Data-Driven Stability and Optimality Certificates for Hybrid Dynamical Systems

Carlos A. Montenegro G.[1], Santiago J. Leudo, Ricardo G. Sanfelice

University of California, Santa Cruz, 606 Eng. Loop, Santa Cruz, 95064 CA, USA

Abstract

In this paper, we develop a learning-based approach for hybrid systems with a twofold purpose: (i) the construction of Lyapunov functions and (ii) the computation of upper bounds on the cost of system solution. Our approach enforces stability and cost conditions on a finite collection of states and then, by exploiting regularity properties of the system dynamics and stage-cost maps, extends these conditions to all relevant unsampled states. Neural networks are used to learn both a Lyapunov function surrogate and a value-like function whose regularity properties allow us to certify stability (in a *practical* sense) of a compact set or to find an upper bound on the cost of solutions, which are not required to be unique. We illustrate the approach on nonlinear continuous-time and discrete-time systems, as well as on a hybrid system modeled by set-valued dynamics.

Keywords: Hybrid Systems, Data-Driven Certificates, Lyapunov Stability, Cost Evaluation

---

1. Introduction

Results on sufficient conditions to guarantee the satisfaction of dynamical properties, such as stability, safety, and optimality rely on pointwise conditions involving certificates, e.g., Lyapunov functions, barrier functions, and value functions. Though such conditions are sufficient to characterize the behavior of a system, synthesizing the certificate to satisfy the required conditions is an open research area, especially when the system dynamics are nonlinear.

Email address: camonten@ucsc.edu (Carlos A. Montenegro G.)

[1]Corresponding author.

On the one hand, different approaches have been considered to synthesize Lyapunov functions for continuous-time systems with specific dynamics, e.g., sum of squares for polynomial systems [1, 2]. In [3], the authors structure the Lyapunov candidate function such that it inherently yields a provable stability certificate. In [4], the authors propose a framework for learning dynamical systems with stable inference dynamics [5, 6]. In [7], a neural network structure is proposed to provably overcome the curse of dimensionality in the synthesis of Lyapunov functions for continuous-time systems with nonlinear dynamics, whereas in [8] a quadratic Lyapunov function is optimized to provide stability guarantees. Furthermore, in [9], a counterexample-guided approach is proposed using finitely many points, as well as an approach to extend the results to a subset of the state space using satisfiability modulo theories (SMT). A similar approach is proposed in [10], where the authors opted for a mixed-integer linear program (MILP) rather than SMT.

On the other hand, the interconnection of physical systems with computational and communication devices, such as analog-to-digital converters, sample-and-hold devices, quantizers, or coder/decoders, etc., and the presence of discrete behavior such as timers that expire, resets, and impacts, give rise to dynamical systems with both continuous and discrete behavior, namely, *hybrid systems*. Such hybrid dynamics impose additional challenges to the construction of certificates to guaranteeing a desired dynamical property. In recent works, synthesizing Lyapunov function using LMI solvers inside a counter-example guided inductive system framework is shown to be feasible for switched systems [11]. In [12], an approach to learn a Lyapunov function given a parametric form with unknown coefficients, based on a system of linear inequality constraints is proposed. Finally, in [13], the authors tackle the problem of finding a Lyapunov function for nonlinear continuous-time systems using transformes. Though impactful, these approaches are not general enough to cover the behavior exhibited by hybrid systems.

To close this gap, in this work, for the broad class of hybrid systems in [14], we propose methods for neural network-based synthesis of certificates for asymptotic stability and optimality. Specifically, we present results for the synthesis of Lyapunov functions and for the construction of upper bounds on the cost associated to a solution to a hybrid system. The hybrid systems modeling framework in [14] is rich enough to cover switched systems, impulsive systems, algebraic differential equations, and hybrid automata. The main contribution of our paper is summarized as follows:

- We present a learning-based approach to synthesize a Lyapunov function that provably guarantees asymptotic stability of a compact set for a class of hybrid systems, by training a neural network through an optimization program.
- We present an approach for synthesizing an upper bound on the cost of solutions to a class of hybrid systems by training a neural network through an optimization program. Under additional sufficient conditions, we further show that this framework also guarantees asymptotic stability of a compact set. These results also cover cases

in which the strict-decrease requirement on the Lyapunov surrogate during can be relaxed.

- Complementary results for Lyapunov neural networks (LNNs) for a class of hybrid systems are also presented: we derive sufficient conditions to guarantee Lipschitz continuity of the decrease of LNNs under set-valued dynamics.

To the best of our knowledge, this paper presents the first approach able to synthesize Lyapunov functions and to find upper bounds on the cost of solutions to hybrid systems modeled as in [14] with set-valued dynamics, by only relying on finitely many samples from the state space. The most related work we are aware of is [15], where the authors propose the use of a neural network to learn a control barrier function and guarantee safety of a set for hybrid systems with single-valued dynamics.

The remainder of the paper is organized as follows. In Section 2, we present preliminary material. In Section 3, we present the data-driven design of Lyapunov functions for hybrid inclusions. Proposition 3.9 and Theorem 3.11 provide the main results of this section, focusing on how to extend sufficient Lyapunov conditions from finitely many points to a given bounded set of interest to certify asymptotic stability in a *practical* sense. Sufficient conditions to find an upper bound on the cost of solutions to hybrid inclusions are presented in Section 4, together with a data-driven approach to construct cost upper bounds for hybrid inclusions. This work expands our preliminary results in the conference paper [16] not only to include study cases of learning certificates for stability and cost evaluation for constrained dynamical systems in continuous, discrete, and hybrid time, but also to provide complete proofs of the results in [16] extended to the case where the hybrid system is modeled via inclusions. Specifically, in Section 3, we present additional numerical examples to certify asymptotic stability of a set for nonlinear continuous- and discrete-time systems (as special cases of a hybrid system) using our learning-based approach and, in Section 5, we present new results that provide sufficient conditions for simultaneous cost evaluation and asymptotic stability for hybrid inclusions, including data-driven conditions.

Notation. Let $\mathbb{N} := \{0, 1, 2, \dots\}$, $\mathbb{R}_{\geq 0} := [0, \infty)$, and $\mathbb{R}_{>0} := (0, \infty)$. Given $d \in \mathbb{N} \setminus \{0\}$, the shorthand $[d] := \{1, 2, \dots, d\}$ is used. For a vector $x \in \mathbb{R}^n$, $x^\top$ denotes its transpose, and let $|x|$ denote the Euclidean norm of x. Given two vectors, $x, y \in \mathbb{R}^n$, we write $(x, y) = [x^\top \ y^\top]^\top$, and $\langle x, y \rangle$ denotes the Euclidean inner product. Given a vector $x \in \mathbb{R}^n$ and a closed nonempty set $\mathcal{A} \subset \mathbb{R}^n$, the distance from x to $\mathcal{A}$ is defined as $|x|_{\mathcal{A}} := \inf_{y \in \mathcal{A}} |x - y|$. We denote by $\operatorname{int} \mathcal{A}$ the interior of $\mathcal{A}$, by $\overline{\mathcal{A}}$ its closure, by $\partial \mathcal{A}$ its boundary, and $\operatorname{co} \mathcal{A}$ represents the convex hull of $\mathcal{A}$. We represent by $\mathbb{B}$ (resp., $\mathbb{B}^\circ$) the closed (resp., open) Euclidean unit ball in $\mathbb{R}^n$; given $x \in \mathbb{R}^n$ and $\varepsilon > 0$, we write $x + \varepsilon\mathbb{B}$ (resp., $x + \varepsilon\mathbb{B}^\circ$) for the corresponding closed (resp., open) Euclidean ball of radius ε centered at x. Let I_n be the identity matrix of size n. We define the set of real symmetric matrices $\mathbb{S}^n := \{A \in \mathbb{R}^{n \times n} : A = A^\top\}$, and use $\mathbb{S}^n_{>0}$ and $\mathbb{S}^n_{\geq 0}$ for the set of real symmetric positive definite and semidefinite matrices of dimension n, respectively.

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $Y \subset \mathbb{R}^m$. Then, the preimage of Y under f, denoted $f^{-1}(Y)$, is the set of all elements of $\mathbb{R}^n$ that map to elements in Y under f. Namely,

$$f^{-1}(Y) := \{x \in \mathbb{R}^n : f(x) \in Y\}.$$

Given a nonempty set $U \subset \mathbb{R}^n$, the function $f : U \to \mathbb{R}^m$ is said to be of differentiability class $\mathcal{C}^k$ if the derivatives up to order $k \in \mathbb{N}$ exist and are continuous on U. For a function $\mathcal{C}^1 \ni f : \mathbb{R} \to \mathbb{R}$, f' represents it derivative . If $\mathcal{C}^1 \ni f : \mathbb{R}^n \to \mathbb{R}$, then $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ denotes its gradient, whereas if $\mathcal{C}^1 \ni f : \mathbb{R}^n \to \mathbb{R}^m$, then we write its Jacobian as $J_f : \mathbb{R}^n \to \mathbb{R}^{n\times m}$.

The notation $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ denotes a set-valued mapping (or map) where, for each $x \in \mathbb{R}^n$, $F(x) \subset \mathbb{R}^m$. For a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, its domain is $\operatorname{dom} F := \{x \in \mathbb{R}^n : F(x) \neq \emptyset\}$ and its graph is $\operatorname{gph} F := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : y \in F(x)\}$.

Given a nonempty set $\mathcal{A} \subset \mathbb{R}^n$, a function $V : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ is said to be positive definite with respect to $\mathcal{A}$, also written $V \in \mathcal{PD}(\mathcal{A})$, if $V(\mathbb{R}^n \setminus \mathcal{A}) \subset (0, \infty)$ and $V(\mathcal{A}) = \{0\}$. A function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a class-$\mathcal{K}$ function, also written $\alpha \in \mathcal{K}$, if α is zero at zero, continuous, and strictly increasing. A function $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a class-$\mathcal{KL}$ function, also written $\beta \in \mathcal{KL}$, if it is nondecreasing in its first argument, nonincreasing in its second argument, $\lim_{r\to 0} \beta(r, s) = 0$ for each $s \in \mathbb{R}_{\geq 0}$, and $\lim_{s\to\infty} \beta(r, s) = 0$ for each $r \in \mathbb{R}_{\geq 0}$.

Finally, given $x : I \to \mathbb{R}^n$, with $I \subset \mathbb{R}_{\geq 0}$, the notation $\dot{x}(t)$ denotes the time derivative of x at $t \in I$. Given $x : J \to \mathbb{R}^n$, with $J \subset \mathbb{N}$ and $\{j, j + 1\} \subset J$, the notation $x^+(j)$ denotes $x(j + 1)$.

Probability notions. For $k \in \mathbb{N} \setminus \{0\}$, $\mathcal{B}(\mathbb{R}^k)$ denotes the Borel σ-algebra[2] on $\mathbb{R}^k$. Let $(\Omega, \mathcal{F}, \gamma)$ be a probability space and $X : \Omega \to \mathbb{R}^k$ be $\mathcal{F}$-measurable, that is, $X^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{B}(\mathbb{R}^k)$. Then X is called a random vector on $(\Omega, \mathcal{F}, \gamma)$.

Definition 1.1. (Expected value [17, Def. 6.2.8]) *Let X be a random vector on $(\Omega, \mathcal{F}, \gamma)$. The expected value of X, denoted by $\mathbb{E}[X]$, is defined as*

$$\mathbb{E}[X] := \int_\Omega X d\gamma, \tag{1}$$

provided the integral is well-defined.

The push-forward measure $\gamma \circ X^{-1}$ is called the distribution of X. For a given probability measure $\bar{\gamma}$, we write $X \sim \bar{\gamma}$ if $\bar{\gamma} = \gamma \circ X^{-1}$ and say that X has distribution $\bar{\gamma}$. Given $\mu \in \mathbb{R}^k$ and $\Sigma \in \mathbb{S}^k_{>0}$, $\mathcal{N}(\mu, \Sigma)$ denotes the standard Gaussian measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ with

[2]See [17, Def. 1.1.2, 1.1.3, and 1.1.4].

parameters (μ, Σ). If $X \sim \mathcal{N}(\mu, \Sigma)$, we say that $X : \Omega \to \mathbb{R}^n$ is Gaussian. We use $\mathcal{L}(\mathbb{R}^k)$ to denote the collection of Lebesgue measurable sets[3] in $\mathbb{R}^k$ and $\lambda : \mathcal{L}(\mathbb{R}^k) \to [0, \infty]$ represents the standard k-dimensional Lebesgue measure (see [18, p. 426] or [19, Thm. 1.55]).

2. Preliminaries

2.1. Set-Valued Mappings

We start this section by recalling the following continuity notions for set-valued maps.

Definition 2.1. (Outer semicontinuity of set-valued maps [20, Def. A.30]) *Consider a set-valued map $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. The map F is said to be outer semicontinuous (osc) at $x \in \mathbb{R}^n$ if, for every convergent sequence $x_i \to x$ and any convergent sequence $y_i \in F(x_i)$, one has $y_i \to y \in F(x)$. It is said to be osc if it is osc for all $x \in \mathbb{R}^n$. Given a set $S \subset \mathbb{R}^n$, F is osc relative to S if the set-valued mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ defined by $F(x)$ for $x \in S$ and $\emptyset$ for $x \notin S$ is osc at each $x \in S$.*

Definition 2.2. (Locally bounded set-valued maps [20, Def. A.32]) *A set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is locally bounded at $x \in \mathbb{R}^n$ if there exists a neighborhood $\mathcal{O}$ of x such that $F(\mathcal{O}) := \bigcup_{x \in \mathcal{O}} F(x)$ is bounded. It is said to be locally bounded if it is locally bounded at each $x \in \mathbb{R}^n$. Given a set $S \subset \mathbb{R}^n$, F is locally bounded relative to S if the set-valued mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ defined by $F(x)$ for $x \in S$ and $\emptyset$ for $x \notin S$ is locally bounded at each $x \in S$.*

Next, we will define the Pompeiu-Hausdorff distance between two sets.

Definition 2.3. (Pompeiu-Hausdorff distance [21, Ex. 4.13]) *For $C, D \subset \mathbb{R}^n$ closed and nonempty, the* Pompeiu-Hausdorff distance *between C and D is the quantity*

$$d_H(C, D) := \sup_{x \in C \cup D} \left| |x|_C - |x|_D \right|, \tag{2}$$

which is equivalent to

$$\begin{aligned} d_H(C, D) &= \inf \{\eta \geq 0 : C \subset D + \eta\mathbb{B},\ D \subset C + \eta\mathbb{B}\} \\ &= \max \left\{ \sup_{x \in C} |x|_D, \sup_{x \in D} |x|_C \right\}. \end{aligned}$$

[3] See [18, p. 426].

Definition 2.4. (Lipschitz continuity [21, Def. 9.26]) *A set-valued mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *is Lipschitz continuous on* $S \subset \mathbb{R}^n$ *if it is nonempty closed-valued on* S *and there exists* $L \in \mathbb{R}_{\geq 0}$ *such that*

$$F(x') \subset F(x) + L|x' - x|\mathbb{B} \qquad \forall\, x, x' \in S.$$

Equivalently, F *is Lipschitz continuous on* S *if there exists* $L \in \mathbb{R}_{\geq 0}$ *such that*

$$d_H\left(F(x), F(x')\right) \leq L|x - x'| \qquad \forall\, x, x' \in S.$$

Notice that, for any single-valued mapping $F : \mathbb{R}^n \to \mathbb{R}^m$, viewed as a special case of a set-valued mapping, Definition 2.4 is equivalent to the standard definition of Lipschitz continuity for single-valued functions (see [21, Def. 9.1]).

2.2. Modeling Hybrid Inclusions

This paper considers hybrid systems that will be modeled based on the framework in [14]. In such a framework, the continuous dynamics of the system are modeled by differential inclusions with constraints, while the discrete dynamics are modeled by difference inclusions with constraints. A hybrid inclusion $\mathcal{H}$ is defined by

$$\mathcal{H} \;:\; \left\{ \begin{array}{lcll} \dot{x} & \in & F(x) & \quad x \in C \\ x^+ & \in & G(x) & \quad x \in D \end{array} \right. \tag{3}$$

where $x \in \mathbb{R}^n$ is the state. The *flow map* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ captures the continuous evolution of the system, when the state is in the *flow set* C. The *jump map* $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ describes the discrete evolution of the system when the state is in the *jump set* D.

Since solutions to $\mathcal{H}$ as in (3) can exhibit both continuous and discrete behavior, we use ordinary time $t \in \mathbb{R}_{\geq 0}$ to determine the amount of flow elapsed and a counter $j \in \mathbb{N}$ that keeps track of the number of jumps that have occurred. Based on this parametrization of time, the concept of hybrid time domain, over which solutions to $\mathcal{H}$ are defined, is as follows.

Definition 2.5. (Hybrid time domain) *A set* $\tilde{E} \subset \mathbb{R}_{\geq 0} \times \mathbb{N}$ *is a compact hybrid time domain if there exists* $J \in \mathbb{N}$ *such that*

$$\tilde{E} = \bigcup_{j=0}^{J} \left([t_j, t_{j+1}] \times \{j\}\right) \tag{4}$$

for some finite sequence of times $\{t_j\}_{j=0}^{J+1}$ *satisfying* $0 = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_J \leq t_{J+1}$. *A set* $E \subset \mathbb{R}_{\geq 0} \times \mathbb{N}$ *is a* hybrid time domain *if it is the union of a nondecreasing sequence* $E_1 \subset E_2 \subset E_3 \subset \ldots$ *of compact hybrid time domains.*

A hybrid signal is a function defined on a hybrid time domain. Given a hybrid signal ϕ and $j \in \mathbb{N}$, we define $I_\phi^j := \{t : (t,j) \in \operatorname{dom}\phi\}$ and $\operatorname{rge}\phi := \{\phi(t,j) : (t,j) \in \operatorname{dom}\phi\}$.

Definition 2.6. (Hybrid arc) *A hybrid signal $\phi : \operatorname{dom}\phi \to \mathbb{R}^n$ is called a* hybrid arc *if, for each $j \in \mathbb{N}$, the function $t \mapsto \phi(t,j)$ is locally absolutely continuous*[4] *on the interval I_ϕ^j. A hybrid arc ϕ is said to be compact if $\operatorname{dom}\phi$ is compact.*

A solution to the hybrid system $\mathcal{H}$ is defined as follows.

Definition 2.7. (Solution to $\mathcal{H}$) *A hybrid arc $\phi : \operatorname{dom}\phi \to \mathbb{R}^n$ defines a* solution *to $\mathcal{H}$ if*

1. *$\phi(0,0) \in \overline{C}$ or $\phi(0,0) \in D$;*
2. *For each $j \in \mathbb{N}$ such that I_ϕ^j has a nonempty interior $\operatorname{int} I_\phi^j$, we have, for all $t \in \operatorname{int} I_\phi^j$,*
$$\phi(t,j) \in C$$
and, for almost all $t \in I_\phi^j$,
$$\frac{d\phi}{dt}(t,j) \in F(\phi(t,j));$$
3. *For all $(t,j) \in \operatorname{dom}\phi$ such that $(t,j+1) \in \operatorname{dom}\phi$,*
$$\begin{aligned} \phi(t,j) &\in D \\ \phi(t,j+1) &\in G(\phi(t,j)). \end{aligned}$$

A solution ϕ to $\mathcal{H}$ from $\xi \in \mathbb{R}^n$ is complete if $\operatorname{dom}\phi$ is unbounded. It is maximal if there is no solution φ from ξ such that $\phi(t,j) = \varphi(t,j)$ for all $(t,j) \in \operatorname{dom}\phi$ and $\operatorname{dom}\phi$ is a proper subset of $\operatorname{dom}\varphi$. We denote by $\widehat{\mathcal{S}}_{\mathcal{H}}(M)$ the set of solutions ϕ to $\mathcal{H}$ with $\phi(0,0) \in M$. The set $\mathcal{S}_{\mathcal{H}}(M) \subset \widehat{\mathcal{S}}_{\mathcal{H}}(M)$ denotes all maximal solution from M. We define

$$\begin{aligned} \sup\nolimits_t \operatorname{dom}\phi &:= \sup\{t \in \mathbb{R}_{\geq 0} : \exists j \in \mathbb{N} \text{ s.t. } (t,j) \in \operatorname{dom}\phi\} \\ \sup\nolimits_j \operatorname{dom}\phi &:= \sup\{j \in \mathbb{N} : \exists t \in \mathbb{R}_{\geq 0} \text{ s.t. } (t,j) \in \operatorname{dom}\phi\}. \end{aligned}$$

Well-posed hybrid systems refer to a class of hybrid systems where the solutions satisfy very useful structural properties [14]. A hybrid system $\mathcal{H}$ as in (3) is well-posed if the following conditions hold.

[4] See [20, Def. A.20].

Assumption 2.8. (Hybrid basic conditions) *A hybrid system* $\mathcal{H} = (C, F, D, G)$ *satisfies the* hybrid basic conditions *if*

1) *C and D are closed subsets of $\mathbb{R}^n$;*
2) *$F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is osc and locally bounded relative to C, $C \subset \operatorname{dom} F$, and $F(x)$ is convex for each $x \in C$;*
3) *$G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is osc and locally bounded relative to D, and $D \subset \operatorname{dom} G$.*

2.3. Stability for Hybrid Dynamical Systems

The following definition provides the notion of pre-asymptotic stability of a closed set of interest for $\mathcal{H}$ as in (3).

Definition 2.9. (Local pre-asymptotic stability (LpAS)) *Given a hybrid system* $\mathcal{H} = (C, F, D, G)$ *as in* (3)*, a closed set* $\mathcal{A} \subset \mathbb{R}^n$ *is said to be*

- *stable for $\mathcal{H}$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that every $\phi \in \widehat{\mathcal{S}}_{\mathcal{H}}(\mathcal{A} + \delta\mathbb{B})$ satisfies $\operatorname{rge} \phi \subset \mathcal{A} + \varepsilon\mathbb{B}$;*
- *locally pre-attractive for $\mathcal{H}$ if there exists $\mu > 0$ such that every $\phi \in \mathcal{S}_{\mathcal{H}}(\mathcal{A} + \mu\mathbb{B})$ is bounded and, if ϕ is complete, then*

$$\lim_{\substack{(t,j)\in \operatorname{dom}\phi \\ t+j\to\infty}} |\phi(t,j)|_{\mathcal{A}} = 0; \tag{5}$$

- *locally pre-asymptotically stable (LpAS) for $\mathcal{H}$ if it is both stable and locally pre-attractive for $\mathcal{H}$.*

Sufficient conditions guaranteeing LpAS of $\mathcal{A}$ for $\mathcal{H}$ without computing solutions to $\mathcal{H}$ rely on Lyapunov functions.

Definition 2.10. (Lyapunov function candidate [20, Def. 3.17]) *The sets $\mathcal{O}, \mathcal{A} \subset \mathbb{R}^n$ and the function $V : \operatorname{dom} V \to \mathbb{R}$ define a Lyapunov function candidate on $\mathcal{O}$ with respect to $\mathcal{A}$ for $\mathcal{H} = (C, F, D, G)$ if the following conditions hold:*

1) *$(\overline{C} \cup D \cup G(D)) \cap \mathcal{O} \subset \operatorname{dom} V$;*

2) $\mathcal{O}$ contains an open neighborhood of $\mathcal{A}$;

3) V is continuously differentiable on an open set containing $\overline{C} \cap \mathcal{O}$;

4) there exist $\alpha_1, \alpha_2 \in \mathcal{K}$ such that[5]

$$\alpha_1(|x|_{\mathcal{A}}) \leq V(x) \leq \alpha_2(|x|_{\mathcal{A}}) \qquad \forall\, x \in C \cup D \cup G(D). \tag{6}$$

With the definition of a Lyapunov function candidate, the following result gives sufficient conditions to certify LpAS of a set for a hybrid system.

Theorem 2.11. (Hybrid Lyapunov theorem [20, Thm. 3.19]) *Consider $\mathcal{O}, \mathcal{A} \subset \mathbb{R}^n$ and suppose that V is a Lyapunov function candidate on $\mathcal{O}$ with respect to $\mathcal{A}$ for $\mathcal{H} = (C, F, D, G)$. If $\mathcal{A}$ is compact, $\mathcal{H}$ satisfies Assumption 2.8, and*

$$\dot{V}(x) := \max_{f \in F(x)} \langle \nabla V(x), f \rangle < 0 \qquad \forall\, x \in (C \cap \mathcal{O}) \setminus \mathcal{A} \tag{7a}$$

$$\Delta V(x) := \max_{g \in G(x)} V(g) - V(x) < 0 \qquad \forall\, x \in (D \cap \mathcal{O}) \setminus \mathcal{A}, \tag{7b}$$

then $\mathcal{A}$ is LpAS for $\mathcal{H}$ on $\mathcal{O}$.

If the Lyapunov function candidate V on $\mathcal{O}$ with respect to $\mathcal{A}$ satisfies (7a) and (7b), then it is said to be a Lyapunov function on $\mathcal{O}$ with respect to $\mathcal{A}$ for $\mathcal{H}$.

3. Learning-Based Lyapunov Functions for Hybrid Inclusions

In this section, our main objective is to design a Lyapunov function that guarantees asymptotic stability of a compact set $\mathcal{A}$ for $\mathcal{H}$ as in (3) via learning-based methods. Specifically, we solve an optimization program at finitely many points satisfying sufficient stability pointwise conditions. Then, using continuity of such Lyapunov function, together with an estimate of how dense the data used for training is, we provide sufficient conditions to guarantee the decrease of the Lyapunov functions during flows and at jumps, even for points not used in the optimization. To achieve this, we use $\varepsilon-$nets, as defined next.

Definition 3.1. (ε-Nets [22, Def. 4.2.1]) *Consider a set $K \subset \mathbb{R}^n$ and let $\varepsilon > 0$. A subset $K_\varepsilon := \{x_1, x_2, \ldots, x_q\} \subset K$, $q \in \mathbb{N} \setminus \{0\}$, is called an* ε-net *of K if*

$$\forall\, x \in K \quad \text{there exists} \quad x' \in K_\varepsilon \quad \text{such that} \quad |x - x'| \leq \varepsilon.$$

[5]Notice that, if (6) holds, then $V \in \mathcal{PD}(\mathcal{A})$ on $C \cup D \cup G(D)$.

Namely, K_ε is an ε-net of K if K can be covered by closed balls with centers in K_ε and radii ε. In particular,

$$K \subset \bigcup_{x' \in K_\varepsilon} x' + \varepsilon \mathbb{B}$$

*The smallest possible cardinality of an ε-net of K is called the ε-*covering number *of K, and it is denoted $N(K, \varepsilon)$.*

Notice that, given $\varepsilon > 0$ and a closed set $K \subset \mathbb{R}^n$, if K_ε is an ε-net of K, then $d_H(K, K_\varepsilon) \leq \varepsilon$.

3.1. Learning a Lyapunov Function Surrogate

To provably guarantee asymptotic stability properties of a set $\mathcal{A}$ for $\mathcal{H}$ using learning-based surrogates for a Lyapunov function, we introduce the following neural network architecture that accounts for the specific properties of Lyapunov function candidates (see Section 2.3).

Definition 3.2. (Lyapunov neural network) *Consider $d \in \mathbb{N} \setminus \{0\}$ and $q_0 := n \in \mathbb{N} \setminus \{0\}$. Define, for each $m \in [d]$,*

$$z^{(m)}(w) := \left(z_1^{(m)}(w), z_2^{(m)}(w), \ \ldots \ , z_{q_m}^{(m)}(w)\right) \qquad \forall w \in \mathbb{R}^{q_{m-1}}, \tag{8}$$

with $q_m \in \mathbb{N} \setminus \{0\}$ and, given a function $\sigma : \mathbb{R} \to \mathbb{R}$ such that $\sigma(x) = 0$ if and only if $x = 0$, we define, for all $w \in \mathbb{R}^{q_{m-1}}$,

$$z_i^{(m)}(w) := \sigma\left(w^\top \theta_i^{(m)}\right) \qquad \forall\, i \in [q_m] \tag{9}$$

where, for each $i \in [q_m]$, the weight vector $\theta_i^{(m)} \in \mathbb{R}^{q_{m-1}}$ is such that

$$\operatorname{rank} \begin{bmatrix} \theta_1^{(m)} & \theta_2^{(m)} & \cdots & \theta_{q_m}^{(m)} \end{bmatrix} = q_{m-1}. \tag{10}$$

Thus, given $\eta : \mathbb{R}^n \to \mathbb{R}^n$, we define a Lyapunov neural network (LNN) *$\widehat{V} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, with parameters (θ, σ, η), as*

$$\widehat{V}(x) := \left|\left(z^{(d)} \circ z^{(d-1)} \circ \cdots \circ z^{(1)}\right)(\eta(x))\right|^2 \tag{11}$$

where

$$\theta := \left(\theta_1^{(1)}, \ \ldots \ , \theta_{q_1}^{(1)}, \theta_1^{(2)}, \ \ldots \ , \theta_{q_2}^{(2)}, \ \ldots \ , \theta_1^{(d)}, \ \ldots \ , \theta_{q_d}^{(d)}\right) \in \mathbb{R}^r \tag{12}$$

with $r = \sum_{m \in [d]} q_{m-1} q_m$.

The following result shows that an LNN $\widehat{V}$ with parameters (θ, σ, η) is positive definite with respect to a given compact set $\mathcal{A} \subset \mathbb{R}^n$ for any parameter θ.

Lemma 3.3. (Positive definiteness of $\widehat{V}$) *Consider a closed and nonempty set $\mathcal{X} \subset \mathbb{R}^n$, a compact and nonempty set $\mathcal{A} \subset \mathcal{X}$, and suppose that $\widehat{V}$ is an LNN with parameters (θ, σ, η). Then, there exist $\alpha_1, \alpha_2 \in \mathcal{K}$ such that*

$$\alpha_1(|x|_{\mathcal{A}}) \leq \widehat{V}(x) \leq \alpha_2(|x|_{\mathcal{A}}) \qquad \forall\, x \in \mathcal{X} \tag{13}$$

if and only if σ and η are continuous, $\eta(\mathcal{A}) = \{0\}$, and $\eta(x) \neq 0$ for all $x \notin \mathcal{A}$.

Proof.

($\Rightarrow$) Pick $x \in \mathcal{A}$ and let $w_0 := \eta(x)$. Then, $w_0 = 0$. For each $m \in [d]$, set $w_m := z^{(m)}(w_{m-1}) \in \mathbb{R}^{q_m}$. Since $\sigma(s) = 0$ if and only if $s = 0$, from (8) and (9), we get

$$w_0 = 0 \quad \Longrightarrow \quad z_i^{(1)}(w_0) = 0 \qquad \forall i \in [q_1].$$

Inductively, for every $m \in [d]$, $w_{m-1} = 0$ implies $z_i^{(m)}(w_{m-1}) = 0$ for all $i \in [q_m]$; hence $w_d = 0$. Then, from (11), $\widehat{V}(x) = 0$.

($\Leftarrow$) To show that $\widehat{V}(x) = 0$ implies $x \in \mathcal{A}$, let us rewrite (11) as

$$\widehat{V}(x) = \sum_{i=1}^{q_d} \left(z_i^{(d)} \left(z^{(d-1)} \left(\dots z^{(2)} \left(z^{(1)} \left(g(x) \right) \right) \dots \right) \right) \right)^2 \qquad \forall x \in \mathcal{X}. \tag{14}$$

Similarly, let $w_0 := \eta(x)$ and, for each $m \in [d]$, set $w_m := z^{(m)}(w_{m-1}) \in \mathbb{R}^{q_m}$. Suppose that $\widehat{V}(x) = 0$. Since squares are nonnegative, from (14), it must follow that

$$\widehat{V}(x) = 0 \quad \Longrightarrow \quad w_d = 0 \quad \Longrightarrow \quad z_i^{(d)}(w_{d-1}) = 0 \qquad \forall i \in [q_d].$$

Recall that, by definition, $\sigma(s) = 0$ if and only if $s = 0$. Thus, from (8) and (9), we have that

$$z_i^{(d)}(w_{d-1}) = 0 \quad \Longrightarrow \quad \left\langle \theta_i^{(d)}, w_{d-1} \right\rangle = 0 \qquad \forall i \in [q_d].$$

From (10), we conclude that $w_{d-1} = 0$. Therefore, inductively, for every $m \in \{d, d-1, \dots, 1\}$, $w_m = 0$ implies $z_i^{(m)}(w_{m-1}) = 0$ for all $i \in [q_m]$; hence, $w_0 = \eta(x) = 0$. With $\eta(\mathcal{A}) = \{0\}$ and $\eta(z) \neq 0$ for all $z \notin \mathcal{A}$, we conclude that $x \in \mathcal{A}$.

Thus, we have that $\widehat{V} \in \mathcal{PD}(\mathcal{A})$ on $\mathcal{X}$. Next, from [23, Prop. 2] there exists a continuous $\rho_c \in \mathcal{PD}(\{0\})$ such that

$$\rho_c(|x'|_{\mathcal{A}}) \leq \widehat{V}(x') \qquad \forall\, x' \in \mathcal{X}.$$

Define $\tilde{\alpha}_1 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ as

$$\tilde{\alpha}_1(s) := \inf_{s' \geq s} \rho_c(s')$$

which is continuous, zero at zero, strictly positive on $\mathbb{R}_{>0}$, and nondecreasing. Then, by [24, Lem. 1], there exists $\alpha_1 \in \mathcal{K}$ such that $\alpha_1(s) \leq \tilde{\alpha}_1(s)$ for all $s \geq 0$, and, as a consequence, we have that

$$\alpha_1(|x'|_{\mathcal{A}}) \leq \tilde{\alpha}_1(|x'|_{\mathcal{A}}) \leq \widehat{V}(x') \qquad \forall\, x' \in \mathcal{X}.$$

Finally, since $\mathcal{X}$ is closed and nonempty, the existence of $\alpha_2 \in \mathcal{K}$ follows directly from [23, Lem. 2]. ■

Notice that, given $\mathcal{A} \subset \mathbb{R}^n$ and $\mathcal{H} = (C, F, D, G)$, under the conditions in Lemma 3.3, we have that there exist $\alpha_1, \alpha_2 \in \mathcal{K}$ such that (13) holds on $\overline{C \cup D \cup G(D)}$ for a Lyapunov neural network $\widehat{V}$ with parameters (θ, σ, η). This is true regardless of the choice of the weight parameters θ. Previous results in the literature, such as [25] and [26], fix $\alpha_1, \alpha_2 \in \mathcal{K}$ and $\widehat{V}$ is designed such that a bound like (13) is obtained.

Furthermore, from the discussion above, an LNN $\widehat{V}$ with parameters (θ, σ, η) is a Lyapunov function candidate on $\mathcal{O}$ with respect to $\mathcal{A}$ for $\mathcal{H}$ (see Definition 2.10) if, in addition, the activation function σ and η in (11) are continuously differentiable on an open set containing $\overline{C} \cap \mathcal{O}$. This motivates the next assumption and the use of $\widehat{V}$ to certify that $\mathcal{A}$ is LpAS for $\mathcal{H}$ on $\mathcal{O}$ following Theorem 2.11.

Assumption 3.4. (Basic conditions for LNNs) *Consider the sets $\mathcal{O}, \mathcal{A} \subset \mathbb{R}^n$, the hybrid system $\mathcal{H} = (C, F, D, G)$, and let $\widehat{V}$ be an LNN with parameters (θ, σ, η). Suppose that*

1) $\mathcal{A}$ is compact;

2) $\mathcal{O}$ contains an open neighborhood of $\mathcal{A}$;

3) $\sigma \in \mathcal{C}^1$ and $\eta \in \mathcal{C}^1$ on an open set containing $\overline{C} \cap \mathcal{O}$;

4) $\eta(\mathcal{A}) = \{0\}$ and $\eta(x) \neq 0$ for all $x \notin \mathcal{A}$.

Notice that Assumption 3.4 does not impose constraints on θ but on the functions σ and η that define an LNN. Thus, to guarantee that $\mathcal{A}$ is LpAS for $\mathcal{H}$ using $\widehat{V}$, now our goal is to *tune* θ such that (7) is satisfied. This is known as *training*, and it is typically done using optimization-based approaches. In particular, let γ be the standard Gaussian measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, $\delta \sim \mathcal{N}(0, I_n)$ be a random vector on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \gamma)$, and suppose that $F(x)$ (resp., $G(x)$) is compact for each $x \in C$ (resp., $x \in D$) and that $\nabla\widehat{V}$ is Borel measurable[6]

[6]Notice that, from Assumption 3.4, $\sigma \in \mathcal{C}^1$ and $\eta \in \mathcal{C}^1$ on an open set containing $\overline{C} \cap \mathcal{O}$. Thus, by function composition, $\nabla\widehat{V}$ is continuous on an open set containing $\overline{C} \cap \mathcal{O}$ and, as a result, it is Boreal measurable on such a set [18, Prop. 3].

on an open set containing $\overline{C} \cap \mathcal{O}$. Thus, we propose to pick the parameters of the LNN as

$$\begin{aligned} \theta^* \in \arg\min_{\theta \in \mathbb{R}^r} \quad & \int_{(C \cap \mathcal{O}) \setminus \mathcal{A}} \mathbb{E}\left[\left|\nabla \widehat{V}(x+\delta)\right|\right] d\lambda(x) \\ \text{subject to} \quad & \max_{f \in F(x)} \left\langle \nabla \widehat{V}(x), f \right\rangle < 0 \qquad \forall x \in (C \cap \mathcal{O}) \setminus \mathcal{A} \\ & \max_{g \in G(x)} \widehat{V}(g) - \widehat{V}(x) < 0 \qquad \forall x \in (D \cap \mathcal{O}) \setminus \mathcal{A}. \end{aligned} \tag{15}$$

Notice that, for each $x \in (C \cap \mathcal{O}) \setminus \mathcal{A}$, $\delta \mapsto \left|\nabla \widehat{V}(x+\delta)\right|$ is a random vector on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \gamma)$ as it is the composition of Borel measurable functions. In addition, notice that $\left|\nabla \widehat{V}(x + \delta(\omega))\right| \geq 0$ for each $\omega \in \mathbb{R}^n$ and each $x \in (C \cap \mathcal{O}) \setminus \mathcal{A}$. Thus, the expected value in the objective function in (15) is well-defined according to [17, Def. 6.2.8]. This objective function is known in the literature as the *gradient penalty* [27], and it enforces smoothness of $\widehat{V}$ on neighborhoods of the training data. Furthermore, if Assumption 2.8 and Assumption 3.4 hold, from Lemma 3.3, there exist $\alpha_1, \alpha_2 \in \mathcal{K}$ such that (13) holds on $C \cup D \cup G(D)$. This, together with the constraints in (15) enforced at training, imply that $\mathcal{A}$ is LpAS for $\mathcal{H}$ using Theorem 2.11.

Training the LNN $\widehat{V}$ using (15) requires constraints satisfaction for (possibly) infinitely many points in the flow and jump sets, which is computationally intractable. Therefore, we propose a tractable approximation to the optimization program in (15) through a scenario[7] program in which only finitely many constraints are considered. To accomplish this, given $0 < \varepsilon < \mu$, let $\mathcal{M}_\star$ be an ε-net of $(\star \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ)$ for each $\star \in \{C, D\}$, and pick the parameters of the LNN as

$$\begin{aligned} \theta^* \in \arg\min_{\theta \in \mathbb{R}^r} \quad & \sum_{x' \in \mathcal{M}_C} \mathbb{E}\left[\left|\nabla \widehat{V}(x'+\delta)\right|\right] \lambda\left((x' + \varepsilon\mathbb{B}) \cap \mathcal{O}\right) \\ \text{subject to} \quad & \max_{f \in F(x')} \left\langle \nabla \widehat{V}(x'), f \right\rangle \leq -\tau_C \qquad \forall x' \in \mathcal{M}_C \\ & \max_{g \in G(x')} \widehat{V}(g) - \widehat{V}(x') \leq -\tau_D \qquad \forall x' \in \mathcal{M}_D, \end{aligned} \tag{16}$$

where $\tau_C, \tau_D > 0$ are slack variables. Once the LNN $\widehat{V}$ is trained using (16), we extend the Lyapunov decrease conditions along flows and at jumps to points that were not used at training. Thus, observe that if we allow $\mu \leq \varepsilon$, generalizing such conditions to every $\varepsilon-$ball with center in $\mathcal{M}_C \cup \mathcal{M}_D$ may impose strict decrease of $\widehat{V}$ on $\mathcal{A}$. This justifies enforcing the constraints only outside a $\mu-$neighborhood of $\mathcal{A}$. Naturally, this does not entail a cost-free implementation, and its implications are included in Theorem 3.11.

---

[7]Referring to the fact that (15) will be solved at finitely many state values [28].

3.2. Generalizing Lyapunov Conditions from ε-Nets

As discussed previously, we aim to generalize the conditions enforced at the nets of the flow and jump sets to states that were not used at training. In particular, we can leverage the slack variables τ_C and τ_D in (16) together with sufficient regularity properties of the LNN $\widehat{V}$ to guarantee that the Lyapunov conditions (7a) and (7b) hold at all points in $(C \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ)$ and $(D \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ)$, respectively.

3.2.1. Regularity Properties of Lyapunov Neural Networks

Sufficient conditions to guarantee Lipschitz continuity of a Lyapunov neural network $\widehat{V}$ are presented next.

Definition 3.5. (Slope restricted nonlinearity [29, Def. 1]) *Given $0 \leq \alpha < \beta < \infty$, a function $\sigma : \mathbb{R} \to \mathbb{R}$ is said to be slope restricted on $[\alpha, \beta]$ if*

$$\alpha \leq \frac{\sigma(y) - \sigma(x)}{y - x} \leq \beta \qquad \forall\, x, y \in \mathbb{R} : x \neq y.$$

For ease of notation, given a Lyapunov neural network $\widehat{V}$ with parameters (θ, σ, η), define

$$A_\theta := \begin{bmatrix} \left[\theta_k^{(1)}\right]_{k=1}^{q_1} & 0 & \cdots & 0 & 0 \\ 0 & \left[\theta_k^{(2)}\right]_{k=1}^{q_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \left[\theta_k^{(d)}\right]_{j=1}^{q_d} & 0 \end{bmatrix}^\top, \qquad B := \begin{bmatrix} 0 & I_{q_1} & 0 & \cdots & 0 \\ 0 & 0 & I_{q_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & I_{q_d} \end{bmatrix}, \tag{17}$$

where, for each $m \in [d]$, q_m denotes the number of neurons in the m-th layer and, for each $k \in [q_m]$, $\theta_k^{(m)} \in \mathbb{R}^{q_{m-1}}$ is the weight vector of the k-th neuron in the m-th layer. Furthermore, for each $p \in \mathbb{N} \setminus \{0\}$, consider the set $\mathcal{T}_p \subset \mathbb{S}^p_{\geq 0}$ defined by

$$\mathcal{T}_p := \left\{ T \in \mathbb{R}^{p \times p} : T = \sum_{i \in [p]} \vartheta_{ii} e_i e_i^\top,\ \vartheta_{ii} \geq 0\ \forall i \in [p] \right\}, \tag{18}$$

which is a convex cone used to encode pairwise constraints between neurons in a neural network [30]. Now, we introduce the following result which presents a sufficient condition for the Lipschitz continuity of an LNN.

Lemma 3.6. (Lipschitz continuity of $\widehat{V}$) *Consider a bounded set $\mathcal{O} \subset \mathbb{R}^n$ and an LNN $\widehat{V}$ with parameters (θ, σ, η). Suppose that*

1) for some $0 \le \alpha < \beta < \infty$, σ is slope restricted on $[\alpha, \beta]$;

2) η is L_η-Lipschitz continuous on $\overline{\mathcal{O}}$;

3) there exist $(\rho, T) \in \mathbb{R}_{\ge 0} \times \mathcal{T}_p$, with $\mathcal{T}_p$ as in (18), such that the following condition holds:

$$M(\rho, T) := \begin{bmatrix} A_\theta \\ B \end{bmatrix}^\top \begin{bmatrix} -2\alpha\beta T & (\alpha+\beta)T \\ (\alpha+\beta)T & -2T \end{bmatrix} \begin{bmatrix} A_\theta \\ B \end{bmatrix} + \begin{bmatrix} -\rho I_n & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_{q_d} \end{bmatrix} \preccurlyeq 0 \qquad (19)$$

with A_θ and B given in (17), $p := \sum_{m \in [d]} q_m$ and, for each $m \in [d]$, $d \in \mathbb{N} \setminus \{0\}$, q_m denotes the number of neurons of the m-th layer (see Definition 3.2).

Then, there exists $\widehat{\alpha} \in \mathbb{R}_{\ge 0}$ such that $\widehat{V}$ is Lipschitz on $\mathcal{O}$ with constant

$$0 \le L_{\widehat{V}} \le \widehat{\alpha}\rho. \qquad (20)$$

Proof. Let us define $h : \mathbb{R}^n \to \mathbb{R}^{q_d}$ as

$$h(y) := \left(z^{(d)} \circ z^{(d-1)} \circ \cdots \circ z^{(1)}\right)(y) \qquad \forall y \in \mathbb{R}^n \qquad (21)$$

where, for each $m \in [d]$, the map $w \mapsto z^{(m)}(w)$ is given by (8). Suppose that (19) holds for some $(\rho, T) \in \mathbb{R}_{\ge 0} \times \mathcal{T}_p$. Then, thanks to [30, Thm. 1], it follows that

$$|h(y) - h(y')| \le \sqrt{\rho}\, |y - y'| \qquad \forall\, y, y' \in \mathbb{R}^n. \qquad (22)$$

Next, notice that we can rewrite $\widehat{V}$ in (11) as $x \mapsto \widehat{V}(x) = |h(\eta(x))|^2$. Thus, for all $x, x' \in \mathcal{O}$, we have that

$$\begin{aligned} |\widehat{V}(x) - \widehat{V}(x')| &= \left| |h(\eta(x))|^2 - |h(\eta(x'))|^2 \right| \\ &= \left| |h(\eta(x))|^2 + \langle h(\eta(x)), h(\eta(x')) \rangle - \langle h(\eta(x)), h(\eta(x')) \rangle - |h(\eta(x'))|^2 \right| \\ &= \left| \left\langle h(\eta(x)) + h(\eta(x')), h(\eta(x)) - h(\eta(x')) \right\rangle \right|. \end{aligned}$$

Now, using Cauchy-Schwarz and the triangle inequality, together with Lipschitz continuity of h and of η, we observe that

$$\begin{aligned} |\widehat{V}(x) - \widehat{V}(x')| &\le |h(\eta(x)) + h(\eta(x'))|\; |h(\eta(x)) - h(\eta(x'))| \\ &\le L_\eta \sqrt{\rho} \big(|h(\eta(x))| + |h(\eta(x'))| \big) |x - x'| \end{aligned}$$

$$\leq 2L_\eta\sqrt{\rho}\, \sup_{y\in\overline{\mathcal{O}}} |h(\eta(y))|\, |x-x'|.$$

Finally, from (22), together with the fact that $\sigma(x)=0$ if and only if $x=0$ (see Definition 3.2) and continuity[8] of σ and η, it readily follows that

$$\sup_{y\in\overline{\mathcal{O}}} |h(\eta(y))| \leq \sqrt{\rho}\sup_{y\in\overline{\mathcal{O}}} |\eta(y)| < \infty$$

which gives a Lipschitz constant $L_{\widehat{V}}$ satisfying

$$0 \leq L_{\widehat{V}} \leq 2L_\eta \rho \sup_{y\in\overline{\mathcal{O}}} |\eta(y)|.$$

■

Notice that, when the function $x \mapsto \sigma(x)$ is slope-restricted on the interval $[\alpha,\beta]$, for some $0\leq\alpha<\beta<\infty$, it is also Lipschitz continuous with constant $L_\sigma := \beta$. This, together with the Lipschitz continuity of $x\mapsto\eta(x)$ on some compact set $\overline{\mathcal{O}}$, implies that the LNN $\widehat{V}$ with parameters (θ,σ,η) is also Lipschitz on $\mathcal{O}$. According to [31], a naive approach is given by the product of the norms of the weight matrices. Though theoretically sound, this bound is known to be quite loose [29, 30]. Thus, the importance of Lemma 3.6 is that it allows us to compute a tighter estimate of the Lipschitz constant of $\widehat{V}$ by solving the following semidefinite program (SDP):

$$\text{minimize} \quad \rho \quad \text{subject to} \quad M(\rho,T)\preccurlyeq 0, \quad \rho\geq 0, \quad T\in\mathcal{T}_p \tag{23}$$

where $\mathcal{T}_p$ is defined as in (18) for $p := \sum_{m\in[d]} q_m$.

The next result shows that the gradient of the Lyapunov neural network is Lipschitz on bounded sets.

Lemma 3.7. (Lipschitz continuity of the gradient of $\widehat{V}$) *Consider a bounded set $\mathcal{O}\subset\mathbb{R}^n$ and an LNN $\widehat{V}$ with parameters (θ,σ,η). Suppose that*

1) $\sigma\in\mathcal{C}^1$ has $L_{\sigma'}$-Lipschitz derivative σ' and, for some $0\leq\alpha<\beta<\infty$, σ is slope-restricted on $[\alpha,\beta]$;

2) $\eta\in\mathcal{C}^2$ on an open set containing $\overline{\mathcal{O}}$ and η is L_η-Lipschitz continuous on $\overline{\mathcal{O}}$;

3) there exists $(\rho,T)\in\mathbb{R}_{\geq 0}\times\mathcal{T}_p$, where $\mathcal{T}_p$ is defined in (18) with[9] $p:=\sum_{m\in[d]} q_m$, such that (19) holds.

[8] Notice that σ is Lipschitz with constant β, thus it is continuous. This implies, by function composition, that h in (21) is continuous.

[9] For each $m\in[d]$, $d\in\mathbb{N}\setminus\{0\}$, q_m denotes the number of neurons of the m-th layer, see Definition 3.2.

Then, there exist $\widehat{\alpha}_1, \widehat{\alpha}_2 \in \mathbb{R}_{\geq 0}$ such that $\nabla \widehat{V}$ is Lipschitz continuous on $\mathcal{O}$ with constant

$$0 \leq L_{\nabla \widehat{V}} \leq \widehat{\alpha}_1 \rho + \widehat{\alpha}_2 \sqrt{\rho}. \tag{24}$$

Proof. Let us define $\mathbb{R}^n \ni x \mapsto h(x) := h_{nn}(\eta(x)) \in \mathbb{R}^{q_d}$ with $h_{nn} := z^{(d)} \circ z^{(d-1)} \circ \cdots \circ z^{(1)}$ and, for each $m \in [d]$, the map $w \mapsto z^{(m)}(w)$ is given by (8). Notice that $x \mapsto \widehat{V}(x) = |h(x)|^2$, which, in turn, implies that $\nabla \widehat{V} : \mathbb{R}^n \to \mathbb{R}^n$ is given by

$$\nabla \widehat{V}(x) := 2 J_h(x)^\top h(x) = 2 J_h(x)^\top h(x) = 2 J_\eta(x)^\top J_{h_{nn}}(\eta(x))^\top h(x) \qquad \forall x \in \mathbb{R}^n$$

with $J_{h_{nn}} : \mathbb{R}^n \to \mathbb{R}^{q_d \times n}$ defined as

$$J_{h_{nn}}(x) := J_{z^{(d)}}\left(z^{(d-1)}\left(\cdots z^{(1)}(x) \cdots\right)\right) J_{z^{(d-1)}}\left(z^{(d-2)}(\cdots)\right) \cdots J_{z^{(1)}}(x) \qquad \forall x \in \mathbb{R}^n.$$

First, thanks to [30, Thm. 1], it follows that

$$|h_{nn}(x) - h_{nn}(x')| \leq \sqrt{\rho}\, |x - x'| \qquad \forall\, x, x' \in \mathbb{R}^n, \tag{25}$$

and thus

$$|J_{h_{nn}}(x)| \leq \sqrt{\rho} \qquad \forall x \in \mathbb{R}^n. \tag{26}$$

Also, from Lipschitz continuity of h_{nn}, together with the fact that $\sigma(s) = 0$ if and only if $s = 0$ (see Definition 3.2), it holds that

$$|h_{nn}(x)| \leq \sqrt{\rho}\, |x| \qquad \forall x \in \mathbb{R}^n. \tag{27}$$

Next, for each $m \in [d]$, notice that, for all $w, w' \in \mathbb{R}^{q_{m-1}}$,

$$\begin{aligned} |J_{z^{(m)}}(w) - J_{z^{(m)}}(w')| &\leq \sqrt{\sum_{i \in [q_m]} \left|\sigma'\left(w^\top \theta_i^{(m)}\right) - \sigma'\left(w'^\top \theta_i^{(m)}\right)\right|^2 \left|\theta_i^{(m)}\right|^2} \\ &\leq L_{\sigma'} \sqrt{\sum_{i \in [q_m]} \left|\theta_i^{(m)}\right|^4}\, |w - w'| \\ &\leq L_{\sigma'} \max_{i \in [q_m]} \left|\theta_i^{(m)}\right| \sqrt{\sum_{i \in [q_m]} \left|\theta_i^{(m)}\right|^2}\, |w - w'| \end{aligned}$$

which, by induction over $m \in [d]$, give us that $J_{h_{nn}}$ is Lipschitz with constant $L_{J_{h_{nn}}}$ satisfying

$$L_{J_{h_{nn}}} \leq L_{\sigma'} \beta^{d-1} \left(\prod_{m \in [d]} \sqrt{\sum_{i \in [q_m]} \left|\theta_i^{(m)}\right|^2}\right) \left(\sum_{m \in [d]} \beta^{m-1} \max_{i \in [q_m]} \left|\theta_i^{(m)}\right| \prod_{\ell \in [m-1]} \sqrt{\sum_{i \in [q_\ell]} \left|\theta_i^{(\ell)}\right|^2}\right). \tag{28}$$

In addition, since η is L_η-Lipschitz continuous on $\mathcal{O}$ then

$$|J_\eta(x)| \leq L_\eta \qquad \forall x \in \mathcal{O} \tag{29}$$

and, from $\eta \in \mathcal{C}^2$ on an open set containing $\overline{\mathcal{O}}$, note that

$$|J_\eta(x) - J_\eta(x')| \leq \left(\sup_{y \in \overline{\mathcal{O}}} \sqrt{\sum_{i \in [n]} \left|\nabla^2 \eta_i(y)\right|^2} \right) |x - x'| \qquad \forall x, x' \in \mathcal{O}. \tag{30}$$

Finally, for all $x, x' \in \mathcal{O}$, we obtain that

$$\begin{aligned}
&\left|\nabla \widehat{V}(x) - \nabla \widehat{V}(x')\right| \\
&\quad = 2\left|J_\eta(x)^\top J_{h_{nn}}(\eta(x))^\top h(x) - J_\eta(x')^\top J_{h_{nn}}(\eta(x'))^\top h(x')\right| \\
&\quad \leq 2\left|\left(J_\eta(x) - J_\eta(x')^\top\right) J_{h_{nn}}(\eta(x))^\top h(x)\right| + 2\left|J_\eta(x')^\top \left(J_{h_{nn}}(\eta(x)) - J_{h_{nn}}(\eta(x'))\right)^\top h(x)\right| \\
&\qquad + 2\left|J_\eta(x')^\top J_{h_{nn}}(\eta(x'))^\top \left(h(x) - h(x')\right)^\top\right| \\
&\quad \leq 2\rho \left(\sup_{y \in \overline{\mathcal{O}}} \sqrt{\sum_{i \in [n]} \left|\nabla^2 \eta_i(y)\right|^2} \right) \left(\sup_{y \in \overline{\mathcal{O}}} |\eta(y)| \right) |x - x'| \qquad \text{from (26), (27), and (30)} \\
&\qquad + 2\sqrt{\rho}\, L_\eta^2 L_{J_{h_{nn}}} \left(\sup_{y \in \overline{\mathcal{O}}} |\eta(y)| \right) |x - x'| \qquad \text{from (25), (27), (28), and (29)} \\
&\qquad + 2\rho L_\eta^2 |x - x'| \qquad \text{from (25), (26), and (29)}
\end{aligned}$$

and this completes the proof.

■

There are common activation functions σ used in deep learning that are Lipschitz on bounded sets, $\sigma(x) = 0$ only if $x = 0$, slope-restricted, and sufficiently smooth. In particular, consider

$$\sigma(z) = \overline{\alpha} \tanh \overline{\beta} z \qquad \overline{\alpha}, \overline{\beta} > 0 \qquad \text{Hyperbolic tangent} \tag{31}$$

$$\sigma(z) = \frac{z}{1 + |z|}, \qquad \text{Softsign activation} \tag{32}$$

$$\sigma(z) = \operatorname{arcsinh}(z) \qquad \text{Inverse hyperbolic sine} \tag{33}$$

$$\sigma(z) = \mathbb{1}_{\{z \geq 0\}} z + \mathbb{1}_{\{z < 0\}} (e^z - 1) \qquad \text{Exponential linear unit} \tag{34}$$

to name a few. Notice, however, that other typically used activations, such as the rectified linear unit (ReLU) and others related, do not have continuous derivatives or are zero even when their argument is different from zero, which prevents us from guaranteeing, for example, the Lipschitz continuity of $\dot{\widehat{V}}$ or positive definiteness of $\widehat{V}$ with respect to a compact set $\mathcal{A}$.

Finally, we will leverage these results to show that the functions defining the change of an LNN during flows and at jumps are Lipschitz continuous.

Proposition 3.8. (Lipschitz $\dot{\widehat{V}}$ and $\Delta \widehat{V}$) *Consider a hybrid system $\mathcal{H} = (C, F, D, G)$, a bounded set $\mathcal{O} \subset \mathbb{R}^n$, and an LNN $\widehat{V}$ with parameters (θ, σ, η). Suppose that F (resp., G)*

is osc and locally bounded relative to C (resp., D), that $\sigma \in \mathcal{C}^1$ and $\eta \in \mathcal{C}^2$ on an open set containing $\overline{\mathcal{O}}$, and define

$$\dot{\widehat{V}}(x) := \max_{f \in F(x)} \left\langle \nabla \widehat{V}(x), f \right\rangle \qquad \forall x \in C \cap \mathcal{O} \tag{35}$$

$$\Delta \widehat{V}(x) := \max_{g \in G(x)} \widehat{V}(g) - \widehat{V}(x) \qquad \forall x \in D \cap \mathcal{O}. \tag{36}$$

Furthermore, suppose that

1) *σ has $L_{\sigma'}$-Lipschitz derivative σ' and, for some $0 \le \alpha < \beta < \infty$, σ is slope-restricted on $[\alpha, \beta]$;*
2) *η is L_η-Lipschitz continuous on $\overline{\mathcal{O}}$;*
3) *there exists $(\rho, T) \in \mathbb{R}_{\ge 0} \times \mathcal{T}_p$, where $\mathcal{T}_p$ is defined in (18) with[10] $p := \sum_{m \in [d]} q_m$, such that (19) holds.*

The following hold:

1) *if C is closed and F is L_F-Lipschitz on $C \cap \mathcal{O}$, then $\dot{\widehat{V}}$ is Lipschitz on $C \cap \mathcal{O}$ with constant $L_{\dot{\widehat{V}}}$ satisfying*
$$0 \le L_{\dot{\widehat{V}}} \le \overline{\gamma}_1 \rho + \overline{\gamma}_2 \sqrt{\rho}, \tag{37}$$
for some $\overline{\gamma}_1, \overline{\gamma}_2 \in \mathbb{R}_{\ge 0}$;
2) *if D is closed, G is L_G-Lipschitz on $D \cap \mathcal{O}$, and η is L'_η-Lipschitz on $G(D \cap \mathcal{O})$, then $\Delta \widehat{V}$ is Lipschitz on $D \cap \mathcal{O}$ with constant $L_{\Delta \widehat{V}}$ satisfying*
$$0 \le L_{\Delta \widehat{V}} \le \overline{\gamma}_3 \rho, \tag{38}$$
for some $\overline{\gamma}_3 \in \mathbb{R}_{\ge 0}$.

Proof. Notice that, from the conditions stated, we have that $\widehat{V}$ is $L_{\widehat{V}}$-Lipschitz on $\mathcal{O}$ from Lemma 3.6. To prove the rest of the results, we will use the following claim.

Claim: Let a set-valued map $S : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be locally bounded relative to a closed set $X \subset \mathbb{R}^n$ and consider a bounded set $\mathcal{O} \subset \mathbb{R}^n$. Then, there exists $\varrho > 0$ such that $|s| < \varrho$ for all $s \in S(x)$ and all $x \in X \cap \mathcal{O}$.

[10] For each $m \in [d]$, $d \in \mathbb{N} \setminus \{0\}$, q_m denotes the number of neurons of the m-th layer, see Definition 3.2.

Proof: From local boundedness of S relative to X, for every $x \in X \cap \overline{\mathcal{O}}$, there exists an open neighborhood V_x of x and $\varrho_x > 0$ such that $S(V_x \cap X) \subset \varrho_x \mathbb{B}$. Thus, it follows that

$$\bigcup_{x \in X \cap \overline{\mathcal{O}}} S(V_x \cap X) \subset \bigcup_{x \in X \cap \overline{\mathcal{O}}} \varrho_x \mathbb{B}.$$

In addition, notice that the set $X \cap \overline{\mathcal{O}}$ is compact and, as a consequence, we have that

$$X \cap \overline{\mathcal{O}} \subset \bigcup_{j \in [p]} V_{x_j}$$

for some $p \in \mathbb{N} \setminus \{0\}$ and $\{x_j\}_{j \in [p]} \subset X \cap \overline{\mathcal{O}}$. Then,

$$S\left(X \cap \overline{\mathcal{O}}\right) := \bigcup_{x \in X \cap \overline{\mathcal{O}}} S(x) \subset \bigcup_{j \in [p]} S(V_{x_j} \cap X) \subset \bigcup_{j \in [p]} \varrho_{x_j} \mathbb{B} \subset \max_{j \in [p]} \varrho_{x_j} \mathbb{B}$$

and it is sufficient to pick $\varrho := \max_{j \in [p]} \varrho_{x_j}$. ■

Proof of Proposition 3.8 (Continued): Notice that $D \cap \mathcal{O}$ is bounded, then from the claim above, it follows that $G(D \cap \mathcal{O})$ is also bounded, as D is closed and G is locally bounded relative to D. Thus, using the assumptions in the statement of this result, it follows that $\widehat{V}$ is also Lipschitz continuous on $G(D \cap \mathcal{O})$ from Lemma 3.6 with constant satisfying

$$0 \leq L'_{\widehat{V}} \leq 2\rho L'_{\eta} \sup_{y \in \overline{G(D \cap \mathcal{O})}} |\eta(y)|.$$

In addition, since $\sigma \in \mathcal{C}^1$ and σ' is $L_{\sigma'}$-Lipschitz, and $\eta \in \mathcal{C}^2$ on $\overline{\mathcal{O}}$, we have that $\nabla \widehat{V}$ is $L_{\nabla \widehat{V}}$-Lipschitz on $\mathcal{O}$ thanks to Lemma 3.7.

To prove that $\dot{\widehat{V}}$ is Lipschitz, notice that C is closed and F is locally bounded relative to C and $C \cap \mathcal{O}$ is bounded. Then, from the claim above, there exists $\varrho > 0$ such that $F(C \cap \mathcal{O}) \subset \varrho \mathbb{B}$. Next, pick any $x, x' \in C \cap \mathcal{O}$ and notice that

$$\begin{aligned}
\left|\dot{\widehat{V}}(x) - \dot{\widehat{V}}(x')\right| &= \left| \max_{f \in F(x)} \langle \nabla \widehat{V}(x), f \rangle - \max_{f' \in F(x')} \langle \nabla \widehat{V}(x'), f' \rangle \right| \\
&\leq \left| \max_{f \in F(x)} \langle \nabla \widehat{V}(x) - \nabla \widehat{V}(x'), f \rangle \right| + \left| \max_{f \in F(x)} \langle \nabla \widehat{V}(x'), f \rangle - \max_{f' \in F(x')} \langle \nabla \widehat{V}(x'), f' \rangle \right| \\
&\leq \left(\max_{f \in F(x)} |f| \right) \left| \nabla \widehat{V}(x) - \nabla \widehat{V}(x') \right| + \max_{x \in \overline{C \cap \mathcal{O}}} |\nabla \widehat{V}(x)|\, d_H\left(F(x), F(x')\right) \\
&\leq \left(\varrho L_{\nabla \widehat{V}} + L_F \max_{x \in \overline{C \cap \mathcal{O}}} |\nabla \widehat{V}(x)| \right) |x - x'|
\end{aligned}$$

where the second to last inequality comes from Lemma A.1. Now, to show that $\Delta \widehat{V}$ is Lipschitz continuous, pick any $y, y' \in D \cap \mathcal{O}$ and observe that

$$\left|\Delta \widehat{V}(y) - \Delta \widehat{V}(y')\right| = \left| \max_{g \in G(y)} \widehat{V}(g) - \widehat{V}(y) - \max_{g' \in G(y')} \widehat{V}(g') + \widehat{V}(y') \right|$$

$$\leq \left| \max_{g \in G(y)} \widehat{V}(g) - \max_{g' \in G(y')} \widehat{V}(g') \right| + L_{\widehat{V}} |y - y'|.$$

where $L_{\widehat{V}}$ satisfies (20). Using the fact that $\widehat{V}$ is $L'_{\widehat{V}}$-Lipschitz on $G\,(D \cap \mathcal{O})$, together with the symmetry of the Pompeiu-Hausdorff distance d_H, gives

$$\begin{aligned} \left|\Delta\widehat{V}(y) - \Delta\widehat{V}(y')\right| &\leq L'_{\widehat{V}} d_H\left(G(y), G(y')\right) + L_{\widehat{V}}|x - x'| \\ &\leq \left(L'_{\widehat{V}} L_G + L_{\widehat{V}}\right)|x - x'|. \end{aligned}$$

■

We are ready to present the main result of this section, which shows that, using the regularity properties of an LNN $\widehat{V}$ and ε-nets, we can extend sufficient Lyapunov conditions enforced at points in $\mathcal{M}_\star$, $\star \in \{C, D\}$, to $(C \cup D) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ)$, for some $\mu > \varepsilon$.

Proposition 3.9. (Generalized Lyapunov conditions) *Consider a compact set $\mathcal{A} \subset \mathbb{R}^n$, a bounded set $\mathcal{O} \supset \mathcal{A}$, a hybrid system $\mathcal{H} = (C, F, D, G)$ with $F(x)$ (resp., $G(x)$) compact for each $x \in C$ (resp., $x \in D$). Given $0 < \varepsilon < \mu$, let $\mathcal{M}_\star$ be an ε-net of $(\star \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ)$ for each $\star \in \{C, D\}$. Furthermore, let $\widehat{V}$ be an LNN with parameters (θ, σ, η) and suppose that $\dot{\widehat{V}} : C \to \mathbb{R}$ is $L_{\dot{\widehat{V}}}$-Lipschitz on $C \cap \mathcal{O}$ and that $\Delta\widehat{V} : D \to \mathbb{R}$ is $L_{\Delta\widehat{V}}$-Lipschitz on $D \cap \mathcal{O}$. The following holds:*

- S1) *if there exists $\tau_C > L_{\dot{\widehat{V}}}\varepsilon$ such that $\dot{\widehat{V}}(x') \leq -\tau_C$ for all $x' \in \mathcal{M}_C$, then $\dot{\widehat{V}}(x) < 0$ for all $x \in (C \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ)$;*
- S2) *if there exists $\tau_D > L_{\Delta\widehat{V}}\varepsilon$ such that $\Delta\widehat{V}(x') \leq -\tau_D$ for all $x' \in \mathcal{M}_D$, then $\Delta\widehat{V}(x) < 0$ for all $x \in (D \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ)$.*

Proof. Given that $\mathcal{M}_C$ is an ε-net of $(C \cap O) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ)$, we have that, for each $x \in (C \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ)$, there exists $x' \in \mathcal{M}_C$ such that

$$|\dot{\widehat{V}}(x) - \dot{\widehat{V}}(x')| \leq L_{\dot{\widehat{V}}}|x - x'| \leq L_{\dot{\widehat{V}}}\varepsilon$$

which implies

$$\dot{\widehat{V}}(x) \leq \dot{\widehat{V}}(x') + L_{\dot{\widehat{V}}}\varepsilon \leq -\tau_C + L_{\dot{\widehat{V}}}\varepsilon.$$

Thus, to show that $\dot{\widehat{V}}(x) < 0$, it is sufficient to pick $\tau_C = L_{\dot{\widehat{V}}}\varepsilon + \delta$ for any $\delta > 0$. Proving the case at jumps follows the same steps; it is therefore omitted for brevity. ■

Proposition 3.9 extends the Lyapunov conditions by imposing appropriate regularity requirements on $\dot{\widehat{V}}$ and $\Delta\widehat{V}$, for which Proposition 3.8 provides sufficient criteria. In addition, notice that Proposition 3.9 implies that, as the chosen μ decreases, the number of closed balls needed to cover $(C \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ)$ and $(D \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ)$ increases and, as a result, the right-hand side in conditions $\tau_C > L_{\dot{\widehat{V}}}\varepsilon$ and $\tau_D > L_{\Delta\widehat{V}}\varepsilon$ become smaller. In addition, notice that conditions S1) and S2) impose additional constraints on the training of an LNN such that LpAS of $\mathcal{A}$ for $\mathcal{H}$ can be certified.

Remark 3.10. (Training LNNs with Lipschitz continuity constraints) *Given an LNN $\widehat{V}$ with parameters (θ, σ, η), observe that, although (23) provides a tight estimate of the Lipschitz constant of $\widehat{V}$, it is a post-hoc certification tool. Following [30], we instead train $\widehat{V}$ directly regularizing its Lipschitz constant. In particular, for any $\varepsilon > 0$, the conditions S1) and S2) in Proposition 3.9 yield an upper bound on $\rho \geq 0$ that defines the Lipschitz constant of $\widehat{V}$. More precisely, from (37), we have that there exist $\overline{\gamma}_1, \overline{\gamma}_2 \in \mathbb{R}_{\geq 0}$ such that $0 \leq L_{\dot{\widehat{V}}} \leq \overline{\gamma}_1\rho + \overline{\gamma}_2\sqrt{\rho}$, and, from (38), there exists $\overline{\gamma}_3 \in \mathbb{R}_{\geq 0}$ such that $0 \leq L_{\Delta\widehat{V}} \leq \overline{\gamma}_3\rho$. Using the conditions in Proposition 3.9, it is therefore sufficient to find $\tau_C, \tau_D \in \mathbb{R}_{>0}$ such that $(\overline{\gamma}_1\rho + \overline{\gamma}_2\sqrt{\rho})\varepsilon < \tau_C$ and $\overline{\gamma}_3\rho\varepsilon < \tau_D$, or, equivalently,*

$$0 \leq \rho \leq \min\left\{ \left(\frac{-\overline{\gamma}_2 + \sqrt{\overline{\gamma}_2^2 + 4\overline{\gamma}_1\tau_C/\varepsilon}}{2\overline{\gamma}_1} \right)^2, \frac{\tau_D}{\overline{\gamma}_3\varepsilon} \right\}. \tag{39}$$

Consequently, to meet the sufficient conditions in Proposition 3.9, we augment (16) by incorporating τ_C and τ_D as decision variables and enforcing (39), consistent with [30, Rmk. 4]. This constraint is imposed jointly with the conditions ensuring decrease of $\widehat{V}$ along flows and at jumps in (16), and these are solved using a multi-block ADMM scheme[11].

Following Remark 3.10, notice that including τ_C and τ_D as decision variables in (16) creates a structural tension between the need for the LNN to be sufficiently "steep" to satisfy decrease along flows and at jumps, and sufficiently "smooth" to satisfy the generalization condition associated with the sampling density ε. For instance, infeasibility may arise when the sampling is coarse (large ε) and the required decrease of the LNN is hard to satisfy without steep gradients, which can happen if the flow or jump map has a large Lipschitz constant. This may violate the smoothness constraints enforced by ρ in (39). A systematic feasibility analysis of (16), together with Remark 3.10, is beyond the scope of this paper, but we recognize it is promising direction for future work.

Furthermore, the smoothness constraints on the LNN relate directly to the choice of activation functions. The approach in [30] targets activation functions that are slope restricted on the interval $[0, \beta]$ for some $\beta > 0$. As acknowledged therein, this is a conservative choice for certain activations, yet many common functions satisfy this slope restriction, including those in (31)–(34).

[11] We refer the reader to [32] for an in-depth study on convergence of the ADMM method.

Given a hybrid system $\mathcal{H} = (C, F, D, G)$, a compact set $\mathcal{A} \subset \mathbb{R}^n$, and a bounded set $\mathcal{O} \supset \mathcal{A}$, in this section, we show that under suitable assumptions, the solution to (16), satisfying the conditions in Proposition 3.9, allows us to use an LNN $\widehat{V}$ to guarantee *practical* pre-asymptotic stability of $\mathcal{A}$ for $\mathcal{H}$ on $\mathcal{O}$. This is shown in the next result.

Theorem 3.11. (Practical LpAS) *Given the sets $\mathcal{O}, \mathcal{A} \subset \mathbb{R}^n$ and the hybrid system $\mathcal{H} = (C, F, D, G)$, let V be a Lyapunov function candidate for $\mathcal{H}$ with respect to $\mathcal{A}$ on $\mathcal{O}$. Suppose that $\mathcal{A}$ is closed and define $\mathcal{X} := (C \cup D) \cap \mathcal{O}$. Pick $r > 0$ such that $\mathcal{A} + r\mathbb{B} \subset \mathcal{X}$ and $0 < \mu < \alpha_2^{-1}(\alpha_1(r))$, where $\alpha_1, \alpha_2 \in \mathcal{K}$ are such that (6) holds. If $G(D) \subset \mathcal{X}$ and*

$$\sup_{\substack{x \in C \cap \mathcal{O} \\ |x|_{\mathcal{A}} \geq \mu}} \dot{V}(x) < 0 \quad \text{and} \quad \sup_{\substack{x \in D \cap \mathcal{O} \\ |x|_{\mathcal{A}} \geq \mu}} \Delta V(x) < 0, \tag{40}$$

then there exist $\beta \in \mathcal{KL}$ and $T \geq 0$ such that, for every $\phi \in \mathcal{S}_{\mathcal{H}}\big(\mathcal{A} + \alpha_2^{-1}(\alpha_1(r))\mathbb{B}\big)$,

$$\begin{aligned} |\phi(t,j)|_{\mathcal{A}} &\leq \beta(|\phi(0,0)|_{\mathcal{A}}, t+j) && \forall (t,j) \in \operatorname{dom}\phi : t+j \leq T \\ |\phi(t,j)|_{\mathcal{A}} &\leq \alpha_1^{-1}(\alpha_2(\mu)) && \forall (t,j) \in \operatorname{dom}\phi : t+j > T. \end{aligned}$$

Proof. Pick $r > 0$ such that $\mathcal{A} + r\mathbb{B} \subset \mathcal{X}$ and $0 < \mu < \alpha_2^{-1}(\alpha_1(r))$, and define

$$\chi_C := -\sup_{\substack{x \in C \cap \mathcal{O} \\ |x|_{\mathcal{A}} \geq \mu}} \dot{V}(x) \quad \text{and} \quad \chi_D := -\sup_{\substack{x \in D \cap \mathcal{O} \\ |x|_{\mathcal{A}} \geq \mu}} \Delta V(x).$$

For each $\star \in \{C, D\}$, define

$$\tilde{\alpha}_\star(s) := \chi_\star \min\left\{1, \frac{s}{\mu}\right\} \qquad \forall s \geq 0$$

and notice that

$$\alpha_\star(s) := \chi_\star\left(1 - \exp\left(-\frac{s}{\mu}\right)\right) \leq \tilde{\alpha}_\star(s) \qquad \forall s \geq 0, \tag{41}$$

where $\alpha_\star \in \mathcal{K}$. Using (40) and (41) we have that

$$\begin{aligned} \dot{V}(x) &\leq -\alpha_C(|x|_{\mathcal{A}}) && \forall x \in (C \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ) \\ \Delta V(x) &\leq -\alpha_D(|x|_{\mathcal{A}}) && \forall x \in (D \cap \mathcal{O}) \setminus \mathcal{A} + \mu\mathbb{B}^\circ). \end{aligned} \tag{42}$$

Next, pick $\xi \in \mathcal{A} + \alpha_2^{-1}(\alpha_1(r))\mathbb{B}$ and let $\varsigma := \alpha_1(r)$. Then $\alpha_2(\mu) < \varsigma$ and $\alpha_2(|\xi|_{\mathcal{A}}) \leq \varsigma$. Let $\vartheta := \alpha_2(\mu)$ and define the sets $\Omega_\vartheta := \{x \in \mathcal{X} : V(x) \leq \vartheta\}$ and $\Omega_\varsigma := \{x \in \mathcal{X} : V(x) \leq \varsigma\}$. Then

$$\mathcal{A} + \mu\mathbb{B} \subset \Omega_\vartheta \subset \{x \in \mathcal{X} : \alpha_1(|x|_{\mathcal{A}}) \leq \vartheta\} \subset \{x \in \mathcal{X} : \alpha_1(|x|_{\mathcal{A}}) \leq \varsigma\} = \mathcal{A} + r\mathbb{B} \subset \mathcal{X}$$

and

$$\Omega_\vartheta \subset \Omega_\varsigma \subset \mathcal{A} + r\mathbb{B} \subset \mathcal{X}.$$

Claim: Let $\phi \in \mathcal{S}_{\mathcal{H}}(\xi)$ with $\xi \in \mathbb{R}^n$, and define $\Omega_c := \{x \in \mathcal{X} : V(x) \leq c\}$ for some $\alpha_2(\mu) \leq c \leq \alpha_1(r)$. If there exists some $(t, j) \in \operatorname{dom}\phi$ such that $\phi(t, j) \in \Omega_c$, then $\phi(t', j') \in \Omega_c$ for all $(t', j') \in \operatorname{dom}\phi$ such that $t' + j' \geq t + j$.

Proof: The proof follows similar ideas from [33, Prop. 4.1] and [34, Thm. 1].

Proceeding by contradiction, suppose that ϕ leaves the set Ω_c. The following cases are possible:

- *The solution leaves Ω_c via a jump.* Using $G(D) \subset \mathcal{X}$, we have that there exist $t \in \mathbb{R}_{\geq 0}$ and $j \in \mathbb{N}$ satisfying

$$\begin{aligned} (t, j) &\in \operatorname{dom}\phi, & \phi(t, j) &\in D \cap \Omega_c && \text{(43a)} \\ (t, j+1) &\in \operatorname{dom}\phi, & \phi(t, j+1) &\in \mathcal{X} \setminus \Omega_c. && \text{(43b)} \end{aligned}$$

  From $x \in \mathcal{X} \setminus \Omega_c$, we have $V(x) > c$, and (43) implies

$$V(\phi(t, j)) \leq c < V(\phi(t, j+1)).$$

  However, from (42), we have that

$$V(\phi(t, j+1)) \leq V(\phi(t, j)) - \alpha_D(|\phi(t, j)|_{\mathcal{A}}) \leq V(\phi(t, j)),$$

  which is a contradiction.

- *The solution leaves Ω_c by flowing.* Then, there exist $t_1, t_2 \in \mathbb{R}_{\geq 0}$ with $t_1 < t_2$, and $j \in \mathbb{N}$ satisfying

$$\begin{aligned} (t_1, j) &\in \operatorname{dom}\phi, & \phi(t_1, j) &\in C \cap \partial\Omega_c && \text{(44a)} \\ (t_2, j) &\in \operatorname{dom}\phi, & \phi(t_2, j) &\in C \setminus \Omega_c. && \text{(44b)} \end{aligned}$$

  From $\phi(t_1, j) \in \partial\Omega_c$, we have $V(\phi(t_1, j)) = c$, and (42) implies

$$V(\phi(t_2, j)) - c = \int_{t_1}^{t_2} \left\langle \nabla V(\phi(\tau, j)), \dot{\phi}(\tau, j) \right\rangle d\tau \leq -\int_{t_1}^{t_2} \alpha_C\left(|\phi(\tau, j)|_{\mathcal{A}}\right) < 0,$$

  which contradicts (44b).

Thus, we conclude that ϕ cannot leave Ω_c. ∎

Proof of Theorem 3.11 (Continued): Notice that, from the claim above, the sets Ω_ϑ and Ω_ς are forward pre-invariant for $\mathcal{H}$. In addition, observe that the following holds:

$$\begin{aligned} \dot{V}(x) &\leq -\alpha_C(\mu) < 0 \qquad \forall x \in (\Omega_\varsigma \cap C) \setminus \Omega_\vartheta \\ \Delta V(x) &\leq -\alpha_D(\mu) < 0 \qquad \forall x \in (\Omega_\varsigma \cap D) \setminus \Omega_\vartheta. \end{aligned}$$

The following cases are considered:

1) Let $\mu < |\xi|_{\mathcal{A}}$ and pick $\phi \in \mathcal{S}_{\mathcal{H}}(\xi)$. Let $t(j)$ denote least time $t \in \mathbb{R}_{\geq 0}$ such that $(t, j) \in \operatorname{dom}\phi$ and $j(t)$ denote the least index $j \in \mathbb{N}$ such that $(t, j) \in \operatorname{dom}\phi$, and suppose that $(\tau, J) \in \operatorname{dom}\phi$. Then

$$\begin{aligned} V(\phi(\tau, J)) - V(\phi(0,0)) &= \int_0^\tau \frac{dV}{dt}(\phi(t, j(t)))\, dt \\ &\qquad\qquad + \sum_{j=1}^{J} \left[V\big(\phi(t(j), j)\big) - V\big(\phi(t(j), j-1)\big)\right] \\ &\leq -\min\{\alpha_C(\mu), \alpha_D(\mu)\}(\tau + J) \end{aligned}$$

which, in turn, implies that

$$\begin{aligned} V(\phi(\tau, J)) &\leq V(\phi(0,0)) - \min\{\alpha_C(\mu), \alpha_D(\mu)\}(\tau + J) \\ &\leq \varsigma - \min\{\alpha_C(\mu), \alpha_D(\mu)\}(\tau + J). \end{aligned}$$

Thus, we have that

$$\phi(t', j') \in \Omega_\vartheta \quad \forall (t', j') \in \operatorname{dom}\phi \quad \text{such that} \quad t' + j' \geq \frac{\varsigma - \vartheta}{\min\{\alpha_C(\mu), \alpha_D(\mu)\}} =: T.$$

For ease of notation, let $s \mapsto \alpha(s) := \min\{\alpha_C(s), \alpha_D(s)\} \in \mathcal{K}$. For all $(t, j) \in \operatorname{dom}\phi$ such that $t + j \leq T$

- if I_ϕ^j has a nonempty interior $\operatorname{int} I_\phi^j$, then V satisfies

$$\frac{dV}{dt}(\phi(s, j)) \leq -\alpha\left(V(\phi(s, j))\right) \qquad \text{for almost all } s \in \operatorname{int} I_\phi^j;$$

- if $(t, j+1) \in \operatorname{dom}\phi$, then V satisfies

$$V\left(\phi(t, j+1)\right) - V\left(\phi(t, j)\right) \leq -\alpha\left(V\left(\phi(t, j)\right)\right).$$

Thus, from Lemma A.2, we have that there exists $\beta \in \mathcal{KL}$ such that

$$V(\phi(t, j)) \leq \beta\left(V(\phi(0,0)), t + j\right) \qquad \forall\, (t, j) \in \operatorname{dom}\phi : t + j \leq T.$$

This implies that, for all $(t,j) \in \operatorname{dom}\phi$ such that $t+j \leq T$,

$$V(\phi(t,j)) \leq \beta\left(V(\phi(0,0)), t+j\right) \leq \beta\left(\alpha_2\left(|\phi(0,0)|_{\mathcal{A}}\right), t+j\right)$$

and, consequently,

$$|\phi(t,j)|_{\mathcal{A}} \leq \alpha_1^{-1}\left(\beta\left(\alpha_2\left(|\phi(0,0)|_{\mathcal{A}}\right), t+j\right)\right)$$

where $(p,s) \mapsto \alpha_1^{-1}\left(\beta\left(\alpha_2(p), s\right)\right) \in \mathcal{KL}$. Finally, for all $(t',j') \in \operatorname{dom}\phi$ such that $t'+j' \geq T$, notice that

$$\phi(t',j') \in \Omega_\vartheta \quad \implies \quad |\phi(t',j')|_{\mathcal{A}} \leq \alpha_1^{-1}(\vartheta) = \alpha_1^{-1}(\alpha_2(\mu)).$$

2) Let $|\xi|_{\mathcal{A}} \leq \mu$ and pick $\phi \in \mathcal{S}_{\mathcal{H}}(\xi)$. Then, from the previous analysis, $T=0$ and we have that

$$\phi(t,j) \in \Omega_\vartheta \quad \implies \quad |\phi(t,j)|_{\mathcal{A}} \leq \alpha_1^{-1}(\vartheta) = \alpha_1^{-1}(\alpha_2(\mu)) \qquad \forall (t,j) \in \operatorname{dom}\phi.$$

■

Notice that if the sets $\mathcal{O}$ and $\mathcal{A}$, and an LNN $\widehat{V}$ with parameters (θ, σ, η) together satisfy Assumption 3.4, then $\widehat{V}$ is a Lyapunov function candidate on $\mathcal{O}$ with respect to $\mathcal{A}$ for $\mathcal{H}$. Proposition 3.8 gives sufficient conditions such that $\dot{\widehat{V}}$ is Lipschitz on $C \cap \mathcal{O}$ and $\Delta\widehat{V}$ is Lipschitz on $D \cap \mathcal{O}$. This is then leveraged in Proposition 3.9, together with additional sufficient conditions, to show that

$$\begin{aligned} \dot{\widehat{V}}(x) &< 0 \qquad \forall x \in (C \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ) \\ \Delta\widehat{V}(x) &< 0 \qquad \forall x \in (D \cap \mathcal{O}) \setminus (\mathcal{A} + \mu\mathbb{B}^\circ) \end{aligned} \tag{45}$$

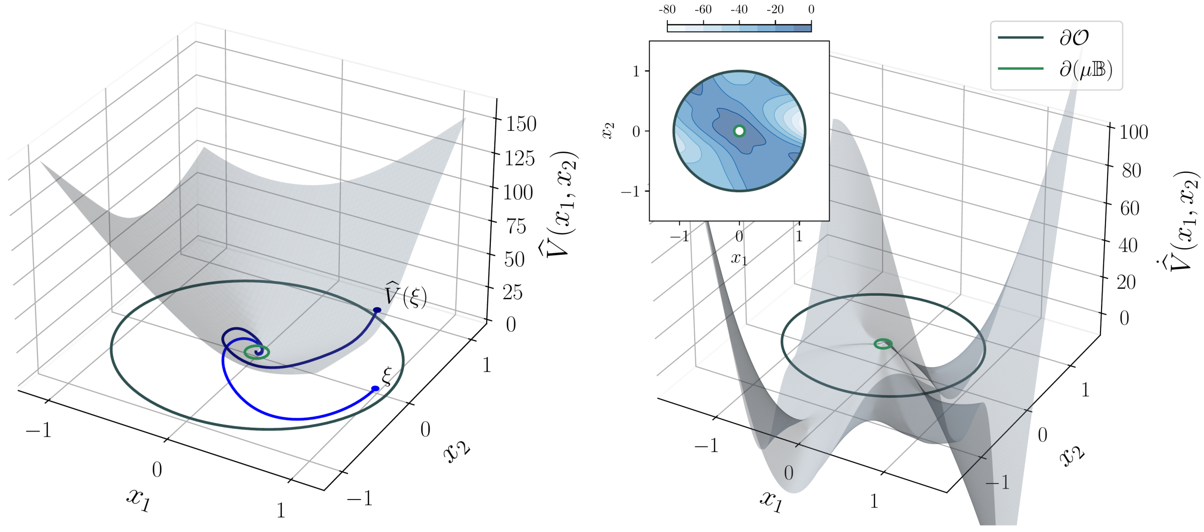
where $\mathcal{O}$ is assumed to be bounded and $\mu > 0$. If, in addition, C, D, and $\mathcal{O}$ are closed, then (45) implies (40) and $\widehat{V}$ can be used to certify that $\mathcal{A}$ is practically LpAS for $\mathcal{H}$ on $\mathcal{O}$.

Remark 3.12. (Connection to the literature) *Given the sets $\mathcal{O}, \mathcal{A} \subset \mathbb{R}^n$ and the hybrid systems $\mathcal{H}$, let V be a Lyapunov function candidate for $\mathcal{H}$ with respect to $\mathcal{A}$ on $\mathcal{O}$. Suppose that the assumptions in Theorem 3.11 are satisfied. Then, notice that given constants $\epsilon, \gamma > 0$ such that, for every $\delta \in (0,\gamma)$, there exists $T \geq 0$ such that for every $\phi \in \mathcal{S}_{\mathcal{H}}(\mathcal{A}+\delta\mathbb{B})$ it follows that*[12]

$$|\phi(t,j)|_{\mathcal{A}} \leq \epsilon \qquad \forall (t,j) \in \operatorname{dom}\phi : t+j \geq T.$$

This notion in the literature is known as ultimate boundedness *of solutions to $\mathcal{H}$. In particular, this introduced in [5, Def. 4.6] and in [35, Def. 4.4] for the case when $\operatorname{dom}\phi \subset \mathbb{R}_{\geq 0} \times \{0\}$, and sufficient conditions are given in [5, Thm. 4.18] and in [35, Cor. 4.4]. When $\operatorname{dom}\phi \subset \{0\} \times \mathbb{N}$, ultimate boundedness is defined in [35, Def. 13.9] and sufficient conditions can be found in [35, Cor. 13.7].*

[12]From Theorem 3.11, notice that $\epsilon = \alpha_1^{-1}(\alpha_2(\mu))$ and $\gamma = \alpha_2^{-1}(\alpha_1(r))$.



(a) (b)

Figure 1: We show the surface plots of the learned LNN $\widehat{V}$ and its time derivative $\dot{\widehat{V}}(x)$ for $\mathcal{H}_C$ as in (46) with respect to $\mathcal{A} = \{(0,0)\}$. In (a) we show that $\widehat{V}((0,0)) = 0$ and $\widehat{V}(\mathbb{R}^2 \setminus \{(0,0)\}) \subset \mathbb{R}_{>0}$. A solution $\phi \in \mathcal{S}_{\mathcal{H}_C}(\mathcal{O})$, with $\mathcal{O}$ as in (47), is also shown. Notice that $\operatorname{dom}\phi \subset \mathbb{R}_{\geq 0} \times \{0\}$. In (b) we show that $\dot{\widehat{V}}(x) < 0$ for all $x \in \mathcal{O} \setminus \mu\mathbb{B}$, as guaranteed by Proposition 3.9.

3.4. *Numerical Examples*

In this section, we present three different examples to show practical LpAS of a set $\mathcal{A}$ for different types of dynamical systems.

3.4.1. *Continuous-time Dynamical Systems*

As a special case of a hybrid system $\mathcal{H}$ as in (3), consider the system $\mathcal{H}_C = (\mathbb{R}^2, F, \emptyset, \star)$, where $\star$ indicates an arbitrary map, given by

$$\mathcal{H}_C : \dot{x} \in F(x) := \left\{ \begin{bmatrix} x_2 \\ -x_1 + \gamma_C(1 - x_1^2)x_2 \end{bmatrix} \right\} \qquad x \in \mathbb{R}^2 \tag{46}$$

with $\gamma_C < 0$. Let $\mathcal{A} := \{(0,0)\}$, and consider the sampling set $\mathcal{O}$ defined by

$$\mathcal{O} := \left\{ x \in \mathbb{R}^2 : x^\top \begin{pmatrix} |\gamma_C| & 0 \\ 0 & 1 \end{pmatrix} x \leq 1 \right\} \tag{47}$$

Notice that F is Lipschitz continuous on $\mathcal{O}$ with constant satisfying

$$0 \leq L_F \leq \max_{x \in \mathcal{O}} \left| \begin{pmatrix} 0 & 1 \\ -(1 + 2\gamma_C x_1 x_2) & \gamma_C(1 - x_1^2) \end{pmatrix} \right|_F \leq \sqrt{1 + \left(1 + \sqrt{|\gamma_C|}\right)^2 + (|\gamma_C| + 1)^2}$$

and, in addition, since F is continuous, then it is osc and locally bounded [21, Cor. 5.20], viewed as a special case of a set-valued mapping. Thus, $F(x)$ is compact[13] for each $x \in \mathbb{R}^2$. Next, consider an LNN $\widehat{V}$ with $x \mapsto \eta(x) := x$, $d = 2$, layers' width $(q_1, q_2) = (16, 32)$, and activation function $z \mapsto \sigma(z) := \mathbb{1}_{\{z \geq 0\}} z + \mathbb{1}_{\{z<0\}}(e^z - 1)$. Notice that:

- σ is continuously differentiable, slope restricted on the interval $[0, 1]$, and $\sigma(z) = 0 \iff z = 0$;
- σ' is 1-Lipschitz continuous;
- η is smooth and 1-Lipschitz continuous on $\mathbb{R}^2$.

For training, the samples are chosen to form an ε-net of $\mathcal{O} \setminus \mu\mathbb{B}^\circ$ with $\varepsilon = 0.08$ and $\mu = 1.1\varepsilon$, and we solve (16) following Remark 3.10. As a result, we have that $\tau_C = 8.4$ and the SDP condition in (19) holds with $\rho = 0.803$ and some $T \in \mathcal{T}_{48}$. Then, by Proposition 3.8, we have that $\dot{\widehat{V}}$ is Lipschitz on $\mathcal{O}$ with constant $0 \leq L_{\dot{\hat{V}}} \leq 79.08$. Leveraging the regularity conditions of $\widehat{V}$ and properties of the ε-net of $\mathcal{O} \setminus \mu\mathbb{B}$, notice that $\tau_C > L_{\dot{\hat{V}}}\varepsilon$, and following Proposition 3.9, it follows that $\dot{\widehat{V}}(x) < 0$ for all $x \in \mathcal{O} \setminus \mu\mathbb{B}$. This is illustrated in Figure 1. Since $\mathcal{O}$ is closed, we have that (40) is satisfied and, thanks to Theorem 3.11, we conclude that $\mathcal{A} = \{(0, 0)\}$ is practically LpAS for $\mathcal{H}_C$.

3.4.2. Discrete-time Dynamical Systems

Similarly, consider the system $\mathcal{H}_D = (\emptyset, \star, \mathbb{R}^2, G)$, where $\star$ indicates an arbitrary map, with data given by

$$\mathcal{H}_D : x^+ = G(x) := \left\{ \begin{bmatrix} x_1 + T x_2 \\ \underline{\varrho} T \sin x_1 + x_2(1 - \bar{\varrho} T) \end{bmatrix} \right\} \qquad x \in \mathbb{R}^2 \tag{48}$$

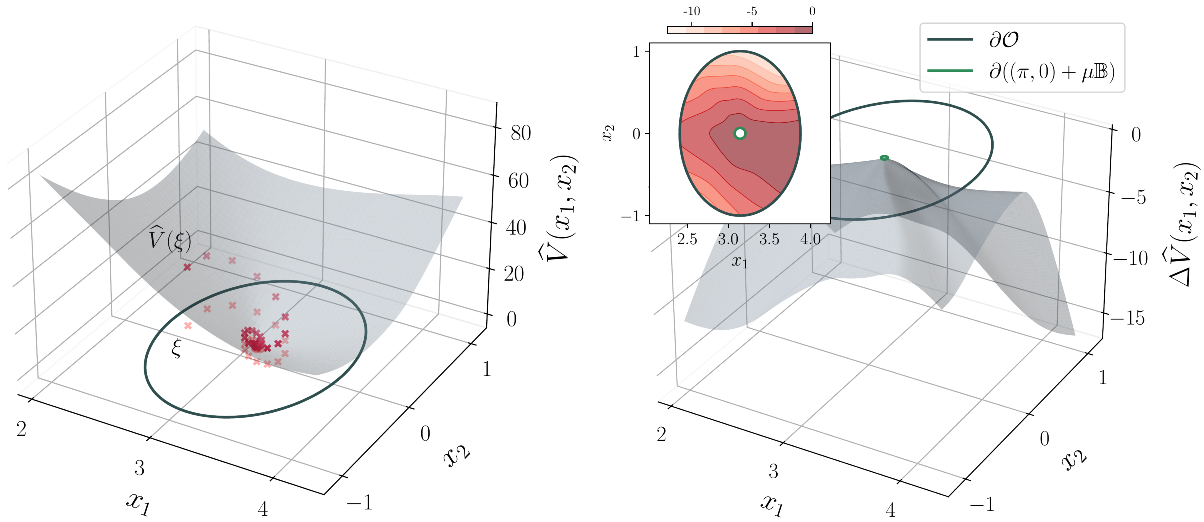
where $\underline{\varrho} > 0$, $0 < T < \sqrt{\frac{2}{\underline{\varrho}}}$, and $\underline{\varrho} T < \bar{\varrho} < \frac{2}{T}$. In particular, pick $(\underline{\varrho}, \bar{\varrho}, T) = (1, 1, 0.5)$. Let $\mathcal{A} := \{(\pi, 0)\}$, and consider the following sampling set

$$\mathcal{O} := \left\{ x \in \mathbb{R}^2 : \frac{\underline{\varrho}}{6\underline{\varrho}}(x_1 - \pi)^2 + x_2^2 \leq 1 \right\} \tag{49}$$

Notice that G is Lipschitz continuous on $\mathcal{O}$ with constant satisfying

$$0 \leq L_G \leq \max_{x \in \mathcal{O}} \left\| \begin{bmatrix} 1 & T \\ \underline{\varrho} T \cos x_1 & 1 - \bar{\varrho} T \end{bmatrix} \right\|_F \leq \sqrt{2 + T^2(1 + \underline{\varrho}^2 + \bar{\varrho}^2) - 2\bar{\varrho} T}$$

[13] Notice that this is straightforward to conclude from the fact that $F(x)$ is a singleton for each $x \in \mathbb{R}^2$.



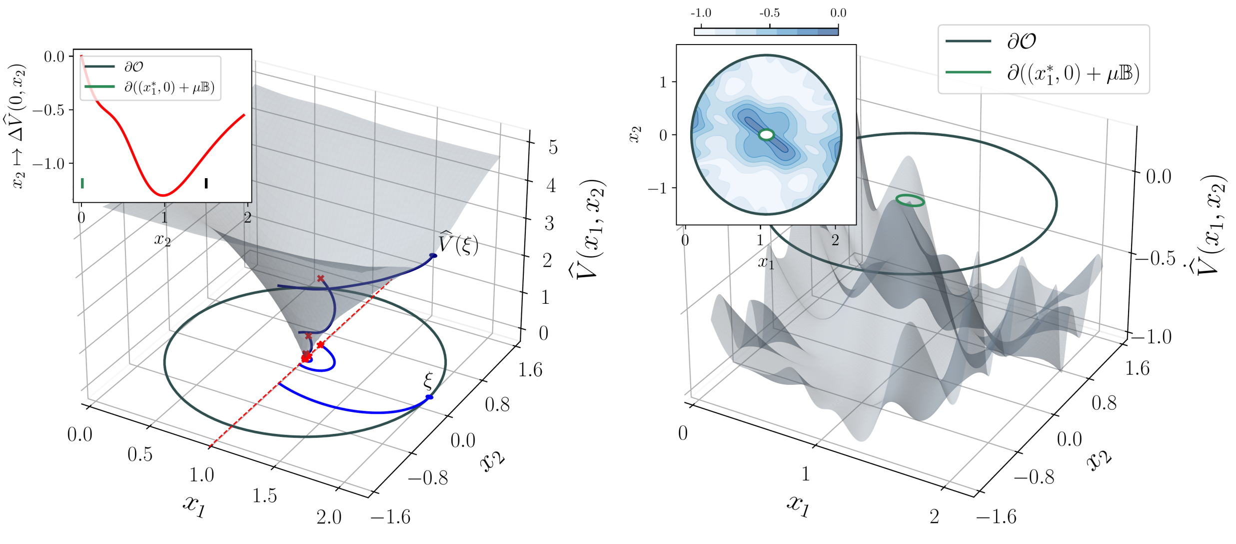
(a) (b)

Figure 2: We show the surface plots of the learned $\widehat{V}$ and $\Delta\widehat{V}$ for $\mathcal{H}_D$ as in (48) with respect to $\mathcal{A} = \{(\pi, 0)\}$. In (a) we show that $\widehat{V}((\pi, 0)) = 0$ and $\widehat{V}(\mathbb{R}^2 \setminus \{(\pi, 0)\}) \subset \mathbb{R}_{>0}$. A solution $\phi \in \mathcal{S}_{\mathcal{H}_D}(\mathcal{O})$, with $\mathcal{O}$ as in (49), is also shown. Notice that $\operatorname{dom}\phi \subset \{0\} \times \mathbb{N}$. In (b) we show that $\Delta\widehat{V}(x) < 0$ for all $x \in \mathcal{O} \setminus ((\pi, 0) + \mu\mathbb{B})$, as guaranteed by Proposition 3.9.

and, given that $G(x)$ is a singleton for each $x \in \mathbb{R}^2$, G is compact valued. Next, we parametrize the LNN $\widehat{V}$ as in with $x \mapsto \eta(x) := x - (\pi, 0)$, $d = 2$, layers' width $(q_1, q_2) = (16, 32)$, and activation function $z \mapsto \sigma(z) := \operatorname{arcsinh}(z)$. Notice that:

- σ is smooth, slope-restricted on the interval $[0, 1]$, and $\sigma(z) = 0 \iff z = 0$;
- σ' is $2\sqrt{3}/9$-Lipschitz continuous;
- η is smooth and 1-Lipschitz continuous on $\mathbb{R}^2$.

For training, the samples are chosen from an ε-cover of $\mathcal{O} \setminus ((\pi, 0) + \mu\mathbb{B}^\circ)$ with $\varepsilon = 0.02$ and $\mu = 1.05\varepsilon$, and we solve (16) following Remark 3.10. As a result, we have that $\tau_D = 0.161$ and the SDP condition in (19) holds with $\rho = 3.461$ and some $T \in \mathcal{T}_{48}$. Then, by Proposition 3.8, we have that $\Delta\widehat{V}$ is Lipschitz on $\mathcal{O}$ with constant $0 \leq L_{\Delta\widehat{V}} \leq 2.94$. In addition, by Proposition 3.9, it follows that $\Delta\widehat{V}(x) < 0$ for all $x \in \mathcal{O} \setminus ((\pi, 0) + \mu\mathbb{B})$, which is illustrated in Figure 2. Since $\mathcal{O}$ is closed, we have that (40) is satisfied and, thanks to Theorem 3.11, we conclude that $\mathcal{A} = \{(\pi, 0)\}$ is practically LpAS for $\mathcal{H}_C$.



(a) (b)

Figure 3: We show the surface plots of the learned LNN $\widehat{V}$ and its time derivative $\dot{\widehat{V}}$ for the example in Section 3.4.3. In (a) we show that $\widehat{V}((x_1^*, 0)) = 0$ and $\widehat{V}(\mathbb{R}^2 \setminus \{(x_1^*, 0)\}) \subset \mathbb{R}_{>0}$. A solution $\phi \in \mathcal{S}_{\mathcal{H}}(\mathcal{O})$, with $\mathcal{O}$ as in (52), is presented. Notice that $\operatorname{dom}\phi \subset \mathbb{R}_{\geq 0} \times \mathbb{N}$. Also, we show that $\Delta\widehat{V}((D \cap \mathcal{O}) \setminus ((x_1^*, 0) + \mu\mathbb{B}^\circ)) \subset \mathbb{R}_{<0}$. In (b) we show that $\dot{\widehat{V}}(x) < 0$ for all $x \in (C \cap \mathcal{O}) \setminus ((x_1^*, 0) + \mu\mathbb{B}^\circ)$.

3.4.3. Hybrid Dynamical Systems

Consider a robotic manipulator interacting with an environment and consisting of a point-mass end effector driven by a controllable force input [20, Ex. 9.1]. The environment is defined as a contact surface at the origin of the coordinate system. The mass is constrained to move horizontally and, during its motion, it may come into contact with the surface. The position and the velocity of the mass are denoted with x_1 and x_2, respectively. When the impact velocity is lower than or equal to a certain threshold, denoted as $\bar{x}_2 > 0$, a compliant impact model is used for the interaction between the end effector and the surface. Assuming unitary mass for the sake of simplicity, the system is described by

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = u_C - f_C(x)$$

where $u_C \in \mathbb{R}$ denotes the force input available for control and f_C is the contact force given by the (discontinuous) function

$$f_C(x) := \begin{cases} k_C x_1 + b_C x_2 & \text{if } x_1 > 0 \\ 0 & \text{if } x_1 \leq 0 \end{cases}$$

where $k_C, b_C > 0$. Note that the Filippov regularization of f_C is given by

$$f_C^r(x) := \begin{cases} k_C x_1 + b_C x_2 & \text{if } x_1 > 0 \\ \overline{\text{co}}\,\{0, b_C x_2\} & \text{if } x_1 = 0 \\ 0 & \text{if } x_1 < 0. \end{cases}$$

The contact condition is modeled as $x_1 = 0$ and $x_2 \geq \bar{x}_2$, and the state variables after the impact are described by the reset law $x_1^+ = x_1$ and $x_2^+ = -\lambda_D x_2$, where $\lambda_D \in [0,1]$. Therefore, the mechanical system of interest is described by means of the following hybrid plant

$$\mathcal{H}_P \; : \; \begin{cases} \dot{x} \; \in \; F_P(x, u_C) := \begin{bmatrix} x_2 \\ u_C - f_C^r(x) \end{bmatrix} & (x, u_C) \in C_P \\ x^+ \; \in \; G(x) := \left\{ \begin{bmatrix} x_1 \\ -\lambda_D x_2 \end{bmatrix} \right\} & x \in D \end{cases} \tag{50}$$

where $C_P := \big\{(x, u_C) \in \mathbb{R}^2 \times \mathbb{R} : x_1 \geq 0\big\} \cup \big\{(x, u_C) \in \mathbb{R}^2 \times \mathbb{R} : x_1 \leq 0,\ x_2 \leq \bar{x}_2\big\}$, and $D := \big\{x \in \mathbb{R}^2 : x_1 = 0,\ x_2 \geq \bar{x}_2\big\}$. Next, consider the following feedback law

$$\kappa_C(x) := \begin{Bmatrix} k_C x_1 - k_p(x_1 - x_1^*) - k_1 x_2 & \text{if } x_1 > 0 \\ -k_p(x_1 - x_1^*) - k_1 x_2 & \text{if } x_1 \leq 0 \end{Bmatrix} \tag{51}$$

for some $k_p, k_1 > 0$ and $x_1^* := 1$. Our objective is to study the stability properties of $\mathcal{A} := \{(x_1^*, 0)\}$ for the closed-loop system $\mathcal{H}$ resulting from assigning $u_C = \kappa_C(x)$. To this end, consider the following sampling set

$$\mathcal{O} := \left\{ x \in \mathbb{R}^2 : \frac{(x_1 - x_1^*)^2}{h_0^2} + \frac{x_2^2}{v_0^2} \leq 1 \right\}. \tag{52}$$

for some $h \in (0, x_1^*]$ and $v_0 > 0$. For starters, notice that gph F_P is closed and, as a result, F_P is osc [21, Thm. 5.7], and it can be shown that F_P is locally bounded relative to C_P. These properties, together with continuity of κ_C, imply that $F(x) := F_P(x, \kappa_C(x))$ is also osc and locally bounded (thereby compact valued) relative to $C := \{x \in \mathbb{R}^2 : (x, \kappa_C(x)) \in C_P\}$. In addition, it can also be shown that F is Lipschitz on $C \cap \mathcal{O}$ with constant satisfying $0 \leq L_F \leq \sqrt{1 + k_p^2 + (k_1 + b_C)^2}$. Similarly, G is compact valued relative to D and Lipschitz continuous with constant $L_G = \lambda_D$.

Consider now an LNN $\widehat{V}$ with $x \mapsto \eta(x) := x - (x_1^*, 0)$, $d = 3$, layer's width $(q_1, q_2, q_3) = (16, 32, 64)$, and activation function $z \mapsto \sigma(z) := z/(1 + |z|)$. Notice that:

- σ is continuously differentiable ($\sigma \in \mathcal{C}^1$), slope-restricted on the interval $[0,1]$, and $\sigma(z) = 0 \iff z = 0$;
- σ' is 2-Lipschitz continuous;
- η smooth and 1-Lipschitz continuous on $\mathbb{R}^2$.

Additionally, given $\varepsilon = 0.01$ and $\mu = 1.1\varepsilon$, for each $\star \in \{C, D\}$, let $\mathcal{M}_\star$ be an $\varepsilon-$net of $(\star \cap \mathcal{O}) \setminus ((x_1^*, 0) + \mu\mathbb{B}^\circ)$. The LNN is trained following Remark 3.10 yielding $\tau_C = 0.037$ and $\tau_D = 0.049$ and, equally important, the SDP condition in (19) holds with $\rho = 0.121$

and with some $T \in \mathcal{T}_{112}$. Then, by Proposition 3.8, we have that $\dot{\widehat{V}}$ is Lipschitz continuous on $C \cap \mathcal{O}$ with constant $0 \leq L_{\dot{\widehat{V}}} \leq 2.22$ and $\Delta\widehat{V}$ is Lipschitz continuous on $D \cap \mathcal{O}$ with constant $0 \leq L_{\Delta\widehat{V}} \leq 1.59$. Given that $\tau_C > L_{\dot{\widehat{V}}}\varepsilon$ and $\tau_D > L_{\Delta\widehat{V}}\varepsilon$, from Proposition 3.9 it follows that $\dot{\widehat{V}}(x) < 0$ for all $x \in (C \cap \mathcal{O}) \setminus ((x_1^*, 0) + \mu\mathbb{B}^\circ)$ and that $\Delta\widehat{V}(x) < 0$ for all $x \in (D \cap \mathcal{O}) \setminus ((x_1^*, 0) + \mu\mathbb{B}^\circ)$. As a result, by Theorem 3.11, we conclude that $\mathcal{A} = \{(x_1^*, 0)\}$ is practically LpAS for $\mathcal{H}$. This is illustrated in Figure 3.

4. Cost Upper Bound for Hybrid Systems

Following the approach in [36, 37, 38], in this section, we derive an upper bound on the cost associated to a solution to a hybrid system $\mathcal{H}$ as in (3) without computing the solution itself.

4.1. Sufficient Conditions for a Cost Upper Bound

Given $\xi \in C \cup D$, the stage cost for flows $\mathcal{L}_C : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, the stage cost for jumps $\mathcal{L}_D : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, and the terminal cost $\varpi : \mathbb{R}^n \to \mathbb{R}$, we define the cost associated to the solutions to $\mathcal{H}$ from the initial condition $\xi \in \mathbb{R}^n$ as

$$\mathcal{J}(\xi) := \sup_{\phi \in \mathcal{S}_{\mathcal{H}}(\xi)} \tilde{\mathcal{J}}(\phi) \tag{53}$$

with

$$\phi \mapsto \tilde{\mathcal{J}}(\phi) := \sum_{j=0}^{\sup_j \operatorname{dom}\phi} \int_{t_j}^{t_{j+1}} \mathcal{L}_C(\phi(t,j))\, dt + \sum_{j=0}^{\sup_j \operatorname{dom}\phi - 1} \mathcal{L}_D(\phi(t_{j+1}, j)) + \limsup_{\substack{t+j \to \sup_t \operatorname{dom}\phi + \sup_j \operatorname{dom}\phi \\ (t,j)\in\operatorname{dom}\phi}} \varpi(\phi(t,j)),$$

where $\{t_j\}_{j=0}^{\sup_j \operatorname{dom}\phi}$ is a nondecreasing sequence associated to the definition of the hybrid time domain of ϕ, see Definition 2.5.

In the next result, following [36], we present sufficient conditions to compute an upper bound on the cost associated to a solution to $\mathcal{H}$. As a difference to [36], and similar to [37], note that (53) includes a terminal cost.

Proposition 4.1. (Cost upper bound) *Consider a hybrid system $\mathcal{H} = (C, F, D, G)$, stage costs $\mathcal{L}_C : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ and $\mathcal{L}_D : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, terminal cost $\varpi : \mathbb{R}^n \to \mathbb{R}$, and the set $\mathcal{O} \subset \mathbb{R}^n$.*

Suppose that, for each $x \in C$ (resp., $x \in D$), $F(x)$ (resp., $G(x)$) is compact and that there exists a function $V : \operatorname{dom} V \to \mathbb{R}$, with $\operatorname{dom} V \supset (\overline{C} \cup D \cup G(D)) \cap \mathcal{O}$, that is continuously differentiable on an open set containing $\overline{C} \cap \mathcal{O}$ such that

$$\mathcal{L}_C(x) + \max_{f \in F(x)} \langle \nabla V(x), f \rangle \leq 0 \qquad \forall x \in C \cap \mathcal{O} \tag{54a}$$

$$\mathcal{L}_D(x) + \max_{g \in G(x)} V(g) - V(x) \leq 0 \qquad \forall x \in D \cap \mathcal{O}. \tag{54b}$$

Given $\xi \in (\overline{C} \cup D) \cap \mathcal{O}$, let ϕ^* be a solution[14] *to*

$$\mathcal{H}_{\max} \ : \begin{cases} \dot{x} \in \operatorname*{arg\,max}_{f \in F(x)} \ \langle \nabla V(x), f \rangle & x \in C \cap \mathcal{O} \\ x^+ \in \operatorname*{arg\,max}_{g \in G(x)} \ V(g) & x \in D \cap \mathcal{O} \end{cases} \tag{55}$$

from ξ, and suppose that the map $(t, j) \mapsto V(\phi^*(t, j))$ is bounded on $\operatorname{dom} \phi^*$ and

$$\limsup_{\substack{t+j \to \sup_t \operatorname{dom} \phi^* + \sup_j \operatorname{dom} \phi^* \\ (t,j) \in \operatorname{dom} \phi^*}} V(\phi^*(t, j)) = \limsup_{\substack{t+j \to \sup_t \operatorname{dom} \phi^* + \sup_j \operatorname{dom} \phi^* \\ (t,j) \in \operatorname{dom} \phi^*}} \varpi(\phi^*(t, j)). \tag{56}$$

Then, it follows that

$$\mathcal{J}(\xi) = \tilde{\mathcal{J}}(\phi^\star) \leq V(\xi) \qquad \forall \xi \in (\overline{C} \cup D) \cap \mathcal{O}. \tag{57}$$

Proof. The proof follows similar arguments as in [39, Prop. 3.7].

Pick any $\xi \in (\overline{C} \cup D) \cap \mathcal{O}$ and $\phi^\star \in \mathcal{S}_{\mathcal{H}_{\max}}(\xi)$. Thanks to [39, Lem. 1.3], $\phi^\star$ is also a solution to $\mathcal{H}$ as in (3). For each $j \in \mathbb{N}$ such that $I^j_{\phi^*}$ has a nonempty interior $\operatorname{int} I^j_{\phi^*}$, we have from (54a)

$$-\mathcal{L}_C(\phi^*(t, j)) \geq \max_{f \in F(\phi^*(t,j))} \langle \nabla V(\phi^*(t, j)), f \rangle =: \frac{dV}{dt}(\phi^*(t, j)) \quad \text{for almost all } t \in I^j_{\phi^*}. \tag{58}$$

Next, from (54b), we have that for every $(t, j) \in \operatorname{dom} \phi^*$ such that $(t, j + 1) \in \operatorname{dom} \phi^*$:

$$\begin{aligned} -\mathcal{L}_D(\phi^*(t, j)) &\geq \max_{g \in G(\phi^*(t,j))} V(g) - V(\phi^*(t, j)) \\ &= V(\phi^*(t, j + 1)) - V(\phi^*(t, j)). \end{aligned} \tag{59}$$

Now, thanks to (58), (59), and (56), together with [39, Prop. 1.1], we have that $\mathcal{J}(\phi^*) \leq V(\xi)$.

Finally, from [39, Prop. 3.6], for any arbitrary $\phi \in \mathcal{S}_{\mathcal{H}}(\xi)$, we have that

$$\tilde{\mathcal{J}}(\phi) \leq \tilde{\mathcal{J}}(\phi^*) \leq V(\xi)$$

[14] Sufficient conditions for the existence of nontrivial solutions to $\mathcal{H}_{\max}$ can be found in [14, Prop. 6.10].

which, in turn, implies that the largest cost of solutions from ξ satisfies

$$\mathcal{J}(\xi) \leq V(\xi).$$

■

Thus, if one can design a function V that satisfies the conditions in Proposition 4.1, it is possible to provide an upper bound on the cost $\mathcal{J}$, without needing to explicitly compute solutions to $\mathcal{H}$, which is obtained by evaluating V at the initial condition ξ.

4.2. Sampled-Based Cost Upper Bound Conditions via Learning

With the aim of designing an upper bound on the cost $\mathcal{J}$ associated to solutions to $\mathcal{H}$ that stay on a set $\mathcal{O} \subset \mathbb{R}^n$, we propose an optimization program with conditions (54a) and (54b) as constraints, enforced at finitely many points. By properly choosing these, we guarantee a provable extension of the aforementioned conditions to all points in $(C \cup D) \cap \mathcal{O}$.

More precisely, consider $\mathcal{H} = (C, F, D, G)$ as in (3), where, for each $x \in C$ (resp., $x \in D$), $F(x)$ (resp., $G(x)$) is compact, and a nonempty and bounded set $\mathcal{O} \subset \mathbb{R}^n$. Let the stage costs for flows and for jumps be given by $\mathcal{L}_C : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ and $\mathcal{L}_D : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, respectively. Given an LNN $\widehat{V}$ with parameters (θ, σ, η), suppose that $\nabla \widehat{V}$ is Borel measurable on an open set containing $\overline{C} \cap \mathcal{O}$ and let us pick the parameters of the LNN as

$$\begin{aligned}
\theta^* \in \quad & \underset{\theta \in \mathbb{R}^r}{\arg\min} && \int_{C \cap \mathcal{O}} \mathbb{E}\left[\left|\nabla \widehat{V}(x + \delta)\right|\right] d\lambda(x) && \\
& \text{subject to} && \max_{f \in F(x)} \left\langle \nabla \widehat{V}(x), f \right\rangle + \mathcal{L}_C(x) \leq 0 && \forall x \in C \cap \mathcal{O} \\
& && \max_{g \in G(x)} \widehat{V}(g) - \widehat{V}(x) + \mathcal{L}_D(x) \leq 0 && \forall x \in D \cap \mathcal{O}
\end{aligned} \tag{60}$$

where $\delta \sim \mathcal{N}(0, I_n)$ is a random vector on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \gamma)$, with γ the standard Gaussian measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. As mentioned in Section 3.1, notice that the constraints in (60) need to be solved at (likely) infinitely many points in $(C \cup D) \cap \mathcal{O}$, which is computationally intractable. Therefore, we propose solving a relaxed version of (60) given by

$$\begin{aligned}
\theta^* \in \underset{\theta \in \mathbb{R}^r}{\arg\min} \quad & \sum_{x' \in \mathcal{Q}_C} \mathbb{E}\left[\left|\nabla \widehat{V}(x' + \delta)\right|\right] \lambda\left((x' + \varepsilon \mathbb{B}) \cap \mathcal{X}\right) && \\
\text{subject to} \quad & \max_{f \in F(x')} \left\langle \nabla \widehat{V}(x'), f \right\rangle + \mathcal{L}_C(x') \leq -\varkappa_C && \forall x' \in \mathcal{Q}_C \\
& \max_{g \in G(x')} \widehat{V}(g) - \widehat{V}(x') + \mathcal{L}_D(x') \leq -\varkappa_D && \forall x' \in \mathcal{Q}_D
\end{aligned} \tag{61}$$

where, for each $\star \in \{C, D\}$, $\mathcal{Q}_\star$ is an ε-net of $\star \cap \mathcal{O}$ and $\varkappa_\star > 0$. In the next section, we provide sufficient conditions to guarantee that if (61) is feasible, then we can provide an upper bound on the cost associated to a solution that starts and remains in $(\overline{C} \cup D) \cap \mathcal{O}$, which can be obtained without computing solutions to $\mathcal{H}$.

4.3. Sufficient Conditions for Design of Learning-Based Cost Upper Bound

We extend the conditions enforced at points in $\mathcal{Q}_C$ and $\mathcal{Q}_D$, to every point in $(C \cup D) \cap \mathcal{O}$. The slack variables $\varkappa_C$ and $\varkappa_D$ in (61), together with sufficient regularity properties of the LNN $\widehat{V}$, allow us to guarantee that (54a) and (54b) hold at all points in $C \cap \mathcal{O}$ and $D \cap \mathcal{O}$, respectively.

Proposition 4.2. (Generalized cost upper bound conditions) *Let $\mathcal{O} \subset \mathbb{R}^n$ be bounded and consider the hybrid system $\mathcal{H} = (C, F, D, G)$ where, for each $x \in C$ (resp., $x \in D$), $F(x)$ (resp., $G(x)$) is compact. Given $\varepsilon > 0$, for each $\star \in \{C, D\}$, let $\mathcal{Q}_\star$ be an ε-net of $\star \cap \mathcal{O}$. Let $\widehat{V}$ be an LNN with parameters (θ, σ, η) and suppose that*

1) *$\widehat{V}$ is $L_{\widehat{V}}$-Lipschitz continuous on $\mathcal{O}$;*
2) *$\dot{\widehat{V}} : C \to \mathbb{R}$ is $L_{\dot{\widehat{V}}}$-Lipschitz continuous on $C \cap \mathcal{O}$;*
3) *$\Delta\widehat{V} : D \to \mathbb{R}$ is $L_{\Delta\widehat{V}}$-Lipschitz continuous on $D \cap \mathcal{O}$.*

Further, consider an L_C-Lipschitz continuous stage cost for flows $\mathcal{L}_C : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ and an L_D-Lipschitz continuous stage cost for jumps $\mathcal{L}_D : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$. The following holds:

1) *if there exists $\varkappa_C \geq \varepsilon(L_C + L_{\dot{\widehat{V}}})$ such that $\dot{\widehat{V}}(x') + \mathcal{L}_C(x') \leq -\varkappa_C$ for all $x' \in \mathcal{Q}_C$, then it follows that $\dot{\widehat{V}}(x) + \mathcal{L}_C(x) \leq 0$ for all $x \in C \cap \mathcal{O}$;*
2) *if there exists $\varkappa_D \geq \varepsilon(L_D + L_{\Delta\widehat{V}})$ such that $\Delta\widehat{V}(x') + \mathcal{L}_D(x') \leq -\varkappa_D$ for all $x' \in \mathcal{Q}_D$, then it follows that $\Delta\widehat{V}(x) + \mathcal{L}_D(x) \leq 0$ for all $x \in D \cap \mathcal{O}$.*

Proof. Given that the set $\mathcal{Q}_C$ is an ε−net of $C \cap \mathcal{O}$, we have that for each $x \in C \cap \mathcal{O}$ there exists $x' \in \mathcal{Q}_C$ such that

$$\begin{aligned} |\dot{\widehat{V}}(x) + \mathcal{L}_C(x) - \dot{\widehat{V}}(x') - \mathcal{L}_C(x')| &\leq |\dot{\widehat{V}}(x) - \dot{\widehat{V}}(x')| + |\mathcal{L}_C(x) - \mathcal{L}_C(x')| \\ &\leq (L_{\dot{\widehat{V}}} + L_C)|x - x'| \\ &\leq (L_{\dot{\widehat{V}}} + L_C)\varepsilon. \end{aligned}$$

Thus, we have that

$$\dot{\hat{V}}(x) + \mathcal{L}_C(x) \leq \dot{\hat{V}}(x') + \mathcal{L}_C(x') + (L_{\dot{\hat{V}}} + L_C)\varepsilon \leq -\varkappa_C + (L_{\dot{\hat{V}}} + L_C)\varepsilon. \tag{62}$$

Similarly, given that the set $\mathcal{Q}_D$ is an $\varepsilon-$net over $D \cap \mathcal{O}$, we have that for each $x \in D \cap \mathcal{O}$ there exists $x' \in \mathcal{Q}_D$ such that

$$\begin{aligned}
|\Delta\widehat{V}(x) + \mathcal{L}_D(x) - \Delta\widehat{V}(x') - \mathcal{L}_D(x')| &\leq |\Delta\widehat{V}(x) - \Delta\widehat{V}(x')| + |\mathcal{L}_D(x) - \mathcal{L}_D(x')| \\
&\leq (L_D + L_{\Delta\widehat{V}})|x - x'| \\
&\leq (L_D + L_{\Delta\widehat{V}})\varepsilon.
\end{aligned}$$

and

$$\Delta\widehat{V}(x) + \mathcal{L}_D(x) - \Delta\widehat{V}(x') - \mathcal{L}_D(x') \leq (L_D + L_{\Delta\widehat{V}})\varepsilon$$

From the conditions in the statement of this result,

$$\begin{aligned}
\Delta\widehat{V}(x) + \mathcal{L}_D(x) &\leq \Delta\widehat{V}(x') + \mathcal{L}_D(x') + (L_D + L_{\Delta\widehat{V}})\varepsilon \\
&\leq -\varkappa_D + (L_D + L_{\Delta\widehat{V}})\varepsilon.
\end{aligned} \tag{63}$$

Therefore, from (62) and (63), notice that

$$\begin{aligned}
\varkappa_C \geq \varepsilon(L_C + L_{\dot{\hat{V}}}) \quad &\implies \quad \sup_{f \in F(x)} \langle \nabla\widehat{V}(x), f\rangle + \mathcal{L}_C(x) \leq 0 \qquad \forall x \in C \cap \mathcal{O} \\
\varkappa_D \geq \varepsilon(L_D + L_{\Delta\widehat{V}}) \quad &\implies \quad \sup_{g \in G(x)} \widehat{V}(g) - \widehat{V}(x) + \mathcal{L}_D(x) \leq 0 \quad \forall x \in D \cap \mathcal{O}.
\end{aligned}$$

■

Training an LNN $\widehat{V}$ such that the conditions on the Lipschitz constants in Proposition 4.2 are satisfied follows a similar procedure as in Remark 3.10.

Corollary 4.3. (Learning-based cost upper bound) *Consider a bounded set $\mathcal{O} \subset \mathbb{R}^n$, a hybrid system $\mathcal{H} = (C, F, D, G)$, and an LNN $\widehat{V}$ with parameters (θ, σ, η). Suppose that the assumptions in Proposition 4.2 hold. Let $\phi^* : \operatorname{dom}\phi^* \to \mathbb{R}^n$ be a solution to $\mathcal{H}_{\max}$ as in (55) from $\xi \in (\overline{C} \cup D) \cap \mathcal{O}$ and suppose that $(t, j) \mapsto \widehat{V}(\phi^*(t, j))$ is bounded on $\operatorname{dom}\phi^*$ and that (56) holds. Then*

$$\mathcal{J}(\xi) \leq \widehat{V}(\xi) \qquad \forall \xi \in (\overline{C} \cup D) \cap \mathcal{O}.$$

Proof. This is a straightforward application of Proposition 4.2 together with Proposition 4.1. ■

5. Learning-Based Cost Upper Bounds and Lyapunov Functions for Hybrid Inclusions

Finally, in this section, we present learning-based sufficient conditions such that an LNN $\widehat{V}$ upper bounds the cost $\mathcal{J}$ associated to solutions to $\mathcal{H}$, but also certifies asymptotic stability of a nonempty and compact set $\mathcal{A}$ for $\mathcal{H}$ on a bounded set $\mathcal{O} \supset \mathcal{A}$.

Based on [36], we present a result that connects cost evaluation and asymptotic stability for hybrid systems.

Theorem 5.1. (Cost evaluation under the existence of an LNN) *Let $\mathcal{O} \subset \mathbb{R}^n$ be bounded and consider hybrid system $\mathcal{H} = (C, F, D, G)$ where, for each $x \in C$ (resp., $x \in D$), $F(x)$ (resp., $G(x)$) is compact. Given $\varepsilon > 0$, for each $\star \in \{C, D\}$, let $\mathcal{Q}_\star$ be an ε-net of $\star \cap \mathcal{O}$. Let $\widehat{V}$ be an LNN with parameters (θ, σ, η) and suppose that*

1) $\widehat{V}$ is $L_{\widehat{V}}$-Lipschitz continuous on $\mathcal{O}$;

2) $\dot{\widehat{V}} : C \to \mathbb{R}$ is $L_{\dot{\widehat{V}}}$-Lipschitz continuous on $C \cap \mathcal{O}$;

3) $\Delta\widehat{V} : D \to \mathbb{R}$ is $L_{\Delta\widehat{V}}$-Lipschitz continuous on $D \cap \mathcal{O}$.

Further, consider the L_C-Lipschitz continuous stage cost for flows $\mathcal{L}_C : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, the L_D-Lipschitz continuous stage cost for jumps $\mathcal{L}_D : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, and the terminal cost $\varpi : \mathbb{R}^n \to \mathbb{R}$. Suppose that

1) there exists $\varkappa_C \geq \varepsilon(L_C + L_{\dot{\widehat{V}}})$ such that $\dot{\widehat{V}}(x') + \mathcal{L}_C(x') \leq -\varkappa_C$ for all $x' \in \mathcal{Q}_C$;

2) there exists $\varkappa_D \geq \varepsilon(L_D + L_{\Delta\widehat{V}})$ such that $\Delta\widehat{V}(x') + \mathcal{L}_D(x') \leq -\varkappa_D$ for all $x' \in \mathcal{Q}_D$.

Additionally, let $\phi^* : \operatorname{dom}\phi^* \to \mathbb{R}^n$ be a solution to $\mathcal{H}_{\max}$ as in (55) from $\xi \in (\overline{C} \cup D) \cap \mathcal{O}$ and suppose that $(t, j) \mapsto \widehat{V}(\phi^*(t, j))$ is bounded on $\operatorname{dom}\phi^*$ and that (56) holds. Then

$$\mathcal{J}(\xi) \leq \widehat{V}(\xi) \qquad \forall \xi \in (\overline{C} \cup D) \cap \mathcal{O}. \tag{64}$$

In addition, consider a compact set $\mathcal{A} \subset \mathcal{O}$ and suppose that $\widehat{V}$ satisfies Assumption 3.4. If one of the following conditions holds

1) $\mathcal{L}_C \in \mathcal{PD}(\mathcal{A})$ and $\mathcal{L}_D \in \mathcal{PD}(\mathcal{A})$;

2) $\mathcal{L}_D \in \mathcal{PD}(\mathcal{A})$ and there exists a continuous function $\nu \in \mathcal{PD}(\{0\})$ such that $\mathcal{L}_C(x) \geq \nu(|x|_{\mathcal{A}})$ for all $x \in C \cap \mathcal{O}$;

3) $\mathcal{L}_C \in \mathcal{PD}(\mathcal{A})$ and there exists a continuous function $\nu \in \mathcal{PD}(\{0\})$ such that $\mathcal{L}_D(x) \geq \nu(|x|_{\mathcal{A}})$ for all $x \in D \cap \mathcal{O}$;

4) $x \mapsto \mathcal{L}_C(x) = 0$, $\mathcal{L}_D \in \mathcal{PD}(\mathcal{A})$, and, for each $r > 0$, there exist $\gamma_r \in \mathcal{K}_\infty$ and $N_r \geq 0$ such that for every solution $\phi \in \mathcal{S}_{\mathcal{H}}(\xi)$

$$|\phi(0,0)|_{\mathcal{A}} \in (0,r],\ (t,j) \in \operatorname{dom}\phi,\ t+j \geq T \quad \Longrightarrow \quad j \geq \gamma_r(T) - N_r;$$

5) $\mathcal{L}_C \in \mathcal{PD}(\mathcal{A})$, $x \mapsto \mathcal{L}_D(x) = 0$, and, for each $r > 0$, there exist $\gamma_r \in \mathcal{K}_\infty$ and $N_r \geq 0$ such that for every solution $\phi \in \mathcal{S}_{\mathcal{H}}(\xi)$

$$|\phi(0,0)|_{\mathcal{A}} \in (0,r],\ (t,j) \in \operatorname{dom}\phi,\ t+j \geq T \quad \Longrightarrow \quad t \geq \gamma_r(T) - N_r;$$

6) there exist $(\lambda_C, \lambda_D) \in \mathbb{R}^2$ such that $\mathcal{L}_C(x) \geq -\lambda_C V(x)$ for all $x \in C \cap \mathcal{O}$ and $\mathcal{L}_D(x) \geq (1 - e^{\lambda_D})V(x)$ for all $x \in D \cap \mathcal{O}$, and there exist $\gamma > 0$ and $M > 0$ such that for each solution $\phi \in \mathcal{S}_{\mathcal{H}}(\xi)$ that remains in $\mathcal{O}$

$$(t,j) \in \operatorname{dom}\phi \quad \Longrightarrow \quad \lambda_C t + \lambda_D j \leq M - \gamma(t+j); \tag{65}$$

then $\mathcal{A}$ is LpAS for $\mathcal{H}$ on $\mathcal{O}$.

Proof. The bound (64) is a straightforward application of Corollary 4.3. In addition, notice that $\widehat{V}$ is a Lyapunov function candidate on $\mathcal{O}$ for $\mathcal{H}$ (see Definition 2.10) and suppose that:

a) Item 1) above holds. Define

$$\rho(x) := \begin{cases} \mathcal{L}_C(x) & \text{if } x \in (C \cap \mathcal{O}) \setminus D \\ \min\{\mathcal{L}_C(x), \mathcal{L}_D(x)\} & \text{if } x \in C \cap D \cap \mathcal{O} \\ \mathcal{L}_D(x) & \text{if } x \in (D \cap \mathcal{O}) \setminus C \end{cases} \qquad \forall\, x \in (C \cup D) \cap \mathcal{O}. \tag{66}$$

Thus, from Proposition 4.2, it follows that

$$\begin{aligned} \max_{f \in F(x)} \left\langle \nabla \widehat{V}(x), f \right\rangle &\leq -\rho(x) < 0 \qquad \forall x \in (C \cap \mathcal{O}) \setminus \mathcal{A} \\ \max_{g \in G(x)} \widehat{V}(g) - \widehat{V}(x) &\leq -\rho(x) < 0 \qquad \forall x \in (D \cap \mathcal{O}) \setminus \mathcal{A} \end{aligned}$$

and, thanks to Theorem 2.11, the set $\mathcal{A}$ is LpAS for $\mathcal{H}$ on $\mathcal{O}$.

b) Item 2) above holds. Define

$$\rho(x) := \begin{cases} \nu(|x|_{\mathcal{A}}) & \text{if } x \in (C \cap \mathcal{O}) \setminus D \\ \min\{\nu(|x|_{\mathcal{A}}), \mathcal{L}_D(x)\} & \text{if } x \in C \cap D \cap \mathcal{O} \\ \mathcal{L}_D(x) & \text{if } x \in (D \cap \mathcal{O}) \setminus C \end{cases} \qquad \forall\, x \in (C \cup D) \cap \mathcal{O};$$

and $\mathcal{H}$ has $\mathcal{A}$ LpAS on $\mathcal{O}$ following similar arguments as in item a);

c) Item 3) above holds. Then $\mathcal{A}$ is LpAS for $\mathcal{H}$ on $\mathcal{O}$ following similar steps as in item b), which are omitted for brevity;

d) Item 4) above holds. Then $\mathcal{A}$ is LpAS for $\mathcal{H}$ on $\mathcal{O}$ thanks to [14, Prop. 3.24] with ρ as in (66) considering $x \mapsto \mathcal{L}_C(x) = 0$.

e) Item item 5) above holds. Then $\mathcal{A}$ is LpAS for $\mathcal{H}$ thanks to [14, Prop. 3.27] with ρ as in (66) considering $x \mapsto \mathcal{L}_D(x) = 0$.

f) Item item 6) above holds. The proof hinges upon [20, Thm. 3.19]. Let $X := C \cup D \cup G(D)$ and pick any $\varepsilon' > 0$ such that $\tilde{\mathcal{A}} + 2\varepsilon' \mathbb{B} \subset \mathcal{O}$, where $\tilde{\mathcal{A}} := \mathcal{A} \cap X$. Let

$$r' := \min\left\{ \widehat{V}(x) : x \in \operatorname{dom} \widehat{V},\ |x|_{\tilde{\mathcal{A}}} = \varepsilon' \right\} \tag{67}$$

which is positive because $\varepsilon' > 0$ and $\widehat{V}$ is positive definite with respect to $\mathcal{A}$. Pick $r_1 \in (0, r')$ small enough so that

$$Z := \left\{ x \in \operatorname{dom} \widehat{V} \cap X : \widehat{V}(x) \le r_1 \right\}$$

is compact. Note that $Z \subset \operatorname{int}\left(\tilde{\mathcal{A}} + 2\varepsilon' \mathbb{B}\right)$; hence $Z \subset \mathcal{O}$. Given M as in (65), choose $r_0 > 0$ such that $\exp(M)\, r_0 \le r_1$. Pick any solution ϕ to $\mathcal{H}$ with $\phi(0,0)$ satisfying $\widehat{V}\big(\phi(0,0)\big) \le r_0$. Then $\operatorname{rge} \phi \subset \mathcal{O}$. Indeed, using (65) together with

$$\begin{aligned} \max_{f \in F(x)} \left\langle \nabla \widehat{V}(x), f \right\rangle &\le -\rho(x) \le -\mathcal{L}_C(x) \le \lambda_C \widehat{V}(x) \qquad && \forall x \in C \cap \mathcal{O} \\ \max_{g \in G(x)} \widehat{V}(g) - \widehat{V}(x) &\le -\mathcal{L}_D(x) \le (e^{\lambda_D} - 1)\widehat{V}(x) \qquad && \forall x \in D \cap \mathcal{O} \end{aligned}$$

we have that

$$\widehat{V}(\phi(t,j)) \le \exp(\lambda_C t + \lambda_D j)\widehat{V}(\phi(0,0)) \le \exp(M)\exp(-\gamma(t+j))\widehat{V}(\phi(0,0)) \tag{68}$$

for all $(t,j) \in \operatorname{dom} \phi$. Because γ is positive and $\exp(M - \gamma(t+j))\widehat{V}(\phi(0,0)) \le r_1$ for every (t,j) and $r_1 \le r'$, then ϕ remains in $\mathcal{O}$. Now, given $\varepsilon > 0$, take r' as in (67) with $\varepsilon' \in (0, \varepsilon/2)$. Choose $r_1 \in (0, r')$ so that Z is compact, and pick $\delta \in (0, \varepsilon')$ satisfying

$$\exp(M) \max\left\{ \widehat{V}(x) : x \in \operatorname{dom} \widehat{V},\ |x|_{\tilde{\mathcal{A}}} \le \delta \right\} \ \le\ r_1.$$

For any solution ϕ to $\mathcal{H}$ with $|\phi(0,0)|_{\tilde{\mathcal{A}}} \leq \delta$ we have $\exp(M)\widehat{V}(\phi(0,0)) \leq r_1$, hence $\widehat{V}(\phi(t,j)) \leq r_1$ by (68), so ϕ stays in Z. Because $Z \subset \tilde{\mathcal{A}} + 2\varepsilon' \mathbb{B} \subset \mathcal{A} + \varepsilon\mathbb{B}$, it follows that $\operatorname{rge}\phi \subset \mathcal{A} + \varepsilon\mathbb{B}$. Thus $\mathcal{A}$ is stable for $\mathcal{H}$. Pre-attractivity follows with $\mu := \delta$ (as chosen above), invoking (68) together with the fact that Z is compact.

■

5.1. Numerical Examples

In classical control theory, the output of a controller of a continuous-time plant evolves continuously in time. *Reset control systems* differ from these traditional controllers as their output experiences jumps caused by resets of the controller state. These resets may depend on the value of the controller inputs. In some scenarios, in comparison to (non-reset) classical controllers, reset controllers[15] lead to improved system performance [14, Ex. 1.6]. Usually, modeling this type of reset controllers lead to linear hybrid systems with periodic jumps, and (possibly) conic flow and/or jumps sets, such as in [36, Sec. III.A], where the state is $x := (x_p, \tau) \in \mathbb{R}^n \times [0,T]$, for some $T > 0$, and hybrid dynamics given by

$$\mathcal{H}_P \; : \; \begin{cases} \dot{x} \;\in\; F(x) := \left\{ \begin{bmatrix} A_C x_p \\ 1 \end{bmatrix} \right\} & x \in C := \mathbb{R}^n \times [0,T] \\ x^+ \;\in\; G(x) := \left\{ \begin{bmatrix} A_D x_p \\ 0 \end{bmatrix} \right\} & x \in D := \mathbb{R}^n \times \{T\} \end{cases} \tag{69}$$

with $A_C, A_D \in \mathbb{R}^{n\times n}$. Let the stage cost for flows be given by $x \mapsto \mathcal{L}_C(x) := x_p^\top Q_C x_p$ and the stage cost for jumps be given by $x \mapsto \mathcal{L}_D(x) := x_p^\top Q_D x_p$ where $Q_C, Q_D \in \mathbb{S}^n_{>0}$. Next, pick $r > 0$ and $\zeta \geq T/2$. Thus, we can consider the sampling set as

$$\mathcal{O} := \left\{ x \in \mathbb{R}^n \times [0,T] : \frac{|x_p|^2}{r^2} + \frac{\left(\tau - \dfrac{T}{2}\right)^2}{\zeta^2} \leq 1 \right\} \tag{70}$$

and notice that $\mathcal{A} := \{0\} \times [0,T] \subset \mathcal{O}$. From this choice, it follows that $\mathcal{L}_C$ is Lipschitz on $C \cap \mathcal{O}$ with constant satisfying $0 \leq L_C \leq 2r|Q_C|$ and $\mathcal{L}_D$ is Lipschitz on $D \cap \mathcal{O}$ with constant satisfying $0 \leq L_D \leq 2r|Q_D|\sqrt{1-(T/(2\zeta))^2}$. Similarly, it is clear that $F(x)$ (resp. $G(x)$) is compact for each $x \in C$ (resp., $x \in D$), and that F and G are Lipschitz.

[15]The first reset controller that appeared in the literature is the so-called *Clegg integrator*, a single-input/single-output linear controller that resets its output to zero when its input and output do not have the same sign.

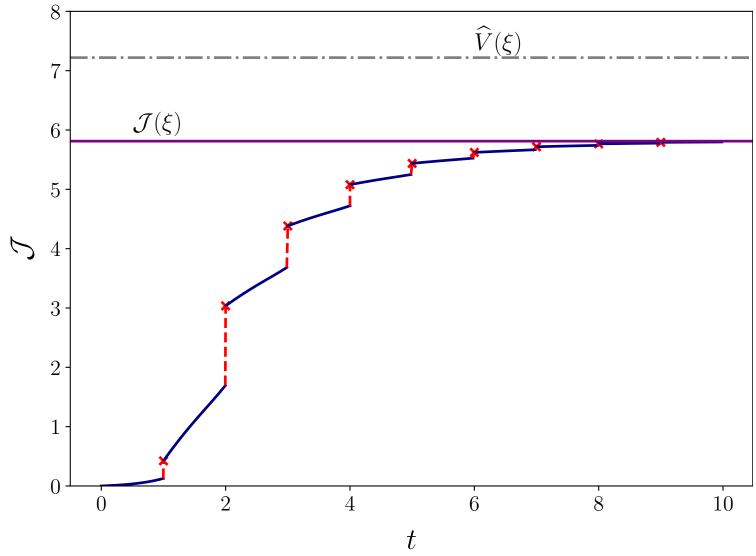


Figure 4: We show the evolution of $\phi \mapsto \tilde{\mathcal{J}}(\phi)$ for a solution $\phi \in \mathcal{S}_{\mathcal{H}}(\xi)$ with $\xi = (0.2, 0, 0)$. Notice that $\operatorname{dom}\phi \subset \mathbb{R}_{\geq 0} \times \mathbb{N}$. For comparison purposes, the exact cost $\mathcal{J}(\xi)$ is shown (obtained using [36, Prop. 3]), and, as expected from Theorem 5.1, we have that $\mathcal{J}(\xi) \leq \widehat{V}(\xi)$.

To compare our results to those in the existing literature [36, Ex. 2], consider the following choice of parameters

$$A_C = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad A_D = \begin{bmatrix} 1 & 0 \\ -2 & 0 \end{bmatrix}, \quad Q_C = Q_D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad T = 1.$$

Next, let $\widehat{V}$ be an LNN with $x \mapsto \eta(x) := (1, 0)^\top x$, $d = 2$, layers' width $(q_1, q_2) = (32, 64)$, and activation function $z \mapsto \sigma(z) := \tanh(z)$. Notice that:

- σ is smooth, slope-restricted on the interval $[0, 1]$, and $\sigma(z) = 0 \iff z = 0$;
- σ' is $4\sqrt{3}/9$-Lipschitz continuous;
- η smooth and 1-Lipschitz continuous on $\mathbb{R}^2 \times [0, T]$.

Additionally, given $\varepsilon = 0.01$, for each $\star \in \{C, D\}$, let $\mathcal{Q}_\star$ be an ε-net of $\star \cap \mathcal{O}$. We train the LNN following Remark 3.10 to enforce the constraints on the Lipschitz constants[16] of $\dot{\widehat{V}}$ and $\Delta\widehat{V}$ in Theorem 5.1, and, by setting the terminal cost $x \mapsto \varpi(x) := \widehat{V}(x)$, from Theorem 5.1, we conclude that $\mathcal{J}(\xi) \leq \widehat{V}(\xi)$ for all $\xi \in (C \cup D) \cap \mathcal{O}$. This is confirmed by Figure 4, where, for comparison purposes, the exact cost $\mathcal{J}(\xi)$ is also presented (obtained using [36, Prop. 3]). Finally, using the fact that $\mathcal{L}_C \in \mathcal{PD}(\mathcal{A})$ and that $\mathcal{L}_D \in \mathcal{PD}(\mathcal{A})$, we conclude that $\mathcal{A}$ is LpAS for $\mathcal{H}$ on $\mathcal{O}$.

[16] See Proposition 3.8 for sufficient conditions.

6. Conclusions and Future Work

In this work, we present a data-driven framework for learning stability and optimality certificates for hybrid systems. Our approach trains a neural network on a finite set of sampled data such that sufficient-decrease conditions of a Lyapunov neural network along flows and at jumps are satisfied, while simultaneously regularizing the network's Lipschitz constant during training. This Lipschitz regularization allows us to extend the decrease property to states that were not included in the training set. Thus, we can conclude about the stability properties of a compact set for a hybrid system $\mathcal{H}$ and further evaluate the associated cost of its solutions (which need not be unique) without explicitly computing any trajectories. In addition, we present complementary results for LNNs for hybrid inclusions. Specifically, sufficient conditions for positive definiteness and Lipschitz continuity of the change of the LNN during flows and at jumps.

Future work includes an in-depth study on how the choice of activation function affects the approximation accuracy of the resulting LNN, along with the derivation of sufficient conditions guaranteeing the feasibility of (16) or (61). To reduce the computational burden introduced by the use of ε-nets, we also plan to explore alternative approximation and sampling-reduction techniques, for example, randomized coverings, adaptive meshing, and sparse verification. Finally, an additional line of research is to find the tightest possible upper bound on $\mathcal{J}$ obtainable with LNNs, while certifying local asymptotic stability.

7. Acknowledgments

This research has been partially supported by NSF Grants no. CNS-2039054 and CNS-2111688, by AFOSR Grants nos. FA9550-19-1-0169, FA9550-20-1-0238, FA9550-23-1-0145, and FA9550-23-1-0313, by AFRL Grant nos. FA8651-22-1-0017 and FA8651-23-1-0004, by ARO Grant no. W911NF-20-1-0253, and by DoD Grant no. W911NF-23-1-0158.

References

[1] P. Giesl, S. Hafstein, Review on Computational Methods for Lyapunov Functions, Discrete and Continuous Dynamical Systems-B 20 (8) (2015) 2291–2331.

[2] A. Papachristodoulou, S. Prajna, On the Construction of Lyapunov Functions using the Sum of Squares Decomposition, in: Proceedings of the 41st IEEE Conference on Decision and Control, Vol. 3, IEEE, 2002, pp. 3482–3487.

[3] S. M. Richards, F. Berkenkamp, A. Krause, The Lyapunov Neural Network: Adaptive Stability Certification for Safe Learning of Dynamical Systems, in: Proceedings of Machine Learning Research, 2018, pp. 466–476.

[4] I. D. Jimenez Rodriguez, A. D. Ames, Y. Yue, LyaNet: A Lyapunov Framework for Training Neural ODEs, in: Proceedings of the International Conference on Machine Learning, 2022, pp. 18687–18703.

[5] H. K. Khalil, Nonlinear systems, 3rd Edition, Prentice Hall, 2002.

[6] A. D. Ames, K. Galloway, K. Sreenath, J. W. Grizzle, Rapidly Exponentially Stabilizing Control Lyapunov Functions and Hybrid Zero Dynamics, IEEE Transactions on Automatic Control 59 (4) (2014) 876–891.

[7] L. Grüne, Computing Lyapunov Functions using Deep Neural Networks, Journal of Computational Dynamics 8 (2) (2021) 131–152.

[8] S. Mohammad Khansari-Zadeh, A. Billard, Learning Control Lyapunov Function to Ensure Stability of Dynamical System-based Robot Reaching Motions, Robotics and Autonomous Systems 62 (6) (2014) 752–765.

[9] A. Abate, D. Ahmed, M. Giacobbe, A. Peruffo, Formal Synthesis of Lyapunov Neural Networks, IEEE Control Systems Letters 5 (3) (2020) 773–778.

[10] H. Dai, B. Landry, M. Pavone, R. Tedrake, Counter-Example Guided Synthesis of Neural Network Lyapunov Functions for Piecewise Linear Systems, in: Proceedings of the 59th IEEE Conference on Decision and Control, 2020, pp. 1274–1281.

[11] H. Ravanbakhsh, S. Sankaranarayanan, Counter-Example Guided Synthesis of Control Lyapunov Functions for Switched Systems, Proceedings of the 54th IEEE Conference on Decision and Control (2015) 4232–4239.

[12] S. Sankaranarayanan, X. Chen, et al., Lyapunov Function Synthesis using Handelman Representations, IFAC Proceedings Volumes 46 (23) (2013) 576–581.

[13] A. Alfarano, F. Charton, A. Hayat, Global Lyapunov Functions: A Long-Standing Open Problem in Mathematics, with Symbolic Transformers, in: 38th Conference on Neural Information Processing Systems, 2024.

[14] R. Goebel, R. G. Sanfelice, A. R. Teel, Hybrid Dynamical Systems: Modeling, Stability, and Robustness, Princeton University Press, New Jersey, 2012.

[15] L. Lindemann, H. Hu, A. Robey, H. Zhang, D. V. Dimarogonas, S. Tu, N. Matni, Learning Hybrid Control Barrier Functions from Data, in: 4th Conference on Robot Learning, 2020.

[16] C. A. Montenegro G., S. Leudo, R. G. Sanfelice, A Data-Driven Approach for Certifying Asymptotic Stability and Cost Evaluation for Hybrid Systems, in: Proceedings of the 27th ACM International Conference on Hybrid Systems: Computation and Control, 2024.

[17] K. B. Athreya, S. N. Lahiri, Measure Theory and Probability Theory, Vol. 19, Springer-Verlag, 2006.

[18] H. L. Royden, P. M. Fitzpatrick, Real Analysis, 4th Edition, Prentice Hall, 2010.

[19] A. Klenke, Probability Theory: A Comprehensive Course, 3rd Edition, Springer, Cham, 2020.

[20] R. G. Sanfelice, Hybrid Feedback Control, Princeton University Press, 2021.

[21] R. T. Rockafellar, R. J.-B. Wets, Variational Analysis, Springer, Berlin, Heidelberg, 1998.

[22] R. Vershynin, High-Dimensional Probability : An Introduction with Applications in Data Science, Cambridge University Press, 2018.

[23] P. K. Wintz, R. G. Sanfelice, Relaxed Lyapunov Conditions for Compact Sets in Dynamical Systems, in: Proceedings of the IEEE American Control Conference, 2025.

[24] C. Kellett, A Compendium of Comparison Function Results, Mathematics of Control, Signals, and Systems 26 (3) (2014) 339–374.

[25] Y. C. Chang, N. Roohi, S. Gao, Neural Lyapunov Control, in: Conference on Neural Information Processing Systems, Proceedings of Machine Learning Research, 2019.

[26] L. Grüne, K. Worthmann, A Deep Neural Network Approach for Computing Lyapunov Functions for Nonlinear Systems, Journal of Computational Dynamics 8 (3) (2021) 233–259.

[27] H. Gouk, E. Frank, B. Pfahringer, M. J. Cree, Regularisation of Neural Networks by Enforcing Lipschitz Continuity, Machine Learning 110 (2021) 393–416.

[28] P. M. Esfahani, T. Sutter, J. Lygeros, Performance Bounds for the Scenario Approach and an Extension to a Class of Non-Convex Programs, IEEE Transactions on Automatic Control 60 (1) (2015) 46–58.

[29] M. Fazlyab, A. Robey, H. Hassani, M. Morari, G. J. Pappas, Efficient and Accurate Estimation of Lipschitz Constants for Deep Neural Networks, Curran Associates Inc., Red Hook, NY, USA, 2019.

[30] P. Pauli, A. Koch, J. Berberich, P. Kohler, F. Allgöwer, Training Robust Neural Networks Using Lipschitz Bounds, IEEE Control Systems Letters 6 (2022) 121–126.

[31] L. Lindemann, H. Hu, A. Robey, H. Zhang, D. Dimarogonas, S. Tu, N. Matni, Learning Hybrid Control Barrier Functions from Data, in: J. Kober, F. Ramos, C. Tomlin (Eds.), Proceedings of the 2020 Conference on Robot Learning, Vol. 155 of Proceedings of Machine Learning Research, PMLR, 2021, pp. 1351–1370.

[32] Y. Wang, W. Yin, J. Zeng, Global Convergence of ADMM in Nonconvex Nonsmooth Optimization, Journal of Scientific Computing 78 (2018) 29–63.

[33] J. Chai, R. G. Sanfelice, On Notions and Sufficient Conditions for Forward Invariance of Sets for Hybrid Dynamical Systems, in: Proceedings of the 54th IEEE Conference on Decision and Control, 2015, pp. 2869–2874.

[34] M. Maghenem, R. G. Sanfelice, Sufficient Conditions for Forward Invariance and Contractivity in Hybrid Inclusions using Barrier Functions, Automatica 124 (2021) 109328.

[35] W. M. Haddad, V. Chellaboina, Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach, Princeton University Press, 2011.

[36] F. Ferrante, R. G. Sanfelice, Cost Evaluation for Hybrid Inclusions: A Lyapunov Approach, in: Proceedings of the 57th IEEE Conference on Decision and Control, 2018, pp. 855–860.

[37] S. Jimenez Leudo, R. G. Sanfelice, Sufficient Conditions for Optimality and Asymptotic Stability in Two-Player Zero-Sum Hybrid Games, in: Proceedings of the 25th International Conference on Hybrid Systems: Computation and Control, 2022, pp. 1–11.

[38] S. Jimenez Leudo, R. G. Sanfelice, Sufficient Conditions for Optimality in Finite-Horizon Two-Player Zero-Sum Hybrid Games, in: Proceedings of the 61st IEEE Conference on Decision and Control, 2022, pp. 3268–3273.

[39] S. Jimenez Leudo, F. Ferrante, R. G. Sanfelice, On the Optimal Cost and Asymptotic Stability in Two-Player Zero-Sum Set-Valued Hybrid Games, in: Proceedings of the IEEE American Control Conference, 2024.

[40] C. Cai, A. R. Teel, Characterizations of Input-to-State Stability for Hybrid Systems, Systems and Control Letters 58 (1) (2009) 47–53.

[41] Z.-P. Jiang, Y. Wang, A Converse Lyapunov Theorem for Discrete-time Systems with Disturbances, Systems and Control Letters 45 (1) (2002) 49–58.

Appendix A. Complementary Results

Given a set $D \subset \mathbb{R}^n$, let us define the support function of D as

$$\nu_D(x) := \sup_{v \in D} \langle v, x \rangle \qquad \forall\, x \in \mathbb{R}^n.$$

The next result relates support functions and the Pompeiu-Hausdorff distance.

Lemma A.1. (Support function) *Let $A, B \subset \mathbb{R}^n$ be nonempty. Then, for any $x \in \mathbb{R}^n$ the following holds*

$$\big|\nu_A(x) - \nu_B(x)\big| \leq |x|\, d_H(A, B).$$

Proof. Pick $x \in \mathbb{R}^n$ and suppose that $\nu_A(x) \geq \nu_B(x)$. Then

$$\nu_A(x) - \nu_B(x) = \sup_{a \in A} \inf_{b \in B} \langle x, a - b \rangle.$$

By the Cauchy-Schwarz inequality:

$$\sup_{a \in A} \inf_{b \in B} \langle x, a - b \rangle \leq |x| \sup_{a \in A} |a|_B.$$

Since $\sup_{a \in A} |a|_B \leq d_H(A, B)$ by (2), we have:

$$\nu_A(x) - \nu_B(x) \leq |x|\, d_H(A, B).$$

Finally, it can be shown that the case where $\nu_B(x) \geq \nu_A(x)$ yields the same upper bound since d_H is symmetric, and this completes the proof. ■

Following [40, Lemma C.1.], and the comparison principle for systems in continuous time [5, Lemmas 3.4, 4.4], and for discrete-time systems [41, Lemma 4.3], we propose a comparison principle for hybrid systems that establishes a $\mathcal{KL}$ bound instead of a $\mathcal{KLL}$ bound.

Lemma A.2. (Hybrid comparison lemma) *Let a class-$\mathcal{K}$ function α defined on and a hybrid arc $y : \operatorname{dom} y \mapsto \mathbb{R}_{\geq 0}$ satisfy*

- *for each $j \in \mathbb{N}$ such that I_y^j has a nonempty interior $\operatorname{int} I_y^j$, we have, for almost all $t \in \operatorname{int} I_y^j$, $\dot{y}(t, j) \leq -\alpha(y(t, j))$,*
- *for all $(t, j) \in \operatorname{dom} y$ such that $(t, j+1) \in \operatorname{dom} y$, $y(t, j+1) - y(t, j) \leq -\alpha(y(t, j))$.*

Then, there exists a class-$\mathcal{KL}$ function β with $\beta(r, 0) = r$ and $\beta(r, t_0 + t_1 + j_0 + j_1) = \beta(\beta(r, t_0 + j_0), t_1 + j_1)$ such that $y(t, j) \leq \beta(y(0, 0), t + j)$ for all $(t, j) \in \operatorname{dom} y$.

Proof. The proof is developed based on the proof of [40, Lemma C.1.]. Without loss of generality, assume $\alpha(r) \leq r$ for all $r \geq 0$. Define $\eta_p : \mathbb{R}_{>0} \to \mathbb{R}$ by

$$\eta_p(a) := -\int_1^a \frac{1}{\alpha(r)} dr \qquad \forall a \in \mathbb{R}_{>0}.$$

This is a strictly decreasing differentiable function on $\mathbb{R}_{>0}$. Moreover, $\lim_{a \to 0} \eta_p(a) = \infty$. For any $y(0,0)$, the solution $y(t,j)$, where $\{t_j\}_{j=0}^{\sup_j \operatorname{dom} y}$ is a nondecreasing sequence associated with the hybrid time domain of y, satisfies

$$\int_{y(0,0)}^{y(t,j)} \frac{1}{\alpha(r)} dr = \sum_{k=0}^{j} \int_{y(t_k,k)}^{y(t_{k+1},k)} \frac{dr}{\alpha(r)} + \sum_{k=1}^{j} \int_{y(t_{k+1},k)}^{y(t_{k+1},k+1)} \frac{dr}{\alpha(r)} \tag{A.1}$$

By integrating over any $I_y^k = [t_k, t_{k+1}]$, with $k \in \{0, 1, 2, \ldots, j\}$, that has a nonempty interior intI_y^k, we have

$$\int_{t_k}^{t_{k+1}} \frac{\dot{y}(\tau, k)}{\alpha(y(\tau, k))} d\tau \leq -\int_{t_k}^{t_{k+1}} d\tau = -(t_{k+1} - t_k),$$

and by changing variables, the integral yields

$$\int_{y(t_k,k)}^{y(t_{k+1},k)} \frac{dr}{\alpha(r)} \leq -(t_{k+1} - t_k).$$

Also, by combining $y(t_{k+1}, k+1) - y(t_{k+1}, k) \leq -\alpha(y(t_{k+1}, k))$ and the monotone property of α, for any $k \in \{0, 1, 2, \ldots, j\}$, we obtain

$$\int_{y(t_{k+1},k)}^{y(t_{k+1},k+1)} \frac{dr}{\alpha(r)} \leq \int_{y(t_{k+1},k)}^{y(t_{k+1},k+1)} \frac{dr}{\alpha(y(t_{k+1}, k))} \leq -1.$$

Thus, from (A.1), it follows

$$-(\eta_p(y(t,j)) - \eta_p(y(0,0))) \leq -t - j \Leftrightarrow \eta_p(y(t,j)) - \eta_p(y(0,0)) \geq t + j.$$

Hence, since η_p is strictly decreasing,

$$y(t,j) \leq \eta_p^{-1}(\eta_p(y(0,0)) + t + j)$$

Notice that, if $y(0,0) = 0$, then $y(t,j) \equiv 0$, because $y = 0$ is an equilibrium point. For $(r,s) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$, define

$$\beta(r,s) := \begin{cases} 0 & \text{for } r = 0 \\ \eta_p^{-1}(\eta_p(r) + s) & \text{for } r > 0, \end{cases}$$

which is continuous. Then, $y(t,j) \leq \beta(y(0,0), t+j)$ for all $(t,j) \in \operatorname{dom} y$. Also, note that $\beta(r, t_0 + j_0 + t_1 + j_1) = \beta(\beta(r, t_0 + j_0), t_1 + j_1)$ and $\beta(r, 0) = r$. The function β is strictly increasing in r for each fixed s, because

$$\frac{\partial}{\partial r}\beta(r,s) = \frac{\alpha(\beta(r,s))}{\alpha(r)} > 0$$

and strictly decreasing in s for each fixed r, because

$$\frac{\partial}{\partial s}\beta(r,s) = -\alpha(\beta(r,s)) < 0.$$

Furthermore, $\beta(r,s) \to 0$ as $s \to \infty$. ■