Towards Learning and Verifying Maximal Neural Lyapunov Functions

Jun Liu, Yiming Meng, Maxwell Fitzsimmons, and Ruikun Zhou

Abstract— The search for Lyapunov functions is a crucial task in the analysis of nonlinear systems. In this paper, we present a physics-informed neural network (PINN) approach to learning a Lyapunov function that is nearly maximal for a given stable set. A Lyapunov function is considered nearly maximal if its sub-level sets can be made arbitrarily close to the boundary of the domain of attraction. We use Zubov's equation to train a maximal Lyapunov function defined on the domain of attraction. Additionally, we propose conditions that can be readily verified by satisfiability modulo theories (SMT) solvers for both local and global stability. We provide theoretical guarantees on the existence of maximal Lyapunov functions and demonstrate the effectiveness of our computational approach through numerical examples.

Index Terms— Learning, formal verification, neural networks, nonlinear systems, stability analysis, Zubov's theorem

I. INTRODUCTION

Recent advancements in neural networks and machine learning have revolutionized the landscape of computational research. With the availability of these tools, researchers have developed sophisticated machine learning models that can perform complex tasks sometimes beyond the reach of traditional methods. These models are used in various fields, including image and speech recognition, natural language processing, protein structure prediction and design, and even drug discovery.

Systems and control can potentially leverage the availability of these computational tools to aid the analysis and design of control systems. One of the longstanding challenges in nonlinear control is the construction of Lyapunov functions for stability analysis and controller design. Since Lyapunov's original work [13] over a hundred years ago, researchers have been searching for constructive approaches to designing Lyapunov functions, and both analytical [9], [18] and computational approaches [5], [6] have been investigated.

In this paper, we propose a neural network approach to constructing Lyapunov functions. In contrast to prior work, we aim to: 1) compute neural Lyapunov functions that can approximate the entire domain of attraction for an asymptotically stable compact set, and 2) formally verify the satisfaction of Lyapunov conditions, both locally and globally. Together, the proposed algorithms provide verified

regions of attraction that are close to the domain of attraction, as illustrated by the numerical examples.

A. Related Work

The computation of Lyapunov functions has a long history, and while it is impossible to mention all related work, we refer readers to a recent survey on computational methods for Lyapunov functions [6], as well as to a more recent survey [3] on the learning of neural Lyapunov functions.

We would like to highlight a few works that are closely related to ours and have inspired our work. The work in [11] proposed a data-driven approach for estimating the domain of attraction using Zubov's equation, which is a partial differential equation (PDE) that characterizes the maximal Lyapunov function [19] defined on the domain of attraction [21]. However, the approach taken in [11] does not directly solve Zubov's PDE. Instead, a purely data-driven approach is taken, which relies on simulations of trajectories.

In contrast, the work in [7] uses an approach closer to physics-informed neural networks (PINNs) [17] for approximating a solution to Zubov's equation. However, the approach in [7] is local in nature, and the Lyapunov conditions used to train the Lyapunov functions are essentially conditions for local exponential stability. Both papers [11] and [7] discuss the potential of using neural networks to break the curse of dimensionality in approximating Lyapunov functions, and our work shares the same optimism [16] without discussing this aspect. Our work shares the same optimism [16] but does not delve into this aspect. A potential future research area could be to leverage the compositional structure of the Lyapunov function for efficient verification.

The work in [10], even though focusing on sums of squares (SOS) approaches for approximating Lyapunov functions, is closely related to our work. The partial differential inequality constraint that the authors used to optimize the polynomial Lyapunov functions takes the form of Zubov's PDEs, although not explicitly mentioned as such in the paper.

Neural Lyapunov functions for control are investigated in [2], where the authors use satisfiability modulo theories (SMT) solvers to verify that Lyapunov conditions are met by neural Lyapunov functions. The work was extended to cover systems with unknown dynamics in [20], while still offering stability guarantees. The search for Lyapunov functions in [2], [20], similar to [7], is local in nature. Furthermore, in both [2], [20], SMT verification for Lyapunov stability conditions omits a region around the origin. We formally address this issue here as a side result.

II. PRELIMINARIES

A. Set Stability

Consider an autonomous nonlinear system of the form

$$\dot{x} = f(x), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is a locally Lipschitz function. We are interested in characterizing uniform asymptotic stability of any compact invariant set for system (1) using Lyapunov functions approximated by neural networks. We denote the unique solution to (1) from the initial condition $x(0) = x_0$ by $\phi(t, x_0)$ for $t \in J$, where J is the maximal interval of existence for ϕ.

Definition 1 (Invariant set): A set $\mathcal{A} \subseteq \mathbb{R}^n$ is said to be a *positively invariant set* of (1) if all solutions of (1) starting in $\mathcal{A}$ remain in $\mathcal{A}$ in positive time, i.e., $x_0 \in \mathcal{A}$ implies $\phi(t, x_0) \in \mathcal{A}$ for all $t \geq 0$.

Since we only consider positively invariant sets in this paper, we will simply refer to them as invariant sets. The distance of a point $x \in \mathbb{R}^n$ to a set $\mathcal{A} \subseteq \mathbb{R}^n$ is defined as $\|x\|_{\mathcal{A}} := \inf_{y \in \mathcal{A}} \|x - y\|$.

Definition 2 (Set stability): A set $\mathcal{A} \subseteq \mathbb{R}^n$ is said to be *uniformly asymptotically stable* (UAS) for (1) if the following two conditions are met:

1) For every $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|x_0\|_{\mathcal{A}} < \delta$ implies $\|\phi(t, x_0)\|_{\mathcal{A}} < \varepsilon$ for all $t \geq 0$; and
2) There exists some $\rho > 0$ such that, for every $\varepsilon > 0$, there exists some $T > 0$ such that $\|\phi(t, x_0)\|_{\mathcal{A}} < \varepsilon$ whenever $\|x_0\|_{\mathcal{A}} < \rho$ and $t \geq T$.

Note that if a closed set $\mathcal{A}$ is stable (i.e., satisfying item 1) in Definition 2), then it has to be positively invariant [1, Theorem 1.6.6]. While the definition above is for any set $\mathcal{A}$, it is of more interest to consider stability of closed sets, because stability may be trivially satisfied if the set $\mathcal{A}$ is not closed (see [1, Remark 1.6.7 and Example 1.6.8]).

The definition above requires that there exists a neighborhood of $\mathcal{A}$ that are uniformly attracted to $\mathcal{A}$. To capture the largest region that is attracted to a uniformly asymptotically stable set $\mathcal{A}$, the domain of attraction of $\mathcal{A}$ is defined as follows.

Definition 3 (Domain of Attraction): Given a set $\mathcal{A} \subseteq \mathbb{R}^n$ that is uniformly asymptotically stable for (1), the *domain of attraction* of $\mathcal{A}$ with respect to (1) is defined as

$$\mathcal{D}(\mathcal{A}) := \left\{ x \in \mathbb{R}^n : \lim_{t \to \infty} \|\phi(t, x)\|_{\mathcal{A}} = 0 \right\}.$$

Any invariant subset of $\mathcal{D}(\mathcal{A})$ is called a *region of attraction*.

We know that the domain of attraction of any set is an open set [1, Theorem 1.6.31].

We shall use the following standing assumption of the paper.

Assumption 1: We have a compact set $\mathcal{A}$ that is uniformly asymptotically stable for (1).

B. Maximal Lyapunov Function and Zubov's Theorem

The domain of attraction can be characterized by a maximal Lyapunov function [19] as described in the following theorem.

Theorem 1: Let $D \subseteq \mathbb{R}^n$ be an open set. Suppose that there exists a continuous function $V : D \to \mathbb{R}$ such that $\mathcal{A} \subseteq D$ and the following conditions hold:

1) V is positive definite on D with respect to $\mathcal{A}$, i.e., $V(x) = 0$ for all $x \in \mathcal{A}$ and $V(x) > 0$ for all $x \in D \setminus \mathcal{A}$;
2) the derivative of V along solutions of (1), given by

$$\dot{V}(x) := \lim_{t \to 0+} \frac{V(\phi(t, x)) - V(x)}{t},$$

is well defined for all $x \in D$ and satisfies

$$\dot{V}(x) = -\Phi(x), \tag{2}$$

where $\Phi : D \to \mathbb{R}$ continuous and positive definite with respect to $\mathcal{A}$;
3) $V(x) \to \infty$ as $x \to \partial D$ or $\|x\| \to \infty$. Then $D = \mathcal{D}(\mathcal{A})$.

Under mild assumptions, the existence of a maximal Lyapunov function on $\mathcal{D}(\mathcal{A})$ is also necessary [19], following a converse Lyapunov argument [14]. Furthermore, if f is Lipschitz continuous on $\mathcal{D}(A)$, then such a V can be chosen to be continuously differentiable[1].

Theorem 1 is closely related, and in fact equivalent, to Zubov's theorem [21] stated below.

Theorem 2: Let $D \subseteq \mathbb{R}^n$ be an open set containing $\mathcal{A}$. Then $D = \mathcal{D}(\mathcal{A})$ if and only if there exists two continuous functions $W : D \to \mathbb{R}$ and $\Psi : D \to \mathbb{R}$ such that the following conditions hold:

1) $0 < W(x) < 1$ for all $x \in D \setminus \mathcal{A}$ and $W(x) = 0$ for all $x \in \mathcal{A}$;
2) Ψ is positive definite on D with respect to $\mathcal{A}$;
3) for any sufficiently small $c_3 > 0$, there exist two positive real numbers c_1 and c_2 such that $\|x\|_{\mathcal{A}} \geq c_3$ implies $W(x) > c_1$ and $\Psi(x) > c_2$;
4) $W(x) \to 1$ as $x \to y$ for any $y \in \partial D$;
5) W and Ψ satisfy

$$\dot{W}(x) = -\Psi(x)(1 - W(x)), \tag{3}$$

where $\dot{W}$ is the derivative of W along solutions of (1) as defined in item 2 of Theorem 1.

Remark 1: Theorem 1 and Theorem 2 can be related by the following equation

$$W(x) = 1 - \exp(-\alpha V(x)), \tag{4}$$

for some constant $\alpha > 0$. It is easy to verify that V satisfying (2) implies

$$\begin{aligned}
\dot{W}(x) &= \alpha \exp(-\alpha V(x)) \dot{V}(x) \\
&= -\alpha \exp(-\alpha V(x)) \Phi(x) = -\alpha(1 - W(x)) \Phi(x),
\end{aligned}$$

---

[1]While the results in [19] were proved for case of an asymptotically stable equilibrium point for (1), they can be readily extended to set stability.

which verifies (3) with $\Psi = \alpha\Phi(x)$. Conversely, one can take

$$V(x) = -\ln(1 - W(x))/\alpha \qquad (5)$$

to verify (2) using (3). Clearly, the choice of such transformation from V to W, and vice versa, is not unique. For example, we can also define

$$W(x) = \tanh(\alpha V(x)). \qquad (6)$$

Then we also have

$$\begin{aligned}\dot{W}(x) &= \alpha(1 - \tanh^2(\alpha V(x)))\dot{V} \\ &= -\alpha(1 - W(x))(1 + W(x))\Phi(x),\end{aligned}$$

which verifies (3) with $\Psi = \alpha(1 + W(x))\Phi(x)$. It is obvious that there are infinitely many such scalar transformations (see Section III-B). Hence, Zubov's PDE (3) is not unique and we need to specify Ψ before solving it.

Remark 2: One potential advantage of Zubov's construction of a Lyapunov function, namely $W(x)$, is that it is bounded and its value approaches one as x approaches the boundary, whereas the function $V(x)$ in Theorem 1 has values approaching infinity as x tends to the boundary. A bounded function can offer potential advantages in numerical approximations, especially when we are interested in solving a PDE to find a Lyapunov function, as will be described later in this paper. The fact that $W(x)$ approaches a finite value as x approaches the boundary of its domain also makes it possible to extend its domain to the entire space $\mathbb{R}^n$ (or any desired set on which computation takes place).

Let $W : D \to \mathbb{R}$ be a function satisfying the conditions in Zubov's theorem, and then $D = \mathcal{D}(\mathcal{A})$. Suppose that the definition of W is extended to a set X containing $\mathcal{D}(\mathcal{A})$ with $W(x) = 1$ for all $x \in X \setminus \mathcal{D}(\mathcal{A})$. Clearly, the domain of attraction of $\mathcal{A}$ is completely characterized by the strict sublevel-1 set of W, i.e.,

$$\mathcal{D}(\mathcal{A}) = \{x \in X : W(x) < 1\}.$$

C. Problem Statement

The objective of this paper is to compute neural network approximations of a maximal Lyapunov function in the form of a function satisfying Zubov's PDE (3), and use these neural Lyapunov functions to certify regions of attraction for $\mathcal{A}$, which are characterized by sublevel sets of the neural Lyapunov functions. We rely on satisfiability modulo theories (SMT) solvers to verify that the Lyapunov conditions are met. Since verification is done over bounded domains and neural approximations offer guarantees over compact domains, we put forward the following assumption.

Assumption 2: We have $\mathcal{D}(\mathcal{A}) \subseteq X$ for some compact set X.

It is well known that the boundary of $\mathcal{D}(\mathcal{A})$ consists of complete trajectories of (1) [1]. For instance, in a two-dimensional system, a period orbit can bound a domain of attraction for a stable equilibrium point.

If the set $\mathcal{D}(\mathcal{A})$ is unbounded and suppose that $\mathcal{A}$ is contained in the interior of X, the proposed algorithms can still certify regions of attractions inside X, although losing guarantees of approximating $\mathcal{D}(\mathcal{A})$.

The main problem to be addressed is stated next.

Problem 1: Given a uniformly asymptotically stable set $\mathcal{A}$, compute neural Lyapunov functions that can certify regions of attraction for $\mathcal{A}$ with provable guarantees.

III. A CONSTRUCTIVE MAXIMAL CONVERSE LYAPUNOV FUNCTION

In this section, we use ideas from converse Lyapunov function theory to explicity construct a maximal Lyapunov-barrier function for a given uniformly asymptotically set $\mathcal{A}$ and establish its properties.

A. Construction of V

Let $\omega : \mathbb{R}^n \to \mathbb{R}$ be a continuous function that is positive definite with respect to $\mathcal{A}$. Define

$$V(x) = \int_0^\infty \omega(\phi(t, x))dt, \quad x \in \mathbb{R}^n, \qquad (7)$$

where if the integral diverges, we let $V(x) = \infty$.

The following is assumed for the function ω and the resulting function V defined by (7).

Assumption 3: The following items hold true:
1) For any $\delta > 0$, there exists $c > 0$ such that $\omega(x) > c$ for all $\|x\|_{\mathcal{A}} > \delta$.
2) There exists some $\rho > 0$ such that the integral $V(x)$ defined by (7) converges for all x such that $\|x\|_{\mathcal{A}} < \rho$.
3) For any $\varepsilon > 0$, there exists $\delta > 0$ such that $\|x\|_{\mathcal{A}} < \delta$ implies $V(x) < \varepsilon$.

Remark 3: We can show that Assumption 3 holds if the set $\mathcal{A}$ is locally exponentially stable and ω is locally Lipschitz. A typical choice for ω is given by $\omega(x) = \|x\|_{\mathcal{A}}$.

Under this assumption, we can verify that V satisfies the conditions for a maximal Lyapunov function stated in Theorem 1.

Proposition 1: The function $V : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ defined by (7) satisfies the following:
1) $V(x) < \infty$ if and only if $x \in \mathcal{D}(\mathcal{A})$;
2) $V(x) \to \infty$ as $x \to \partial\mathcal{D}(\mathcal{A})$;
3) V is positive definite with respect to $\mathcal{A}$;
4) V is continuous on $\mathcal{D}(\mathcal{A})$ and

$$\dot{V}(x) = -\omega(x), \qquad (8)$$

for all $x \in \mathcal{D}(\mathcal{A})$.

Proof: 1) Suppose $x \in \mathcal{D}(\mathcal{A})$. Let $\rho > 0$ be from Assumption 3. There exists some $T_x > 0$ such that $\|\phi(t, x)\|_{\mathcal{A}} < \rho$ for all $t \geq T_x$. It follows that

$$\begin{aligned}V(x) &= \int_0^{T_x} \omega(\phi(t, x))dt + \int_{T_x}^\infty \omega(\phi(t, x))dt \\ &= \int_0^{T_x} \omega(\phi(t, x))dt + \int_0^\infty \omega(\phi(t, \phi(T_x, x)))dt \\ &= \int_0^{T_x} \omega(\phi(t, x))dt + V(\phi(T_x, x)) < \infty,\end{aligned}$$

where we used finiteness of the first integral and Assumption 3 to conclude $V(\phi(T_x, x)) < \infty$. Now suppose that $x \notin \mathcal{D}(\mathcal{A})$. Then $\phi(t, x) \notin \mathcal{D}(\mathcal{A})$ for any $t \geq 0$. Since $\mathcal{D}(\mathcal{A})$ is open and $\mathcal{A}$ is compact, there exists some $\delta > 0$ such that $\|\phi(t, x)\|_{\mathcal{A}} > \delta$ for all $t \geq 0$. By Assumption 3, $V(x) = \infty$.

2) Let $\{x_n\}$ be a sequence such that $x_n \to y$ for some $y \in \partial \mathcal{D}(\mathcal{A})$. Choose $\delta \in (0, \rho)$, where $\rho > 0$ is from the definition of set stability of $\mathcal{A}$. Let T_n be the first time that $\|\phi(t, x_n)\|_{\mathcal{A}} \leq \delta$. Then by continuity, $\|\phi(T_n, x_n)\|_{\mathcal{A}} = \delta$ and for all $t \in [0, T_n)$, we have $\|\phi(t, x_n)\|_{\mathcal{A}} > \delta$ and $\omega(\phi(t, x_n)) > c$ for some $c > 0$ because of Assumption 3. Hence

$$V(x_n) = \int_0^\infty \omega(\phi(t, x_n)) dt \geq \int_0^{T_n} \omega(\phi(t, x_n)) dt \geq c T_n.$$

We can conclude $V(x_n) \to \infty$ if $T_n \to \infty$. Suppose that this is not the case. Then $\{T_n\}$ contains a bounded subsequence, still denoted by $\{T_n\}$, that converges to T. It follows that $\phi(T_n, x_n) \to \phi(T, y)$ and $\|\phi(T, y)\|_{\mathcal{A}} = \delta < \rho$. Hence $\phi(T, y) \in \mathcal{D}(\mathcal{A})$. It follows that $y \in \mathcal{D}(\mathcal{A})$. Since $\mathcal{D}(\mathcal{A})$ is open, this is a contradiction. We must have $V(x_n) \to \infty$ as $n \to \infty$.

3) Positive definiteness of V follows from positive definiteness of ω and by a continuity argument.

4) For $x \in \mathcal{D}(\mathcal{A})$, we have

$$\begin{aligned}
\dot{V}(x) &= \lim_{t \to 0+} \frac{V(\phi(t, x)) - V(x)}{t} \\
&= \lim_{t \to 0+} \frac{\int_0^\infty \omega(\phi(s, \phi(t, x))) ds - \int_0^\infty \omega(\phi(s, x)) ds}{t} \\
&= \lim_{t \to 0+} \frac{\int_0^\infty \omega(\phi(s + t, x)) ds - \int_0^\infty \omega(\phi(s, x)) ds}{t} \\
&= \lim_{t \to 0+} \frac{\int_t^\infty \omega(\phi(s, x)) ds - \int_0^\infty \omega(\phi(s, x)) ds}{t} \\
&= \lim_{t \to 0+} \frac{\int_t^0 \omega(\phi(s, x)) ds}{t} = -\omega(x).
\end{aligned}$$

The proof is complete. ∎

B. Construction of W

Let $\beta : [0, \infty) \to \mathbb{R}$ satisfy

$$\dot{\beta} = (1 - \beta) \psi(\beta) \quad \beta(0) = 0, \tag{9}$$

where $\psi : [0, \infty) \to \mathbb{R}$ is a locally Lipschitz function satisfying $\psi(s) > 0$ for $s > 0$.

Lemma 1: Any function satisfying (9) is continuously differentiable, strictly increasing, and satisfies $\beta(0) = 0$ and $\beta(s) \to 1$ as $s \to \infty$.

With V defined by (7) and satisfying the conditions in Proposition 1, let

$$W(x) = \begin{cases} \beta(V(x)), & \text{if } V(x) < \infty, \\ 1, & \text{otherwise,} \end{cases} \tag{10}$$

where $\beta : [0, \infty) \to \mathbb{R}$ satisfies (9).

We can verify the following properties for W.

Proposition 2: The function $W : \mathbb{R}^n \to \mathbb{R}$ defined by (10) satisfies the conditions in Theorem 2 on $\mathcal{D}(\mathcal{A})$. Furthermore, W is continuous on $\mathbb{R}^n$.

Proof: Items (1), (2), (4) of Theorem 2 directly follow from Proposition 1 and Lemma 1. To verify item (3), by Assumption 3, we know for any $\delta > 0$, there exists $c > 0$ such that $\omega(x) > c$ for all $\|x\|_{\mathcal{A}} > \delta$. By local Lipschitz of f, it takes at least $\tau > 0$ time for any solutions to reach from the level set $\{\|x\|_{\mathcal{A}} = 2\delta\}$ to $\{\|x\|_{\mathcal{A}} = \delta\}$. Hence for $V(x) \geq c\tau$ for all x such that $\|x\|_{\mathcal{A}} > 2\delta$. By the properties of β, item (3) of Theorem 2 holds for W. Continuity of W follows from its definition outside $\mathcal{D}(\mathcal{A})$ and item (4) of Theorem 2.

To check (5), taking the derivative of $W(x)$ along solutions of (1) gives

$$\begin{aligned}
\dot{W}(x) &= \dot{\beta} \dot{V} = -(1 - \beta(V(x))) \psi(\beta(V)) \omega(x) \\
&= -(1 - W(x)) \Phi(x),
\end{aligned}$$

where $\Phi(x) = \psi(\beta(V)) \omega(x)$. ∎

Remark 4: The constructions in Remark 1 are special cases of Proposition 2. In particular, (4) is given by setting $\psi(s) = \alpha$ for some constant $\alpha > 0$ and (6) is given by $\psi(s) = \alpha(1 + s)$.

Example 3.1 (A concrete example of W): Consider the scalar system $\dot{x} = -x + x^3$. It has three equilibrium points at $\{0, \pm 1\}$. The equilibrium point $\mathcal{A} = \{0\}$ is uniformly asymptotically stable with $\mathcal{D}(\mathcal{A}) = (-1, 1)$. Consider

$$V(x) = \int_0^\infty |\phi(t, x)|^2 dt,$$

where $\beta > 0$. By Proposition 1, we have

$$\dot{V}(x) = V'(x)(-x + x^3) = -x^2.$$

For $x \in (0, 1)$, assuming differentiability of V, we have

$$V'(x) = \frac{x}{1 - x^2}.$$

Integrating this with the condition $V(0) = 0$ gives

$$V(x) = -\frac{1}{2} \ln(1 - x^2).$$

Taking $W(x) = 1 - \exp(-\alpha V(x))$ as in (4), we obtain

$$W(x) = 1 - (1 - x^2)^{\frac{\alpha}{2}}, \quad \alpha > 0.$$

One can easily verify that $W(x)$ satisfies conditions in Zubov's theorem.

IV. TRAINING LYAPUNOV FUNCTIONS WITH PHYSICS-INFORMED NEURAL NETWORKS

Given the constructions of $V(x)$ and $W(x)$ in Section III, we propose a physics-informed neural network (PINN) approach to train neural Lyapunov functions that can approximately capture the domain of attraction. Furthermore, the PINN approach can be augmented with a data-driven approach to improve its performance.

In this work, a neural network is trained to solve the Zubov's PDE with a particular choice of $\Psi(x)$. For the numerical examples in this paper, we choose

$$\Psi(x) = \alpha(1 + W(x))\Phi(x), \qquad (11)$$

with $\Phi(x) = \|x\|_{\mathcal{A}}^2$. The details of constructing the loss function are discussed next.

Let $W_N(x; \theta)$ be a neural approximation for solving Zubov's PDE (3). The loss function consists of three terms:

$$\mathcal{L}(\theta) = L_r(\theta) + L_b(\theta) + L_d(\theta), \qquad (12)$$

where L_r is the residual error of the PDE, evaluated over a set of collocation points $S = \{x_i\}_{i=1}^{N} \subseteq X$, where X is assumed to be a compact set on which training takes place. For example, we can choose it to be the mean-square error defined by

$$L_r = \frac{1}{N} \sum_{i=1}^{N} (\nabla_x W_N(x_i; \theta) f(x_i) + \Psi(x_i)(1 - W_N(x_i; \theta)))^2. \qquad (13)$$

The loss L_b captures the boundary conditions. In this case, we want $W(x) = 1$ for $x \notin \mathcal{D}(\mathcal{A})$ and $W(x) = 0$ for all $x \in A$. While it may not be able to precisely define the boundaries, some approximation knowledge suffices. For example, if an over-approximation U of $\mathcal{D}(\mathcal{A})$ is known, the loss can be defined as

$$L_b = \frac{1}{N_b} \sum_{x_i \in S \cap (X \setminus U)} (W_N(x_i; \theta) - 1)^2, \qquad (14)$$

where N_b is the number of points in $S \cap X \setminus U$. If $\mathcal{A}$ is an equilibrium point, adding $W(0) = (0)$ alone may not be ideal. If the equilibrium point is exponentially stable, we know that locally $V(x)$ constructed in (7) with $\omega(x) = \|x\|^2$ is a Lyapunov function for exponential stability in the sense that

$$c_1 \|x\|^2 \leq V(x) \leq c_2 \|x\|^2, \qquad (15)$$

and

$$\dot{V}(x) \leq -c_3 \|x\|^3,$$

for x in a small neighborhood of the origin. Hence it is without loss of generality to assume that (15) holds locally, which translates to a boundary constraint to W as

$$\beta(c_1 \|x\|^2) \leq W(x) \leq \beta(c_2 \|x\|^2), \qquad (16)$$

where β is given by (9). As discussed in Remark 1, we can simply choose $\beta(s) = (\tanh(\alpha s))$ or $\beta(s) = 1 - \exp(-\alpha s)$ for some $\alpha > 0$, which needs to be consistent with the PDE that is being solved. For the numerical examples in this paper, we choose $\beta(s) = (\tanh(\alpha s))$, which is consistent with the PDE (3) with Ψ given by (11).

Remark 5: Inequality constraints can be encoded in the cost function by defining a loss term $L_+ = \max(h, 0)$ or $L_- = \min(h, 0)$. It is easy to see that $L_+ = 0$ if and only if $h \leq 0$, and similarly, $L_- = 0$ if and only if $h \geq 0$. It is suggested in [7] that relaxing a Zubov-type equation $\dot{V}(x) = -\|x\|^2$ to an inequality $\dot{V}(x) \leq -\|x\|^2$ can

improve the process of training V. However, in our examples, we did not notice any issues using physics-informed neural networks to solve the Zubov PDE directly. Solving the PDE more globally, as opposed to locally in the examples in [7], seems to help the training process. Moreover, to accurately capture the domain of attraction, solving Zubov's PDE, in one way or another, seems unavoidable.

Finally, L_d is a data loss defined by an optional set of data points $S_d = \{y_i\}_i^{N_d}$ on which approximate values $\hat{W}(y_i)$ for $W(y_i)$ can be obtained. According to the analysis in Section III, $W(y_i)$ can be approximated by simulating $\phi(t, y_i)$ for a sufficiently long period of time until it converges to the set $\mathcal{A}$, or when the value of $V(y_i)$ defined by the integral in (7) exceeds a certain threshold. To be consistent with the PDE that is being solved, we should use (6) to evaluate $\hat{W}(y_i)$ based on simulated data for $V(y_i)$, if $\beta(s) = (\tanh(\alpha s))$ is used to define the PDE loss in (13) and boundary constraints in (16). Based on the values of $\left\{\hat{W}(y_i)\right\}_i^{N_d}$, the data loss L_d can be defined as

$$L_d = \frac{1}{N_d} \sum_{i=1}^{N_d} (W_N(y_i; \theta) - \hat{W}(y_i))^2. \qquad (17)$$

Our numerical examples indicate that the data loss term is optional. Adding it, however, can potentially improve the accuracy of approximating the domain of attraction.

The training process is standard, and we omitted a detailed description of the algorithm. Interested readers can refer to an earlier preprint of the paper [12] for details.

V. FORMAL VERIFICATION OF NEURAL LYAPUNOV FUNCTIONS

One perceived weakness of neural network approximation is its lack of formal guarantees. To overcome this, formal verification of the learned neural Lyapunov functions is conducted using satisfiability modulo theories (SMT) solvers. In this section, we will discuss the details of verifying neural Lyapunov functions using SMT solvers, and for simplicity, we will limit our discussion to the case of a stable equilibrium point at the origin. We discuss verification of both local and global stabiltiy.

A. Verification of Local Stability and Region of Attraction

In this section, we assume that f is continuously differentiable. We further assume that the origin is exponentially stable, i.e., (1) admits a linearization

$$\dot{x} = Ax, \qquad (18)$$

where A is a Hurwitz matrix. Now write the nonlinear system (1) as

$$\dot{x} = Ax + g(x), \qquad (19)$$

where $g(x) = f(x) - Ax$ satisfies $\lim_{x \to 0} \frac{\|g(x)\|}{\|x\|} = 0$. The linear system (18) has a quadratic Lyapunov function $V(x) = x^T P x$ given by solving P from the Lyapunov equation

$$PA + A^T P = -Q, \qquad (20)$$

for any positive definite matrix Q. Consider two sets

$$S_r := \left\{ x \in \mathbb{R}^n : 2x^T P g(x) \leq r \|x\|^2 \right\}$$

and

$$\Omega_c := \left\{ x \in \mathbb{R}^n : x^T P x \leq c \right\}.$$

Suppose that we can choose $c > 0$ and $r > 0$ such that $\Omega_c \subseteq S_r$. Then for all $x \in \Omega_c$, we have

$$\begin{aligned} \dot{V}(x) &= x^T (PA + A^T P)x + 2x^T P g(x) \\ &= -x^T Q x + 2x^T P g(x) \\ &\leq (-\lambda_{\min}(Q) + r) \|x\|^2. \end{aligned}$$

If $r < \lambda_{\min}(Q)$, then we verified local stability of the origin and obtained Ω_c as a verified invariant set that is attracted to the origin. To summarize, the conditions required for verifying local stability of the origin and Ω_c being a local region of attraction are

$$r < \lambda_{\min}(Q), \tag{21}$$

$$x^T P x \leq c \Longrightarrow 2x^T P g(x) \leq r \|x\|^2. \tag{22}$$

The inequality (21) can be readily verified by computing the eigenvalue of Q and picking r accordingly. To verify (22), one could use sums of squares (SOS) or SMT verification. In this paper, we emphasize the use of SMT solvers because it can potentially handle a broader class of nonlinear functions [4]. However, at a glance, inequality (22) may not be amenable to SMT verification because the identity holds trivially at the origin. Hence δ-decidable tools such as dReal [4] may have trouble verifying (22) directly, as the origin always satisfies a δ-weakened version of (22) for any $\delta > 0$. Fortunately, a closer examination of the inequality (22) and the assumption on g reveals that, since g is continuously differentiable, we can leverage this fact to help verify (22). Indeed, by the mean value theorem, we have

$$Pg(x) = Pg(x) - Pg(0) = \int_0^1 P \cdot Dg(tx)dt \cdot x, \tag{23}$$

where Dg is the Jacobian of g given by $Dg = Df - A$. Clearly, $Dg(0) = Df(0) - A = 0$ by construction. Hence

$$\|Pg(x)\| \leq \sup_{0 \leq t \leq 1} \|P \cdot Dg(tx)\| \|x\|, \tag{24}$$

which implies

$$2x^T P g(x) \leq 2 \sup_{0 \leq t \leq 1} \|P \cdot Dg(tx)\| \|x\|^2.$$

As a result, to verify (22), we just need to verify

$$x^T P x \leq c \Longrightarrow 2 \sup_{0 \leq t \leq 1} \|P \cdot Dg(tx)\| \leq r. \tag{25}$$

Since $Dg(0) = 0$ and Dg is continuous, for any $r > 0$, one can always choose $c > 0$ sufficiently small such that (25) can be verified. While SMT solvers may not be able to directly handle matrix 2-norm in (25), other easily computable norms, such as the Frobenious norm, can be used to over-approximate $\|P \cdot Dg(tx)\|$.

Example 5.1 (Reversed Van der Pol): Consider the reversed Van der Pol oscillator

$$\dot{x}_1 = -x_2, \quad \dot{x}_2 = x_1 - (1 - x_1^2)x_2,$$

which has a stable equilibrium point at the origin. The linearization at the origin is given by (18) with $A = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$. Solving the Lyapunov equation (20) with $Q = I$ gives $P = \begin{pmatrix} 1.5 & -0.5 \\ -0.5 & 1 \end{pmatrix}$. With $r = 0.9999$, we can verify (25) with $c = 0.29$. The ellipsoid $\Omega_c = \left\{ x \in \mathbb{R}^n : x^T P x \leq c \right\}$ provides a verified local region of attraction for the origin.

B. Verification of Regions of Attraction in the Large

Having verified a local ROA, we can use the learned neural Lyapunov function W_N to enlarge the verified ROA. To do this, we can verifiy the following conditions using SMT:

$$c_1 \leq W_N(x) \leq c_2 \Longrightarrow \dot{W}_N \leq -\varepsilon, \tag{26}$$

$$W_N(x) \leq c_1 \Longrightarrow x^T P x \leq c, \tag{27}$$

where $\varepsilon > 0$, $c_2 > c_1 > 0$, and c and P are from Section V-A. We can immediately show the following result.

Proposition 3: If (26) and (27) hold for all $x \in X$ and the set $\mathcal{W}_{c_2} = \{x \in X : W_N(x) \leq c_2\}$ does not intersect with the boundary of X, then $\mathcal{W}_{c_2}$ is contained in the ROA of the origin.

Proof: A solution starting from $\mathcal{W}_{c_2}$ remains in $\mathcal{W}_{c_2}$ as long as the solution do not leave X. However, to leave X it has to cross the boundary of $\mathcal{W}_{c_2}$ first. This is impossible because of (26). Within $\mathcal{W}_{c_2}$, solutions converge to $\Omega_c = \left\{ x \in \mathbb{R}^n : x^T P x \leq c \right\}$ in finite time, which is a verified region of attraction. ∎

VI. NUMERICAL EXAMPLES

In this section, we provide numerical examples to demonstrate the proposed approach for learning and verifying neural Lyapunov functions. For consistency, the reported computation times were recorded on a MacBook Pro equipped with a 2GHz Quad-Core Intel Core i5 processor and 16GB of memory. However, computations testing the feasibility of the work were performed on clusters provided by the Digital Research Alliance of Canada. Code for the numerical examples can be found at `https://github.com/j49liu/cdc23-zubov-pinn`.

Example 6.1 (Reversed Van der Pol, continued): Revisit the reversed Van der Pol oscillator in Example 5.1. We know the real domain of attraction of the origin is bounded by a period orbit. To compute a neural Lyapunov function that can approximately capture the true ROA, we use a data-driven mechanism to generate data for the integral

$$V(x) = \int_0^\infty \|\phi(t, x)\|^2 \, dt$$

as described in Section IV. A total of $90,000$ uniformly spaced samples on the domain $X = [-2.5, 2.5] \times [-3.5, 3.5]$ are generated. The total time for data generation is $2,156$ seconds. The set X contains the domain of attraction as

a subset[2]. The terminating condition for numerical integration to determine V is either $\|\phi(T, x)\| \leq 10^{-3}$ or $\int_0^T \|\phi(t, x)\|^2 \, dt \geq 200$. We then set $W(x) = \tanh(\alpha V(x))$ with $\alpha = 0.1$. The parameters chosen are similar to those in [11], except that we choose a larger set of samples to demonstrate the limit of the data-driven approach compared to a physics-informed neural network approach that incorporates Zubov's PDE in the training process. The ellipsoidal local region of attraction verified in Example 5.1 is shown with the dashed line.

We trained neural networks of different sizes as detailed in Table I. A batch size of 32 collocation points was used, and the learning rate was chosen to be 10^{-3} after some testing. We terminated training when the mean-square training loss was smaller than 10^{-5} or after 200 epochs. For the purely data-driven approach, we run full gradient descent for 200 steps. The learned neural Lyapunov functions were verified using dReal with its Python binding [4]. The computation times for training and verification are recorded in Table I. Figure 1 depicts the result of training and verification with a learned Lyapunov function and a verified region of attraction close to the boundary of the domain of attraction.

In Table I, a "data-driven" approach refers to a network trained with 90,000 data samples without using Zubov's equation explicitly, even though as shown in III, the construction results in a solution to Zubov's equation. This is close to the method proposed in [11]. A PINN approach refers to learning a neural Lyapunov function only through evaluations of the residual errors at 90,000 collocation points. No data is required except for the evaluation of loss at these points. A "PINN + data" approach refers to a neural network trained with PINN and Zubov's PDE, but also incorporated a data loss in the training process. The number of data points chosen is 900, which is 1% of the total number of samples in a data-driven approach. It is evident from the results in this table that a PINN+data approach performs the best in terms of obtaining a larger certified region of attraction.

Example 6.2 (Polynomial system): We consider the polynomial system in [15]:

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -2x_1 + \frac{1}{3}x_1^3 - x_2.$$

The origin of this system is known to have an unbounded domain of attraction, delimited by the stable manifolds of the saddle equilibrium points at $(\pm\sqrt{6}, 0)$. We restricted our computations to the domain $[-6, 6] \times [-6, 6]$ and similarly generated 90,000 samples. The computational results for training and verification are summarized in Table II. Figure
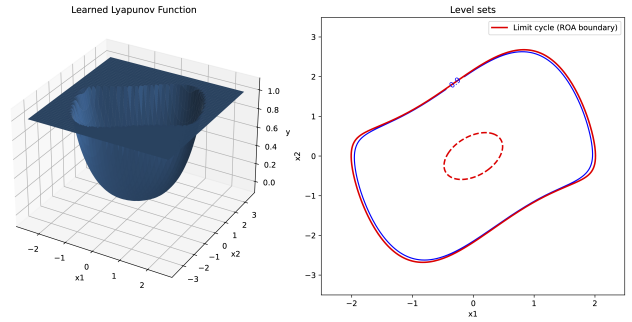

Fig. 1. A neural network, consisting of three hidden layers with 10 neurons per layer and trained using Zubov's PDE (3), is capable of approximately representing a verifiable Lyapunov function that closely approximates the domain of attraction for the stable equilibrium point of the Van der Pol equation. The dashed lines represent a local verified region of attraction computed in Example 5.1.

2 depicts the learned Lyapunov function, and the verified region of attraction using a neural network with three hidden layers of width 10. It can be seen that even though the actual domain of attraction is unbounded, the proposed algorithm approximates the largest region of attraction that can be certified by the choice of neural Lyapunov function. Similar to Example 5.1, we were able to verify a local region of attraction of the form $\Omega_c = \{x \in \mathbb{R}^n : x^T P x \leq c\}$, using the SMT solver dReal [4] by verifying (21) and (25) with $P = \begin{pmatrix} 1.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$, $Q = I$, $r = 0.9999$, and $c = 1.05$. The region Ω_c is depicted with a dashed line in Figure 2.
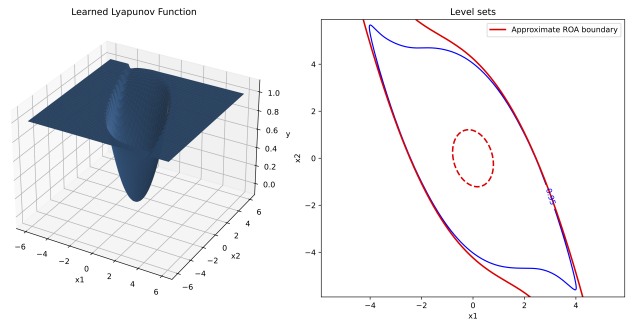

Fig. 2. A neural network, consisting of three hidden layers with 10 neurons per layer and trained using Zubov's PDE (3), is capable of representing a verifiable Lyapunov function that captures a region of attraction for the stable equilibrium point of the polynomial system in Example 6.2 inside the domain $[-6, 6] \times [-6, 6]$, even though the actual domain of attraction for the origin is unbounded. The dashed lines represent a local verified region of attraction.

---

[2]The domain of attraction can be estimated roughly by numerical simulations or observational data. If there is no prior knowledge of $\mathcal{D}(\mathcal{A})$, we can still arbitrarily specify a set X containing $\mathcal{A}$, assuming we know $\mathcal{A}$. Then we run into a similar situation as in an unbounded domain of attraction. As shown in Example 6.2, although we cannot approximate the entire domain of attraction, we can still obtain verified regions of attraction inside X.

[3]Training results may vary due to randomization. Note that, although not reported here, both data generation and dReal verification can be easily done in parallel to significantly reduce time. Furthermore, we only used CPU for training, and training time can be significantly improved by using GPUs. A bisection search can be used to determine the largest c_2 such that the level set $W_N(x) \leq c_2$ is a verified ROA.

VII. CONCLUSIONS

In this paper, we present a physics-informed neural network approach for learning nearly maximal Lyapunov functions using Zubov's equation. Our work provides theoretical justification on the existence of such Lyapunov functions that can be approximated by neural networks and verified using SMT solvers. The proposed algorithms allow for the computation of Lyapunov functions that can approximate

TABLE I

Approaches	Layer	Width	No. of params.	Data gen. time	Training time	Epochs	Final loss	Verification time	Verified level $(W_N(x) \le c_2)$	Volume %
Data-driven	2	10	151	2,156 (s)	326 (s)	200	9×10^{-5}	46 (s)	0.7	84.93%
PINN	2	10	151	0	636 (s)	200	11×10^{-5}	24 (s)	0.84	87.20%
PINN + data (1%)	2	10	151	216	943 (s)	200	56×10^{-3}	60 (s)	0.77	**87.45%**
Data-driven	2	30	1,051	2,156 (s)	363 (s)	200	3×10^{-5}	3,207	0.82	92.52%
PINN	2	30	1051	0	729 (s)	200	36×10^{-5}	538 (s)	0.86	91.55%
PINN + data (1%)	2	30	1051	216 (s)	1,293 (s)	200	15×10^{-5}	842 (s)	0.86	**94.22%**
Data-driven	3	10	261	2,156 (s)	381 (s)	200	10^{-5}	51,301 (s)	0.86	95.00%
PINN	3	10	261	0	129 (s)	32	9×10^{-6}	8,416 (s)	0.88	93.42%
PINN + data (1%)	3	10	261	216 (s)	593 (s)	107	8×10^{-6}	24,533 (s)	0.90	**96.31%**

TABLE II

Approaches	Layer	Width	No. of params.	Data gen. time	Training time	Epochs	Final loss	Verification time	Verified level $(W_N(x) \le c_2)$	Volume %
PINN + data	2	10	151	97 (s)	989 (s)	200	37×10^{-5}	16 (s)	0.91	74.15%
PINN + data	2	30	1,051	97 (s)	1,075 (s)	200	2×10^{-5}	295 (s)	0.95	**85.03%**
PINN + data	3	10	261	97 (s)	725 (s)	123	6×10^{-6}	2,990 (s)	0.95	84.92%

the entire domain of attraction for an asymptotically stable compact set. We also provide formal verification of the Lyapunov candidate generated by the neural network, both locally and globally, which allows us to find verified regions of attraction that are close to the domain of attraction. Our work demonstrates that combining a physics-informed neural network approach with a small number of data points can improve the results of learning and verification.

One potential limitation of the work is its scalability to high-dimensional systems, as is the case with other computational approaches to constructing Lyapunov functions. Future work could focus on investigating the scalability of the proposed approach. It is believed that a compositional structure can help break the curse of dimensionality in learning function approximations using neural networks [16]. Both [11] and [7], [8] have discussed this in the context of approximating Lyapunov functions. It would be interesting to investigate whether such compositionality can alleviate the curse of dimensionality in the formal verification of Lyapunov functions as well. Another interesting direction is to provide theoretical guarantees of learning a solution to Zubov's equation if one exists. Combining computation and verification of neural Lyapunov functions with a policy iteration scheme to provide verifiable guarantees of neural policy iteration is another promising research direction.

REFERENCES

[1] Nam P Bhatia and George P Szegö. *Dynamical Systems: Stability Theory and Applications*, volume 35. Springer, 1967.

[2] Ya-Chien Chang, Nima Roohi, and Sicun Gao. Neural Lyapunov control. In *NeurIPS*, pages 3245–3254, 2019.

[3] Charles Dawson, Sicun Gao, and Chuchu Fan. Safe control with learned certificates: A survey of neural Lyapunov, barrier, and contraction methods for robotics and control. *IEEE Transactions on Robotics*, pages 1–19, 2023.

[4] Sicun Gao, Soonho Kong, and Edmund M Clarke. dreal: an SMT solver for nonlinear theories over the reals. In *Proc. of CADE*, 2013.

[5] Peter Giesl. *Construction of Global Lyapunov Functions Using Radial Basis Functions*, volume 1904. Springer.

[6] Peter Giesl and Sigurdur Hafstein. Review on computational methods for Lyapunov functions. *Discrete & Continuous Dynamical Systems-B*, 20(8):2291, 2015.

[7] Lars Grüne. Computing Lyapunov functions using deep neural networks. *Journal of Computational Dynamics*, 8(2), 2021.

[8] Lars Grüne. Overcoming the curse of dimensionality for approximating Lyapunov functions with deep neural networks under a small-gain condition. *IFAC-PapersOnLine*, 54(9):317–322, 2021.

[9] Wassim M Haddad and VijaySekhar Chellaboina. *Nonlinear Dynamical Systems and Control: A Lyapunov-based Approach*. Princeton University Press, 2008.

[10] Morgan Jones and Matthew M Peet. Converse Lyapunov functions and converging inner approximations to maximal regions of attraction of nonlinear systems. In *Proc. of CDC*, pages 5312–5319. IEEE, 2021.

[11] Wei Kang, Kai Sun, and Liang Xu. Data-driven computational methods for the domain of attraction and Zubov's equation. *arXiv preprint arXiv:2112.14415*, 2021.

[12] Jun Liu, Yiming Meng, Maxwell Fitzsimmons, and Ruikun Zhou. Towards learning and verifying maximal neural Lyapunov functions. *arXiv preprint arXiv:2304.07215v1*, 2023.

[13] A. M. Lyapunov. The general problem of the stability of motion. *International Journal of Control*, 55(3):531–534, 1992.

[14] José L Massera. Contributions to stability theory. *Annals of Mathematics*, pages 182–206, 1956.

[15] Alexandre Mauroy and Igor Mezić. Global stability analysis using the eigenfunctions of the koopman operator. *IEEE Transactions on Automatic Control*, 61(11):3356–3369, 2016.

[16] Tomaso Poggio, Hrushikesh Mhaskar, Lorenzo Rosasco, Brando Miranda, and Qianli Liao. Why and when can deep-but not shallow-networks avoid the curse of dimensionality: a review. *International Journal of Automation and Computing*, 14(5):503–519, 2017.

[17] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

[18] Rodolphe Sepulchre, Mrdjan Jankovic, and Petar V Kokotovic. *Constructive Nonlinear Control*. Springer, 2012.

[19] Anthony Vannelli and Mathukumalli Vidyasagar. Maximal Lyapunov functions and domains of attraction for autonomous nonlinear systems. *Automatica*, 21(1):69–80, 1985.

[20] Ruikun Zhou, Thanin Quartz, Hans De Sterck, and Jun Liu. Neural Lyapunov control of unknown nonlinear systems with stability guarantees. In *NeurIPS*, 2022.

[21] V. I. Zubov. *Methods of A. M. Lyapunov and Their Application*. Noordhoff, 1964.