# Stability Analysis Using Quadratic Constraints for Systems With Neural Network Controllers

He Yin, Peter Seiler, and Murat Arcak, *Fellow, IEEE*

***Abstract*—A method is presented to analyze the stability of feedback systems with neural network controllers. Two stability theorems are given to prove asymptotic stability and to compute an ellipsoidal innerapproximation to the region of attraction (ROA). The first theorem addresses linear time-invariant systems, and merges Lyapunov theory with local (sector) quadratic constraints to bound the nonlinear activation functions in the neural network. The second theorem allows the system to include perturbations such as unmodeled dynamics, slope-restricted nonlinearities, and time delay, using integral quadratic constraint (IQCs) to capture their input/output behavior. This in turn allows for off-by-one IQCs to refine the description of activation functions by capturing their slope restrictions. Both results rely on semidefinite programming to approximate the ROA. The method is illustrated on systems with neural networks trained to stabilize a nonlinear inverted pendulum as well as vehicle lateral dynamics with actuator uncertainty.**

***Index Terms*—LMIs, neural networks, stability of linear systems, uncertain systems.**

## I. Introduction

The paradigm of stabilizing dynamical systems with neural network (NN) controllers [1] has been revived following recent development in deep NNs, e.g., policy gradient [2]–[5] and behavioral cloning [6]. However, feedback systems with NN controllers suffer from lack of stability and safety certificates due to the complexity of the NN structure. Specifically, NNs have various types of nonlinear activation functions, potentially numerous layers, and a large number of hidden neurons, making it difficult to apply classical analysis methods, e.g., Lyapunov theory [7]. Monte Carlo simulations can be used to investigate stability but lack formal guarantees, which are important in safety-critical applications.

Several works propose using quadratic constraints (QCs) to bound the nonlinear activation functions. This approach is used to outerbound the outputs of an (static) NN given a set of inputs in [8] and upperbound the Lipschitz constant of NNs in [9] and [10]. The work [11] uses this idea for finite-time reachability analysis of a system with an NN controller. The work [12] performs stability analysis by constructing QCs from the bounds of partial gradients of NN controllers. Kim *et al.* [13] assess global asymptotic stability of dynamic NN models using QCs and Lyapunov theory.

This article presents two main stability results for a feedback system with an NN controller. Theorem 1 provides a condition to prove stability and to innerapproximate the region of attraction (ROA) for a linear time-invariant (LTI) plant. It uses Lyapunov theory, and local (sector) QCs to bound the nonlinear activation functions in the NN. Theorem 2 allows the plant to include perturbations such as unmodeled dynamics, slope-restricted nonlinearities, and time delay, characterizing them with integral quadratic constraints (IQCs) [14], [15]. This in turn allows for the use of off-by-one IQCs [16] to capture the slope restrictions of activation functions. Both results rely on semidefinite programming to approximate the ROA.

The specific contributions of this article are three-fold. First, our nominal analysis with LTI plants and NN controllers uses *offset* local sector QCs that are centered around the equilibrium inputs to the NN activation functions and allow for analyzing stability around a nonzero equilibrium point. Second, our analysis of uncertain plants and NN controllers provides robustness guarantees for the feedback system. The uncertain plant is modeled as an interconnection of the nominal plant and perturbations that are described by IQCs. The use of IQCs also allows for plants that are not necessarily LTI. Third, the proposed framework allows for local (dynamic) off-by-one IQCs to further sharpen the description of activation functions by capturing their slope restrictions. This differs from [8]–[12], which derive only static QCs for activation functions.

Local (static) sector IQCs have been used in the stability analysis of linear systems with actuator saturation [17], [18], and unbounded nonlinearities [19]. The description of these nonlinearties is refined by incorporating soft (dynamic) IQCs in the stability analysis framework for linear systems [20], and polynomial systems [21]. Compared with these works, this article is specialized to NN-controlled systems, it exploits the specific properties of NNs and uses the interval bound propagation method [22] to derive nonconservative static and dynamic local IQCs to describe NN controllers; and it also allows for the analysis of NN-controlled nonlinear systems by accommodating perturbations.

This article is organized as follows. Section II presents the problem formulation and the nominal stability analysis when the plant is LTI. Section III addresses uncertain systems using IQCs. Section IV provides numerical examples, including a nonlinear inverted pendulum and an uncertain vehicle model.

*Notation:* $\mathbb{S}^n$ denotes the set of $n$-by-$n$ symmetric matrices. $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ denote the sets of $n$-by-$n$ symmetric, positive semidefinite, and positive definite matrices, respectively. $\mathbb{RL}_\infty$ is the set of rational functions with real coefficients and no poles on the unit circle. $\mathbb{RH}_\infty \subset \mathbb{RL}_\infty$ contains functions that are analytic in the closed exterior of the unit disk in the complex plane. $\ell_2^{n_x}$ is the set of sequences $x : \mathbb{N} \to \mathbb{R}^{n_x}$ with $\|x\|_2 := \sqrt{\sum_{k=0}^{\infty} x(k)^\top x(k)} < \infty$. When applied to vectors, the orders $>, \leq$ are applied elementwise. For $P \in \mathbb{S}^n_{++}$, $x_* \in \mathbb{R}^n$, define

Manuscript received January 26, 2021; accepted March 23, 2021. Date of publication March 29, 2021; date of current version March 29, 2022. This work was supported in part by the Air Force Office of Scientific Research under Grant FA9550-18-1-0253, in part by the Office of Naval Research under Grant N00014-18-1-2209, and in part by the U.S. National Science Foundation under Grant ECCS-1906164. Recommended by Associate Editor A. Macchelli. *(Corresponding author: He Yin.)*

He Yin and Murat Arcak are with the University of California, Berkeley, Berkeley, CA 94010 USA (e-mail: he_yin@berkeley.edu; arcak@eecs.berkeley.edu).

Peter Seiler is with the University of Michigan, Ann Arbor,, MI 48109 USA (e-mail: pseiler@umich.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2021.3069388.

Digital Object Identifier 10.1109/TAC.2021.3069388

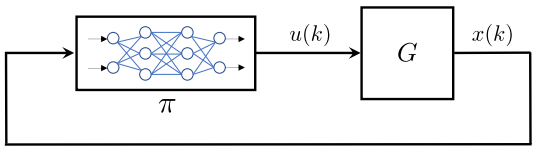


Fig. 1. Feedback system with plant $G$ and NN $\pi$.

the ellipsoid

$$\mathcal{E}(P, x_*) := \{x \in \mathbb{R}^n : (x - x_*)^\top P(x - x_*) \leq 1\}. \tag{1}$$

## II. Nominal Stability Analysis

### A. Problem Formulation

Consider the feedback system consisting of a plant $G$ and state-feedback controller $\pi$, as shown in Fig. 1. As a first step, we assume the plant $G$ is an LTI system defined by the following discrete-time model:

$$x(k+1) = A_G\, x(k) + B_G\, u(k) \tag{2}$$

where $x(k) \in \mathbb{R}^{n_G}$ is the state, $u(k) \in \mathbb{R}^{n_u}$ is the input, $A_G \in \mathbb{R}^{n_G \times n_G}$ and $B_G \in \mathbb{R}^{n_G \times n_u}$. The controller $\pi : \mathbb{R}^{n_G} \to \mathbb{R}^{n_u}$ is an $\ell$-layer, feed-forward NN defined as

$$w^0(k) = x(k) \tag{3a}$$

$$w^i(k) = \phi^i\left(W^i w^{i-1}(k) + b^i\right),\ i = 1, \ldots, \ell \tag{3b}$$

$$u(k) = W^{\ell+1} w^\ell(k) + b^{\ell+1} \tag{3c}$$

where $w^i \in \mathbb{R}^{n_i}$ are the outputs (activations) from the $i$th layer and $n_0 = n_G$. The operations for each layer are defined by a weight matrix $W^i \in \mathbb{R}^{n_i \times n_{i-1}}$, bias vector $b^i \in \mathbb{R}^{n_i}$, and activation function $\phi^i : \mathbb{R}^{n_i} \to \mathbb{R}^{n_i}$. The activation function $\phi^i$ is applied element-wise, i.e.,

$$\phi^i(v) := \left[\varphi(v_1), \ldots, \varphi(v_{n_i})\right]^\top \tag{4}$$

where $\varphi : \mathbb{R} \to \mathbb{R}$ is the (scalar) activation function selected for the NN. Common choices for the scalar activation function include $\varphi(\nu) := \tanh(\nu)$, sigmoid $\varphi(\nu) := \frac{1}{1+e^{-\nu}}$, ReLU $\varphi(\nu) := \max(0, \nu)$, and leaky ReLU $\varphi(\nu) := \max(a\nu, \nu)$ with $a \in (0, 1)$. We assume the activation $\varphi$ is identical in all layers; this can be relaxed with minor changes to the notation.

The state vector $x_*$ is an equilibrium point of the feedback system with input $u_*$ if the following conditions hold:

$$x_* = A_G\, x_* + B_G\, u_* \tag{5a}$$

$$u_* = \pi(x_*). \tag{5b}$$

Let $\chi(k; x_0)$ denote the solution to the feedback system at time $k$ from initial condition $x(0) = x_0$. Our goal is to analyze asymptotic stability of the equilibrium point and to find the largest estimate of the ROA, defined in the following, using an ellipsoidal innerapproximation.

*Definition 1:* The ROA of the feedback system with plant $G$ and NN $\pi$ is defined as

$$\mathcal{R} := \{x_0 \in \mathbb{R}^{n_G} : \lim_{k \to \infty} \chi(k; x_0) = x_*\}. \tag{6}$$

### B. NN Representation: Isolation of Nonlinearities

It is useful to isolate the nonlinear activation functions from the linear operations of the NN as done in [8] and [13]. Define $v^i$ as the input to the activation function $\phi^i$

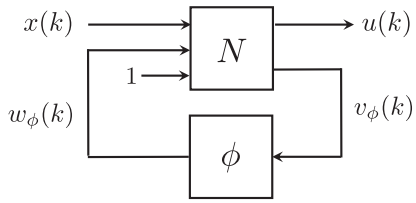


Fig. 2. NN representation to isolate the nonlinearities $\phi$.

$$v^i(k) := W^i w^{i-1}(k) + b^i,\ i = 1, \ldots, \ell. \tag{7}$$

The nonlinear operation of the $i$th layer (3b) is, thus, expressed as $w^i(k) = \phi^i(v^i(k))$. Gather the inputs and outputs of all activation functions

$$v_\phi := \begin{bmatrix} v^1 \\ \vdots \\ v^\ell \end{bmatrix} \in \mathbb{R}^{n_\phi} \text{ and } w_\phi := \begin{bmatrix} w^1 \\ \vdots \\ w^\ell \end{bmatrix} \in \mathbb{R}^{n_\phi} \tag{8}$$

where $n_\phi := n_1 + \cdots + n_\ell$, and define the combined nonlinearity $\phi : \mathbb{R}^{n_\phi} \to \mathbb{R}^{n_\phi}$ by stacking the activation functions

$$\phi(v_\phi) := \begin{bmatrix} \phi^1(v^1) \\ \vdots \\ \phi^\ell(v^\ell) \end{bmatrix}. \tag{9}$$

Thus, $w_\phi(k) = \phi(v_\phi(k))$, where the scalar activation function $\varphi$ is applied element-wise to each entry of $v_\phi$. Finally, the NN control policy $\pi$ defined in (3) can be rewritten as

$$\begin{bmatrix} u(k) \\ v_\phi(k) \end{bmatrix} = N \begin{bmatrix} x(k) \\ w_\phi(k) \\ 1 \end{bmatrix} \tag{10a}$$

$$w_\phi(k) = \phi(v_\phi(k)). \tag{10b}$$

The matrix $N$ depends on the weights and biases as follows, where the vertical and horizontal bars partition $N$ compatibly with the inputs $(x, w_\phi, 1)$ and outputs $(u, v_\phi)$

$$N := \left[\begin{array}{c|ccccc|c} 0 & 0 & 0 & \cdots & W^{\ell+1} & b^{\ell+1} \\ \hline W^1 & 0 & \cdots & 0 & 0 & b^1 \\ 0 & W^2 & \cdots & 0 & 0 & b^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & W^\ell & 0 & b^\ell \end{array}\right] \tag{11a}$$

$$:= \begin{bmatrix} N_{ux} & N_{uw} & N_{ub} \\ N_{vx} & N_{vw} & N_{vb} \end{bmatrix}. \tag{11b}$$

This decomposition of the NN, depicted in Fig. 2, isolates the activation functions in preparation for the stability analysis.

Suppose $(x_*, u_*)$ satisfies (5). Then, $x_*$ can be propagated through the NN to obtain equilibrium values $v_*^i$, $w_*^i$ for the inputs/outputs of each activation function ($i = 1, \ldots, \ell$), yielding $(v_\phi, w_\phi) = (v_*, w_*)$. Thus, $(x_*, u_*, v_*, w_*)$ is an equilibrium point of (2) and (3) if

$$x_* = A_G\, x_* + B_G\, u_* \tag{12a}$$

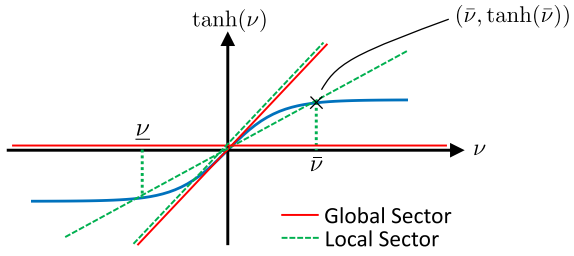


Fig. 3. Sector constraints on tanh.

$$\begin{bmatrix} u_* \\ v_* \end{bmatrix} = N \begin{bmatrix} x_* \\ w_* \\ 1 \end{bmatrix} \tag{12b}$$

$$w_* = \phi(v_*). \tag{12c}$$

### C. QCs: Scalar Activation Functions

The stability analysis relies on QCs to bound the activation function. A typical constraint is the sector bound as defined in the following.

*Definition 2:* Let $\alpha \le \beta$ be given. The function $\varphi : \mathbb{R} \to \mathbb{R}$ lies in the (global) sector $[\alpha, \beta]$ if

$$(\varphi(\nu) - \alpha\nu) \cdot (\beta\nu - \varphi(\nu)) \ge 0 \quad \forall \nu \in \mathbb{R}. \tag{13}$$

The interpretation of the sector $[\alpha, \beta]$ is that $\varphi$ lies between lines passing through the origin with slope $\alpha$ and $\beta$. Many activation functions are bounded in the sector [0,1], e.g., tanh and ReLU. Fig. 3 illustrates $\varphi(\nu) = \tanh(\nu)$ (blue solid) and the global sector defined by [0,1] (red solid lines).

The global sector constraint is often too coarse for stability analysis, and a local sector constraint provides tighter bounds.

*Definition 3:* Let $\alpha$, $\beta$, $\underline{\nu}$, $\bar{\nu} \in \mathbb{R}$ with $\alpha \le \beta$ and $\underline{\nu} \le 0 \le \bar{\nu}$. The function $\varphi : \mathbb{R} \to \mathbb{R}$ satisfies the local sector $[\alpha, \beta]$ if

$$(\varphi(\nu) - \alpha\,\nu) \cdot (\beta\,\nu - \varphi(\nu)) \ge 0 \quad \forall \nu \in [\underline{\nu}, \bar{\nu}]. \tag{14}$$

As an example, $\varphi(\nu) := \tanh(\nu)$ restricted to the interval $[-\bar{\nu}, \bar{\nu}]$ satisfies the local sector bound $[\alpha, \beta]$ with $\alpha := \tanh(\bar{\nu})/\bar{\nu} > 0$ and $\beta := 1$. As shown in Fig. 3 (green dashed lines), the local sector provides a tighter bound than the global sector. These bounds are valid for a symmetric interval around the origin with $\underline{\nu} = -\bar{\nu}$; nonsymmetric intervals ($\underline{\nu} \ne -\bar{\nu}$) can be handled similarly.

The local and global sector constraints mentioned above were defined to be centered at the point $(\nu, \varphi(\nu)) = (0, 0)$. The stability analysis will require offset sectors centered around an arbitrary point $(\nu_*, \varphi(\nu_*))$ on the function. For example, $\varphi(\nu) = \tanh(\nu)$ satisfies the global sector bound (red solid) around the point $(\nu_*, \tanh(\nu_*))$ with $[\alpha, \beta] = [0, 1]$, as shown in Fig. 4. It satisfies a tighter local sector bound (green dashed) when the input is restricted to $\nu \in [\underline{\nu}, \bar{\nu}]$. An explicit expression for this local sector is $\beta = 1$ and

$$\alpha := \min\left( \frac{\tanh(\bar{\nu}) - \tanh(\nu_*)}{\bar{\nu} - \nu_*}, \frac{\tanh(\nu_*) - \tanh(\underline{\nu})}{\nu_* - \underline{\nu}} \right).$$

The local sector upper bound $\beta$ can be tightened further. This leads to the following definition of an offset local sector.

*Definition 4:* Let $\alpha$, $\beta$, $\underline{\nu}$, $\bar{\nu}$, $\nu_* \in \mathbb{R}$ be given with $\alpha \le \beta$ and $\underline{\nu} \le \nu_* \le \bar{\nu}$. The function $\varphi : \mathbb{R} \to \mathbb{R}$ satisfies the offset local sector $[\alpha, \beta]$ around the point $(\nu_*, \varphi(\nu_*))$ if

$$(\Delta\varphi(\nu) - \alpha\,\Delta\nu) \cdot (\beta\,\Delta\nu - \Delta\varphi(\nu)) \ge 0 \quad \forall \nu \in [\underline{\nu}, \bar{\nu}] \tag{15}$$

where $\Delta\varphi(\nu) := \varphi(\nu) - \varphi(\nu_*)$ and $\Delta\nu := \nu - \nu_*$.

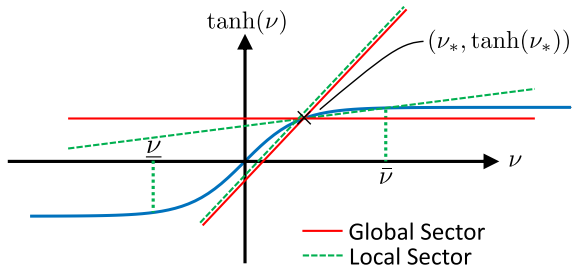


Fig. 4. Offset local sector constraint on tanh.

### D. QCs: Combined Activation Functions

Offset local sector constraints can also be defined for the combined nonlinearity $\phi$, given by (9). Let $\underline{v}$, $\bar{v}$, $v_* \in \mathbb{R}^{n_\phi}$ be given with $\underline{v} \le v_* \le \bar{v}$. Assume that the activation input $v_\phi \in \mathbb{R}^{n_\phi}$ lies, element-wise, in the interval $[\underline{v}, \bar{v}]$, and the $i$th input/output pair is $w_{\phi,i} = \varphi(v_{\phi,i})$. Further assume the scalar activation function satisfies the local sector $[\alpha_i, \beta_i]$ around the point $v_{*,i}$ with the input restricted to $v_{\phi,i} \in [\underline{v}_i, \bar{v}_i]$ for $i = 1, \ldots, n_\phi$. The local sector bounds can be computed for $\varphi$ on the given interval either analytically (as above for tanh) or numerically. These local sectors can be stacked into vectors $\alpha_\phi, \beta_\phi \in \mathbb{R}^{n_\phi}$ that provide QCs satisfied by the combined nonlinearity $\phi$.

*Lemma 1:* Let $\alpha_\phi$, $\beta_\phi$, $\underline{v}$, $\bar{v}$, $v_* \in \mathbb{R}^{n_\phi}$ be given with $\alpha_\phi \le \beta_\phi$, $\underline{v} \le v_* \le \bar{v}$, and $w_* := \phi(v_*)$. Assume $\phi$ satisfies the offset local sector $[\alpha_\phi, \beta_\phi]$ around the point $(v_*, w_*)$ element-wise for all $v_\phi \in [\underline{v}, \bar{v}]$. If $\lambda \in \mathbb{R}^{n_\phi}$ with $\lambda \ge 0$ then

$$\begin{bmatrix} v_\phi - v_* \\ w_\phi - w_* \end{bmatrix}^\top \Psi_\phi^\top M_\phi(\lambda) \Psi_\phi \begin{bmatrix} v_\phi - v_* \\ w_\phi - w_* \end{bmatrix} \ge 0$$

$$\forall v_\phi \in [\underline{v}, \bar{v}],\ w_\phi = \phi(v_\phi)$$

$$\text{where } \Psi_\phi := \begin{bmatrix} diag(\beta_\phi) & -I_{n_\phi} \\ -diag(\alpha_\phi) & I_{n_\phi} \end{bmatrix} \tag{16}$$

$$M_\phi(\lambda) := \begin{bmatrix} 0_{n_\phi} & diag(\lambda) \\ diag(\lambda) & 0_{n_\phi} \end{bmatrix}. \tag{17}$$

*Proof:* For any $v_\phi \in \mathbb{R}^{n_\phi}$ and $w_\phi = \phi(v_\phi)$

$$\begin{bmatrix} v_\phi - v_* \\ w_\phi - w_* \end{bmatrix}^\top \Psi_\phi^\top M_\phi(\lambda) \Psi_\phi \begin{bmatrix} v_\phi - v_* \\ w_\phi - w_* \end{bmatrix}$$

$$= \sum_{i=1}^{n_\phi} \lambda_i (\Delta w_i - \alpha_i\,\Delta v_i) \cdot (\beta_i\,\Delta v_i - \Delta w_i)$$

where $\Delta w_i := \varphi(v_{\phi,i}) - \varphi(v_{*,i})$ and $\Delta v_i := v_{\phi,i} - v_{*,i}$. If $v_\phi \in [\underline{v}, \bar{v}]$ then each term in the sum is non-negative by the offset local sector constraints and $\lambda \ge 0$. ■

In order to apply the local sector and slope bounds in the stability analysis, we must first compute the bounds $\underline{v}, \overline{v} \in \mathbb{R}^{n_\phi}$ on the activation input $v_\phi$. The process to compute the bounds is briefly discussed here with more details provided in [22]. Let $v_*^1$ be the equilibrium value at the first NN layer. Select $\underline{v}^1$, $\bar{v}^1 \in \mathbb{R}^{n_1}$ with $\underline{v}^1 \le v_*^1 \le \bar{v}^1$. The assumed bounds on $v^1$ can be used to compute an interval $[\underline{w}^1, \bar{w}^1]$ for the output $w^1 = \phi^1(v^1)$[1] which can then be used to compute bounds $[\underline{v}^2, \bar{v}^2]$ on

[1]For example, if $\varphi(\nu) = \tanh(\nu)$ then the input bound $\nu \in [-\bar{\nu}, \bar{\nu}]$ implies the output bound $\varphi(\nu) \in [-\tanh(\bar{\nu}), \tanh(\bar{\nu})]$.

the input $v^2$ to the next activation function.[2] The intervals computed for $w^1$ and $v^2$ will contain their equilibrium value $w_*^1$ and $v_*^2$. This process can be propagated through all layers of the NN to obtain the bounds $\underline{v}, \overline{v} \in \mathbb{R}^{n_\phi}$ for the activation function input $v_\phi$. The remainder of this article will assume the local sector bounds have been computed as briefly summarized in the following property.

*Property 1:* Let $v_* \in \mathbb{R}^{n_\phi}$ be an equilibrium value of the activation input and $v_*^1 \in \mathbb{R}^{n_1}$ be the corresponding value at the first layer. Let $\underline{v}^1$, $\bar{v}^1 \in \mathbb{R}^{n_1}$ with $v_*^1 \in [\underline{v}^1, \bar{v}^1]$ and their corresponding activation input bounds $\underline{v}, \overline{v}$ be given. There exist $\alpha_\phi$, $\beta_\phi \in \mathbb{R}^{n_\phi}$ such that $\phi$ satisfies the offset local sector around the point $(v_*, \phi(v_*))$ for all $v_\phi \in [\underline{v}, \bar{v}]$.

### E. Lyapunov Condition

This section uses a Lyapunov function and the offset local sector to compute an innerapproximation for the ROA of the feedback system of $G$ and $\pi$. To simplify notation, the interval bound on $v^1$ is assumed to be symmetrical about $v_*^1$, i.e., $\underline{v}^1 = 2v_*^1 - \bar{v}^1$ so that $\bar{v}^1 - v_*^1 = v_*^1 - \underline{v}^1$. This can be relaxed to handle nonsymmetrical intervals with minor notational changes.

*Theorem 1:* Consider the feedback system of plant $G$ in (2) and NN $\pi$ in (3) with equilibrium point $(x_*, u_*, v_*, w_*)$ satisfying (12). Let $\bar{v}^1 \in \mathbb{R}^{n_1}$, $\underline{v}^1 := 2v_*^1 - \bar{v}^1$, and $\alpha_\phi, \beta_\phi \in \mathbb{R}^{n_\phi}$ be given vectors satisfying Property 1 for the NN. Denote the $i$th row of the first weight $W^1$ by $W_i^1$ and define matrices

$$R_V := \begin{bmatrix} I_{n_G} & 0_{n_G \times n_\phi} \\ N_{ux} & N_{uw} \end{bmatrix}, \text{ and } R_\phi := \begin{bmatrix} N_{vx} & N_{vw} \\ 0_{n_\phi \times n_G} & I_{n_\phi} \end{bmatrix}.$$

If there exists a matrix $P \in \mathbb{S}_{++}^{n_G}$, and vector $\lambda \in \mathbb{R}^{n_\phi}$ with $\lambda \geq 0$ such that

$$R_V^\top \begin{bmatrix} A_G^\top P A_G - P & A_G^\top P B_G \\ B_G^\top P A_G & B_G^\top P B_G \end{bmatrix} R_V + R_\phi^\top \Psi_\phi^\top M_\phi(\lambda) \Psi_\phi R_\phi < 0 \tag{18}$$

$$\begin{bmatrix} (\bar{v}_i^1 - v_{*,i}^1)^2 & W_i^1 \\ W_i^{1\top} & P \end{bmatrix} \geq 0, \quad i = 1, \ldots, n_1 \tag{19}$$

then, i) the feedback system consisting of $G$ and $\pi$ is locally stable around $x_*$, and ii) the set $\mathcal{E}(P, x_*)$, defined by (1), is an innerapproximation to the ROA.

*Proof:* By Schur complements, (21) is equivalent to

$$W_i^1 P^{-1} W_i^{1\top} \leq (\bar{v}_i^1 - v_{*,i}^1)^2, \; i = 1, \ldots, n_1. \tag{20}$$

It follows from [23, Lemma 1] that

$$\mathcal{E}(P, x_*) \subseteq \{x \in \mathbb{R}^{n_G} : \underline{v}^1 - v_*^1 \leq W^1(x - x_*) \leq \bar{v}^1 - v_*^1\}.$$

Finally, use $v^1 - v_*^1 = W^1(x - x_*)$ to rewrite this as

$$\mathcal{E}(P, x_*) \subseteq \{x : \underline{v}^1 \leq v^1 \leq \bar{v}^1\}.$$

[2]The next activation input is $v^2 := W^2 w^1 + b^2$. The largest value of the $i$th entry of this vector is obtained by solving the following optimization:

$$\bar{v}_i^2 := \max_{\underline{w}^1 \leq w^1 \leq \bar{w}^1} y^\top w^1 + b_i^2 \tag{21}$$

where $y^\top$ is the $i^{th}$ row of $W^2$. Define $c := \frac{1}{2}(\bar{w}^1 + \underline{w}^1)$ and $r := \frac{1}{2}(\bar{w}^1 - \underline{w}^1)$. The optimization can be rewritten as:

$$\bar{v}_i^2 := \left(y^\top c + b_i^2\right) + \max_{-r \leq \delta \leq r} y^\top \delta. \tag{22}$$

This has the explicit solution $\bar{v}_i^2 = y^\top c + b_i^2 + \sum_{j=1}^{n_1} |y_j r_j|$. Similarly, the minimal value is $\underline{v}_i^2 = y^\top c + b_i^2 - \sum_{j=1}^{n_1} |y_j r_j|$.

To summarize, feasibility of (21) verifies that if $x(k) \in \mathcal{E}(P, x_*)$ then $v^1(k) \in [\underline{v}^1, \bar{v}^1]$, and hence, the offset local sector conditions are valid.

Next, since the LMI in (20) is strict, there exists $\epsilon > 0$ such that left/right multiplication of the LMI by $\begin{bmatrix} (x(k) - x_*)^\top & (w_\phi(k) - w_*)^\top \end{bmatrix}$ and its transpose yields

$$[\star]^\top \begin{bmatrix} A_G^\top P A_G - P & A_G^\top P B_G \\ B_G^\top P A_G & B_G^\top P B_G \end{bmatrix} \begin{bmatrix} x(k) - x_* \\ u(k) - u_* \end{bmatrix} + [\star]^\top \Psi_\phi^\top M_\phi(\lambda) \Psi_\phi \begin{bmatrix} v_\phi(k) - v_* \\ w_\phi(k) - w_* \end{bmatrix} \leq -\epsilon \|x(k) - x_*\|^2$$

where the entries denoted by $\star$ can be inferred from symmetry. Define the Lyapunov function $V(x) := (x - x_*)^\top P (x - x_*)$ and use (2) and (12) to show

$$V(x(k+1)) - V(x(k)) + [\star]^\top \Psi_\phi^\top M_\phi(\lambda) \Psi_\phi \begin{bmatrix} v_\phi(k) - v_* \\ w_\phi(k) - w_* \end{bmatrix} \leq -\epsilon \|x(k) - x_*\|^2. \tag{23}$$

Assume $x(k) \in \mathcal{E}(P, x_*)$ for some $k \geq 0$, i.e., $V(x(k)) \leq 1$. As noted earlier, $x(k) \in \mathcal{E}(P, x_*)$ implies the offset local sector $[\alpha_\phi, \beta_\phi]$ around $v_*$. Then, by Lemma 1, the final term on the left side of (23) is $\geq 0$, and thus, from (23), we have $V(x(k+1)) \leq 1$, i.e., $x(k+1) \in \mathcal{E}(P, x_*)$. By induction, we have that $\mathcal{E}(P, x_*)$ is forward invariant, i.e., $x(0) \in \mathcal{E}(P, x_*) \Rightarrow x(k) \in \mathcal{E}(P, x_*) \; \forall k \geq 0$. As a result, if $x(0) \in \mathcal{E}(P, x_*)$, then the final term on the left side of (23) is $\geq 0$ for all $k \geq 0$, and $V(x(k+1)) - V(x(k)) \leq -\epsilon \|x(k) - x_*\|^2$ for all $k \geq 0$. It follows from a Lyapunov argument, e.g., [7, Th. 4.1] that $x_*$ is an asymptotically stable equilibrium point and $\mathcal{E}(P, x_*)$ is an innerapproximation of the ROA. ■

*Remark 1:* Note that $\bar{v}^1$ should be chosen with care as it affects the size of ROA innerapproximations directly: decreasing $(\bar{v}^1 - v_*^1)$ gives rise to sharper local sector bounds, which is beneficial on ROA estimation, but also restricts the region where ROA innerapproximations lie in; increasing $(\bar{v}^1 - v_*^1)$ leads to a larger region that contains ROA innerapproximations, but also provides looser local sector bounds. A possible way of choosing $\bar{v}^1$ is to parameterize $(\bar{v}^1 - v_*^1)$ as $\bar{v}^1 - v_*^1 = \delta_v \times 1_{n_1 \times 1}$ with $\delta_v \in \mathbb{R}_{++}$, grid the interval $[0, \overline{\delta}_v]$[3] where $\delta_v$ lies in, innerapproximate the ROA on the grid, and choose $\delta_v$ that leads to the largest innerapproximation.

*Remark 2:* In this article, the NN controller is assumed to be state-feedback. For the output-feedback case, i.e., $u = \pi(Cx)$, where $C \in \mathbb{R}^{n_y \times n_G}$, the stability analysis can be performed similarly, using a new $N_{vx}$ defined as $N_{vx} := \begin{bmatrix} W^1 C \\ 0_{(n_2 + \cdots + n_\ell) \times n_G} \end{bmatrix}$.

## III. Robust Stability Analysis

### A. Problem Formulation

Consider the uncertain feedback system in Fig. 5, consisting of an uncertain plant $F_u(G, \Delta)$ and a NN controller $\pi$ as defined by (3). The uncertain plant $F_u(G, \Delta)$ is an interconnection of a nominal plant $G$ and a perturbation $\Delta$. The nominal plant $G$ is defined by the following equations:

$$x(k+1) = A_G \, x(k) + B_{G1} \, q(k) + B_{G2} \, u(k) \tag{24a}$$

$$p(k) = C_G \, x(k) + D_{G1} \, q(k) + D_{G2} \, u(k) \tag{24b}$$

[3]$\overline{\delta}_v$ is the largest value such that (20) and (21) stay feasible.

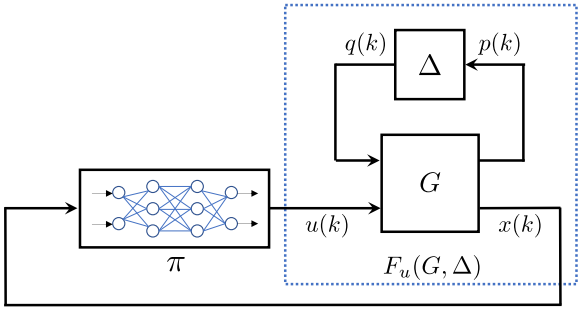


Fig. 5. Feedback system with uncertain plant $F_u(G,\Delta)$ and NN controller $\pi$.

where $x(k) \in \mathbb{R}^{n_G}$ is the state, $u(k) \in \mathbb{R}^{n_u}$ is the control input, $p(k) \in \mathbb{R}^{n_p}$ and $q(k) \in \mathbb{R}^{n_q}$ are the input and output of $\Delta$, $A_G \in \mathbb{R}^{n_G \times n_G}$, $B_{G1} \in \mathbb{R}^{n_G \times n_q}$, $B_{G2} \in \mathbb{R}^{n_G \times n_u}$, $C_G \in \mathbb{R}^{n_p \times n_G}$, $D_{G1} \in \mathbb{R}^{n_p \times n_q}$, and $D_{G2} \in \mathbb{R}^{n_p \times n_u}$. The perturbation is a bounded, causal operator $\Delta : \ell_{2e}^{n_p} \to \ell_{2e}^{n_q}$. The nominal plant $G$ and perturbation $\Delta$ form the interconnection $F_u(G,\Delta)$ through the constraint

$$q(\cdot) = \Delta(p(\cdot)). \tag{25}$$

Denote the set of perturbations to be considered as $\mathcal{S}$.

*Assumption 1:* In this section, we assume (i) the equilibrium point $(x_*, u_*, v_*, w_*, p_*, q_*)$ of the feedback system is at the origin, and (ii) $0 = \Delta(0)$ for all $\Delta \in \mathcal{S}$. Note that $\Delta$ is modeled as an operator mapping inputs to outputs. If $\Delta$ has an internal state then there is an implicit assumption that it has zero initial condition.

Let $\chi(k; x_0, \Delta)$ denote the solution to the feedback system of $F_u(G,\Delta)$ and $\pi$ with $\Delta \in \mathcal{S}$ at time $k$ from the initial condition $x(0) = x_0$.[4] Define the robust ROA associated with $x_*$ as follows.

*Definition 5:* The robust ROA of the feedback system with the uncertain plant $F_u(G,\Delta)$ and NN $\pi$ is defined as

$$\mathcal{R} := \{x_0 \in \mathbb{R}^{n_G} : \lim_{k\to\infty} \chi(k; x_0, \Delta) = x_* \quad \forall \Delta \in \mathcal{S}\}. \tag{26}$$

The objective is to prove the uncertain feedback system is asymptotically stable and, if so, to find the largest estimate of the robust ROA using an ellipsoidal innerapproximation.

### B. Integral Quadratic Constraints

The perturbation can represent various types of uncertainty [14], [15], including saturation, time delay, unmodeled dynamics, and slope-restricted nonlinearities. The input–output relationship of $\Delta$ is characterized with an IQC, which consists of a "virtual" filter $\Psi_\Delta$ applied to the input $p$ and output $q$ of $\Delta$ and a constraint on the output $r$ of $\Psi_\Delta$. The filter $\Psi_\Delta$ is an LTI system of the form

$$\psi(k+1) = A_\Psi\, \psi(k) + B_{\Psi 1}\, p(k) + B_{\Psi 2}\, q(k) \tag{27a}$$

$$r(k) = C_\Psi\, \psi(k) + D_{\Psi 1}\, p(k) + D_{\Psi 2}\, q(k) \tag{27b}$$

$$\psi(0) = 0 \tag{27c}$$

where $\psi(k) \in \mathbb{R}^{n_\psi}$ is the state, $r(k) \in \mathbb{R}^{n_r}$ is the output, and $A_\Psi$ is a Schur matrix. The state matrices have compatible dimensions. The dynamics of $\Psi_\Delta$ can be compactly denoted by $\left[\begin{array}{c|cc} A_\Psi & B_{\Psi 1} & B_{\Psi 2} \\ \hline C_\Psi & D_{\Psi 1} & D_{\Psi 2} \end{array}\right]$. By $(p_*, q_*) = 0$ from Assumption 1, the equilibrium state $\psi_* \in \mathbb{R}^{n_\psi}$ of (27) is also zero.

The Lyapunov analysis in the following section makes use of time-domain IQCs as defined next:

*Definition 6:* Let $\Psi_\Delta \in \mathbb{RH}_\infty^{n_r \times (n_p+n_q)}$ and $M_\Delta \in \mathbb{S}^{n_r}$ be given. A bounded, causal operator $\Delta : \ell_{2e}^{n_p} \to \ell_{2e}^{n_q}$ satisfies the time domain IQC defined by $(\Psi_\Delta, M_\Delta)$ if the following inequality holds for all $p \in \ell_{2e}^{n_p}$, $q = \Delta(p)$, and for all $N \geq 0$:

$$\sum_{k=0}^{N} r(k)^\top M_\Delta r(k) \geq 0. \tag{28}$$

The notation $\Delta \in \mathrm{IQC}(\Psi_\Delta, M_\Delta)$ indicates that $\Delta$ satisfies the IQC defined by $\Psi_\Delta$ and $M_\Delta$. Therefore, the precise relation (25), for analysis, is replaced by the constraint (28) on $r$. The QC proposed in Lemma 1 is a special instance of a time-domain IQCs. Specifically, Lemma 1 defines a QC that holds at each time step $k$ and, hence, the inequality also holds summing over any finite horizons. This is referred to as the offset local sector IQC.

The time-domain IQCs, as defined here, hold on any finite horizon $N \geq 0$. These are typically called "hard IQCs" [14]. IQCs can also be defined in the frequency domain and equivalently expressed as time-domain constraints over an infinite horizon ($N = \infty$). These are called soft IQCs. Although this article focuses on the use of hard IQCs, it is possible to also incorporate soft IQCs [20], [21], [24], [25].

### C. Lyapunov Condition

Let $\zeta := \left[\begin{smallmatrix} x \\ \psi \end{smallmatrix}\right] \in \mathbb{R}^{n_\zeta}$ define the extended state vector, $n_\zeta = n_G + n_\psi$, whose dynamics are

$$\zeta(k+1) = \mathcal{A}\,\zeta(k) + \mathcal{B}\begin{bmatrix} q(k) \\ u(k) \end{bmatrix} \tag{29a}$$

$$r(k) = \mathcal{C}\,\zeta(k) + \mathcal{D}\begin{bmatrix} q(k) \\ u(k) \end{bmatrix} \tag{29b}$$

$$u(k) = \pi(x(k)) \tag{29c}$$

where the state-space matrices are

$$\mathcal{A} = \begin{bmatrix} A_G & 0 \\ B_{\Psi 1} C_G & A_\Psi \end{bmatrix},\ \mathcal{B} = \begin{bmatrix} B_{G1} & B_{G2} \\ B_{\Psi 1} D_{G1} + B_{\Psi 2} & B_{\Psi 1} D_{G2} \end{bmatrix}$$

$$\mathcal{C} = \begin{bmatrix} D_{\Psi 1} C_G & C_\Psi \end{bmatrix},\ \mathcal{D} = \begin{bmatrix} D_{\Psi 1} D_{G1} + D_{\Psi 2} & D_{\Psi 1} D_{G2} \end{bmatrix}.$$

Let $\zeta_* := \left[\begin{smallmatrix} x_* \\ \psi_* \end{smallmatrix}\right] = 0$ define the equilibrium point of the extended system (29). Since IQCs implicitly constrain the input $p$ of the extended system (29), the response of the extended system subject to IQCs "covers" the behaviors of the original uncertain feedback system. The following theorem provides a method for innerapproximating the robust ROA by performing analysis on the extended system subject to IQCs.

*Theorem 2:* Consider the feedback system of an uncertain plant $F_u(G,\Delta)$ in (24) and (25), and the NN $\pi$ in (3) with zero equilibrium point $(\zeta_*, u_*, v_*, w_*, p_*, q_*)$. Assume $\Delta \in \mathrm{IQC}(\Psi_\Delta, M_\Delta)$ with $\Psi_\Delta$ and $M_\Delta$ given. Let $\bar{v}^1 \in \mathbb{R}^{n_1}$, $\underline{v}^1 := -\bar{v}^1$, and $\alpha_\phi, \beta_\phi \in \mathbb{R}^{n_\phi}$ be given vectors satisfying Property 1 for the NN, and define matrices

$$R_V = \left[\begin{array}{ccc} I_{n_\zeta} & 0 & 0 \\ \hline 0 & 0 & I_{n_q} \\ N_{u\zeta} & N_{uw} & 0 \end{array}\right],\ N_{u\zeta} = [N_{ux}, 0_{n_u \times n_\psi}]$$

[4] An input/output model is used for the perturbation $\Delta$ so that its internal state and initial condition is not explicitly considered.

$$R_\phi = \begin{bmatrix} N_{v\zeta} & N_{vw} & 0 \\ 0 & I_{n_\phi} & 0 \end{bmatrix},\ N_{v\zeta} = [N_{vx}, 0_{n_\phi \times n_\psi}]$$

$$\mathcal{W}_i^1 = \begin{bmatrix} W_i^1 & 0_{1\times n_\psi} \end{bmatrix},\ W_i^1 \text{ is the } i\text{th row of } W^1.$$

If there exists a matrix $P \in \mathbb{S}_{++}^{n_\zeta}$, and vector $\lambda \in \mathbb{R}^{n_\phi}$ with $\lambda \geq 0$ such that

$$R_V^\top \begin{bmatrix} \mathcal{A}^\top P \mathcal{A} - P & \mathcal{A}^\top P \mathcal{B} \\ \mathcal{B}^\top P \mathcal{A} & \mathcal{B}^\top P \mathcal{B} \end{bmatrix} R_V + R_\phi^\top \Psi_\phi^\top M_\phi(\lambda) \Psi_\phi R_\phi$$

$$+ R_V^\top \begin{bmatrix} \mathcal{C} & \mathcal{D} \end{bmatrix}^\top M_\Delta \begin{bmatrix} \mathcal{C} & \mathcal{D} \end{bmatrix} R_V < 0 \tag{30a}$$

$$\begin{bmatrix} (\bar{v}_i^1)^2 & \mathcal{W}_i^1 \\ \mathcal{W}_i^{1\top} & P \end{bmatrix} \geq 0,\ i = 1, \ldots, n_1 \tag{30b}$$

then, (i) the feedback system comprising $F_u(G, \Delta)$ and $\pi$ is locally stable around $x_*$ for any $\Delta \in \text{IQC}(\Psi_\Delta, M_\Delta)$, and (ii) the intersection of $\mathcal{E}(P, \zeta_*)$ with the hyperplane $\psi = 0$, i.e., $\mathcal{E}(P_x, x_*)$ where $P_x \in \mathbb{R}^{n_G \times n_G}$ is the upper left block of $P$, is an innerapproximation to the robust ROA.

*Proof:* As in the proof of Theorem 1, feasibility of (30b) implies that if $\zeta(k) \in \mathcal{E}(P, \zeta_*)$ then $v^1(k) \in [\underline{v}^1, \bar{v}^1]$, and hence, the offset local sectors conditions are valid. Since the LMI in (30a) is strict, there exists $\epsilon > 0$ such that left/right multiplication of the LMI by $[(\zeta(k) - \zeta_*)^\top, (w_\phi(k) - w_*)^\top, (q(k) - q_*)^\top]$ and its transpose yields

$$\begin{bmatrix} \star \end{bmatrix}^\top \begin{bmatrix} \mathcal{A}^\top P \mathcal{A} - P & \mathcal{A}^\top P \mathcal{B} \\ \mathcal{B}^\top P \mathcal{A} & \mathcal{B}^\top P \mathcal{B} \end{bmatrix} \begin{bmatrix} \zeta(k) - \zeta_* \\ \hline q(k) - q_* \\ u(k) - u_* \end{bmatrix}$$

$$+ \begin{bmatrix} \star \end{bmatrix}^\top \begin{bmatrix} \mathcal{C} & \mathcal{D} \end{bmatrix}^\top M_\Delta \begin{bmatrix} \mathcal{C} & \mathcal{D} \end{bmatrix} \begin{bmatrix} \zeta(k) - \zeta_* \\ \hline q(k) - q_* \\ u(k) - u_* \end{bmatrix}$$

$$+ [\star]^\top \Psi_\phi^\top M_\phi(\lambda) \Psi_\phi \begin{bmatrix} v_\phi(k) - v_* \\ w_\phi(k) - w_* \end{bmatrix} \leq -\epsilon \|\zeta(k) - \zeta_*\|^2.$$

Define the Lyapunov function $V(\zeta) := (\zeta - \zeta_*)^\top P (\zeta - \zeta_*)$, and use (29) to show

$$V(\zeta(k+1)) - V(\zeta(k)) + r(k)^\top M_\Delta r(k)$$

$$+ [\star]^\top \Psi_\phi^\top M_\phi(\lambda) \Psi_\phi \begin{bmatrix} v_\phi(k) - v_* \\ w_\phi(k) - w_* \end{bmatrix} \leq -\epsilon \|\zeta(k) - \zeta_*\|^2.$$

Sum this inequality from $k = 0$ to any finite time $N \geq 0$. The third and fourth term on the left side will be $\geq 0$ by the local sector conditions and the IQC. This yields

$$V(\zeta(N+1)) - V(\zeta(0)) \leq -\sum_{k=0}^{N} \epsilon \|\zeta(k) - \zeta_*\|^2.$$

Thus, if $\zeta(0) \in \mathcal{E}(P, \zeta_*)$ then $\zeta(k) \in \mathcal{E}(P, \zeta_*)$ for all $k \geq 0$. Moreover, this inequality implies that $\zeta(N) \to \zeta_*$ as $N \to \infty$. The initial condition for the virtual filter is $\psi(0) = 0$ so that $\zeta(0) \in \mathcal{E}(P, \zeta_*)$ is equivalent to $x(0) \in \mathcal{E}(P_x, x_*)$. Hence, $\mathcal{E}(P_x, x_*)$ is an innerapproximation for the ROA. ■

For a particular perturbation $\Delta$, there is typically a class of valid time-domain IQCs defined by a fixed filter $\Psi_\Delta$ and a matrix $M_\Delta$ drawn from a constraint set $\mathcal{M}_\Delta$. Therefore, when formulating an optimization problem, along with $P$ and $\lambda$, we can treat $M_\Delta \in \mathcal{M}_\Delta$ as an additional decision variable to reduce conservatism. In this article, the set $\mathcal{M}_\Delta$ is restricted to one that is described by LMIs [15]. Using $\text{trace}(P_x)$ as the cost function to minimize along with the LMIs developed before, we have the following optimization to compute the "largest" ROA innerapproximation:

$$\min_{P \in \mathbb{S}_{++}^{n_\zeta}, \lambda \geq 0, M_\Delta \in \mathcal{M}_\Delta} \text{trace}(P_x) \text{ s.t. (30a)} - \text{(30b) hold} \tag{31}$$

which is convex in $(P, \lambda, M_\Delta)$. The strict inequality in (30a) can be enforced by either replacing $< 0$ with $\leq -\epsilon I$ with $\epsilon = 1 \times 10^{-6}$, or solving (31) with a nonstrict inequality $\leq 0$, and checking if the constraint is active afterward.

### D. IQCs for Combined Activation Functions $\phi$

Now that we have the general framework that merges Lyapunov theory with IQCs, we will revisit the problem of describing the activation functions $\phi$ using more general tools. Recall that offset local sector QCs have been used in Sections II and III to bound activation functions $\phi$. However, these local sectors fail to incorporate slope bounds of $\phi$. In this section, in addition to local sectors, we will use off-by-one IQCs [16] to capture the slope information of $\phi$ to achieve less conservative ROA innerapproximations.

Besides the local sector bound $\alpha_\phi, \beta_\phi$, the bounds $\underline{v}, \overline{v}$ on activation input $v_\phi$ can also be used to compute the local slope bounds $[m_\phi, L_\phi]$ of $\phi$, with $m_\phi, L_\phi \in \mathbb{R}^{n_\phi}$. For example, $\phi_i(v_{\phi,i}) := \tanh(v_{\phi,i})$ restricted to the interval $[\underline{v}_i, \overline{v}_i]$ for $i = 1, \ldots, n_\phi$ satisfies the local slope bound $[m_{\phi,i}, L_{\phi,i}]$ with $m_{\phi,i} := \min(\frac{d\tanh(v_i)}{dv_i}|_{v_i=\underline{v}_i}, \frac{d\tanh(v_i)}{dv_i}|_{v_i=\overline{v}_i})$, and $L_{\phi,i} := 1$. If $w_\phi = \phi(v_\phi)$ with $v_\phi(k) \in [\underline{v}, \overline{v}]$, then $\phi$ also satisfies the hard IQC defined by $(\Psi_{\text{off}}, M_{\text{off}})$, where

$$\Psi_{\text{off}} := \left[\begin{array}{c|cc} 0_{n_\phi} & -\text{diag}(L_\phi) & I_{n_\phi} \\ \hline I_{n_\phi} & \text{diag}(L_\phi) & -I_{n_\phi} \\ 0_{n_\phi} & -\text{diag}(m_\phi) & I_{n_\phi} \end{array}\right]$$

$$M_{\text{off}}(\eta) := \begin{bmatrix} 0_{n_\phi} & \text{diag}(\eta) \\ \text{diag}(\eta) & 0_{n_\phi} \end{bmatrix},\ \text{for all } \eta \in \mathbb{R}^{n_\phi} \text{ with } \eta \geq 0.$$

This is the so-called "off-by-one" IQC [16], which is a special instance of the Zames-Falb IQC [26], [27]. It provides constraints that relate the activation at different time instances, e.g., between $\phi_i(v_{\phi,i}(k))$ and $\phi_i(v_{\phi,i}(k+1))$ for any $i = 1, \ldots, n_\phi$.

The analysis on the feedback system of $F_u(G, \Delta)$ and $\pi$ can be instead performed on the extended system made up by $G$, $\Psi_\Delta$, and $\Psi_{\text{off}}$ with additional constraints that $\Delta \in \text{IQC}(\Psi_\Delta, M_\Delta)$, and $\phi$ satisfies the offset local sector and $\phi \in \text{IQC}(\Psi_{\text{off}}, M_{\text{off}})$. However, since $\Psi_{\text{off}}$ introduces a number of $n_\phi$ states to the extended system, the size of the corresponding Lyapunov matrix $P$ will increase from $\mathbb{S}_{++}^{n_x+n_\psi}$ to $\mathbb{S}_{++}^{n_x+n_\psi+n_\phi}$, which leads to longer computation time. The effectiveness of the off-by-one IQC is demonstrated in Section IV-B.

## IV. Examples

In the following examples, the optimization (31) is solved using MOSEK with CVX.[5]

### A. Inverted Pendulum

Consider the nonlinear inverted pendulum example with mass $m = 0.15$ kg, length $l = 0.5$ m, and friction coefficient $\mu = 0.5$ N $\cdot$ m$\cdot$s /

[5]The code is available at https://github.com/heyinUCB/Stability-Analysis-using-Quadratic-Constraints-for-Systems-with-Neural-Network-Controllers https://github.com/heyinUCB/Stability-Analysis-using-Quadratic-Constraints-for-Systems-with-Neural-Network-Controllers.

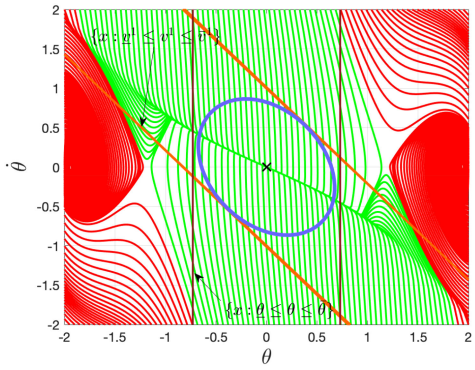


Fig. 6. ROA innerapproximation of the inverted pendulum.

rad. The dynamics are

$$\ddot{\theta}(t) = \frac{mgl\sin(\theta(t)) - \mu\dot{\theta}(t) + \text{sat}(u(t))}{ml^2} \tag{32}$$

where $\theta$ is the angular position (rad) and $u$ is the control input (N $\cdot$ m). The plant state is $x = [\theta, \dot{\theta}]$. The saturation function is defined as $\text{sat}(u) = \text{sgn}(u)\min(|u|, u_{\max})$, with $u_{\max} = 0.7$ N $\cdot$ m. The controller $\pi$ is obtained through a reinforcement learning process using policy gradient [3]–[5]. During training, the agent decision making process is characterized by a probability: $u(k) \sim Pr(u(k) = u \mid x(k) = x)$ for all $u \in \mathbb{R}$ and $x \in \mathbb{R}^2$ where the probability is a Gaussian distribution with mean $\pi(x(k))$, and standard deviation $\sigma$. After training, the policy mean $\pi$ is used as the deterministic controller $u(k) = \pi(x(k))$. The controller $\pi$ is parameterized by a two layer, feedforward NN with $n_1 = n_2 = 32$ and $\tanh$ as the activation function for both layers. The biases in the NN are set to zero during training to ensure that the equilibrium point is $x_* = 0$ and $u_* = 0$. The dynamics used for training are the discretized version of (32) with sampling time $dt = 0.02$ s.

We rearrange (32) into the form

$$\ddot{\theta}(t) = \frac{-mglq(t) + mgl\theta(t) - \mu\dot{\theta}(t) + \text{sat}(u(t))}{ml^2} \tag{33a}$$

$$q(t) = \Delta(\theta(t)) := \theta(t) - \sin(\theta(t)). \tag{33b}$$

The static nonlinearity $\Delta(\theta) = \theta - \sin(\theta)$ is slope-restricted, and sector bounded. If we assume that $\theta(k) \in [\underline{\theta}, \bar{\theta}]$ with $\bar{\theta} = -\underline{\theta} = 0.73$, then the nonlinearity is slope-restricted in $[0, 0.2548]$, and sector bounded in $[0, 0.087]$. We also assume that $v^1 \in [\underline{v}^1, \bar{v}^1]$ with $\bar{v}^1 = -\underline{v}^1 = \delta_v \times 1_{32\times 1}$ using $\delta_v = 0.1$. Both assumptions are verified using the ROA innerapproximation. Two types of IQCs are used to characterize the nonlinearity $\Delta(\cdot)$: an off-by-one IQC to capture the slope information, and a local sector IQC to express the local sector bound. Only the local sector IQC is used to characterize the activation functions $\phi$. The saturation function is static and can also be described using a local sector bound. Let $\bar{u}$ be the largest possible control command from $\pi$ induced from the assumption that $v^1 \in [\underline{v}^1, \bar{v}^1]$. Then, the saturation function satisfies the local sector $[\alpha, \beta]$, where $\alpha := \frac{u_{\max}}{\bar{u}}$ and $\beta := 1$.

Fig. 6 shows the boundaries for the sets $\{x : \underline{v}^1 \le v^1 \le \bar{v}^1\}$ and $\{x : \underline{\theta} \le \theta \le \bar{\theta}\}$ with orange and brown lines, the ROA innerapproximation with a blue ellipsoid, and the phase portrait of the closed-loop system, with green and red curves representing trajectories inside and outside the ROA.

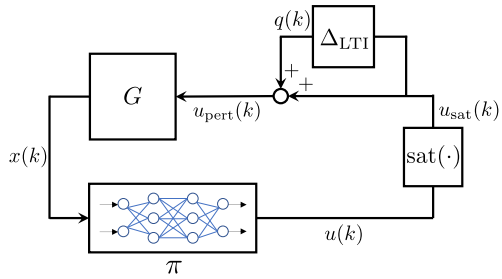


Fig. 7. Uncertain vehicle system with actuator uncertainty.

### B. Vehicle Lateral Control

Consider the vehicle lateral dynamics from [28]

$$\begin{bmatrix} \dot{e} \\ \ddot{e} \\ \dot{e}_\theta \\ \ddot{e}_\theta \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & \frac{C_{\alpha f}+C_{\alpha r}}{mU} & -\frac{C_{\alpha f}+C_{\alpha r}}{m} & \frac{aC_{\alpha f}-bC_{\alpha r}}{mU} \\ 0 & 0 & 0 & 1 \\ 0 & \frac{aC_{\alpha f}-bC_{\alpha r}}{I_z U} & -\frac{aC_{\alpha f}-bC_{\alpha r}}{I_z} & \frac{a^2C_{\alpha f}+b^2C_{\alpha r}}{I_z U} \end{bmatrix} \begin{bmatrix} e \\ \dot{e} \\ e_\theta \\ \dot{e}_\theta \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{C_{\alpha f}}{m} \\ 0 \\ -\frac{aC_{\alpha f}}{I_z} \end{bmatrix} u + \begin{bmatrix} 0 \\ \frac{aC_{\alpha f}-bC_{\alpha r}}{m} - U^2 \\ 0 \\ \frac{a^2C_{\alpha f}+b^2C_{\alpha r}}{I_z} \end{bmatrix} c \tag{34}$$

where $e$ is the perpendicular distance to the lane edge $(m)$, and $e_\theta$ is the angle between the tangent to the straight section of the road and the projection of the vehicle's longitudinal axis (rad). Let $x = [e, \dot{e}, e_\theta, \dot{e}_\theta]^\top$ denote the plant state. The control $u$ is the steering angle of the front wheel (rad), the disturbance $c$ is the road curvature $(1/\text{m})$, and the parameters are as follows: longitudinal velocity $U = 28$ m / s, front cornering stiffness $C_{\alpha f} = -1.232 \times 10^5$ N / rad, rear cornering stiffness $C_{\alpha r} = -1.042 \times 10^5$ N / rad, mass $m = 1.67 \times 10^3$ kg, moment of inertia $I_z = 2.1 \times 10^3$ kg/m$^2$, distances from vehicle center of gravity to front axle $a = 0.99$ m, and rear axle $b = 1.7$ m.

Again, the controller $\pi$ is obtained using policy gradient, and is parameterized by a two layer, feedforward NN, with $n_1 = n_2 = 32$ and $\tanh$ as the activation function for both layers. The training process uses a discretized version of (34) with sampling time $dt = 0.02$ s and draws the curvature $c(k)$ at each time step from an interval $[-1/200, 1/200]$. The control command derived from $u(k) = \pi(x(k))$ enters the vehicle dynamics through a saturation function $\text{sat}(\cdot)$ with $u_{\max} = \pi/6$. Let $u_{\text{sat}} := \text{sat}(\pi(x))$ define the saturated control signal.

The analysis is performed for a constant curvature $c \equiv 0$, resulting in a zero equilibrium state $x_* = 0$. In the analysis problem, on top of saturation, we also add a norm-bounded LTI uncertainty $\Delta_{\text{LTI}} \in \mathbb{RH}_\infty$ with $\|\Delta_{\text{LTI}}\|_\infty \le 0.1$ to the control input. This is used to assess the robustness of the NN controller against actuator uncertainty. As shown in Fig. 7, the actual input to the vehicle dynamics is

$$u_{\text{pert}}(k) = u_{\text{sat}}(k) + q(k), \text{ and } q(\cdot) = \Delta_{\text{LTI}}(u_{\text{sat}}(\cdot)).$$

It is assumed that $v^1 \in [\underline{v}^1, \bar{v}^1]$, where $\bar{v}^1 = -\underline{v}^1 = \delta_v \times 1_{32\times 1}$ with $\delta_v = 0.6$. To show effectiveness of the off-by-one IQC, two experiments were carried out one with only local sector IQC to describe $\phi$, and one with both local sector and off-by-one IQCs. The achieved trace$(P_x)$ for the two experiments is 4.4 and 2.9, respectively. Moreover, the achieved $\det(P_x^{-1})$ (proportional to the volume) for the experiments are $3.2 \times 10^5$, and $1.1 \times 10^6$, respectively. Therefore, with the help of

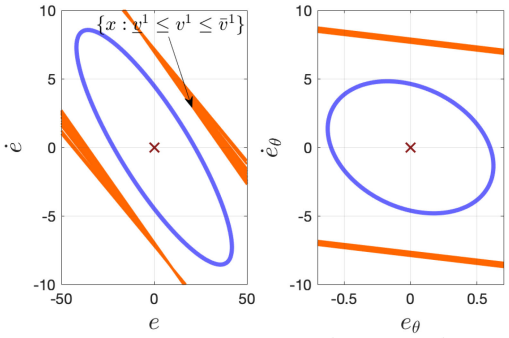


Fig. 8. ROA innerapproximation on the $e$–$\dot{e}$ and $e_\theta$–$\dot{e}_\theta$ spaces using both local sector and off-by-one IQCs with $\delta_v = 0.6$.

off-by-one IQC to sharpen the description of $\phi$, the second experiment achieves a larger ROA innerapproximation. It is also important to note that thanks to the off-by-one IQC, the SDP is able to tolerate looser local sector bounds. The largest value of $\delta_v$ such that the SDP is feasible is 0.67 for the first experiment, and 1.4 for the second experiment.

Fig. 8 shows slices of the ROA innerapproximation from the second experiment on the $e$–$\dot{e}$ and $e_\theta$–$\dot{e}_\theta$ spaces. Specifically, these are intersections of $\mathcal{E}(P_x, x_*)$ with the hyperplanes $(e_\theta, \dot{e}_\theta) = (e_{\theta*}, \dot{e}_{\theta*})$ and $(e, \dot{e}) = (e_*, \dot{e}_*)$, respectively, where $x_* = [e_*, \dot{e}_*, e_{\theta*}, \dot{e}_{\theta*}]^\top$. The slices are shown with blue ellipsoids. The boundary of the polytopic set $\{x : \underline{v}^1 \le v^1 \le \bar{v}^1\}$ is shown with the orange lines. The brown crosses represent the zero equilibrium stfate $x_*$.

## REFERENCES


[1] A. U. Levin and K. S. Narendra, "Control of nonlinear dynamical systems using neural networks: Controllability and stabilization," *IEEE Trans. Neural Netw.*, vol. 4, no. 2, pp. 192–206, Mar. 1993.

[2] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Mach. Learn.*, vol. 8, no. 3-4, pp. 229–256, May 1992.

[3] J. Peters and S. Schaal, "Reinforcement learning of motor skills with policy gradients," *Neural Netw.: Official J.Int. Neural Netw. Soc.*, vol. 21, no. 4, pp. 682–697, 2008.

[4] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," *Int. Conf. Mach. Learn. PMLR*, pp. 1889–1897, Feb. 2015.

[5] S. Kakade, "A natural policy gradient," in *Proc. 14th Int. Conf. Neural Inf. Process. Syst.: Natural Synthetic*, 2001, pp. 1531–1538.

[6] D. A. Pomerleau, *ALVINN: An Autonomous Land Vehicle in a Neural Network*. San Francisco, CA, USA: Morgan Kaufmann, 1989, pp. 305–313.

[7] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[8] M. Fazlyab, M. Morari, and G. J. Pappas, "Safety verification and robustness analysis of neural networks via quadratic constraints and semidefinite programming," *IEEE Trans. Autom. Control*, Mar. 20.

[9] M. Fazlyab, A. Robey, H. Hassani, M. Morari, and G. Pappas, "Efficient and accurate estimation of Lipschitz constants for deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 11 427–11438.

[10] P. Pauli, A. Koch, J. Berberich, P. Kohler, and F. Allgower, "Training robust neural networks using Lipschitz bounds," *IEEE Control Syst. Lett.*, 2021.

[11] H. Hu, M. Fazlyab, M. Morari, and G. J. Pappas, "Reach-SDP: Reachability analysis of closed-loop systems with neural network controllers via semidefinite programming," *2020 59th IEEE Conf. Decision Cont.*, 2020, pp. 5929–5934.

[12] M. Jin and J. Lavaei, "Stability-certified reinforcement learning: A control-theoretic perspective," *IEEE Access*, 2020.

[13] K. K. Kim, E. R. Patrón, and R. D. Braatz, "Standard representation and unified stability analysis for dynamic artificial neural network models," *Neural Netw.*, vol. 98, pp. 251–262, 2018.

[14] A. Megretski and A. Rantzer, "System analysis via integral quadratic constraints," *IEEE Trans. Autom. Control*, vol. 42, no. 6, pp. 819–830, Jun. 1997.

[15] J. Veenman, C. W. Scherer, and H. Köroğlu, "Robust stability and performance analysis based on integral quadratic constraints," *Eur. J. Control*, vol. 3 1, pp. 1–32, 2016.

[16] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM J. Optim.*, vol. 26, no. 1, pp. 57–95, 2016.

[17] H. Fang, Z. Lin, and M. Rotea, "On IQC approach to the analysis and design of linear systems subject to actuator saturation," *Syst. Control Lett.*, vol. 57, no. 8, pp. 611–619, 2008.

[18] S. Tarbouriech, *Stability and Stabilization of Linear Systems With Saturating Actuators*. Berlin, Germany: Springer, 2014.

[19] E. Summers and A. Packard, "L2 gain verification for interconnections of locally stable systems using integral quadratic constraints," in *Proc. 49th IEEE Conf. Decis. Control*, 2010, pp. 1460–1465.

[20] M. Fetzer, C. W. Scherer, and J. Veenman, "Invariance with dynamic multipliers," *IEEE Trans. Autom. Control*, vol. 63, no. 7, pp. 1929–1942, Jul. 2018.

[21] A. Iannelli, P. Seiler, and A. Marcos, "Region of attraction analysis with integral quadratic constraints," *Automatica*, vol. 109, 2019, Art. no. 108543.

[22] S. Gowal *et al.*, "On the effectiveness of interval bound propagation for training verifiably robust models," 2018, *arXiv:1810.12715*.

[23] H. Hindi and S. Boyd, "Analysis of linear systems with saturation using convex optimization," in *Proc. 37th IEEE Conf. Decis. Control*, vol. 1, 1998, pp. 903–908.

[24] H. Yin, P. Seiler, and M. Arcak, "Backward reachability using integral quadratic constraints for uncertain nonlinear systems," *IEEE Control Syst. Lett.*, vol. 5, no. 2, pp. 707–712, Apr. 2021.

[25] H. Yin, A. Packard, M. Arcak, and P. Seiler, "Reachability analysis using dissipation inequalities for uncertain nonlinear systems," *Syst. Control Lett.*, vol. 142, 2020, Art. no. 104736.

[26] G. Zames and P. L. Falb, "Stability conditions for systems with monotone and slope-restricted nonlinearities," *SIAM J. Control*, vol. 6, no. 1, pp. 89–108, 1968.

[27] V. V. Kulkarni and M. G. Safonov, "All multipliers for repeated monotone nonlinearities," *IEEE Trans. Autom. Control*, vol. 47, no. 7, pp. 1209–1212, Jul. 2002.

[28] A. Alleyne, "A comparison of alternative intervention strategies for unintended roadway departure (URD) control," *Veh. Syst. Dyn.*, vol. 27, no. 3, pp. 157–186, 1997.