic-only structure to provide an approximate solution to the HJB equation, which requires identifying fewer free parameters.

Organization

The rest of the paper is organized as follows: Section 2 contains the problem formulation which introduces the system and some key assumptions about the system, Section 3 presents the design of the state observer for the nonlinear output-feedback system with a linear measurement model, Section 4 introduces the concept of robust barrier functions and develops conditions sufficient to guarantee safety despite state measurement errors, Section 5 presents the control design using an observer-based SLOC method, Section 6 contains stability analysis of the developed control architecture, Section 7 presents simulation studies on two systems with distinct safety constraints and we compare the effectiveness of our approach against similar methods in the literature, and Section 8 concludes the paper.

1.1 Notation

In this paper, the symbol $\|\cdot\|$ represents the 2-norm for vectors and the induced 2-norm for matrices. I_n denotes an $n \times n$ identity matrix and 0_n denotes the $n \times n$ zero matrix. For a square matrix $A \in \mathbb{R}^{n \times n}$, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote its maximum and minimum eigenvalues, respectively; $A^\top$ denotes its transpose; $A > 0$ ($A < 0$) indicates that A is a positive definite (negative definite) matrix; and $A \geq 0$ ($A \leq 0$) indicates that A is a positive semi-definite (negative semi-definite) matrix. The total derivative $\frac{\partial f(x)}{\partial x}$ is denoted by $\nabla_x f(x)$ and the partial derivative $\frac{\partial f(x,y)}{\partial x}$ is denoted by $\nabla_x f(x,y)$. The convex hull of a set $\mathcal{O}$ is denoted as $\text{Conv}(\mathcal{O})$. The symbol $(\cdot)_{(\cdot)}$ denotes the elements of the function or variable $(\cdot)$ at the indices indicated by the subscript.

2 Problem Formulation and Preliminaries

2.1 System description

Consider nonlinear dynamical systems of the form

$$\dot{x} = f(x) + g(x)u, \quad y = Cx, \quad (1)$$

where $x \in \mathbb{R}^n$ is the system state, $u \in \mathbb{R}^m$ is the control input, $C \in \mathbb{R}^{q \times n}$ is the output matrix, and $y \in \mathbb{R}^q$ is the measured output. The functions $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times m}$ denote the drift vector and the control effectiveness matrix, respectively. The following assumptions on the functions f and g are needed to facilitate the development and analysis of the method in this paper.

Assumption 1 *The functions f and g are known, differentiable functions. Given any compact ball $\mathcal{X} \subset \mathbb{R}^n$ and a compact set $\mathcal{U} \subset \mathbb{R}^m$ the partial derivatives satisfy the element-wise bounds*

$$(K_{f_1})_{i,j} \leq \nabla_{(x)_j}(f(x))_i \leq (K_{f_2})_{i,j}, \quad (2)$$

$$(K_{g_1})_{i,j} \leq [\nabla_{(x)_j}(g(x))_{i,k}]u_k \leq (K_{g_2})_{i,j}, \quad (3)$$

for all $x \in \mathcal{X} \oplus \mathcal{X}$, $u \in \mathcal{U}$, $i, j = 1, \dots, n$, and $k = 1, \dots, m$, where the operator $\oplus$ denotes the Minkowski sum. Furthermore, the function f satisfies $f(0) = 0$.

Remark 2 Conditions similar to those in Assumption 1 are commonly required in several observer design schemes (see, e.g., [46, 52, 61, 62]). However, while Jacobian bounds are typically assumed to be global in those papers, here the bounds are assumed to be local, which is a less conservative assumption.

To develop the notion of safety, consider the set $\mathcal{S} \subset \mathbb{R}^n$ defined as the zero super-level set of a continuously differentiable function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$\mathcal{S} = \{x \in \mathbb{R}^n : h(x) \geq 0\}, \quad (4)$$

$$\partial\mathcal{S} = \{x \in \mathbb{R}^n : h(x) = 0\}, \quad (5)$$

$$\text{Int}(\mathcal{S}) = \{x \in \mathbb{R}^n : h(x) > 0\}, \quad (6)$$

where $\partial\mathcal{S}$ and $\text{Int}(\mathcal{S})$ represent the boundary and interior of the set $\mathcal{S}$, respectively.

The objective is to design an observer to estimate the system state online, using input-output measurements, and to simultaneously synthesize and utilize a controller that minimizes the cost functional defined in (30), under the saturation constraint $|(u)_k(t)| \leq \bar{u} > 0$ for all $t \geq t_0$ and $k = 1, \dots, m$ (i.e., $u \in \mathcal{U}$ where $\mathcal{U}$ is the m -dimensional hypercube with side length $2\bar{u}$), and the safety constraint $x^u(t; t_0, x_0) \in \mathcal{S}$ for all $t \geq t_0$, and sufficiently small x_0 (see Theorem 12), while ensuring local uniform ultimate boundedness of the trajectories of the closed-loop system.

In the following section, a state observer inspired by the extended Luenberger observer [8, 9, 44, 52] is developed to generate estimates of x .

3 Observer design using Jacobian bounds

The state observer is designed under the assumption that the system state remains within a compact convex set $\mathcal{X}$ and that the control input also remains within a compact set $\mathcal{U}$. CBFs and control saturation are employed in the following sections to ensure that the control signal remains in $\mathcal{U}$ and the system state remains in $\mathcal{X}$, thus ensuring that the Jacobian bounds in Assumption 1 remain valid along the system trajectories.

3.1 State observer

Using the fact that $f(x) = K_{f_1}x + (f(x) - K_{f_1}x)$ and $g(x) = K_{g_1}x + (g(x) - K_{g_1}x)$, the nonlinear dynamics in (1) can be expressed in the form $\dot{x} = Ax + \mathcal{F}(x) +$

$\mathcal{G}_u(x, u)$, where $A := K_{f_1} + K_{g_1} \in \mathbb{R}^{n \times n}$, $\mathcal{F}(x) := -K_{f_1}x + f(x)$, and $\mathcal{G}_u(x, u) := -K_{g_1}x + \sum_{i=1}^m g_i(x)(u)_i$. Under Assumption 1, the derivatives of $\mathcal{F}$ and $\mathcal{G}$ satisfy the element-wise inequalities

$$0 \leq \nabla_{(x)_j}(\mathcal{F}(x))_i \leq (K_{f_2})_{i,j} - (K_{f_1})_{i,j}, \text{ and} \quad (7)$$

$$0 \leq \nabla_{(x)_j}(\mathcal{G}_u(x, u))_i \leq (K_{g_2})_{i,j} - (K_{g_1})_{i,j}, \quad (8)$$

where $i, j = 1, \dots, n$. Since the Jacobian bounds specified in Assumption 1 are local, even if the system state x remains within the compact ball $\mathcal{X}$ for all $t \geq t_0$ through the use of CBFs, there is no guarantee that the state estimate $\hat{x} \in \mathbb{R}^n$ will remain in $\mathcal{X}$ for all $t \geq t_0$. Inspired by the results of [62], a state observer which utilizes the well-known Hilbert projection theorem [36] is designed as

$$\begin{aligned} \dot{\hat{x}} = & f(\mathbf{Pr}(\hat{x}) + L_1(y - C\mathbf{Pr}(\hat{x}))) \\ & + g[\mathbf{Pr}(\hat{x}) + L_2(y - C\mathbf{Pr}(\hat{x}))]u + L_3(y - C\mathbf{Pr}(\hat{x})), \end{aligned} \quad (9)$$

where L_1, L_2 , and $L_3 \in \mathbb{R}^{n \times q}$ are observer gains, $L_1(y - C\mathbf{Pr}(\hat{x}))$ and $L_2(y - C\mathbf{Pr}(\hat{x}))$ are injection terms, $L_3(y - C\mathbf{Pr}(\hat{x}))$ is a linear correction term, and $\mathbf{Pr} : \mathbb{R}^n \rightarrow \mathcal{X}$ is the Hilbert projection, which is defined as $\mathbf{Pr}(\hat{x}) = \arg \min_{s \in \mathcal{X}} \|\hat{x} - s\|$. The following Hilbert projection theorem ensures the existence and uniqueness of $\mathbf{Pr}(\hat{x})$.

Theorem 3 [36, Hilbert projection theorem] *Let $\mathcal{X}$ be a closed convex subset of $\mathbb{R}^n$. For any $\hat{x} \in \mathbb{R}^n$, there exists a unique element $\mathbf{Pr}(\hat{x}) \in \mathcal{X}$ such that $\|\hat{x} - \mathbf{Pr}(\hat{x})\| = \inf_{s \in \mathcal{X}} \|\hat{x} - s\|$.*

Properties of $\mathbf{Pr}(\hat{x})$, highlighted in [62, Lemma 1], are used in this development to guarantee that the local Jacobian bounds remain valid for all $t \geq t_0$, notwithstanding that the bounds on the Jacobian of the functions f and g are not global. The state estimation error is defined as $\tilde{x} := x - \hat{x}$, and the estimation error dynamics can be expressed as $\dot{\tilde{x}} = (A - L_3C)(\mathbf{Pr}(x) - \mathbf{Pr}(\hat{x})) + \mathcal{F}(\mathbf{Pr}(x)) + \mathcal{G}_u(\mathbf{Pr}(x), u) - \mathcal{F}[\mathbf{Pr}(\hat{x}) + L_1C(\mathbf{Pr}(x) - \mathbf{Pr}(\hat{x}))] - \mathcal{G}_u[\mathbf{Pr}(\hat{x}) + L_2C(\mathbf{Pr}(x) - \mathbf{Pr}(\hat{x})), u]$. If $x \in \mathcal{X}$, then $Ax = A\mathbf{Pr}(x)$, $y - C\mathbf{Pr}(\hat{x}) = C(\mathbf{Pr}(x) - \mathbf{Pr}(\hat{x}))$, $\mathcal{F}(x) = \mathcal{F}(\mathbf{Pr}(x))$ and $\mathcal{G}_u(x, u) = \mathcal{G}_u(\mathbf{Pr}(x), u)$. Proposition 1 from [62] implies that there exists a function $(\Phi)_{i,j} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ such that for all $x \in \mathcal{X}$ and $\hat{x} \in \mathbb{R}^n$ $\mathbf{Pr}(x) - \mathbf{Pr}(\hat{x}) = \left[\sum_{i=1}^n \sum_{j=1}^n (\Phi(x, \hat{x}))_{i,j} (H)_{i,j} \right] (x - \hat{x})$, and $|(\Phi(x, \hat{x}))_{i,j}| \leq 1$, where $(H)_{i,j} := \mathbf{e}(i)\mathbf{e}^\top(j)$ and $\mathbf{e}(i) \in \mathbb{R}^n$ denotes the i -th standard basis vector in $\mathbb{R}^n$. Therefore, the observer error dynamics can be equivalently expressed as

$$\dot{\tilde{x}} = \mathcal{A}_c(\theta)\tilde{x} + \Delta_{\mathcal{F}}(x, \hat{x}) + \Delta_{\mathcal{G}_u}(x, \hat{x}, u), \quad (10)$$

where $\mathcal{A}_c(\theta) := (A - L_3C)\theta$ with the matrix parameter $\theta(t) := \sum_{i=1}^n \sum_{j=1}^n (\Phi(x, \hat{x}))_{i,j} (H)_{i,j}$ belonging to the set $\mathcal{O}$ which is defined as $\mathcal{O} := \{M \in \mathbb{R}^{n \times n} : M = \sum_{i=1}^n \sum_{j=1}^n (\Phi(x, \hat{x}))_{i,j} (H)_{i,j}, 0 \leq (\Phi(x, \hat{x}))_{i,j} \leq 1\}$, $\Delta_{\mathcal{F}}(x, \hat{x}) := \mathcal{F}(\mathbf{Pr}(x)) - \mathcal{F}(\mathbf{Pr}(\hat{x}) + L_1C(\mathbf{Pr}(x) - \mathbf{Pr}(\hat{x})))$, and $\Delta_{\mathcal{G}_u}(x, \hat{x}, u) := \mathcal{G}_u(\mathbf{Pr}(x), u) - \mathcal{G}_u[\hat{x} + L_2C(\mathbf{Pr}(x) - \mathbf{Pr}(\hat{x})), u]$.

3.2 Formulation of LMI conditions

The functions $\mathcal{F}$ and $\mathcal{G}_u$ can be equivalently be expressed as $\mathcal{F}(x) = \sum_{i=1}^n \mathbf{e}(i)(\mathcal{F}(x))_i$, and $\mathcal{G}_u(x, u) = \sum_{i=1}^n \mathbf{e}(i)(\mathcal{G}_u(x, u))_i$, respectively. By the DMVT [61, Theorem 2.3], for all $x \in \mathcal{X}$ and $\hat{x} \in \mathbb{R}^n$, there exist $(s_f)_{i,j} \in \text{Conv}(\mathbf{Pr}(x), \mathbf{Pr}(\hat{x}) + L_1C(\mathbf{Pr}(x) - \mathbf{Pr}(\hat{x})))$ such that the difference function $\Delta_{\mathcal{F}}$ can be expressed as $\Delta_{\mathcal{F}}(x, \hat{x}) = \left[\sum_{i=1}^n \sum_{j=1}^n (H)_{i,j} (\varphi_f(x, \hat{x}))_{i,j} \right] (\mathbf{I}_n - L_1C)(x - \hat{x})$, where the function $(\varphi_f)_{i,j} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is defined as

$$(\varphi_f)_{i,j}(x, \hat{x}) := \begin{cases} 0, & \text{if } (x)_j = (\hat{x})_j, \\ \frac{\mathbf{Pr}((x)_j) - \mathbf{Pr}((\hat{x})_j)}{(x)_j - (\hat{x})_j} \nabla_{(x)_j} (\mathcal{F}((s_f)_{i,j}))_i, & \text{if } (x)_j \neq (\hat{x})_j. \end{cases} \quad (12)$$

The subsequent development requires the property that for all $x \in \mathcal{X}$ and $\hat{x} \in \mathbb{R}^n$,

$$\begin{aligned} L_1C(\mathbf{Pr}(x) - \mathbf{Pr}(\hat{x})) &\in \mathcal{X} \oplus \mathcal{X} \text{ and} \\ L_2C(\mathbf{Pr}(x) - \mathbf{Pr}(\hat{x})) &\in \mathcal{X} \oplus \mathcal{X}. \end{aligned} \quad (13)$$

Since $\mathcal{X} \oplus \mathcal{X}$ is a ball, if the induced 2-norm of L_1C and L_2C is less than 1, then L_1C and L_2C map $\mathcal{X} \oplus \mathcal{X}$ to itself, and (13) follows. These norm constraints have been incorporated into the LMI condition developed in Theorem 4 to ensure the validity of (13), and as a result, the inequalities in (15) and (16).

Under Assumption 1, the partial derivative $\nabla_{(x)_j} (\mathcal{F}((s_f)_{i,j}))_i$ is bounded, element-wise, by the bounds in (7) (i.e $0 \leq \nabla_{(x)_j} (\mathcal{F}((s_f)_{i,j}))_i \leq (K_{f_2})_{i,j} - (K_{f_1})_{i,j}, \forall i, j = 1, \dots, n$).¹ Similarly, by the DMVT, for all $x \in \mathcal{X}$ and $\hat{x} \in \mathbb{R}^n$, there exist $(s_g)_{i,j} \in \text{Conv}(\mathbf{Pr}(x), \mathbf{Pr}(\hat{x}) + L_2C(\mathbf{Pr}(x) - \mathbf{Pr}(\hat{x})))$ such that the difference function $\Delta_{\mathcal{G}_u}$ can be expressed as $\Delta_{\mathcal{G}_u}(x, \hat{x}, u) = \left[\sum_{i=1}^n \sum_{j=1}^n (H)_{i,j} (\varphi_g(x, \hat{x}, u))_{i,j} \right] \times (\mathbf{I}_n - L_2C)(x - \hat{x})$, where $(\varphi_g)_{i,j} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is defined as

$$(\varphi_g)_{i,j} := \begin{cases} 0, & \text{if } (x)_j = (\hat{x})_j, \\ \frac{\mathbf{Pr}((x)_j) - \mathbf{Pr}((\hat{x})_j)}{(x)_j - (\hat{x})_j} \nabla_{(x)_j} (\mathcal{G}_u((s_g)_{i,j}, u))_i, & \text{if } (x)_j \neq (\hat{x})_j, \end{cases} \quad (14)$$

¹ The notation $\mathbf{Pr}((x)_j)$ is a slight abuse of notation used to simplify the expression. To be precise, $\mathbf{Pr}((x)_j)$ should be considered as $\mathbf{Pr}((x)_j \mathbf{e}(j))$ and $\mathbf{Pr}((\hat{x})_j)$ should be considered as $\mathbf{Pr}((\hat{x})_j \mathbf{e}(j))$.

and under Assumption 1, the partial derivative $\nabla_{(x)_j} (\mathcal{G}_u((s_g)_{i,j}, u))_i$ is bounded element-wise by the bounds in (8). Using the monotonicity and 1-Lipschitz property of the Hilbert projection [62], $0 \leq \frac{\mathbf{Pr}((x)_j) - \mathbf{Pr}((\hat{x})_j)}{(x)_j - (\hat{x})_j} \leq 1$. The representations in (12) and (14) both leverage the DMVT to ensure the existence of points $(s_f)_{i,j}$ and $(s_g)_{i,j}$, which lie along convex paths between $\mathbf{Pr}(x)$ and affine perturbations of $\mathbf{Pr}(\hat{x})$, determined by the matrices C , L_1 , and L_2 , at which the Jacobian of $\mathcal{F}$ and $\mathcal{G}_u$ are evaluated, respectively.

Using Assumption 1, the DMVT, and the bounds in (7), the difference function $\Delta_{\mathcal{F}}$ is bounded as (cf. [41])

$$\mathbf{0}_n \leq \Delta_{\mathcal{F}}(x, \hat{x}) \leq (K_{f_2} - K_{f_1}) (\mathbf{I}_n - L_1C)(x - \hat{x}), \quad (15)$$

and similarly, using the bounds in (8), the difference function $\Delta_{\mathcal{G}_u}$ is bounded as

$$\mathbf{0}_n \leq \Delta_{\mathcal{G}_u}(x, \hat{x}, u) \leq (K_{g_2} - K_{g_1}) (\mathbf{I}_n - L_2C)(x - \hat{x}). \quad (16)$$

In the following theorem, a bound on the Lie derivative of a candidate Lyapunov function along the flow of the observer error system in (10) is derived. The bound is subsequently used in Theorem 12 to establish local ultimate boundedness of the closed-loop system.

Theorem 4 *Let $V_x(\tilde{x}) := \tilde{x}^\top P \tilde{x}$ be a candidate Lyapunov function where $P \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix. If*

- H1 *Assumption 1 holds and*
- H2 *there exist a symmetric positive definite matrix $P \in \mathbb{R}^{n \times n}$, and observer gains L_1 , L_2 , and $R \in \mathbb{R}^{n \times q}$ that satisfy the constraints $\|L_1C\| \leq 1$, $\|L_2C\| \leq 1$, and the matrix inequality,*

$$M < 0, \quad \forall \theta \in \mathcal{O}, \quad (17)$$

where M is defined in (11) at the top of Page 6, $\alpha \in \mathbb{R}_{>0}$ is a constant learning rate, $A_\theta := A\theta$ and $C_\theta := C\theta$,

then for all $\tilde{x} \in \mathbb{R}^n$, $x \in \mathcal{X}$, and $u \in \mathcal{U}$, the Lie derivative of V_x along the flow of (10) satisfies

$$\dot{V}_x(\tilde{x}, x, u) \leq -2\alpha V_x(\tilde{x}). \quad (18)$$

PROOF. Note that V_x satisfies the inequality $\lambda_{\min}(P) \|\tilde{x}\|^2 \leq V_x(\tilde{x}) \leq \lambda_{\max}(P) \|\tilde{x}\|^2$. Taking the Lie derivative of V_x along the flow of (10) yields $\dot{V}_x(\tilde{x}, x, u) = \tilde{x}^\top (\mathcal{A}_c^\top(\theta)P + \mathcal{A}_c(\theta)P) \tilde{x} + \Delta_{\mathcal{F}}^\top(x, \hat{x})P\tilde{x} + \tilde{x}^\top P\Delta_{\mathcal{F}}(x, \hat{x}) + \Delta_{\mathcal{G}_u}^\top(x, \hat{x}, u)P\tilde{x} + \tilde{x}^\top P\Delta_{\mathcal{G}_u}(x, \hat{x}, u)$. Using the Cauchy-Schwarz inequality and Young's inequality, it can be concluded that $\Delta_{\mathcal{F}}^\top(x, \hat{x})P\tilde{x} +$

$$M := \begin{bmatrix} A_\theta^\top P + P A_\theta - C_\theta^\top R^\top - R C_\theta + 2\alpha P & \sqrt{2}P + (I_n - L_1 C)^\top (K_{f_2} - K_{f_1})^\top + (I_n - L_2 C)^\top (K_{g_2} - K_{g_1})^\top \\ \sqrt{2}P + (K_{f_2} - K_{f_1})(I_n - L_1 C) + (K_{g_2} - K_{g_1})(I_n - L_2 C) & -3I_n \end{bmatrix}. \quad (11)$$

$\tilde{x}^\top P \Delta_{\mathcal{F}}(x, \hat{x}) \leq \|P\tilde{x}\|^2 + \|\Delta_{\mathcal{F}}(x, \hat{x})\|^2$. By applying the bound from (15), it can be concluded that for all $x \in \mathcal{X}$ and $\hat{x} \in \mathbb{R}^n$

$$\Delta_{\mathcal{F}}^\top(x, \hat{x}) P \tilde{x} + \tilde{x}^\top P \Delta_{\mathcal{F}}(x, \hat{x}) \leq \tilde{x}^\top P^2 \tilde{x} + \tilde{x}^\top (I_n - L_1 C)^\top \times (K_{f_2} - K_{f_1})^\top (K_{f_2} - K_{f_1}) (I_n - L_1 C) \tilde{x}. \quad (19)$$

Following the same steps, by using Cauchy-Schwarz, Young's inequality and applying the bound in (16), it can be concluded that for all $x \in \mathcal{X}$, $\hat{x} \in \mathbb{R}^n$ and $u \in \mathcal{U}$,

$$\Delta_{\mathcal{G}_u}^\top(x, \hat{x}, u) P \tilde{x} + \tilde{x}^\top P \Delta_{\mathcal{G}_u}(x, \hat{x}, u) \leq \|P\tilde{x}\|^2 + \|\Delta_{\mathcal{G}_u}(x, \hat{x}, u)\|^2 \leq \tilde{x}^\top P^2 \tilde{x} + \tilde{x}^\top (I_n - L_2 C)^\top \times (K_{g_2} - K_{g_1})^\top (K_{g_2} - K_{g_1}) (I_n - L_2 C) \tilde{x}. \quad (20)$$

Therefore, if the algebraic Ricatti inequality (ARI)

$$\mathcal{A}_c^\top(\theta)P + \mathcal{A}_c(\theta)P + 2P^2 + (I_n - L_1 C)^\top (K_{f_2} - K_{f_1})^\top \times (K_{f_2} - K_{f_1}) (I_n - L_1 C) + (I_n - L_2 C)^\top (K_{g_2} - K_{g_1})^\top \times (K_{g_2} - K_{g_1}) (I_n - L_2 C) < 0 \quad (21)$$

is satisfied, then the Lie derivative is negative definite. Splitting the ARI in (21) into three inequalities and applying the Schur complement Lemma [11] on each inequality, the first quadratic inequality, $\mathcal{A}_c^\top(\theta)P + \mathcal{A}_c(\theta)P + 2P^2 < 0$, can be equivalently expressed in matrix form as

$$\begin{bmatrix} \mathcal{A}_c^\top(\theta)P + \mathcal{A}_c(\theta)P & \sqrt{2}P \\ \sqrt{2}P & -I_n \end{bmatrix} < 0, \quad (22)$$

the second inequality, $(I_n - L_1 C)^\top (K_{f_2} - K_{f_1})^\top (K_{f_2} - K_{f_1}) (I_n - L_1 C) \leq 0$, can be expressed as

$$\begin{bmatrix} \mathbf{0}_n & (I_n - L_1 C)^\top (K_{f_2} - K_{f_1})^\top \\ (K_{f_2} - K_{f_1})(I_n - L_1 C) & -I_n \end{bmatrix} \leq 0, \quad (23)$$

and the third inequality, $(I_n - L_2 C)^\top (K_{g_2} - K_{g_1})^\top (K_{g_2} - K_{g_1}) (I_n - L_2 C) \leq 0$, can be expressed as

$$\begin{bmatrix} \mathbf{0}_n & (I_n - L_2 C)^\top (K_{g_2} - K_{g_1})^\top \\ (K_{g_2} - K_{g_1})(I_n - L_2 C) & -I_n \end{bmatrix} \leq 0. \quad (24)$$

Combining the matrix inequalities in (22), (23), and

(24), it can be concluded that if the LMI in (17) is satisfied for some constant $\alpha \in \mathbb{R}_{>0}$, then the Lie derivative is bounded as described in (18).

Remark 5 The observer gain L_3 can be obtained via the typical variable substitution $L_3 = P^{-1}R$.

4 Safe control using RCBFs

The state-feedback SLOC techniques developed by [14] and [35], which embed the CBF in the cost function of the optimal control problem to enforce safety constraints, fail in the output-feedback case when implemented via naïve certainty equivalence without introducing a robustifying term. Figure 1 illustrates this failure using the system from the simulation study in Section 7.

Inspired by RCBF techniques utilized in [3, 18, 23, 31, 33, 49] to ensure safety in the presence of state estimation errors, this section introduces RLCBFs which will be utilized in this paper to achieve safety objectives. The purpose of this section is to motivate the design of the RLCBF introduced in (33) and to provide the rationale behind its selection. It should be noted that the stability results presented herein are not directly utilized in the control design; they only justify the choice of the final RLCBF used in the control design, as detailed in Section 5.

4.1 Control Barrier Functions

The following definitions from [7] formalize the concept of CBFs, which are employed to enforce the safety constraints.

Definition 6 [7] Given a set $\mathcal{S} \subset \mathcal{X} \subset \mathbb{R}^n$ as defined in (4)–(6) and the set $\mathcal{U} \subset \mathbb{R}^m$, a continuously differentiable function $h : \mathcal{X} \rightarrow \mathbb{R}$ is a CBF for (1), if there exist a class $\mathcal{K}$ function ν such that, $\sup_{u \in \mathcal{U}} \nabla_x h(x) (f(x) + g(x)u) \geq -\nu(h(x))$, $\forall x \in \mathcal{S}$.

The following theorem from [7], stated here for completeness, establishes conditions that must be satisfied for the existence of control policies that guarantee safety.

Theorem 7 [7, Theorem 1] Given the sets $\mathcal{S} \subset \mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{U} \subset \mathbb{R}^m$, along with a continuously differentiable CBF $h : \mathcal{X} \rightarrow \mathbb{R}$ for the system in (1) such that $\nabla h(x) \neq 0$ for all $x \in \partial\mathcal{S}$ and a set defined as

$$\mathcal{K}_{\text{CBF}} := \left\{ u \in \mathcal{U} : \nabla_x h(x) (f(x) + g(x)u) \geq -\nu(h(x)) \right\}, \quad (25)$$

any control policy $\pi_\phi : \mathcal{X} \times \mathbb{R}_{\geq 0} \rightarrow \mathcal{U}$ that is locally Lipschitz continuous in x , piecewise continuous in t , and satisfies $\pi_\phi(x, t) \in \mathcal{K}_{\text{CBF}}$ for all $x \in \mathcal{X}$ and $t \geq t_0$, guarantees forward invariance of the set $\mathcal{S}$.

4.2 Safety using RCBFs

Since state estimates are used for feedback instead of the true state, the performance of CBFs in guaranteeing the safety of the system in (1) may be degraded. A RCBF [33] can provide robustness despite the error resulting from the difference between the exact and approximate controller.

Even if a given control policy satisfies $\pi_\phi(x, t) \in \mathcal{K}_{\text{CBF}}$ for all $(x, t) \in \mathcal{X} \times \mathbb{R}_{\geq 0}$, the implemented controller relies on the estimated state, $u(t) = \pi_\phi(\hat{x}, t)$. Since $\hat{x} \neq x$ in general, the resulting control input may not guarantee the safety of the actual system (1). By using the observer designed in Section 3 and the bound on the state estimation error in (28), an RCBF can be developed to guarantee the safety of the controller $u(t) = \pi_\phi(\hat{x}, t)$ regardless of the state estimation error. To that end, let $\hat{\mathcal{S}}$ denote a set of state estimates defined as

$$\hat{\mathcal{S}}(t) := \{\hat{x} \in \mathbb{R}^n : h_r(\hat{x}, t) \geq 0\}, \quad (26)$$

where the function $h_r \in \mathbb{R}^n \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ is defined as $h_r(\hat{x}, t) = h(\hat{x}) - \ell\xi(t)$, where $\ell \in \mathbb{R}_{>0}$ is a Lipschitz constant that satisfies $|h(x) - h(\hat{x})| \leq \ell \|x - \hat{x}\|$, $\forall x, \hat{x} \in \mathbb{R}^n$, where $\xi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ is a non-increasing robustifying term defined as

$$\xi(t) := \left(\sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} \right) \epsilon_0 \exp(-\alpha t), \quad (27)$$

where $\epsilon_0 \in \mathbb{R}_{>0}$ is a constant such that $\|x(t_0) - \hat{x}(t_0)\| \leq \epsilon_0$. The robustifying term in (27) depends on the symmetric positive definite matrix P , which is obtained by solving the LMI in (17). The robustifying term in (27) is inspired by approaches such as [3, 23, 33, 53], but it is redesigned to integrate with our observer design. The robustifying term is obtained using the fact that if the control signal $u(\cdot)$ and the corresponding trajectory $x(\cdot)$ of the system in (1) satisfy $x(t) \in \mathcal{X}$ and $u(t) \in \mathcal{U}$ for all $t \geq t_0$ such that the bound in (18) holds, then the Comparison Lemma [29, Lemma 3.4] can be invoked to develop the bound

$$\|\hat{x}(t)\| \leq \xi(t), \quad \forall t \geq t_0. \quad (28)$$

The following definition formalizes the concept of an RCBF for nonlinear systems with partial state measurement.

Definition 8 Given a set $\mathcal{S} \subset \mathbb{R}^n$ defined in (26), a

continuously differentiable function $h_r : \mathcal{X} \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ is a RCBF for (1) with an observer (9) of known error bound (28), if there exists a class $\mathcal{K}$ function ν such that, $\sup_{u \in \mathcal{U}} \nabla_x h_r(x, t) (f(x) + g(x)u) + \nabla_t h_r(x, t) \geq \nu(h_r(x, t_0))$, for all $x \in \mathcal{S}$ and $t \geq t_0$.

The RCBF in Definition 8 is motivated by similar definitions in [3, 23, 53].

Lemma 9 Given the set $\mathcal{S} \subset \mathcal{X} \subset \mathbb{R}^n$, an RCBF $h_r : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ that satisfies Definition 8, and the control set $\mathcal{U} \subset \mathbb{R}^m$, if Assumption 1 holds, Hypothesis H2 of Theorem 4 holds, and $x(t) \in \mathcal{X}, \forall t \geq t_0$, then any control policy that ensures $\hat{x}_0 \in \hat{\mathcal{S}}(t_0) \implies h_r(\hat{x}(t), t) \geq 0$ for all $t \geq t_0$ guarantees the safety of the system in (1) with respect to $\mathcal{S}$.

PROOF. By Lipschitz continuity of h and using the triangle inequality $|h(x) - h(\hat{x})| \leq \ell \|x - \hat{x}\| \implies h(\hat{x}) - \ell \|x - \hat{x}\| \leq h(x) \leq h(\hat{x}) + \ell \|x - \hat{x}\|$. Since $h_r(\hat{x}, t) \geq 0$ and $\ell \|x - \hat{x}\| \leq \ell \xi(t)$ for all $x, \hat{x} \in \mathbb{R}^n$, and $t \geq t_0$, it can be concluded that the inequality $0 \leq h(\hat{x}) - \ell \xi(t) \leq \ell \|x - \hat{x}\| \leq h(x)$ holds for all $t \geq t_0$. As a result, $h(x) \geq 0$ for all $t \geq t_0$, which implies that $x \in \mathcal{S}$ from (4) and thus, the system in (1) is safe.

In the following section, a safe control policy that utilizes RCBFs in the cost function of the output-feedback optimal control problem will be designed, so that trajectories of the system in (1) remain in a given set $\mathcal{S}$ despite the lack of full state measurement.

5 Control Design

5.1 ADP-Based Safe and Stabilizing Optimal Control

The developed technique uses a cost function that includes the RCBF to meet the safety constraints. Since the RCBF depends on the robustifying term ξ in (27), so does the value function. To address this dependency, a state augmentation technique similar to [25] is used. Let $\zeta := [x^\top, \xi]^\top \in \mathbb{R}^{n+1}$ be a concatenated vector such that the augmented dynamics are given by

$$\dot{\zeta} = F(\zeta) + G(\zeta)u, \quad (29)$$

where the augmented functions $F : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$ and $G : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{(n+1) \times m}$ are defined as $F(\zeta) := [f^\top(x), -\alpha\xi]^\top$ and $G(\zeta) := [g^\top(x), 0_{m \times 1}]^\top$, respectively.

The optimal control problem is formulated using approximate dynamic programming (ADP)-based design

techniques to develop a control signal, $u : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^m$, online, that minimizes the cost functional defined as

$$J(\zeta, u(\cdot)) := \int_t^\infty r(\zeta^u(t_0; \tau, \zeta_0), u(\tau)) d\tau, \quad (30)$$

over the set of piecewise continuous functions $u : \mathbb{R}_{\geq 0} \rightarrow \mathcal{U}$, where $\zeta^u(t_0; t, \zeta_0)$ is a solution of (29) under control signal $u(\cdot)$ starting from initial condition $\zeta(t_0) = \zeta_0 \in \mathbb{R}^{n+1}$, and $r : \mathbb{R}^{n+1} \times \mathbb{R}^m \rightarrow \mathbb{R}$ is the instantaneous cost defined as

$$r(\zeta, u) := \bar{Q}(\zeta) + U(u) + B_r(\zeta) \quad (31)$$

where the function $\bar{Q} : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ is defined as $\bar{Q}([x^\top, \xi]^\top) := Q(x), \forall x \in \mathcal{X}$ and $\forall \xi \in \mathbb{R}$ with $Q : \mathcal{X} \rightarrow \mathbb{R}$ representing a continuous positive definite function. By [29, Lemma 4.3], Q satisfies $\underline{q}(\|x\|) \leq Q(x) \leq \bar{q}(\|x\|)$, where $\underline{q}, \bar{q} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ are class $\mathcal{K}$ functions. The function $U : \mathbb{R}^m \rightarrow \mathbb{R}$, introduced to enforce the saturation constraint on the control input, is defined as

$$U(u) := 2 \int_0^u \left(\bar{u} \tanh^{-1} \left(\frac{v}{\bar{u}} \right) \right)^\top R dv, \quad (32)$$

where $R := \text{diag}(r_1, \dots, r_m)$. In this development, the function B_r is selected as an RLCBF of the form

$$B_r(\zeta) := \left(b_r(\zeta) - b_r(0) \right)^2, \quad (33)$$

where $b_r(\zeta) := -\ln \left(\frac{\kappa h_r(\zeta)}{\kappa h_r(\zeta) + 1} \right)$ and $\kappa \in \mathbb{R}_{>0}$ is a user-defined constant that determines the magnitude of the barrier penalty as the trajectories approach the boundary of $\mathcal{S}$. Assuming the optimal controller exists, the optimal value function, $V^* : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ can be expressed as

$$V^*(\zeta) := \min_{u(\tau) \in \mathcal{U}, \tau \in \mathbb{R}_{\geq t}} \int_t^\infty r(\zeta^u(t_0; \tau, \zeta_0), u(\tau)) d\tau. \quad (34)$$

Assuming that the optimal value function is continuously differentiable, it can be shown to be the unique positive definite solution of the HJB equation, [28, Theorem 1.5],

$$\min_{u \in \mathcal{U}} \left(\nabla_\zeta V(\zeta) (F(\zeta) + G(\zeta)u) + r(\zeta, u) \right) = 0. \quad (35)$$

Therefore, the optimal stabilizing controller is given by the feedback policy $u(t) = u^*(\zeta^u(t_0; \tau, \zeta_0))$, where u^* is given by

$$u^*(\zeta) := -\bar{u} \tanh(D^*(\zeta)), \quad (36)$$

where $D^*(\zeta) := \frac{R^{-1}G^\top(\zeta)}{2\bar{u}} \nabla_\zeta (V^*(\zeta))^\top \in \mathbb{R}^m$. When the true states x are available, the robustifying term in (27) vanishes because $\epsilon_0 \equiv 0$ when $x = \hat{x}$. In this case, the RLCBF-based SLOC in (36) reduces to the input-constrained version of existing CBF-based SLOCs (e.g., [14, 35]).

5.2 Value Function Approximation

Solving the HJB equation in (35) is generally infeasible for nonlinear systems; hence, to find an approximate solution, estimates of the value function and the control policy are introduced. Let $\mathcal{B}(0, \chi)$ be a closed ball of radius χ containing the origin, where χ , defined as $\chi := \left(\sqrt{\frac{\lambda_{\max}(P)}{\lambda_{\min}(P)}} \right) \epsilon_0$, is obtained from the bound on the state estimation error in (28). The unknown difference between the optimal value function V^* and the barrier function B_r can be expressed using a NN for local parametric approximation over the compact set $\Omega \subset (\mathcal{X} \oplus \mathcal{B}(0, \chi)) \times \mathcal{B}(0, \chi)$ containing the origin as

$$V^*(\zeta) - B_r(\zeta) = W^\top \phi(\zeta) + \epsilon(\zeta), \quad \forall \zeta \in \Omega, \quad (37)$$

where $W \in \mathbb{R}^L$ is an unknown vector of bounded weights, $\phi : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^L$ is a vector of continuously differentiable nonlinear activation functions that satisfy $\phi(0) = 0$ and $\nabla_\zeta \phi(0) = 0$, L is the number of basis functions, and $\epsilon : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ is the reconstruction error. Invoking Stone-Weierstrass Theorem [48, Theorem 1.5], the activation functions ϕ can be selected so that the weights and the approximation errors satisfy $\sup_{\zeta \in \Omega} \|W\| \leq \bar{W}$, $\sup_{\zeta \in \Omega} \|\phi(\cdot)\| \leq \bar{\phi}$, $\sup_{\zeta \in \Omega} \|\nabla_\zeta \phi(\cdot)\| \leq \bar{\nabla} \bar{\phi}$, $\sup_{\zeta \in \Omega} \|\epsilon(\cdot)\| \leq \bar{\epsilon}$ and $\sup_{\zeta \in \Omega} \|\nabla_\zeta \epsilon(\cdot)\| \leq \bar{\nabla} \bar{\epsilon}$, where the notation $(\bar{\cdot}) := \sup_{\zeta \in \Omega} \|\cdot\|$ is a positive constant.

Since the actual state ζ is unknown and the ideal weights W are unknown, let $\hat{\zeta} := [\hat{x}^\top, \hat{\xi}^\top]^\top \in \Omega$ be an estimate of the concatenated state ζ and let the estimates of the optimal value function and the optimal controller denoted as $\hat{V} : \mathbb{R}^{n+1} \times \mathbb{R}^L \rightarrow \mathbb{R}$ and $\hat{u} : \mathbb{R}^{n+1} \times \mathbb{R}^L \rightarrow \mathbb{R}^m$, respectively, be defined as

$$\hat{V}(\hat{\zeta}, \hat{W}_c) := \hat{W}_c^\top \phi(\hat{\zeta}) + B_r(\hat{\zeta}) \quad \text{and} \quad (38)$$

$$\hat{u}(\hat{\zeta}, \hat{W}_c) := -\bar{u} \tanh(\hat{D}(\hat{\zeta}, \hat{W}_c)), \quad (39)$$

for all $\hat{\zeta} \in \Omega$, where $\hat{D}(\hat{\zeta}, \hat{W}_c) := \frac{R^{-1}G(\hat{\zeta})^\top}{2\bar{u}} (\nabla_{\hat{\zeta}} \phi(\hat{\zeta})^\top \hat{W}_c + \nabla_{\hat{\zeta}} B_r^\top(\hat{\zeta})) \in \mathbb{R}^m$ and $\hat{W}_c \in \mathbb{R}^L$ are estimates of the ideal weights W .

5.3 Bellman Error and Simulation of Experience

By substituting the approximations of the optimal value function and optimal controller from (38) and (39), respectively, into the HJB equation in (35), the residual term $\hat{\delta} : \mathbb{R}^{n+1} \times \mathbb{R}^L \rightarrow \mathbb{R}$, referred to as the Bellman error (BE), is obtained as

$$\hat{\delta}(\hat{\zeta}, \hat{W}_c) = \nabla_{\hat{\zeta}} \hat{V}(\hat{\zeta}, \hat{W}_c) \left(F(\hat{\zeta}) + G(\hat{\zeta}) \hat{u}(\hat{\zeta}, \hat{W}_c) \right) + \bar{Q}(\hat{\zeta}) + U(\hat{u}) + B_r(\hat{\zeta}). \quad (40)$$

To accurately approximate the value function, online SLOC methods require the persistence of excitation (PE) condition [26, 37], which is difficult to guarantee in practice. However, through BE extrapolation for excitation via simulation, stability and convergence of online SLOC can be established [26]. To simulate experience using BE extrapolation, select a set of trajectories $\{\zeta_k : \mathbb{R}_{\geq 0} \rightarrow \Omega : k = 1, \dots, N\}$ and extrapolate the BE along these trajectories to yield the extrapolated BEs, $\hat{\delta}_k : \mathbb{R}^{n+1} \times \mathbb{R}^L \rightarrow \mathbb{R}$, given by

$$\hat{\delta}_k(\zeta_k, \hat{W}_c) := \nabla_{\zeta_k} \hat{V}(\zeta_k, \hat{W}_c) \left(F(\zeta_k) + \hat{G}(\zeta_k) \hat{u}(\zeta_k, \hat{W}_c) \right) + \bar{Q}(\zeta_k) + U(\hat{u}) + B_r(\zeta_k). \quad (41)$$

The control objective is achieved by updating the critic NN weights $\hat{W}_c$ online to minimize the BE using an adaptive update law developed from the subsequent stability analysis in Section 6.

5.4 Update laws for Critic weights

To guarantee that the estimated value function weights, $\hat{W}_c$, converge to their ideal values in (37), the estimated value function weights are updated based on the stability analysis in Section 6 as

$$\dot{\hat{W}}_c = -\frac{k_c}{N} \Gamma \sum_{k=1}^N \frac{\omega_k}{\rho_k} \hat{\delta}_k, \quad \dot{\Gamma} = \beta \Gamma - \frac{k_c}{N} \Gamma \sum_{k=1}^N \frac{\omega_k \omega_k^\top}{\rho_k^2} \Gamma, \quad (42)$$

with $\Gamma(t_0) = \Gamma_0$, where $\Gamma : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^{L \times L}$ is a time-varying least-squares gain matrix, $\omega_k := \nabla_{\zeta_k} \phi_k(\zeta_k) (F(\zeta_k) + G(\zeta_k) \hat{u}(\zeta_k, \hat{W}_c))$, $\rho_k := 1 + \gamma_c \omega_k^\top \omega_k$, $\gamma_c \in \mathbb{R}_{>0}$ is a constant normalization gain, $\beta \in \mathbb{R}_{>0}$ is a constant forgetting factor, and $k_c \in \mathbb{R}_{>0}$ is a constant adaptation gain. Note that the update laws in (42) rely on randomly sampled off-policy states ζ_k , not the actual state ζ . To guarantee the safety of the closed-loop system in (1), the control signal is then designed as

$$u(t) = \hat{u}(\hat{\zeta}(t), \hat{W}_c(t)), \quad \forall t \geq t_0. \quad (43)$$

The following section analyzes the convergence properties of the closed-loop system and establishes local uniform ultimate boundedness of its trajectories.

6 Stability Analysis

In this section, stability analysis of the safe observer-based SLOC architecture is carried out using Lyapunov methods. To facilitate the stability analysis, the following rank condition is necessary.

Assumption 10 *There exists a constant $\underline{c}_1$ such that the finite set of trajectories $\{\zeta_k : \mathbb{R}_{\geq 0} \rightarrow \Omega \mid k = 1, \dots, N\}$ satisfies*

$$0 < \underline{c}_1 \leq \inf_{t \in [0, \infty)} \lambda_{\min} \left(\frac{1}{N} \sum_{k=1}^N \frac{\omega_k(t) \omega_k^\top(t)}{\rho_k^2(t)} \right). \quad (44)$$

As described in [34], since ω_k is a function of ζ_k and $\hat{W}_c$, Assumption 10 cannot be guaranteed a priori. However, unlike the PE condition utilized in results such as [50], Assumption 10 can be verified online. Furthermore, since $\lambda_{\min} \left(\sum_{k=1}^N \frac{\omega_k(t) \omega_k^\top(t)}{\rho_k^2(t)} \right)$ is non-decreasing in the number of samples, N , Assumption 10 can be met, heuristically, by increasing the number of extrapolation trajectories. The calculation of a precise bound on the number of extrapolation trajectories required is outside the scope of this paper.

The following lemma presents an alternate form of the BE to facilitate the stability analysis. For notational brevity, let $\hat{F} := F(\hat{\zeta})$, $\hat{G} := G(\hat{\zeta})$, $\nabla_{\hat{\zeta}} \phi := \nabla_{\hat{\zeta}} \phi(\hat{\zeta})$, $\nabla_{\hat{\zeta}} B_r := \nabla_{\hat{\zeta}} B_r(\hat{\zeta})$, and $\nabla_{\hat{\zeta}} \epsilon := \nabla_{\hat{\zeta}} \epsilon(\hat{\zeta})$. Furthermore, the dependence on $\hat{\zeta}$, ζ_k , $\hat{W}_c$, and t is omitted henceforth whenever it is clear from context.

Lemma 11 *Given the critic NN weight estimation error $\tilde{W} := W - \hat{W}_c$, the BE can alternatively be expressed in its unmeasurable form as*

$$\hat{\delta} := -\omega^\top \tilde{W}_c + \Delta, \quad (45)$$

where $\omega := \nabla_{\hat{\zeta}} \phi(\hat{F} + \hat{G} \hat{u}(\hat{\zeta}, \hat{W}_c))$, $\Delta := -\nabla_{\hat{\zeta}} \epsilon(\hat{F} + \hat{G} u^*(\hat{\zeta})) + \nabla_{\hat{\zeta}} B_r(\hat{F} + \hat{G} \hat{u}(\hat{\zeta}, \hat{W}_c)) + \bar{u} W^\top \nabla_{\hat{\zeta}} \phi \hat{G} (\tanh(D^*(\hat{\zeta})) - \tanh(\hat{D}(\hat{\zeta}, \hat{W}_c))) + 2\bar{u}^2 \bar{R} (c_D^* - \hat{c}_D) + \bar{u}^2 \bar{R} (\hat{e}_D - \epsilon_D^*)$, c_D^* and $\hat{c}_D$ are bounded approximation errors defined as $c_D^* := D^*(\hat{\zeta}) \operatorname{sgn}(D^*(\hat{\zeta})) - D^*(\hat{\zeta}) \tanh(D^*(\hat{\zeta}))$ and $\hat{c}_D := \hat{D}(\hat{\zeta}, \hat{W}_c) \operatorname{sgn}(\hat{D}(\hat{\zeta}, \hat{W}_c)) - \hat{D}(\hat{\zeta}, \hat{W}_c) \tanh(\hat{D}(\hat{\zeta}, \hat{W}_c))$, respectively, and ϵ_D^* and $\hat{e}_D$ are errors obtained when

approximating the $\text{sgn}(\cdot)$ function with the $\tanh(\cdot)$ function, which satisfy $\|\epsilon_D^*\| \leq \ln(4)$ and $\|\hat{\epsilon}_D\| \leq \ln(4)$ so that $\ln(\mathbf{1} - \tanh^2(D^*(\hat{\zeta}))) = \ln(4) - 2D^*(\hat{\zeta}) \text{sgn}(D^*(\hat{\zeta})) + \epsilon_D^*$ and $\ln(\mathbf{1} - \tanh^2(\hat{D}(\hat{\zeta}, \hat{W}_c))) = \ln(4) - 2\hat{D}(\hat{\zeta}, \hat{W}_c) \text{sgn}(\hat{D}(\hat{\zeta}, \hat{W}_c)) + \hat{\epsilon}_D$, respectively.

PROOF. Substituting $\hat{\zeta}$ for ζ in (37) and taking its gradient with respect to $\hat{\zeta}$, followed by taking the gradient of (38) with respect to $\hat{\zeta}$, then substituting $\hat{\zeta}$ for ζ in the HJB equation in (35) and subtracting it from (40), and finally substituting the gradients into the resulting expression, yields

$$\begin{aligned} \delta(\hat{\zeta}, \hat{W}_c) &= \hat{W}_c^\top \nabla_{\hat{\zeta}} \phi(\hat{F} + \hat{G}\hat{u}(\hat{\zeta}, \hat{W}_c)) + U(\hat{u}(\hat{\zeta}, \hat{W}_c)) \\ &+ \nabla_{\hat{\zeta}} B_r(\hat{F} + \hat{G}\hat{u}(\hat{\zeta}, \hat{W}_c)) - W^\top \nabla_{\hat{\zeta}} \phi(\hat{F} + \hat{G}u^*(\hat{\zeta})) \\ &- \nabla_{\hat{\zeta}} B_r(\hat{F} + \hat{G}u^*(\hat{\zeta})) - \nabla_{\hat{\zeta}} \epsilon(\hat{F} + \hat{G}u^*(\hat{\zeta})) - U(u^*(\hat{\zeta})). \end{aligned} \quad (46)$$

Substituting equation (36) and (39) in (32) yields alternate representations of the optimal and approximate forms of the function $U(u)$, given by

$$\begin{aligned} U(u^*(\hat{\zeta})) &= \bar{u} \nabla_{\hat{\zeta}} V^*(\hat{\zeta}) G(\hat{\zeta}) \tanh(D^*(\hat{\zeta})) \\ &+ \bar{u}^2 \bar{R} \ln(\mathbf{1} - \tanh^2(D^*(\hat{\zeta}))) \text{ and } \end{aligned} \quad (47)$$

$$\begin{aligned} U(\hat{u}(\hat{\zeta}, \hat{W}_c)) &= \bar{u} \nabla_{\hat{\zeta}} \hat{V}(\hat{\zeta}, \hat{W}_c) G(\hat{\zeta}) \tanh(\hat{D}(\hat{\zeta}, \hat{W}_c)) \\ &+ \bar{u}^2 \bar{R} \ln(\mathbf{1} - \tanh^2(\hat{D}(\hat{\zeta}, \hat{W}_c))), \end{aligned} \quad (48)$$

respectively, where $\bar{R} := [r_1, \dots, r_m] \in \mathbb{R}^{1 \times m}$ and $\mathbf{1} \in \mathbb{R}^m$ denotes a column vector having all of its elements equal to one. Subtracting (47) from (48) yields

$$\begin{aligned} U(\hat{u}(\hat{\zeta}, \hat{W}_c)) - U(u^*(\hat{\zeta})) &= \bar{u} \hat{W}_c^\top \nabla_{\hat{\zeta}} \phi \hat{G} \tanh(\hat{D}(\hat{\zeta}, \hat{W}_c)) \\ &+ 2\bar{u}^2 \bar{R} (D^*(\hat{\zeta}) \text{sgn}(D^*(\hat{\zeta})) - \hat{D}(\hat{\zeta}, \hat{W}_c) \text{sgn}(\hat{D}(\hat{\zeta}, \hat{W}_c))) \\ &+ \bar{u} \nabla_{\hat{\zeta}} B_r \hat{G} \tanh(\hat{D}(\hat{\zeta}, \hat{W}_c)) - \bar{u} W^\top \nabla_{\hat{\zeta}} \phi \hat{G} \tanh(D^*(\hat{\zeta})) \\ &- \bar{u} \nabla_{\hat{\zeta}} B_r \hat{G} \tanh(D^*) - \bar{u} \nabla_{\hat{\zeta}} \epsilon \hat{G} \tanh(D^*(\hat{\zeta})) \\ &+ \bar{u}^2 \bar{R} (\hat{\epsilon}_D - \epsilon_D^*), \end{aligned} \quad (49)$$

where ϵ_D^* and $\hat{\epsilon}_D$ are approximation errors introduced in (45). By using the approximation errors c_D^* and $\hat{c}_D$ obtained when approximating the $\text{sgn}(\cdot)$ function with a $\tanh(\cdot)$ function introduced in (45) (cf. [37]), therefore,

$$\begin{aligned} D^*(\hat{\zeta}) \text{sgn}(D^*(\hat{\zeta})) - \hat{D}(\hat{\zeta}, \hat{W}_c) \text{sgn}(\hat{D}(\hat{\zeta}, \hat{W}_c)) &= \\ \frac{R^{-1}}{2\bar{u}} (W^\top \nabla_{\hat{\zeta}} \phi + \nabla_{\hat{\zeta}} B_r + \nabla_{\hat{\zeta}} \epsilon) \hat{G} \tanh(D^*(\hat{\zeta})) + c_D^* & \\ - \frac{R^{-1}}{2\bar{u}} (\hat{W}_c^\top \nabla_{\hat{\zeta}} \phi + \nabla_{\hat{\zeta}} B_r) \hat{G} \tanh(\hat{D}(\hat{\zeta}, \hat{W}_c)) - \hat{c}_D. \end{aligned} \quad (50)$$

After substituting (50) into (49) and then substituting (38), (39) and (49) into (46), the BE can be expressed as given in (45).

Note that, the residual term Δ in (45) is bounded for all $\hat{\zeta} \in \Omega$ and $\hat{W}_c \in \mathbb{R}^L$ as

$$\begin{aligned} \Delta &\leq (\bar{F} + \bar{u}\bar{G})(\bar{\nabla}\epsilon + \bar{\nabla}B_r) + 2\bar{W}^\top \bar{\nabla}\phi \bar{G} \\ &+ 2\bar{u}^2 \|\bar{R}\| \|(c_D^* - \hat{c}_D)\| + \bar{u}^2 \|\bar{R}\| \|(\hat{\epsilon}_D - \epsilon_D^*)\| =: \bar{\Delta}, \end{aligned} \quad (51)$$

where $\bar{F} \in \mathbb{R}_{>0}$ and $\bar{G} \in \mathbb{R}_{>0}$ are upper bounds satisfying $\sup_{\hat{\zeta} \in \Omega} \|F(\hat{\zeta})\| \leq \bar{F}$ and $\sup_{\hat{\zeta} \in \Omega} \|G(\hat{\zeta})\| \leq \bar{G}$, respectively, which follows since the derivatives of the functions f and g are bounded from Assumption 1 and since B_r is continuously differentiable in $\hat{\zeta} \in \Omega$, its gradient $\nabla_{\hat{\zeta}} B_r$ is both continuous and bounded within the compact set Ω . Hence, by the Extreme Value Theorem, there exist some constant $\bar{\nabla}B_r \in \mathbb{R}_{>0}$ such that $\sup_{\hat{\zeta} \in \Omega} \|\nabla_{\hat{\zeta}} B_r(\hat{\zeta})\| \leq \bar{\nabla}B_r$. For notational brevity, let $F_k := F(\zeta_k)$, $G_k := G(\zeta_k)$, $u_k^* := u^*(\zeta_k)$, $\hat{u}_k := \hat{u}(\zeta_k, \hat{W}_c)$, $D_k^* := D^*(\zeta_k)$, $\hat{D}_k := \hat{D}(\zeta_k, \hat{W}_c)$, $\nabla_{\zeta_k} \phi_k := \nabla_{\zeta_k} \phi(\zeta_k)$, $\nabla_{\zeta_k} B_{r_k} := \nabla_{\hat{\zeta}} B_r(\zeta_k)$, and $\nabla_{\zeta_k} \epsilon_k := \nabla_{\hat{\zeta}} \epsilon(\zeta_k)$. Following the same argument as Lemma 11, by substituting (38) and (39) into (35), and subtracting from (40), the extrapolated BE can be expressed in an unmeasurable form as

$$\delta_k := -\omega_k^\top \tilde{W}_c + \Delta_k, \quad (52)$$

where ω_k is introduced below (42), $\Delta_k := -\nabla_{\zeta_k} \epsilon_k^\top (F_k + G_k u_k^*) + \nabla_{\zeta_k} B_{r_k}^\top (F_k + G_k \hat{u}_k) + \bar{u} W^\top \nabla_{\zeta_k} \phi_k G_k (\tanh(D_k^*) - \tanh(\hat{D}_k)) + 2\bar{u}^2 \bar{R} (c_{D_k}^* - \hat{c}_{D_k}) + \bar{u}^2 \bar{R} (\hat{\epsilon}_{D_k} - \epsilon_{D_k}^*)$, and $c_{D_k}^*$, $\hat{c}_{D_k}$, $\hat{\epsilon}_{D_k}$, and $\epsilon_{D_k}^* \in \mathbb{R}$ are bounded approximation errors. Similar to (51), note that $\sup_{\hat{\zeta} \in \Omega} \|\Delta_k(\cdot)\| \leq \bar{\Delta}_k$ for all $t \geq t_0$, where $\bar{\Delta}_k \in \mathbb{R}_{>0}$ is a constant.

Let $Z \in \mathbb{R}^{2n+L}$ represent the concatenated state vector of the closed-loop system, defined as $Z := \begin{bmatrix} x^\top & \tilde{x}^\top & \tilde{W}_c^\top \end{bmatrix}^\top$, and let $V_L : \mathbb{R}^{2n+L} \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ be a continuously differentiable candidate Lyapunov function defined as

$$V_L(Z, t) := \mathcal{V}(Z, t) + \Theta(\tilde{W}_c, t), \quad (53)$$

where $\Theta : \mathbb{R}^L \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ be defined as $\Theta(\tilde{W}_c, t) := \frac{1}{2} \tilde{W}_c^\top \Gamma^{-1}(t) \tilde{W}_c$ and $V : \mathbb{R}^{2n+L} \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ is defined as $\mathcal{V}(Z, t) = V_t^*(x, t) + \tilde{x}^\top P \tilde{x}$. The function V_t^* is the positive definite non-autonomous form of the optimal function V^* , defined as $V_t^*(x, t) := V^*([x^\top, \xi(t)^\top]^\top)$, which according to [25, Lemma 1] serves as a valid Lyapunov function, being both positive definite and decrescent. In addition, similar to the results of [25, Lemma 2],

the function V_t satisfies the properties $V_t^*(0, t) = 0$ and $\underline{v}(\|x\|) \leq V_t^*(x, t) \leq \bar{v}(\|x\|)$ for all $x \in \mathbb{R}^n$ and $t \in [0, \infty)$, where $\underline{v}, \bar{v} : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ are $\mathcal{K}$ functions. The choice of V_t^* as a candidate Lyapunov function is motivated by the unsuitability of the optimal value function due to its positive semi-definiteness.

Since the candidate Lyapunov function is positive definite, [29, Lemma 4.3] and the bound in (58) can be used to conclude that it is bounded as

$$\underline{v}_l(\|Z\|) \leq V_L(Z, t) \leq \bar{v}_l(\|Z\|), \quad (54)$$

for all $t \geq t_0$ and for all $Z \in \mathbb{R}^{2n+L}$, where $\underline{v}_l, \bar{v}_l : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ are class $\mathcal{K}$ functions.

The Lie derivative of $\mathcal{V}$ is given by

$$\begin{aligned} \dot{\mathcal{V}}(Z, t) &= \nabla_{\zeta} V^*(F + Gu^*(\zeta)) + \dot{V}_x(\tilde{x}, x, u) \\ &\quad + \nabla_{\zeta} V^*G(\hat{u}(\hat{\zeta}, \hat{W}_c) - u^*(\zeta)). \end{aligned} \quad (55)$$

Note that, by the Cauchy-Schwarz inequality, Young's inequality, and the triangle inequality, the norm of the difference between the approximate control policy $\hat{u}$ and the optimal policy u^* is given by,

$$\begin{aligned} \|\hat{u}(\hat{\zeta}, \hat{W}_c) - u^*(\zeta)\| &= \|\bar{u} \tanh(D^*(\zeta)) - \bar{u} \tanh(\hat{D}(\hat{\zeta}, \hat{W}_c))\| \\ &\leq \ell_{g\phi r} \bar{W} \|\tilde{x}\| + \overline{G_{\phi R}} \|\tilde{W}_c\| + \frac{\lambda_{\max}(R^{-1}) \bar{G}}{2} \bar{\nabla} \epsilon, \end{aligned} \quad (56)$$

for all $\zeta, \hat{\zeta} \in \Omega$ and $\hat{W}_c \in \mathbb{R}^L$, where $\sup_{\zeta \in \Omega} G_{\phi R}(\cdot) := \frac{R^{-1}G(\cdot)^{\top}}{2} \nabla(\cdot) \phi(\cdot)^{\top} \leq \overline{G_{\phi R}}$, $\ell_{g\phi r} \in \mathbb{R}_{>0}$ is a constant bound over the set Ω such that $\|G_{\phi R}(\zeta) - G_{\phi R}(\hat{\zeta})\| \leq \ell_{g\phi r} \|\tilde{x}\|$ and $\bar{G} \in \mathbb{R}_{>0}$ is a positive bound such that $\sup_{\zeta \in \Omega} \|G(\zeta)\| \leq \bar{G}$. Thus, provided that the basis functions ϕ are selected such that the weights W , the value function approximation error ϵ , the basis itself ϕ , and their gradients with respect to ζ and $\hat{\zeta}$ are bounded on the compact set Ω as described below (37), by substituting the bounds in (18) and (56), using the fact that $\nabla_{\zeta} V^*(\zeta)(F(\zeta) + G(\zeta)u^*(\zeta)) = -r(\zeta, u^*(\zeta))$ from the substitution of (34) and (36) into (35), and using the positive definiteness of the functions U and B_r , the Lie derivative of the function $\mathcal{V}$ is bounded by

$$\dot{\mathcal{V}}(Z, t) \leq -\underline{q}(\|x\|) - \alpha \lambda_{\min}(P) \|\tilde{x}\|^2 + \varpi_1 \|\tilde{W}_c\| + \iota_1, \quad (57)$$

for all $t \geq t_0$, $x \in \mathcal{X}$, $\tilde{x} \in \mathcal{B}(0, \chi)$, and $\tilde{W}_c \in \mathbb{R}^L$, where $\varpi_1 := \overline{GW \nabla \phi G_{\phi R}} + \overline{G G_{\phi R} \nabla \epsilon}$ and $\iota_1 := \frac{\bar{G}^2 \lambda_{\max}(R^{-1}) \overline{W \nabla \phi \nabla \epsilon}}{2} + \frac{\bar{G}^2 \lambda_{\max}(R^{-1}) \bar{\nabla} \epsilon^2}{2} + \frac{(\ell_{g\phi r} \overline{GW \nabla \phi W} + \overline{G \ell_{g\phi r} W \nabla \epsilon})^2}{4\alpha \lambda_{\min}(P)}$. Note that the bound in (57) relies on the fact that $x \in \mathcal{X}$ and $\tilde{x} \in \mathcal{B}(0, \chi)$ implies

that $\zeta, \hat{\zeta} \in \Omega$.

As shown in [26, Lemma 1], provided (10) holds and $\lambda_{\min}\{\Gamma_0^{-1}\} > 0$, the update laws in (42), ensure that the least squares update law satisfies

$$\underline{\Gamma}_L \leq \Gamma \leq \bar{\Gamma}_L, \quad (58)$$

for all $t \geq t_0$ and some $\bar{\Gamma}, \underline{\Gamma} \in \mathbb{R}_{>0}$, where I_L is a L by L identity matrix. Using the least square gain update law in (42) and the bound in (58), the normalized regressor $\frac{\omega_k}{\rho_k}$ is bounded as $\sup_{t \geq t_0} \|\frac{\omega_k}{\rho_k}\| \leq \frac{1}{2\sqrt{\gamma_c \underline{\Gamma}}}$. The Lie derivative of Θ along the flow of (42) is given by

$$\dot{\Theta}(\tilde{W}_c, t) := -\tilde{W}_c^{\top} \Gamma^{-1} \dot{\tilde{W}}_c - \frac{1}{2} \tilde{W}_c^{\top} \Gamma^{-1} \dot{\Gamma} \Gamma^{-1} \tilde{W}_c. \quad (59)$$

Substituting the update laws in (42) into the derivative and simplifying yields

$$\begin{aligned} \dot{\Theta}(\tilde{W}_c, t) &= -\frac{k_c}{N} \tilde{W}_c^{\top} \sum_{k=1}^N \frac{\omega_k \omega_k^{\top}}{\rho_k} \tilde{W}_c + \frac{k_c}{N} \tilde{W}_c^{\top} \sum_{k=1}^N \frac{\omega_k}{\rho_k} \hat{\delta}_k \\ &\quad - \frac{1}{2} \beta \tilde{W}_c^{\top} \Gamma^{-1} \tilde{W}_c + \frac{k_c}{2N} \tilde{W}_c^{\top} \sum_{k=1}^N \frac{\omega_k \omega_k^{\top}}{\rho_k^2} \tilde{W}_c. \end{aligned} \quad (60)$$

Provided Assumption 10 holds and using the fact that $\frac{\omega_k \omega_k^{\top}}{\rho_k^2} \leq \frac{\omega_k \omega_k^{\top}}{\rho_k}$, the Lie derivative of Θ is bounded, by

$$\dot{\Theta}(\tilde{W}_c, t) \leq -k_c \underline{c} \|\tilde{W}_c\|^2 + \varpi_2 \|\tilde{W}_c\|, \quad (61)$$

for all $t \geq t_0$ and $\tilde{W}_c \in \mathbb{R}^L$, where $\underline{c} \in \mathbb{R}_{>0}$ be a constant defined as $\underline{c} := \frac{\beta}{2\Gamma k_c} + \frac{\underline{c}_1}{2}$, $\varpi_2 := \frac{k_c \bar{\Delta}_k}{2\sqrt{\gamma_c \underline{\Gamma}}}$.

To facilitate the development of Theorem 12, let $\iota \in \mathbb{R}_{>0}$ be a constant defined as $\iota := \frac{\bar{G}^2 \lambda_{\max}(R^{-1}) \bar{\nabla} \epsilon^2}{2} + \frac{\bar{G}^2 \lambda_{\max}(R^{-1}) \overline{W \nabla \phi \nabla \epsilon}}{2} + \frac{(\varpi_1 + \varpi_2)^2}{2k_c \underline{c}} + \frac{(\ell_{g\phi r} \overline{GW \nabla \phi W} + \overline{G \ell_{g\phi r} W \nabla \epsilon})^2}{4\alpha \lambda_{\min}(P)}$ and let $v_l : \mathbb{R}^{2n+L} \rightarrow \mathbb{R}$ be a class $\mathcal{K}$ function defined as

$$v_l := \frac{q(\|x\|)}{2} + \frac{\alpha \lambda_{\min}(P)}{2} \|\tilde{x}\|^2 + \frac{k_c \underline{c}}{4} \|\tilde{W}_c\|^2. \quad (62)$$

The following theorem establishes local uniform ultimate boundedness of the trajectories of the closed-loop system.

Theorem 12 *Let $\bar{\chi} > 0$ be such that $\mathcal{B}(0, \bar{\chi}) \subset \mathcal{X} \times \mathcal{B}(0, \chi) \times \mathbb{R}^L$. Given the system in (29) controlled using the controller designed in (39), if*

- H1: Assumptions 1 and 10 hold,*
- H2: Hypothesis H2 of Theorem 4 holds,*

H3: the state estimate (1) is obtained using the state observer in (9),

H4: the barrier-like function B_r is selected such that it satisfies the properties highlighted above (33),

H5: the control gains are selected large enough based on the sufficient condition²

$$v_l^{-1}(\iota) \leq \bar{v}_l^{-1}(\underline{v}_l(\bar{\chi})), \quad (63)$$

H6: and the weights $\hat{W}_c$ and least squares gain matrix Γ are updated according to the update laws in (42),

then the concatenated state Z is locally uniformly ultimately bounded.

PROOF. The Lie derivative of V_L along the flow of (1), (10), and (42) is given by

$$\dot{V}_L(Z, t) = \dot{V}(Z, t) + \dot{\Theta}(\tilde{W}_c, t) \quad (64)$$

By substituting (18), (57) and (61) into (64) and applying completion of squares, the Lie derivative in (64) can be bounded as

$$\dot{V}_L(Z, t) \leq -q(\|x\|) - \alpha \lambda_{\min}(P) \|\tilde{x}\|^2 - \frac{k_c \underline{c}}{2} \|\tilde{W}_c\|^2 + \iota, \quad (65)$$

for all $t \geq t_0$, $x \in \mathcal{X}$, $\tilde{x} \in \mathcal{B}(0, \chi)$, and $\tilde{W}_c \in \mathbb{R}^L$. Using (62), the Lie derivative can be bounded as

$$\dot{V}_L(Z, t) \leq -v_l(\|Z\|), \quad (66)$$

for all Z such that $Z \in \mathcal{B}(0, \bar{\chi})$ and $\|Z\| \geq v_l^{-1}(\iota)$. Using the sufficient conditions stated in (63) and the bound in (54), [29, Theorem 4.18] can be invoked to conclude that Z is locally uniformly ultimately bounded. In particular, all trajectories starting from initial conditions that satisfy $\|Z(t_0)\| \leq \bar{v}_l^{-1}(\underline{v}_l(\bar{\chi}))$ remain in $\mathcal{B}(0, \bar{\chi})$ for all $t \geq t_0$ and satisfy $\limsup_{t \rightarrow \infty} \|Z(t)\| \leq \bar{v}_l^{-1}(\underline{v}_l(\iota))$. Therefore, provided $\|Z(t_0)\| \leq \bar{v}_l^{-1}(\underline{v}_l(\bar{\chi}))$, the state and the state estimates, under the controller in (39) and the observer in (9), remain within the compact ball $\mathcal{B}(0, \bar{\chi})$.

7 Simulation Results

In this section, a simulation study is performed to demonstrate the effectiveness of the developed method

² Although ι generally increases with increasing ζ , the condition in (63) can be satisfied provided the points for BE extrapolation are selected such that $\underline{c}$, introduced and control gain, k_c is large enough, and the basis for the value function approximation are selected such that $\bar{\epsilon}$ and $\|\nabla \epsilon\|$ are sufficiently small.

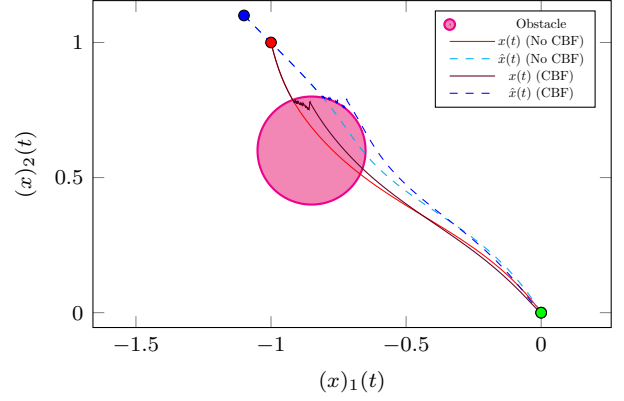


Fig. 1. Trajectories of the state and estimated state for the system in (67) when the SLOC framework from [14] is used to solve the obstacle avoidance problem in Section 7.2. It can be observed that while the state estimate trajectory $\hat{x}(t)$ remains outside the obstacle boundary, the actual state trajectory $x(t)$ breaches it. The failure of the actual system at avoiding the obstacle when using a controller with state estimates highlights the limitations of relying solely on CBFs to guarantee the safety of an output feedback nonlinear system without the augmentation of the CBF with a robustifying term.

using a control-affine nonlinear system of the form (1) with state $x = [(x)_1, (x)_2]^\top$, where

$$f(x) = \begin{bmatrix} -(x)_1 + (x)_2 \\ -\frac{1}{2}(x)_1 - \frac{1}{2}(x)_2(1 - (\cos(2(x)_1) + 2)^2) \end{bmatrix}, \quad (67)$$

$g(x) = [0, \cos(2(x)_1) + 2]$, and $C = [0, 1]$. In this study, the system state is not available for measurement, but the output is measurable. The control objective is to minimize the infinite horizon cost in (30) and to drive the trajectories of the nonlinear system in (67) to the origin while ensuring safety. To demonstrate the efficacy of the developed method and to show the robustness of the controller to errors due to the lack of full state measurement, two simulation studies using the dynamics in (67) are performed.

7.1 Ensuring safety by staying within a given set

The first study shows that the developed method can ensure that trajectories of the system remain within a given safe set $\mathcal{S}$ as defined in (4)–(6) where $h(x) = -(x)_2^2 - (x)_1 + 1$ (cf. [23]). The basis for value function approximation is selected as a polynomial basis given by³ $\phi(\hat{\zeta}) := [(\hat{\zeta})_1^2, (\hat{\zeta})_1(\hat{\zeta})_2, (\hat{\zeta})_2^2, (\hat{\zeta})_1(\hat{\zeta})_3, (\hat{\zeta})_1(\hat{\zeta})_3, (\hat{\zeta})_3^2]^\top$. Initial conditions of the state and the estimated state

³ The selection of a polynomial basis is motivated by the Weierstrass Approximation Theorem (see [48, Theorem 1.5]) and the specific monomials used here are obtained using trial and error.

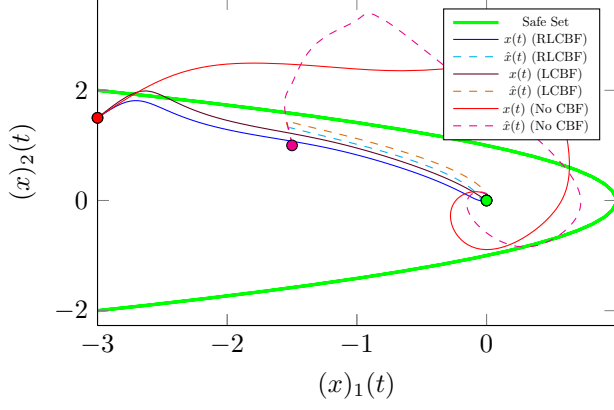


Fig. 2. Ensuring safety by staying within a given set (enclosed by a thick green boundary).

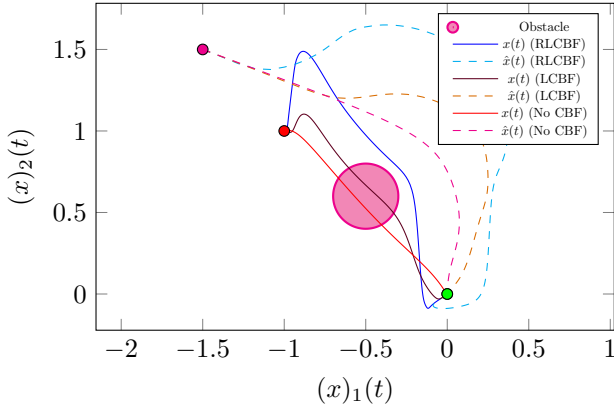


Fig. 3. Ensuring safety by avoiding a circular obstacle.

for the first study are selected as $x(0) = [-3, 1.5]^\top$ and $\hat{x}(0) = [-1.5, 1]^\top$, respectively, with the state and the control penalty in (31) selected as $Q = I_2$ and $R = 1$, respectively. The Jacobian bounds for the functions f and g are selected by assuming that $(x)_1 \in [-3, 3]$ and $(x)_2 \in [-3, 3]$ throughout the simulation, and the barrier function ensures that the bounds remain valid. To obtain the symmetric positive definite matrix, P , and the three observer gains, L_1 , L_2 , and L_3 that satisfy the stability conditions developed in Section 3, the LMI in (17) is solved using **Sedumi** in **YALMIP** [32] with the learning rate $\alpha = 2$ to get the matrices $P = [0.27222, 0.15875; 0.15875, 0.40954]$, $L_3 = [-8.82113, 11.5823]^\top$, $L_1 = [0.14719, 0.14719]^\top$, and $L_2 = [0.045396, 0.045396]^\top$. The gain for the robust barrier function in (33) is selected as $\kappa = 0.01$. Since the true values of the value function weights are unknown, the initial estimate of the weights is selected as $\hat{W}_c(0) = [0.5, 1, 0.8, 0.1, 0.1, 0.1]^\top$ with the initial least squares gain matrix selected as $\Gamma(0) = I_6$. The saturation constraint on the control input is considered as $\bar{u} = 10$, the LMI parameter is selected a $\theta = I_2$, and the learning gains are selected as $k_c = 5$, $\nu = 0.7$, $\beta = 0.01$, $\kappa = 0.1$, $\ell = 0.1$, and $e_0 = 2.5$. The simulation uses 100

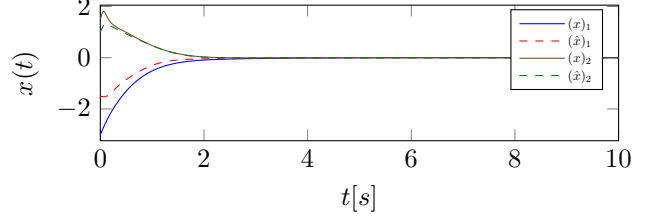


Fig. 4. The trajectories of the actual state x and estimated state $\hat{x}$ for the ensuring safety within a given set.

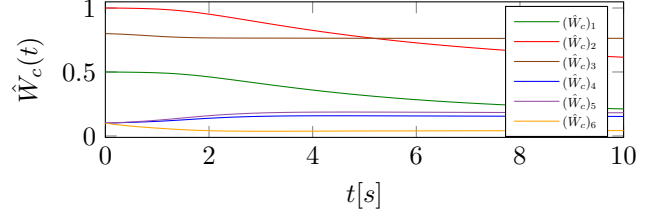


Fig. 5. The trajectories of the estimated critic NN weights for ensuring safety within a given set.

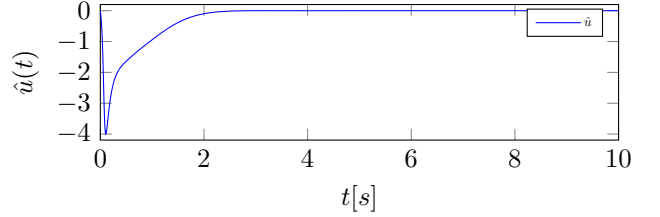


Fig. 6. The trajectory of constrained control input for ensuring safety within a given set.

fixed Bellman error extrapolation points selected from a 1×1 square centered around the origin of the system.

7.2 Obstacle Avoidance

For the second study, an obstacle avoidance problem is considered (33), where the nonlinear system in (67) is tasked with the safety objective of avoiding a circular obstacle while being regulated to the origin using the approximate controller defined in (43), where the safe set in (4)–(6) is defined using $h(x) := \|x - z\|^2 - r^2$, where $z = [(z)_1, (z)_2]^\top$ represents the center of the circular obstacle and $r \in \mathbb{R}_{>0}$ is the radius of the set.

Initial conditions of the state and the estimated state are selected as $x(0) = [-1, 1]^\top$ and $\hat{x}(0) = [-1.5, 1.5]^\top$ with the state and the control penalty in (31) selected as $Q = I_2$ and $R = 1$, respectively. The gain for the robust barrier function in (33) is selected as $\kappa = 2.5$. The Jacobian bounds for the functions f and g are selected by assuming that $(x)_1 \in [-2, 2]$ and $(x)_2 \in [-2, 2]$ throughout the simulation. Using

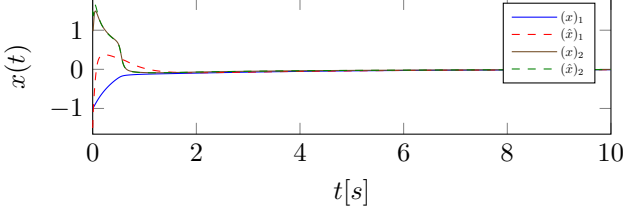


Fig. 7. The trajectories of the actual state x and estimated state $\hat{x}$ for the obstacle avoidance problem.

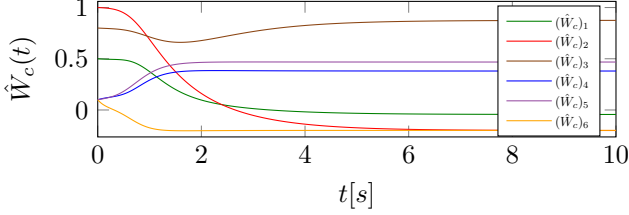


Fig. 8. The trajectories of the estimated critic NN weights for the obstacle avoidance problem.

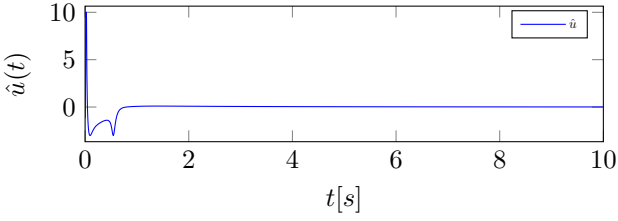


Fig. 9. The trajectory of constrained control input for the obstacle avoidance problem.

this assumption, the values P , L_1 , L_2 , and L_3 are obtained as $P = [0.47897, 1.0306; 1.0306, 2.6555]$, $L_3 = [-99.6211, 41.064]^T$, $L_1 = [0.3956, 0.13187]^T$, and $L_2 = [0.15735, 0.15735]^T$, respectively. The center of the circular obstacle is selected to be $z = [-0.5, 0.6]^T$ and the radius of the obstacle is selected as $r = 0.2$. The simulation uses 100 fixed Bellman error extrapolation points selected from a 2×2 square centered around the origin of the system. The simulation parameters are $\ell = 0.175$, $\epsilon_0 = 0.7$, and the rest of the initial conditions, parameters, and gains for the second simulation study are selected to be the same as the first study in Section 7.1.

7.3 Results

For the first study in Section 7.1, focused on ensuring safety within the set $\mathcal{S}$, it is evident from Figure 2 that the RLCBF-based controller, developed in this study, adheres to the safety constraints even when utilizing the approximate controller designed in (39). This robust performance stems from adding the robustifying term in (27) to account for any difference between the exact and approximate feedback control policy. In contrast, rely-

ing solely on LCBFs, as depicted in Figure 2, proves inadequate in guaranteeing safety. For the obstacle avoidance problem, Figure 3 shows that the LCBF approach also falls short of circumventing the red circular obstacle. Conversely, it can be observed in Figure 3 that the RLCBF-based method ensures that system trajectories in (67) steer clear of the red obstacle while regulating towards the origin.

Figures 4 and 7 show that the trajectories of the actual state and state estimates converge to the same values and are regulated to the origin for both problems, thereby demonstrating the effectiveness of the designed LMI-based observer for state estimation of output-feedback nonlinear systems and validating the stability results developed in Section 3. The weight estimates, as depicted in Figures 5 and 8, are shown to be UUB. While Figures 5 and 8 show that the learning algorithm converges quickly, it cannot be claimed that the estimates of the weights converge to their true values since the true value function and hence the exact critic weights are unknown. Figures 6 and 9 show that the trajectory of the approximate safety-aware controller remains bounded while maintaining the stability of the closed-loop system as it simultaneously learns the optimal value function.

8 Conclusion

In this paper, an observer-based SLOC framework for input-constrained nonlinear systems is developed using an RLCBF-based penalty introduced in the cost function of the optimal control problem. LMIs are formulated to obtain observer gain matrices, and an observer-based safe adaptive near-optimal controller is developed that guarantees safety and maintains stability during the learning phase as it seeks to find an approximate solution to the optimal control problem.

While the simulation results demonstrate the efficacy of our method and validate the results from the Lyapunov-based stability analysis, certain limitations to the developed method exist. The computational complexity introduced by RCBFs, sensitivity to model uncertainties, numerical issues that occur if the LMI is poorly conditioned, and the inherently conservative nature of safe controllers that utilize barrier functions to enforce safety constraints are ongoing areas of investigation. Future research will involve exploring techniques involving tunable CBFs like those in [42] to improve system safety while being sufficiently far from the boundary of the safe set. Additionally, the integration of current adaptive techniques to address model uncertainties holds promise for advancing the robustness of the proposed framework.

References

- [1] Farzaneh Abdollahi, Heidar A Talebi, and Rajnikant V Patel. A stable neural network-based observer with application

- to flexible-joint manipulators. *IEEE Trans. Neural Netw.*, 17(1):118–129, 2006.
- [2] Sander Adam, Lucian Busoniu, and Robert Babuska. Experience replay for real-time reinforcement learning control. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, 42(2):201–212, 2012.
 - [3] Devansh R Agrawal and Dimitra Panagou. Safe and robust observer-controller synthesis using control barrier functions. *IEEE Control Syst. Lett.*, 7:127–132, 2022.
 - [4] Hassan Almubarak, Maitham F. AL-Sunni, Justin T. Dubbin, Nader Sadegh, John M. Dolan, and Evangelos A. Theodorou. Nearly optimal nonlinear safe control with basdre. arXiv:2504.15453, 2025.
 - [5] Hassan Almubarak, Nader Sadegh, and Evangelos A. Theodorou. Barrier states theory for safety-critical multi-objective control. *IEEE Trans. Autom. Control*, pages 1–16, 2025.
 - [6] Hassan Almubarak, Evangelos Theodorou, and Nader Sadegh. HJB based optimal safe control using control barrier functions. In *Proc. IEEE Conf. Decis. Control*, pages 6829–6834. IEEE, 2021.
 - [7] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada. Control barrier function based quadratic programs for safety critical systems. *IEEE Trans. Autom. Control*, 62(8):3861–3876, August 2017.
 - [8] Murat Arcak and Petar Kokotović. Nonlinear observers: a circle criterion design and robustness analysis. *Automatica*, 37(12):1923–1930, 2001.
 - [9] Behçet Açıkmeşe and Martin Corless. Stability analysis with quadratic Lyapunov functions: Some necessary and sufficient multiplier conditions. *Syst. Control Lett.*, 57(1):78–94, 2008.
 - [10] Soutrik Bandyopadhyay and Shubhendu Bhasin. Lagrangian-based online safe reinforcement learning for state-constrained systems. *Automatica*, 179:112458, 2025.
 - [11] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear matrix inequalities in system and control theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, 1994.
 - [12] Yinlam Chow, Ofir Nachum, Edgar Duenez-Guzman, and Mohammad Ghavamzadeh. A Lyapunov-based approach to safe reinforcement learning. *Adv. Neural Inf. Process. Syst.*, 31, 2018.
 - [13] Pawel Cichosz. An analysis of experience replay in temporal difference learning. *Cybern. Syst.*, 30(5):341–363, 1999.
 - [14] M. H. Cohen and C. Belta. Approximate optimal control for safety-critical systems with control barrier functions. In *Proc. IEEE Conf. Decis. Control*, pages 2062–2067, 2020.
 - [15] Max H. Cohen and Calin Belta. Safe exploration in model-based reinforcement learning using control barrier functions. *Automatica*, 147:110684, 2023.
 - [16] Max H Cohen, Zachary Serlin, Kevin Leahy, and Calin Belta. Temporal logic guided safe model-based reinforcement learning: a hybrid systems approach. *Nonlinear Anal. Hybrid Syst.*, 47:101295, 2023.
 - [17] Ryan Cosner, Andrew Singletary, Andrew Taylor, Tamas Molnar, Katherine Bouman, and Aaron Ames. Measurement-robust control barrier functions: Certainty in safety with uncertainty in state. In *IEEE/RSJ Int. Conf. Intel. Robot. Syst.*, pages 6286–6291. IEEE, IEEE, 2021.
 - [18] Sarah Dean, Andrew Taylor, Ryan Cosner, Benjamin Recht, and Aaron Ames. Guaranteeing safety of learned perception modules via measurement-robust control barrier functions. In *Conf. Robot. Learn.*, pages 654–670. PMLR, 2021.
 - [19] P. Deptula, H. Chen, R. A. Licitra, J. A. Rosenfeld, and W. E. Dixon. Approximate optimal motion planning to avoid unknown moving avoidance regions. *IEEE Trans. Robot.*, 36(2):414–430, 2020.
 - [20] Mondher Farza, Amine Sboui, Estelle Cherrier, and Mohammed M’Saad. High-gain observer for a class of time-delay nonlinear systems. *Int. J. Control*, 83(2):273–280, 2010.
 - [21] K. Graichen and N. Petit. Incorporating a class of constraints into the dynamics of optimal control problems. *Optim. Control Appl. Methods*, 30(6):537–561, 2009.
 - [22] M. L. Greene, P. Deptula, S. Nivison, and W. E. Dixon. Sparse learning-based approximate dynamic programming with barrier constraints. *IEEE Control Syst. Lett.*, 4(3):743–748, 2020.
 - [23] Mrdjan Jankovic. Robust control barrier functions for constrained stabilization of nonlinear systems. *Automatica*, 96:359–367, 2018.
 - [24] Woongsun Jeon, Ankush Chakrabarty, Ali Zemouche, and Rajesh Rajamani. LMI-based neural observer for state and nonlinear function estimation. *Int. J. Robust Nonlinear Control*, 34(10):6964–6984, 2024.
 - [25] Rushikesh Kamalapurkar, Huyen T. Dinh, Shubhendu Bhasin, and Warren E. Dixon. Approximate optimal trajectory tracking for continuous-time nonlinear systems. *Automatica*, 51:40–48, January 2015.
 - [26] Rushikesh Kamalapurkar, Joel A. Rosenfeld, and Warren E. Dixon. Efficient model-based reinforcement learning for approximate online optimal control. *Automatica*, 74:247–258, December 2016.
 - [27] Rushikesh Kamalapurkar, Patrick Walters, and Warren E. Dixon. Model-based reinforcement learning for approximate optimal regulation. *Automatica*, 64:94–104, February 2016.
 - [28] Rushikesh Kamalapurkar, Patrick Walters, Joel A. Rosenfeld, and Warren E. Dixon. *Reinforcement learning for optimal feedback control: A Lyapunov-based approach*. Communications and Control Engineering. Springer International Publishing, May 2018.
 - [29] Hassan K. Khalil. *Nonlinear systems*. Prentice Hall, Upper Saddle River, NJ, third edition, 2002.
 - [30] Y. H. Kim, F. L. Lewis, and C. T. Abdallah. A dynamic recurrent neural-network-based adaptive observer for a class of nonlinear systems. *Automatica*, 33:1539–1543, 1997.
 - [31] Shishir Kolathaya and Aaron D. Ames. Input-to-state safety with control barrier functions. *IEEE Control Syst. Lett.*, 3(1):108–113, 2019.
 - [32] Johan. Lofberg. YALMIP : a toolbox for modeling and optimization in MATLAB. In *IEEE Int. Conf. on Robot. and Autom. (IEEE Cat. No.04CH37508)*, pages 284–289, 2004.
 - [33] Brett T Lopez, Jean-Jacques E Slotine, and Jonathan P How. Robust adaptive control barrier functions: An adaptive and data-driven approach to safety. *IEEE Control Syst. Lett.*, 5(3):1031–1036, 2020.
 - [34] S M Nahid Mahmud, Scott A. Nivison, Zachary I. Bell, and Rushikesh Kamalapurkar. Safe model-based reinforcement learning for systems with parametric uncertainties. *Front. Robot. AI*, 8(733104):1–13, December 2021.
 - [35] Zahra Marvi and Bahare Kiumarsi. Safe reinforcement learning: A control barrier function optimization approach. *Int. J. Robust Nonlinear Control*, 31(6):1923–1940, 2021.
 - [36] Tapas Mazumdar. Generalized projection theorem and weak noncoercive evolution problems in Hilbert space. *J. Math. Anal. Appl.*, 46(1):143–168, 1974.

- [37] Hamidreza Modares, Frank L. Lewis, and Mohammad-Bagher Naghibi-Sistani. Adaptive optimal control of unknown constrained-input systems using policy iteration and neural networks. *IEEE Trans. Neural Netw. Learn. Syst.*, 24(10):1513–1525, 2013.
- [38] Hamidreza Modares, Frank L. Lewis, and Mohammad-Bagher Naghibi-Sistani. Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems. *Automatica*, 50(1):193–202, 2014.
- [39] Shivaraj Mohite, Marouane Alma, and Ali Zemouche. Toward enhancing nonlinear observers for Lipschitz system: Exploiting the matrix multipliers-based LMIs. *Int. J. Robust Nonlinear Control*, 34(15):10799–10820, 2024.
- [40] Tochukwu Elijah Ogri, Zachary I. Bell, and Rushikesh Kamalapurkar. State and parameter estimation for affine nonlinear systems. In *Proc. IEEE Conf. Decis. Control*, pages 1517–1522, Singapore, December 2023.
- [41] Tochukwu Elijah Ogri, S. M. Nahid Mahmud, Zachary I. Bell, and Rushikesh Kamalapurkar. Output feedback adaptive optimal control of affine nonlinear systems with a linear measurement model. In *Proc. IEEE Conf. Control Technol. Appl.*, pages 645–650, August 2023.
- [42] Hardik Parwana, Aquib Mustafa, and Dimitra Panagou. Trust-based rate-tunable control barrier functions for non-cooperative multi-agent systems. In *Proc. IEEE Conf. Decis. Control*, pages 2222–2229. IEEE, 2022.
- [43] G Phnomchoeng and Rajesh Rajamani. The bounded Jacobian approach to nonlinear observer design. In *Proc. Am. Control Conf.*, pages 6083–6088, 2010.
- [44] Daniel Quintana, Víctor Estrada-Manzo, and Miguel Bernal. An exact handling of the gradient for overcoming persistent problems in nonlinear observer design via convex optimization techniques. *Fuzzy Sets Syst.*, 416:125–140, 2021. Systems Engineering.
- [45] R. Rajamani. Observers for Lipschitz nonlinear systems. *IEEE Trans. Autom. Control*, 43(3):397–401, March 1998.
- [46] Rajesh Rajamani, Woongsun Jeon, Hamidreza Movahedi, and Ali Zemouche. On the need for switched-gain observers for non-monotonic nonlinear systems. *Automatica*, 114:108814, 2020.
- [47] Matheus F. Reis, A. Pedro Aguiar, and Paulo Tabuada. Control barrier function-based quadratic programs introduce undesirable asymptotically stable equilibria. *IEEE Control Syst. Lett.*, 5(2):731–736, 2021.
- [48] Friedrich Sauvigny. *Partial Differential Equations 1*. Springer, 2012.
- [49] Rin Takano and Masaki Yamakita. Robust constrained stabilization control using control Lyapunov and control barrier function in the presence of measurement noises. In *Proc. IEEE Conf. Control Technol. Appl.*, pages 300–305. IEEE, 2018.
- [50] K. G. Vamvoudakis and F. L. Lewis. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica*, 46(5):878–888, 2010.
- [51] Yan Wang, Rajesh Rajamani, and David M. Bevly. Observer design for differentiable Lipschitz nonlinear systems with time-varying parameters. In *Proc. IEEE Conf. Decis. Control*, pages 145–152, 2014.
- [52] Yan Wang, Rajesh Rajamani, and David M. Bevly. Observer design for parameter varying differentiable nonlinear systems, with application to slip angle estimation. *IEEE Trans. Autom. Control*, 62(4):1940–1945, 2017.
- [53] Yujie Wang and Xiangru Xu. Observer-based control barrier functions for safety critical systems. In *Proc. Am. Control Conf.*, pages 709–714. IEEE, 2022.
- [54] Paweł Wawrzyński. Real-time reinforcement learning by sequential actor-critics and experience replay. *Neural Netw.*, 22(10):1484–1497, 2009.
- [55] Peter Wieland and Frank Allgöwer. Constructive safety using control barrier functions. *IFAC Proc. Vol.*, 40(12):462–467, 2007.
- [56] A.G. Wills and W.P. Heath. A recentred barrier for constrained receding horizon control. In *Proc. Am. Control Conf.*, volume 5, pages 4177–4182, 2002.
- [57] Xiong Yang, Derong Liu, and Yuzhu Huang. Neural-network-based online optimal control for uncertain nonlinear continuous-time systems with control constraints. *IET Control Theory Appl.*, 7(17):2037–2047, 2013.
- [58] Xiong Yang, Derong Liu, and Qinglai Wei. Online approximate optimal control for affine non-linear systems with unknown internal dynamics using adaptive dynamic programming. *IET Control Theory Appl.*, 8(16):1676–1688, 2014.
- [59] Yongliang Yang, Kyriakos G. Vamvoudakis, and Hamidreza Modares. Safe reinforcement learning for dynamical games. *Int. J. Robust Nonlinear Control*, 30(9):3706–3726, 2020.
- [60] Ali Zemouche and Mohamed Boutayeb. On LMI conditions to design observers for Lipschitz nonlinear systems. *Automatica*, 49(2):585–591, 2013.
- [61] Ali Zemouche, Mohamed Boutayeb, and G. Iulia Bara. Observer design for nonlinear systems: An approach based on the Differential Mean Value Theorem. In *Proc. IEEE Conf. Decis. Control*, pages 6353–6358, 2005.
- [62] Ali Zemouche and Rajesh Rajamani. Observer design for non-globally lipschitz nonlinear systems using Hilbert Projection Theorem. *IEEE Control Syst. Lett.*, 6:2581–2586, 2022.
- [63] Huaguang Zhang, Lili Cui, Xin Zhang, and Yanhong Luo. Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method. *IEEE Trans. Neural Netw.*, 22(12):2226–2236, December 2011.