UNIVERSALITY OF AUTONOMOUS NEURAL ODES

Anonymous authors
Paper under double-blind review

ABSTRACT

We show that using composition of autonomous-neural-ODEs, one can universally approximate diffeomorphisms in the identity component defined on the cube with rate $\mathcal{O}(P^{-1/d})$ with P parameters. On the other hand, we show that by using only a single autonomous flow, the class of neural ODEs is nowhere dense on the cube in dimension $d \geqslant 2$. Under a compact-support$_{\mathrm{id}}$ condition on $(0,1)^d$, we show by a composition of at most I_d autonomous-neural-ODEs with the same support$_{\mathrm{id}}$, where I_d depends only on the dimension one can universally approximate compactly supported$_{\mathrm{id}}$ diffeomorphisms on $(0,1)^d$ for any dimension with rate $\mathcal{O}((\frac{P}{\log P})^{-2/d})$, and also compactly supported$_{\mathrm{id}}$ homeomorphisms on $(0,1)^d$ in dimension $d \geqslant 5$. Moreover, we show that the class of single autonomous-neural-ODEs compactly supported$_{\mathrm{id}}$ on $(0,1)^d$ is meagre in the space of compactly supported$_{\mathrm{id}}$ homeomorphisms on $(0,1)^d$ for $d \geqslant 2$. By linearly lifting, we obtain a universal approximation result for Lipschitz functions compactly supported on $(0,1)^d$ with rate $\mathcal{O}(P^{-1/(d+1)})$ with P parameters.

1 INTRODUCTION

Autonomous-flow-based generation has revolutionized areas including computational anatomy and image registration (Thottupattu et al., 2022);Bostelmann et al. (2026), manifold learning Hanson & Raginsky (2025);(Vigouroux et al., 2026), kinematic calibration in robotics Lafferriere & Sussmann (1991); Park & Okamura (1994); Wang et al. (2012); Rana et al. (2020), ResNets and neural ODEs Scagliotti (2023b); Kuehn & Wöhrer (2026); Zhang et al. (2020); Honda (2026). Though there have been some studies on approximation of diffeomorphisms Arango & Gómez (2002); Gelfreich & Vieiro (2018) and in machine learning Scagliotti (2023a) using non-autonomous neural ODE Duan & Cai (2025), no one has proved universality of autonomous-neural-ODE generation for the group of diffeomorphisms or homeomorphisms through autonomous flows and autonomous-flow-based generation.

One can consider flows as solution of Cauchy ODE systems of vector fields (for a standard definition of flow refer to for example Lee (2013, Chapter 9)). As an example, consider the solution of the Cauchy ODE system below:

$$\frac{dx_t^x}{dt} = V(x_t^x), \quad x_0^x = x, \tag{1}$$

where $V : \mathbb{R}^d \to \mathbb{R}^d$ is at least Lipschitz. By the Picard–Lindelöf theorem it has a unique solution. We consider the flow at a fixed time, which we take without loss of generality, to be time-one. The resulting time-one transformation $\varphi \stackrel{\text{def.}}{=} x_1^x$ is then *Lipschitz continuous*. This transformation has a Lipschitz inverse φ^{-1} with infinitesimal generator $-V$.

An *autonomous flow* refers to a flow having a time-independent infinitesimal generator. In this paper we use these facts to analyze the limitations of using single autonomous flows. Namely, we show that single neural ODEs are nowhere dense when considering the cube, and meagre when considering the compact support$_{\mathrm{id}}$ case.

On the other hand, it is well known that one can represent diffeomorphisms as a composition of time-one flows Grong & Schmeding (2025); Lukatskii (1978); Ishikawa et al. (2023); Agrachev & Caponigro (2009). In this paper we prove it for compactly supported$_{\mathrm{id}}$ diffeomorphisms on $(0,1)^d$ and show that, by considering this representation, one can have universality for neural ODEs. Moreover, in this support$_{\mathrm{id}}$, we show that the minimal number of time-one autonomous flows which is

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

needed to represent is bounded by a constant I_d depending only on the dimension d. Furthermore, we use these facts to approximate compactly supported$_{\mathrm{id}}$ homeomorphisms on $(0,1)^d$ when $d \geqslant 5$ using composition of time-one autonomous flows thereby obtaining a universal approximation by neural ODEs. Unfortunately, our approach doesn't extend to the homeomorphisms defined on the cube, as we will show that diffeomorphisms on the cube are not dense in the space of homeomorphisms defined on the cube. Although we will show one can quantitatively approximate diffeomorphisms on the cube, we cannot prove whether there exists any similar constant I_d.

Traditionally, in approximation-theoretic deep learning theorems, they consider the cube $[0,1]^d$ as the domain of the maps Mhaskar & Poggio (2016); Yarotsky (2017); Petersen & Voigtlaender (2018); Elbrächter et al. (2021); Kratsios & Zamanlooy (2022); Zhang et al. (2022); Siegel (2023). This is because analytically, as it is compact and bounded, norms and approximation rates are controlled, and extension and rescaling arguments are available. Moreover, geometrically and computationally, its product structure makes grids, ReLU constructions, triangulations, and coordinate-wise arguments easy. In our paper, to study functions on a compactly supported$_{\mathrm{id}}$ domain defined on $\mathbb{R}^d$, because of technical topological and analytical convenience, (e.g. to use a ReLU bump function, BIP-meagreness etc.) we will consider the quite similar domain $(0,1)^d$. This means, we will study maps φ that are compactly supported$_{\mathrm{id}}$ on $(0,1)^d$; This is also compatible with applications considering group of diffeomorphisms DUPUIS et al. (1998) where they consider open bounded subset of $\mathbb{R}^d$.

Note that by Alexander's trick Alexander (1923) every such homeomorphism must be orientation-preserving. We denote the sets of compactly supported$_{\mathrm{id}}$ orientation-preserving homeomorphisms and diffeomorphisms on $(0,1)^d$ by $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$ and $\mathrm{Diff}^\infty\left(\mathbb{R}^d,(0,1)^d\right)_0$ respectively. Note that these are not vector spaces. It is known that $\mathrm{Diff}^\infty\left(\mathbb{R}^d,(0,1)^d\right)_0$ is a topological $l^2\times\mathbb{R}^\infty$-manifold equipped with the Whitney topology. This is only true for $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$ when $d=2$ Banakh et al. (2011). Note that $\mathbb{R}^\infty$ is a non-metrizable LF-space Banakh & Repovš (2012), which places it outside the scope of the available constructive approximation toolbox Pinkus (1985); Lorentz et al. (1996); Cohen et al. (2022); Petrova & Wojtaszczyk (2023), where they assume a Banach space structure. On the other hand $\mathrm{Diff}^\infty\left(\mathbb{R}^d,(0,1)^d\right)_0$ and $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$ possess a rich and well-studied infinite-dimensional topological group structure Brechner & Kawamura (2001); Mather (1974; 1984); Thurston (1974); Fathi (1980); Banyaga (1997); Haller et al. (2013);(Fukui et al., 2019). Moreover, unlike $\mathrm{Diff}^\infty\left(\mathbb{R}^d,(0,1)^d\right)_0$ which is a smooth regular infinite-dimensional Lie group Michor (2016; 1980) and the global exponential map exists (in the sense of infinite-dimensional Lie groups; cf. Kriegl & Michor (1997); Bauer et al. (2025)), for $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$ we do not have access to a global exponential map to transfer results from linear spaces onto $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$, (as in Kratsios & Bilokopytov (2020); Kratsios & Papon (2022)), to apply signature-based machine learning methods, (Cass & Salvi, 2024);Bayer et al. (2025), rooted in rough path theory Lyons (1998); Gubinelli (2004); Hambly & Lyons (2010). In short, there are currently no available tools to apply classical approximation results from linear spaces onto $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$ using a global exponential map. We are required to develop new approximation-theoretic techniques suited to its non-vectorial structure, where our approximation, or hypothesis, class consists only of (neural network-based) homeomorphisms in $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$.

One might already know that autonomous-flow-based generation can be represented as a single *non-autonomous flow*, i.e., where the vector field in equation 1 is allowed to be time-dependent but undergoes only finitely many changes in direction, which directly allows the conversion of our theory of autonomous-flow-based generation to time-dependent generation. This means one can write any autonomous-flow-based generation as time-one of a time-dependent vector field, and it is common to use a time-dependent vector field in neural ODEs Chen et al. (2018). Moreover, one can write any C^1-diffeomorphism isotopic to the identity as a time-one of a time-dependent vector field Banyaga (1997), so time-dependent vector fields do not lack expressive power. On the other hand, *controlled neural ODEs* Kidger et al. (2020); Cuchiero et al. (2020); Muca Cirone et al. (2023), which mimic controlled rough paths Morrill et al. (2021); Walker et al. (2024) can potentially adapt their vector fields arbitrarily often over time, resulting in a higher parametric complexity but more expressive power. Moreover one can also write any autonomous flow using a one-dimensional control. Now one might wonder:Is there any advantage of considering the autonomous-flow-based generation at all?

The class of functions being used matters; if one only wants to learn or approximate a group of homeomorphisms or diffeomorphisms for example, defining additional controls doesn't help and the machinery in rough analysis is then useless. On the other hand, despite time-dependent vector fields, the autonomous restriction, which produces building blocks, might be easier to consider in practice, for example in kinematic robotics, it is easier to think of autonomous vector fields for each screw motion, but time-dependent vector fields might be too complicated and unnecessary to consider. Moreover, autonomous blocks have much more algebraic structure so one can utilize the tools in Lie groups, diffeomorphism groups, exponential maps, fragmentation, generation of groups, splitting methods, etc. Moreover, for non-autonomous flows, one doesn't have the nice one-parameter group identity, which yields a stationary transport or conservation equation (Cheuk-Kit et al., 2026). In the case of approximating with neural ODEs, the only difference is that the same parameter θ in a time-independent vector field V_θ determines the whole dynamics throughout the entire interval, but for a time-dependent vector field, the network must additionally model the dependence on t. One has to mention that in high dimensions, changing d to $d+1$ is negligible and there is no advantage in considering autonomous vector fields.

We show that autonomous-flow-based generation has enough expressivity and as a consequence we approximate them with neural ODEs, as this is the natural neural Network framework for this context.

1.1 Main Contributions

For Compactly Supported$_{\mathrm{id}}$ case, our negative result (theorem 3.1) shows that in autonomous-flow-based generation for neural ODEs, a smaller compact support$_{\mathrm{id}}$ makes it nowhere dense in the space of the larger support$_{\mathrm{id}}$. Moreover, we prove that for the same support$_{\mathrm{id}}$, it is meagre (theorem 3.2).

Our second result (in theorem 3.3) shows for any diffeomorphism $\varphi \in \mathrm{Diff}^\infty\left(\mathbb{R}^d,(0,1)^d\right)_0$, there exists a representation of it as a composition of at most $I_d \geqslant 1$ time-one autonomous flows such that I_d only depends on the dimension d and doesn't depend on φ. Our result for diffeomorphisms (theorem 3.4) shows that by only considering composition, neural ODEs of the same support$_{\mathrm{id}}$ are universal in the space of $\mathrm{Diff}^\infty\left(\mathbb{R}^d,(0,1)^d\right)_0$ quantitatively. Our result for homeomorphisms (theorem 3.6) shows that again by considering composition, autonomous neural ODEs are universal in the space of $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$ for $d \geqslant 5$. Finally, we prove that one can approximate compactly supported Lipschitz functions between arbitrary dimensions (corollary 3.7).

For defined on the Cube, our first result (theorem 4.1) shows that the class of single autonomous neural ODEs is nowhere dense. Our second result (theorem 4.4) shows that by considering composition, neural ODEs are universal in the space of diffeomorphisms on the cube.

2 Preliminaries

2.1 Background

Isotopy and Orientation-Preserving: We will say that two homeomorphisms are **isotopic** if one can be transformed continuously into the other with homeomorphisms.

Example 2.1. A simple example of an isotopy on $\mathbb{R}^2$ is given by:

$$H_t(x,y) = \begin{pmatrix} \cos(t\theta) & -\sin(t\theta) \\ \sin(t\theta) & \cos(t\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \qquad t \in [0,1].$$

For every $t \in [0,1]$, the map H_t is a homeomorphism. Thus, $(H_t)_{t\in[0,1]}$ is an isotopy from the identity map to a rotation.

A C^k-diffeomorphism is **orientation-preserving** if it preserves local orientations, equivalently, if its Jacobian determinant is positive in oriented coordinates; for homeomorphisms it is formulated using local degree or local orientation classes, cf. Hirsch (1976). In Euclidean space, it can be considered as being isotopic to the identity as the set of orientation-preserving homeomorphisms and the identity component are the same.

Example 2.2 (Orientation-Preserving Homeomorphisms From Computer Vision (Rotations))**.** Standard multidimensional examples arising in rotation invariances in computer vision, e.g. Lui

(2012); Cohen & Welling (2016); (Thomas et al., 2018), include the linear transformation $\varphi : \mathbb{R}^d \ni x \mapsto Ox \in \mathbb{R}^d$ where O is a $d \times d$ orthogonal matrix; in which case φ is orientation-preserving if and only if $\det(O) = 1$ and it fails to be precisely when $\det(O) = -1$.

If O is orientation-preserving; i.e. if $\det(O) = 1$, then any such homeomorphism can be expressed as the solution to an ordinary differential equation (ODE) at time 1; namely, $\varphi(x) = Ox = x_1^x$ where:

$$\begin{aligned} \tfrac{d}{dt}\, x_t^x &= \mathfrak{o} x_t^x \\ x_0^x &= x, \end{aligned} \tag{2}$$

where $O = \exp(\mathfrak{o})$ for some $d \times d$-skew-symmetric matrix $\mathfrak{o}$; where $\exp$ is the matrix exponential. This can be seen by observing that the ODE solution to the ODE equation 2 is given by the curve $x_\cdot^x = (x_t^x)_{t \geqslant 0}$ where

$$x_t^x = \exp(t\mathfrak{o})\, x.$$

This connection is the starting-point of the theory of finite-dimensional Lie groups; cf. Helgason (1978). Indeed, finite-dimensional Lie groups induce the prototypical and simplest classes of "well-behaved" homeomorphisms on $\mathbb{R}^d$ in this way; Note that one can make any rotation compactly $\text{supported}_{\text{id}}$ by multiplying its vector field by a suitable smooth bump function.

2.2 Notation

2.2.1 Used notations and definitions:

For reference, we collect the notation and definitions used in this paper:

- Let $\mathbb{N} \overset{\text{def.}}{=} \{0, 1, 2, \dots, \}$, $\mathbb{N}_+ \overset{\text{def.}}{=} \{n \in \mathbb{N} : n > 0\}$ and for $N \in \mathbb{N}_+$ denote $[N] \overset{\text{def.}}{=} \{1, \dots, N\}$, and let $\mathbb{R}_+$ be the set of positive real numbers.
- For a Lipschitz function f we denote its Lipschitz constant as L^f.
- We denote the i-th component of a function $f : \mathbb{R}^d \to \mathbb{R}^D$ where $f = (f_1, \dots, f_D)$ as $(f)_i \overset{\text{def.}}{=} f_i$ for $i \in [D]$.
- Given I functions $\{f_1, \dots, f_I\}$, the iterated composition operator $\bigcirc$ maps any finite set of composable functions $f_1, \dots, f_I$ to their composition $\bigcirc_{i=1}^I f_i \overset{\text{def.}}{=} f_I \circ \cdots \circ f_1$.
- Given $E \subseteq \mathbb{R}^d$, d-tuple $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_d]^\top \in \mathbb{N}^d$ and functions $f(x) : \mathbb{R}^d \to \mathbb{R}^D$, $g : \mathbb{R}^d \to \mathbb{R}$ and $\text{id} : \mathbb{R}^d \to \mathbb{R}^d$ let:
  - $\|x\|_{l^\infty} \overset{\text{def.}}{=} \max_{i \in [d]} |x_i|$,
  - $\|g\|_{L^\infty(E)} \overset{\text{def.}}{=} \text{ess sup}_{x \in E} |g(x)|$,
  - $\|f\|_{L^\infty(E; l^\infty)} \overset{\text{def.}}{=} \left\| \text{ess sup}_{x \in E} |f_1(x)|, \dots, \text{ess sup}_{x \in E} |f_D(x)| \right\|_{l^\infty} = \max_{i \in [D]} \text{ess sup}_{x \in E} |f_i(x)|$,
  - $\|\cdot\|_2$ denotes the Euclidean norm,
  - $\|\alpha\|_1 \overset{\text{def.}}{=} |\alpha_1| + |\alpha_2| + \cdots + |\alpha_d|$,
  - $\partial^\alpha \overset{\text{def.}}{=} \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}}$,
  - $\|g\|_{C^s(E)} = \max \left\{ \|\partial^\alpha g\|_{L^\infty(E)} : \alpha \in \mathbb{N}^d \text{ with } \|\alpha\|_1 \leqslant s \right\}$.[1]
- Let $I = [0, \infty)$ or $I = [0, T]$ for some $T \geqslant 0$. A function $\omega \colon I \to [0, \infty)$ is called a *modulus of regularity* if ω is monotone increasing, concave and $\omega(0) = 0$.

Given a subset $P \subseteq M$ of a manifold M, for $k \in \mathbb{N}_+ \cup \{\infty\}$:

- We denote the interior of P by $int(P)$.
- Given a function $f : M \to \mathbb{R}^D$, we denote its support if it is well defined, by $\text{supp}(f) \overset{\text{def.}}{=} \overline{\{x \in M : f(x) \neq 0\}}$. For a function $f : M \to M$ its $\text{support}_{\text{id}}$ if it is well defined is denoted as $\text{supp}_{\text{id}}(f) \overset{\text{def.}}{=} \overline{\{x \in M : f(x) \neq x\}}$.

[1] Note that $\|g\|_{C^0} = \|g\|_{L^\infty}$

- Let $\mathcal{H}(M,P)_0, \mathrm{Diff}^k(M,P)_0$ be the set of homeomorphisms and C^k-diffeomorphisms in the identity component of M while compactly supported$_{\mathrm{id}}$ on P. Let $\mathcal{H}(M,P)_+, \mathrm{Diff}^k(M,P)_+$ be the set of orientation-preserving homeomorphisms and C^k-diffeomorphisms of M compactly supported$_{\mathrm{id}}$ on P. For simplicity, when there is no confusion, we might write $\mathcal{H}(P,P)_0$ and $\mathrm{Diff}^k(P,P)_0$ as $\mathcal{H}(P)_0$ and $\mathrm{Diff}^k(P)_0$.
- Let $\mathcal{X}^k(M,P)$ and $\mathcal{X}^0(M,P)$ be the set of C^k and *Lipschitz* continuous vector fields on M compactly supported$_{\mathrm{id}}$ on P respectively. For simplicity, when there is no confusion, we might write $\mathcal{X}^k(P,P)$ and $\mathcal{X}^0(P,P)$ as $\mathcal{X}^k(P)$ and $\mathcal{X}^0(P)$.
- For a given vector field $V \in \mathcal{X}^0(M,P)$, let $\mathrm{Flow}(V)$ denote the time-one map of the corresponding C^1 integral curve γ whose ODE system can be written in coordinates as:

$$\begin{cases} \dot{x}_t^x &= V(x_t^x), \\ x_0^x &= x. \end{cases} \tag{3}$$

- Fix $I \in \mathbb{N}_+$, $k \in \mathbb{N}$. We say $\varphi \in \mathcal{H}^I(M,P)_0$ if there exist vector fields $V_1, \ldots, V_I \in \mathcal{X}^0(M,P)$ such that:

$$\varphi = \bigcirc_{i=1}^{I} \mathrm{Flow}(V_i). \tag{4}$$

  and I is minimal among all of the possible representations. We say $\varphi \in \mathrm{Diff}^{k,I}(M,P)_0$ if there exist vector fields $V_1, \ldots, V_I \in \mathcal{X}^k(M,P)$ such that:

$$\varphi = \bigcirc_{i=1}^{I} \mathrm{Flow}(V_i). \tag{5}$$

  and I is minimum among all of the possible representations. For simplicity, when there is no confusion, we might write $\mathcal{H}^I(P,P)_0$ and $\mathrm{Diff}^{k,I}(P,P)_0$ as $\mathcal{H}^I(P)_0$ and $\mathrm{Diff}^{k,I}(P)_0$.
- A function $f : M \to M$ is called **flowable** if there exists a vector field V on M such that $f = \mathrm{Flow}(V)$.

Let (M,g) be a connected Riemannian manifold:

-
$$d_{C^0}(f,g) \overset{\text{def.}}{=} \sup_{x \in M} d_g\big(f(x), g(x)\big).$$

For additional notations and definitions refer to Appendix

2.2.2 Neural ODEs

Let $\Delta \in \mathbb{N}_+$ and consider a multi-index $\mathbf{d} \overset{\text{def.}}{=} [d_1, \ldots, d_{\Delta+1}]^\top$ such that $d_i \in \mathbb{N}_+$ for $i \in [\Delta+1]$. The class $\mathcal{NN}(\mathbf{d})$ consists of all multilayer perceptrons with a Lipschitz activation function $\sigma \in C(\mathbb{R})$ (σ-MLPs):

$$\Phi : \mathbb{R}^{d_1} \to \mathbb{R}^{d_{\Delta+1}},$$

admitting the following iterative representation:

$$\begin{aligned} \Phi(\mathbf{x}) &= \mathbf{W}^{(\Delta)}\mathbf{x}^{(\Delta)} + \mathbf{b}^{(\Delta)}, \\ \mathbf{x}^{(l+1)} &\overset{\text{def.}}{=} \sigma \bullet \big(\mathbf{W}^{(l)}\mathbf{x}^{(l)} + \mathbf{b}^{(l)}\big), \qquad \text{for } l = 1, \ldots, \Delta - 1, \\ \mathbf{x}^{(1)} &\overset{\text{def.}}{=} \mathbf{x}\,. \end{aligned} \tag{6}$$

Here, for $l \in [\Delta]$, $\mathbf{W}^{(l)}$ is a $d_{l+1} \times d_l$ matrix and $\mathbf{b}^{(l)} \in \mathbb{R}^{d_{l+1}}$, and $\sigma \bullet$ denotes componentwise application of the σ function.

A σ-neural ODE of respective depth and width $\Delta, W \in \mathbb{N}^+$ is a flowable homeomorphism $\Psi \in \mathcal{H}\big(\mathbb{R}^d\big)_0$ for which there exists a σ-MLP $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ of depth Δ and width W such that $\Psi = \mathrm{Flow}(\Phi)$. We denote the class of all σ-neural ODEs compactly supported$_{\mathrm{id}}$ on $P \subseteq \mathbb{R}^d$ as $\mathrm{NODE}_\sigma(\mathbb{R}^d, P)$. For simplicity, when there is no confusion, we might write $\mathrm{NODE}_\sigma(P,P)$ as $\mathrm{NODE}_\sigma(P)$.

3 First Case : Compactly Supported$_{\mathrm{id}}$ on $(0,1)^d$

3.1 Limitations of Autonomous Flows and Autonomous NODEs

3.1.1 Negative Results

We show that the set $\mathrm{NODE}_\sigma(\mathbb{R}^d,(0,1)^d)$ is meagre in $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$. The reason that the following theorem is stated in $\mathbb{R}^d$ and not on a general C^∞ manifold is that $\mathcal{H}\left(\mathbb{R}^d,\mathbb{R}^d\right)_0=\mathcal{H}(\mathbb{R}^d,\mathbb{R}^d)_+$ and lemma A.4 yields an orientation-preserving homeomorphism, not necessarily in the identity component. Note that naturally, time-one flows are in the identity component. The theorem is as follows cf. Bonomo & Varandas (2020, Theorem A):

Theorem 3.1. *Let $d\in\mathbb{N}_+$ with $d>1$. Fix compact sets $K\subseteq Q\subseteq\mathbb{R}^d$ with path connected interiors such that K is compactly contained in int(Q) and K has non-empty interior. Then the subset of flowable $\mathcal{H}(\mathbb{R}^d,K)$ is nowhere dense in $\mathcal{H}\left(\mathbb{R}^d,Q\right)_0$ in C^0-topology.*

Theorem 3.2. *Let $d\in\mathbb{N}_+$ with $d>1$, then the subset of flowable $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$ is meagre in $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$ in C^0-topology.*

This means for any $d\in\mathbb{N}_+$ with $d>1$ and Lipschitz activation function σ, $\mathrm{NODE}_\sigma(\mathbb{R}^d,(0,1)^d)$ which is a subset of flowable $\mathcal{H}(\mathbb{R}^d,(0,1)^d)$ is meagre in $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$. Note that meagreness does not rule out universal approximation. Note that by the above theorem, the subset of flowable $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$ is meagre in itself and as a result, it is not a Baire space. On the other hand, as the complete metric space $\mathcal{H}(\mathbb{R}^d,[0,1]^d)$ is a Baire space Willard (2004, Corollary 25.4), the complement of flowable $\mathcal{H}(\mathbb{R}^d,(0,1)^d)$ is a dense subset Oxtoby (1971, Theorem 9.1.), and thus a potential for universal approximation.

3.2 Universal Approximation Gaurantees:

3.2.1 Quantitative Result:

Using the following theorem, we give an order of approximation for any $\mathrm{Diff}^\infty\left(\mathbb{R}^d,(0,1)^d\right)_0$ with autonomous neural ODEs having the same support$_{\mathrm{id}}$ but using composition in theorem 3.4.

Theorem 3.3 (Finite Composition of Flows for Diffeomorphisms)**.** *One can represent any diffeomorphism $\varphi\in\mathrm{Diff}^\infty\left(\mathbb{R}^d,(0,1)^d\right)_0$ as a composition of at most I_d autonomous flows where $I_d\in\mathbb{N}_+$ only depends on the dimension d.*

Theorem 3.4 (Universal Approximation by Deep Neural ODEs)**.** *Let $\varphi\in\mathrm{Diff}^\infty\left(\mathbb{R}^d,(0,1)^d\right)_0$ and $N,L,k\in\mathbb{N}_+$ with L fixed; then, there exist ReLU neural ODEs $\{\Psi_i=\mathrm{Flow}(\Phi_i)\}_{i=1}^I$ such that $\Psi\overset{\text{def.}}{=}\bigcirc_{i=1}^I\Psi_i$ satisfies the approximation guarantee:*

$$\left\|\varphi-\Psi\right\|_{L^\infty(\mathbb{R}^d;l^\infty)}\in O\big((NL)^{-2k/d}\big). \tag{7}$$

Moreover, $\Psi\in\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$, with Lipschitz constant at most $\prod_{i=1}^I e^{L^{\Phi_i}}$ and $\Phi_1,\dots,\Phi_I$ are ReLU MLPs of width at most $d+2d^2+17k^{d+1}3^dd^2(N+2)\log_2(8N)$ and depth at most $18k^2(L+2)\log_2(4L)+2d+3$. Note that L^Ψ depends on the parameter N. In terms of number of parameters P when L is not fixed, one can write:

$$\|\Psi-\varphi\|_{L^\infty(\mathbb{R}^d;l^\infty)}\in\mathcal{O}\big((\frac{P}{\log P})^{-2k/d}\big). \tag{8}$$

Remark 3.5. *Note that one can also use proposition A.8 instead of proposition A.11 in the above theorem, which gives a Lipschitz constant L^Ψ independent of N but with worse approximation rates $\mathcal{O}(P^{-1/d})$ as in remark 3.8. See also table 1.*

3.2.2 Qualitative Result:

We show that for $d\geqslant 5$, one can approximate $\mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$ with autonomous neural ODEs having the same support$_{\mathrm{id}}$:

Theorem 3.6 (Universal Approximation of Orientation-Preserving Homeomorphisms)**.** *Let $d \in \mathbb{N}_+$ and $d \geqslant 5$. There exists a constant $I_d \in \mathbb{N}_+$ such that for every $\varphi \in \mathcal{H}\left(\mathbb{R}^d, (0,1)^d\right)_0$ and every $\varepsilon > 0$ there exist an $I \leqslant I_d$ and ReLU neural ODEs $\{\Psi_i = \mathrm{Flow}(\Phi_i)\}_{i=1}^I$ such that the homeomorphism $\Psi \overset{\text{def.}}{=} \bigcirc_{i=1}^I \Psi_i$ is Lipschitz and compactly supported$_{\mathrm{id}}$ on $(0,1)^d$ and satisfies the approximation guarantee:*

$$\left\|\varphi - \Psi\right\|_{L^\infty(\mathbb{R}^d; l^\infty)} \leqslant \varepsilon. \tag{9}$$

3.3 Universal Approximation of compactly supported Lipschitz functions between arbitrary dimensions

Using an idea similar to Klee's trick, cf. Klee (1955), we encode the graph of an arbitrary continuous function into a homeomorphism in a higher dimension, in particular, by embedding the function into a *single* simple time-1 autonomous flow.

Before proceeding, we recall that in all of the theorems in this paper, the support can be any bounded open set diffeomorphic to $(0,1)^d$.

Fix $d, D \in \mathbb{N}_+$. Suppose the Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}^D$ is compactly supported on $(0,1)^d$. f induces a compactly supported Lipschitz vector field $V_f : \mathbb{R}^{d+D} \to \mathbb{R}^{d+D}$, defined for each $(x, y) \in \mathbb{R}^{d+D}$ by:

$$V_f(x,y) \overset{\text{def.}}{=} \big(0, f(x)\rho(y)\big), \tag{10}$$

where ρ is a smooth bump function equal to 1 on $\overline{B_{\mathbb{R}^D}(0, \max_{x\in\mathbb{R}^d} \|f(x)\|_2 + 1)}$ and zero outside the ball $B_{\mathbb{R}^D}(0, \max_{x\in\mathbb{R}^d} \|f(x)\|_2 + 2)$. The solution $\Phi : [0,\infty) \times \mathbb{R}^{d+D} \to \mathbb{R}^{d+D}$ to the *autonomous* ODE induced by V_f is defined by:

$$\frac{dx_t^{(x,y)}}{dt} = 0, \qquad \frac{dy_t^{(x,y)}}{dt} = f(x_t^{(x,y)})\rho(y_t^{(x,y)}), \text{ where } (x_0, y_0) = (x, y),$$

can be solved as:

$$\frac{dx_t^{(x,y)}}{dt} = 0 \Rightarrow x_t^{(x,y)} = x \Rightarrow y_t^{(x,y)} = y_0 + f(x)\int_0^t \rho(y_t^{(x,y)})dt.$$

This means that:

$$\mathrm{Flow}\,(V_f)(x, 0) = (x, f(x)). \tag{11}$$

and $\mathrm{Flow}\,(V_f)$ is compactly supported$_{\mathrm{id}}$. Now, let $\pi_D^{d+D} : \mathbb{R}^{d+D} \to \mathbb{R}^D$ denote the canonical (linear) projection sending any $(x, y) \in \mathbb{R}^{d+D}$ to $y \in \mathbb{R}^D$ and let $\iota_d^{d+D} : \mathbb{R}^d \to \mathbb{R}^{d+D}$ denote the (linear) embedding sending any $x \in \mathbb{R}^d$ to $(x, 0) \in \mathbb{R}^{d+D}$; Putting it all together we have that for all $x \in \mathbb{R}^d$:

$$\pi_D^{d+D} \circ \mathrm{Flow}\,(V_f) \circ \iota_d^{d+D}(x) = f(x). \tag{12}$$

Applying theorem 3.6 to $\mathrm{Flow}\,(V_f)$ for $d + D \geqslant 5$, we conclude that by post- and pre-composing, one can uniformly approximate any Lipschitz function compactly supported on $(0,1)^d$.

However, for diffeomorphisms, we note that, the model in equation 12 does not scale well when D is large, since one would obtain an approximation rate of $\mathcal{O}(P^{-1/(d+D)})$, (where P is the number of parameters cf. remark 3.8), which is significantly slower than the "unconstrained" optimal rate achievable by ReLU MLPs; cf. Yarotsky (2018). Instead, a rate nearly equal to the "unconstrained" ReLU MLP rate is possible if we alternatively approximate each component of the target function f independently and then concatenate the resulting lifted flow-based approximations. Denoting $f = (f_1, \ldots, f_D)$, we replace equation 12 with the model:

$$\bigoplus_{i=1}^D \left(\pi_1^{d+1} \circ \mathrm{Flow}\,(V_{f_i}) \circ \iota_d^{d+1}(x)\right) = \big(f_1(x), \ldots, f_D(x)\big) = f(x). \tag{13}$$

The advantage of the representation in equation 13, with its greater width, over the more "naive narrow" version in equation 12, is that each $\mathrm{Flow}\,(V_{f_i})$ performs its approximation in only one extra dimension beyond the physical dimension d. This design minimizes the approximation-theoretic difficulties that typically arise from high dimensionality. Importantly, doing so achieves the minimax

optimal approximation rates cf. Shen et al. (2022, Remark (ii), p.111) and Yarotsky (2017), for the *lifted, and thus higher-dimensional* space $\mathbb{R}^{d+1}$ and nearly achieves the optimal rate on the original low-dimensional domain $\mathbb{R}^d$, at the cost of one extra ambient dimension, which may be inevitable due to the extra invertibility structure of the autonomous models studied herein.

Corollary 3.7 (Approximation of compactly supported Lipschitz Functions by Linear Lifting)**.** *Let $d, D \in \mathbb{N}_+$, and $f : \mathbb{R}^d \to \mathbb{R}^D$ be Lipschitz and compactly supported on $(0,1)^d$. For every $n \in \mathbb{N}_+$ and each $i \in [D]$ there exist ReLU neural ODEs $\Psi_i : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$ such that the **Latent** neural ODE defined as:*

$$\Psi \overset{\text{def.}}{=} \bigoplus_{i=1}^{D} (\pi_1^{d+1} \circ \Psi_i \circ \iota_d^{d+1}), \tag{14}$$

satisfies

$$\|f - \Psi\|_{L^\infty(\mathbb{R}^d; l^\infty)} \in \mathcal{O}\big(\tfrac{d+1}{n}\big).$$

Moreover, Ψ is Lipschitz with a Lipschitz constant independent of n and the ReLU MLPs parameterizing the vector fields defining each Ψ_i have width and parameters at most $\mathcal{O}(D(d+1)^2(n+1)^{d+1})$ and depth at most $\mathcal{O}(\lceil \log_2(d+1) \rceil)$ for each $i \in [D]$.

Remark 3.8. *To write the order in terms of parameters, as the number of parameters $P \in \mathcal{O}\left(D(d+1)^2(n+1)^{d+1}\right)$ then one can write the upper bound in $O\big((d+1)(\frac{D(d+1)}{P})^{1/(d+1)}\big)$ which means for fixed d, D:*

$$\|f - \Psi\|_{L^\infty(\mathbb{R}^d; l^\infty)} \in \mathcal{O}\big(P^{-1/(d+1)}\big). \tag{15}$$

Thus, Corollary 3.7 provides a quantitative and more explicit version of the very recent qualitative result of (Marinis et al., 2026).

4 Second Case : Defined on $[0,1]^d$

4.1 Negative Result:

We show that considering only one single autonomous-neural-ODE is not universal and is in particular nowhere dense.

Theorem 4.1. $\mathrm{NODE}_\sigma([0,1]^d)$ *is nowhere dense in $\mathcal{H}\left([0,1]^d\right)_0$ for $d \geqslant 2$.*

This means only considering a single autonomous flow does not result in universality and one has to consider composition. In the following, we show diffeomorphisms are not dense in the space of homeomorphisms, which is an obstruction for any method using this technique.

Theorem 4.2. *The subset $\mathrm{Diff}^1\left([0,1]^d\right)_0$ is not dense in the space $\mathcal{H}\left([0,1]^d\right)_0$.*

This shows one cannot use the method in the proof of theorem 3.6 to show universality for homeomorphisms.

4.2 Positive Results:

Using the following theorem, we show that composition of autonomous-neural-ODEs uniformly approximate every smooth diffeomorphism of the cube that is smoothly isotopic to the identity through diffeomorphisms of the cube, with error $\mathcal{O}(P^{-1/d})$ for each fixed target, where P is the total number of parameters.

Theorem 4.3 (Finite Composition of Flows for Diffeomorphisms)**.** *One can represent any diffeomorphism $\varphi \in \mathrm{Diff}^\infty\left([0,1]^d\right)_0$ as a composition of autonomous flows.*

Theorem 4.4 (Universal Approximation by Deep Neural ODEs)**.** *Let $\varphi \in \mathrm{Diff}^\infty\left([0,1]^d\right)_0$ and $k \in \mathbb{N}_+$; then, there exist ReLU neural ODEs $\{\Psi_i = \mathrm{Flow}(\Phi_i)\}_{i=1}^I$ such that $\Psi \overset{\text{def.}}{=} \bigcirc_{i=1}^I \Psi_i$ satisfies the approximation guarantee:*

$$\|\varphi - \Psi\|_{L^\infty([0,1]^d; l^\infty)} \in O\big(\frac{d}{n}\big). \tag{16}$$

Moreover, $\Psi \in \mathcal{H}\left([0,1]^d\right)_0$*, with Lipschitz constant at most* $\prod_{i=1}^{I} e^{L^{\Phi_i}}$ *and* $\Phi_1, \ldots, \Phi_i$ *are ReLU MLPs of width at most* $8d^2(n+1)^d$*, depth at most* $\lceil log_2 d \rceil + 4$ *and at most* $16d^2(n+1)^d$ *free parameters. Note that* L^{Ψ} *does not depend on the parameter* n*. In terms of number of parameters* P*, one can write:*

$$\|\Psi - \varphi\|_{L^\infty([0,1]^d;l^\infty)} \in \mathcal{O}\left(P^{-1/d}\right). \tag{17}$$

5 Discussion:

5.1 Negative Result:

5.1.1 Related Works

Researchers have tried to show that the set of autonomous flows is nowhere dense or meagre. For instance, for the class of homeomorphisms in the identity component[2] of a compact smooth manifold, in dimension $d \geqslant 2$, the set of homeomorphisms that can be written as a time-one continuous autonomous flow, is proved to be nowhere dense Bonomo & Varandas (2020). Moreover, it is proved for the class of C^1-diffeomorphisms on closed connected smooth manifolds Bonatti (2009) that the set of flowable C^1-diffeomorphisms is meagre.[3] For the classical compact manifold see Palis (1974). Unfortunately, for C^∞diffeomorphisms, only a subset of them has been proved to be meagre. Examples include those that satisfy Axiom A and the transversality condition for $d \geqslant 2$ Palis (1989), or Axiom A diffeomorphisms with the no cycles property that are non-Anosov for $d \geqslant 3$ Fisher (2008). The full result has been proved only for the circle Chern & Smale (1970).

5.1.2 Comparison of Domains:

Considering only the cube results a stronger negative result (nowhere density) while considering compactly supported$_{\mathrm{id}}$ on $(0,1)^d$ only gives meagreness, which is not an obstruction for universality despite nowhere density.

5.2 Positive Result:

5.2.1 Related Works:

Related controllability-based results include universal interpolation for controlled ODEs established by Cuchiero et al. (2020) and L^p universal approximation, for $1 \leqslant p < \infty$, established by Cheng et al. (2025). These results concern different approximation guarantees and do not directly establish uniform approximation under the support and domain-preservation constraints and upper bound for the composition considered here.

5.2.2 Comparison of Domains:

Considering the compactly supported$_{\mathrm{id}}$ on $(0,1)^d$ case, one can not only approximate diffeomorphisms, but also homeomorphisms. This is possible as one can approximate homeomorphisms in high dimensions by diffeomorphisms. Unfortunately, on the cube, we proved diffeomorphisms are not dense in the space of homeomorphisms so we cannot apply this method.

AI use statement

In this work, we used generative AI tools for aid or polish writing, retrieval and discovery, research ideation or execution and proving math claims. In particular, we used AI to find spelling and math faults, relevant papers for a subject and to find useful lemmas for a method. We have not used generative AI tools for generating synthetic datasets or draft sections of the paper. We have reviewed all AI-assisted work. We checked every proof and references and we take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

[2] Homeomorphisms that are isotopic to identity.

[3] As flows don't have trivial centralizer and the set of diffeomorphisms on that space is completely metrizable so it is a Baire space Willard (2004, Corollary 25.4).

REFERENCES

A. A. Agrachev and M. Caponigro. Controllability on the group of diffeomorphisms. *Annales de l'I.H.P. Analyse non linéaire*, 26(6):2503–2509, 2009. URL http://eudml.org/doc/78944.

J. W. Alexander. On the deformation of an n cell. *Proceedings of the National Academy of Sciences*, 9(12):406–407, 1923. doi: 10.1073/pnas.9.12.406. URL https://www.pnas.org/doi/abs/10.1073/pnas.9.12.406.

J. Arango and A. Gómez. Diffeomorphisms as time-one maps. *Aequationes Mathematicae*, 64(3):304–314, 2002. doi: 10.1007/PL00013195. URL https://doi.org/10.1007/PL00013195.

Taras Banakh and Dušan Repovš. A topological characterization of lf-spaces. *Topology and its Applications*, 159(5):1475–1488, 2012. ISSN 0166-8641. doi: https://doi.org/10.1016/j.topol.2012.01.011. URL https://www.sciencedirect.com/science/article/pii/S0166864112000703.

Taras Banakh, Kotaro Mine, Katsuro Sakai, and Tatsuhiko Yagasaki. Homeomorphism and diffeomorphism groups of non-compact manifolds with the Whitney topology. *Topology Proceedings*, 37:61–93, 2011. ISSN 0146-4124. URL https://topology.nipissingu.ca/tp/reprints/v37/tp37004.pdf.

Augustin Banyaga. *The structure of classical diffeomorphism groups*, volume 400 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1997. ISBN 0-7923-4475-8. doi: 10.1007/978-1-4757-6800-8. URL https://doi.org/10.1007/978-1-4757-6800-8.

Martin Bauer, Philipp Harms, and Peter W. Michor. Regularity and completeness of half-Lie groups. *Journal of the European Mathematical Society*, 2025. doi: 10.4171/JEMS/1587. URL https://doi.org/10.4171/JEMS/1587. Published online first.

Christian Bayer, Luca Pelizzari, and John Schoenmakers. Primal and dual optimal stopping with signatures. *Finance and Stochastics*, 29(4):981–1014, 2025. doi: 10.1007/s00780-025-00570-8. URL https://doi.org/10.1007/s00780-025-00570-8.

Crovisier Sylvain-Wilkinson Amie Bonatti, Christian. The c 1 generic diffeomorphism has trivial centralizer. *Publications Mathématiques de l'IHÉS*, 109:185–244, 2009. URL http://eudml.org/doc/273598.

Wescley Bonomo and Paulo Varandas. Continuous flows generate few homeomorphisms. *Proceedings of the Edinburgh Mathematical Society*, 63(4):971–983, 2020. doi: 10.1017/S0013091520000280. URL https://doi.org/10.1017/S0013091520000280.

Johannes Bostelmann, Ole Gildemeister, and Jan Lellmann. Stationary velocity fields on matrix groups for deformable image registration. *Journal of Mathematical Imaging and Vision*, 68:9, 2026. doi: 10.1007/s10851-026-01284-y. URL https://doi.org/10.1007/s10851-026-01284-y.

Beverly L. Brechner and Kazuhiro Kawamura. On the dimension of a homeomorphism group. *Proceedings of the American Mathematical Society*, 129(2):617–620, 2001. ISSN 00029939, 10886826. URL http://www.jstor.org/stable/2668725.

Dmitri Burago, Sergei Ivanov, and Leonid Polterovich. Conjugation-invariant norms on groups of geometric origin. In *Groups of diffeomorphisms*, volume 52 of *Adv. Stud. Pure Math.*, pp. 221–250. Math. Soc. Japan, Tokyo, 2008. ISBN 978-4-931469-48-8. doi: 10.2969/aspm/05210221. URL https://doi.org/10.2969/aspm/05210221.

Thomas Cass and Cristopher Salvi. Lecture notes on rough paths and applications to machine learning, 2024. URL https://arxiv.org/abs/2404.06583.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf.

Jingpu Cheng, Qianxiao Li, Ting Lin, and Zuowei Shen. Interpolation, approximation, and controllability of deep neural networks. *SIAM Journal on Control and Optimization*, 63(1):625–649, 2025. doi: 10.1137/23M1599744. URL https://doi.org/10.1137/23M1599744.

Shiing-Shen Chern and Stephen Smale (eds.). *Global Analysis*, volume 14 of *Proceedings of Symposia in Pure Mathematics*, Providence, RI, 1970. American Mathematical Society. ISBN 978-0-8218-1414-7. doi: 10.1090/pspum/014. URL https://doi.org/10.1090/pspum/014.

Lee Cheuk-Kit, Florentin Coeurdoux, Yuyuan Chen, Sophia Tang, Peter Potaptchik, Yilun Du, Michael Samuel Albergo, and Eric Vanden-Eijnden. Beckmann transport models: From autonomous flows to one-step maps, 2026. URL https://arxiv.org/abs/2608.01692.

Albert Cohen, Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. Optimal stable nonlinear approximation. *Foundations of Computational Mathematics*, 22(3):607–648, 2022. doi: 10.1007/s10208-021-09494-z. URL https://doi.org/10.1007/s10208-021-09494-z.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In Maria Florina Balcan and Kilian Q. Weinberger (eds.), *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2990–2999, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/cohenc16.html.

Christa Cuchiero, Martin Larsson, and Josef Teichmann. Deep neural networks, generic universal interpolation, and controlled odes. *SIAM Journal on Mathematics of Data Science*, 2(3):901–919, 2020. doi: 10.1137/19M1284117. URL https://doi.org/10.1137/19M1284117.

Yifei Duan and Yongqiang Cai. A minimal control family of dynamical systems for universal approximation. *IEEE Transactions on Automatic Control*, 70(11):7233–7244, 2025. doi: 10.1109/TAC.2025.3569123. URL https://ieeexplore.ieee.org/document/10999048.

PAUL DUPUIS, ULF GRENANDER, and MICHAEL I. MILLER. Variational problems on flows of diffeomorphisms for image matching. *Quarterly of Applied Mathematics*, 56(3):587–600, 1998. ISSN 0033569X, 15524485. URL http://www.jstor.org/stable/43638248.

Dennis Elbrächter, Dmytro Perekrestenko, Philipp Grohs, and Helmut Bölcskei. Deep neural network approximation theory. *IEEE Transactions on Information Theory*, 67(5):2581–2623, 2021. doi: 10.1109/TIT.2021.3062161. URL https://doi.org/10.1109/TIT.2021.3062161.

A. Fathi. Structure of the group of homeomorphisms preserving a good measure on a compact manifold. *Annales scientifiques de l'École Normale Supérieure*, 13(1):45–93, 1980. URL http://eudml.org/doc/82046.

Todd Fisher. Trivial centralizers for axiom a diffeomorphisms. *Nonlinearity*, 21(11):2505, sep 2008. doi: 10.1088/0951-7715/21/11/002. URL https://doi.org/10.1088/0951-7715/21/11/002.

Kazuhiko Fukui, Tomasz Rybicki, and Tatsuhiko Yagasaki. The uniform perfectness of diffeomorphism groups of open manifolds, 2019. URL https://arxiv.org/abs/1905.07664.

V Gelfreich and A Vieiro. Interpolating vector fields for near identity maps and averaging. *Nonlinearity*, 31(9):4263, aug 2018. doi: 10.1088/1361-6544/aacb8e. URL https://doi.org/10.1088/1361-6544/aacb8e.

Helge Glöckner, Erlend Grong, and Alexander Schmeding. Boundary values of diffeomorphisms of simple polytopes, and controllability. *Differential Geometry and its Applications*, 104:102404, 2026. ISSN 0926-2245. doi: https://doi.org/10.1016/j.difgeo.2026.102404. URL https://www.sciencedirect.com/science/article/pii/S0926224526000756.

Paweł Goldstein, Zofia Grochulska, and Piotr Hajłasz. Gluing diffeomorphisms, bi-lipschitz mappings and homeomorphisms. *Expositiones Mathematicae*, 43(4):125681, 2025. ISSN 0723-0869. doi: https://doi.org/10.1016/j.exmath.2025.125681. URL https://www.sciencedirect.com/science/article/pii/S0723086925000362.

Erlend Grong and Alexander Schmeding. Controllability and diffeomorphism groups on manifolds with boundary. *Forum Mathematicum*, 37(4):1291–1308, 2025. doi: doi:10.1515/forum-2024-0160. URL https://doi.org/10.1515/forum-2024-0160.

M Gubinelli. Controlling rough paths. *Journal of Functional Analysis*, 216(1):86–140, 2004. ISSN 0022-1236. doi: https://doi.org/10.1016/j.jfa.2004.01.002. URL https://www.sciencedirect.com/science/article/pii/S0022123604000497.

Stefan Haller, Tomasz Rybicki, and Josef Teichmann. Smooth perfectness for the group of diffeomorphisms, 2013. ISSN 1941-4889. URL https://www.aimsciences.org/article/id/2a013df3-d2a8-476b-baaf-af484cba1212.

Ben Hambly and Terry Lyons. Uniqueness for the signature of a path of bounded variation and the reduced path group. *Annals of Mathematics*, 171(1):109–167, 2010. doi: 10.4007/annals.2010.171.109. URL https://annals.math.princeton.edu/wp-content/uploads/annals-v171-n1-p02-p.pdf.

Joshua Hanson and Maxim Raginsky. Fitting an immersed submanifold to data via Sussmann's orbit theorem. *Mathematics of Control, Signals, and Systems*, 37(3):479–506, 2025. doi: 10.1007/s00498-025-00410-2. URL https://doi.org/10.1007/s00498-025-00410-2.

Sigurdur Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*, volume 80 of *Pure and Applied Mathematics*. Academic Press, New York, 1978. ISBN 978-0-12-338460-7. URL https://www.sciencedirect.com/science/book/9780123384607.

Morris W. Hirsch. *Differential Topology*, volume 33 of *Graduate Texts in Mathematics*. Springer, New York, NY, 1976. ISBN 978-0-387-90148-0. doi: 10.1007/978-1-4684-9449-5. URL https://doi.org/10.1007/978-1-4684-9449-5.

Hirotada Honda. ODE-based learning on manifold and its expressive power. *Advances in Continuous and Discrete Models*, 2026(1):25, 2026. doi: 10.1186/s13662-026-04057-4. URL https://doi.org/10.1186/s13662-026-04057-4.

Ruiyang Hong and Anastasis Kratsios. Bridging the gap between approximation and learning via optimal approximation by relu mlps of maximal regularity, 2025. URL https://arxiv.org/abs/2409.12335.

Isao Ishikawa, Takeshi Teshima, Koichi Tojo, Kenta Oono, Masahiro Ikeda, and Masashi Sugiyama. Universal approximation property of invertible neural networks. *Journal of Machine Learning Research*, 24(287):1–68, 2023. URL http://jmlr.org/papers/v24/22-0384.html.

Johanna Jakob. Diffeomorphism groups of compact locally polyhedral manifolds, 2026. URL https://arxiv.org/abs/2606.14423.

Patrick Kidger, James Morrill, James Foster, and Terry Lyons. Neural controlled differential equations for irregular time series. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6696–6707. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4a5876b450b45371f6cfe5047ac8cd45-Paper.pdf.

V. L. Klee, Jr. Some topological properties of convex sets. *Trans. Amer. Math. Soc.*, 78:30–45, 1955. ISSN 0002-9947,1088-6850. doi: 10.2307/1992947. URL https://doi.org/10.2307/1992947.

Anastasis Kratsios and Ievgen Bilokopytov. Non-euclidean universal approximation. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 10635–10646. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/786ab8c4d7ee758f80d57e65582e609d-Paper.pdf.

Anastasis Kratsios and Léonie Papon. Universal approximation theorems for differentiable geometric deep learning. *Journal of Machine Learning Research*, 23(196):1–73, 2022. URL http://jmlr.org/papers/v23/21-0716.html.

Anastasis Kratsios and Behnoosh Zamanlooy. Do ReLU networks have an edge when approximating compactly-supported functions? *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=sNxNi54B8b.

Andreas Kriegl and Peter W. Michor. Regular infinite dimensional lie groups. *Journal of Lie Theory*, 7(1):61–99, 1997. URL https://www.heldermann-verlag.de/jlt/jlt07/MICHPL.PDF.

Christian Kuehn and Tobias Wöhrer. Tracking finite-time Lyapunov exponents to robustify neural ODEs. *Neurocomputing*, 699:134371, 2026. doi: 10.1016/j.neucom.2026.134371. URL https://doi.org/10.1016/j.neucom.2026.134371.

Anna A. Kwiecińska. On the C^0-closing lemma. *Annales Polonici Mathematici*, 64(2):131–138, 1996. URL http://eudml.org/doc/269953.

Gerardo Lafferriere and Héctor J. Sussmann. Motion planning for controllable systems without drift. In *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*, volume 2, pp. 1148–1153, Sacramento, CA, USA, 1991. IEEE. doi: 10.1109/ROBOT.1991.131763. URL https://doi.org/10.1109/ROBOT.1991.131763.

John M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2013. ISBN 978-1-4419-9981-8. doi: 10.1007/978-1-4419-9982-5. URL https://doi.org/10.1007/978-1-4419-9982-5.

George G. Lorentz, Manfred von Golitschek, and Yuly Makovoz. *Constructive Approximation: Advanced Problems*, volume 304 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 1996. URL https://link.springer.com/book/9783642646102.

Jianfeng Lu, Zuowei Shen, Haizhao Yang, and Shijun Zhang. Deep network approximation for smooth functions. *SIAM Journal on Mathematical Analysis*, 53(5):5465–5506, 2021. doi: 10.1137/20M134695X. URL https://doi.org/10.1137/20M134695X.

Yui Man Lui. Human gesture recognition on product manifolds. *Journal of Machine Learning Research*, 13(106):3297–3321, 2012. URL http://jmlr.org/papers/v13/lui12a.html.

A. M. Lukatskii. Finite generation of groups of diffeomorphisms. *Russian Mathematical Surveys*, 33 (1):207–208, 1978. doi: 10.1070/RM1978v033n01ABEH002248. URL https://doi.org/10.1070/RM1978v033n01ABEH002248.

Terry J. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998. URL http://eudml.org/doc/39555.

Arturo De Marinis, Davide Murari, Elena Celledoni, Nicola Guglielmi, Brynjulf Owren, and Francesco Tudisco. Approximation properties of neural odes, 2026. URL https://arxiv.org/abs/2503.15696.

John N. Mather. Commutators of diffeomorphisms. *Commentarii Mathematici Helvetici*, 49(1):512–528, 1974. doi: 10.1007/BF02566746. URL https://doi.org/10.1007/BF02566746.

John N. Mather. A curious remark concerning the geometric transfer map. *Commentarii Mathematici Helvetici*, 59(1):86–110, 1984. doi: 10.1007/BF02566338. URL https://doi.org/10.1007/BF02566338.

Hrushikesh N Mhaskar and Tomaso Poggio. Deep vs. shallow networks: An approximation theory perspective. *Analysis and Applications*, 14(06):829–848, 2016. doi: 10.1142/S0219530516400042. URL https://doi.org/10.1142/S0219530516400042.

P. Michor. Manifolds of smooth maps, II : the Lie group of diffeomorphisms of a non-compact smooth manifold. *Cahiers de topologie et géométrie différentielle*, 21(1):63–86, 1980. URL https://www.numdam.org/item/CTGDC_1980__21_1_63_0/.

Peter W. Michor. Manifolds of mappings and shapes. In Lizhen Ji, Frans Oort, and Shing-Tung Yau (eds.), *The Legacy of Bernhard Riemann After One Hundred and Fifty Years, Volume II*, volume 35.2 of *Advanced Lectures in Mathematics*, pp. 459–486. Higher Education Press and International Press, Beijing and Somerville, MA, 2016. ISBN 978-1-57146-319-7. URL https://www.mat.univie.ac.at/~michor/convenient-overview.pdf.

John W. Milnor. *Topology from the Differentiable Viewpoint*. Princeton University Press, 1997. ISBN 978-0-691-04833-8.

James Morrill, Cristopher Salvi, Patrick Kidger, and James Foster. Neural rough differential equations for long time series. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 7829–7838. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/morrill21b.html.

Nicola Muca Cirone, Maud Lemercier, and Cristopher Salvi. Neural signature kernels as infinite-width-depth-limits of controlled ResNets. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 25358–25425. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/muca-cirone23a.html.

Stefan Müller. Uniform approximation of homeomorphisms by diffeomorphisms. *Topology and its Applications*, 178:315–319, 2014. ISSN 0166-8641. doi: https://doi.org/10.1016/j.topol.2014.10.003. URL https://www.sciencedirect.com/science/article/pii/S0166864114003915.

Z. Nitecki and M. Shub. Filtrations, decompositions, and explosions. *American Journal of Mathematics*, 97(4):1029–1047, 1975. ISSN 00029327, 10806377. URL http://www.jstor.org/stable/2373686.

John C. Oxtoby. *Measure and Category: A Survey of the Analogies between Topological and Measure Spaces*, volume 2 of *Graduate Texts in Mathematics*. Springer-Verlag, New York–Berlin, 1971. doi: 10.1007/978-1-4615-9964-7. URL https://doi.org/10.1007/978-1-4615-9964-7.

B. G. Pachpatte. *Inequalities for Differential and Integral Equations*, volume 197 of *Mathematics in Science and Engineering*. Academic Press, San Diego, 1998. ISBN 0-12-543430-8. URL https://www.sciencedirect.com/science/book/9780125434300.

Jacob Palis. Vector fields generate few diffeomorphisms. *Bulletin of the American Mathematical Society*, 80:503–505, 1974. URL https://projecteuclid.org/journals/bulletin-of-the-american-mathematical-society-new-series/volume-80/issue-3/Vector-fields-generate-few-diffeomorphisms/bams/1183535529.full.

Yoccoz J.-C. Palis, J. Rigidity of centralizers of diffeomorphisms. *Annales scientifiques de l'École Normale Supérieure*, 22(1):81–98, 1989. URL http://eudml.org/doc/82248.

Frank C. Park and Koichiro Okamura. *Kinematic Calibration and the Product of Exponentials Formula*, pp. 119–128. Springer Netherlands, Dordrecht, 1994. ISBN 978-94-015-8348-0. doi: 10.1007/978-94-015-8348-0_12. URL https://doi.org/10.1007/978-94-015-8348-0_12.

Philipp Petersen and Felix Voigtlaender. Optimal approximation of piecewise smooth functions using deep relu neural networks. *Neural Networks*, 108:296–330, 2018. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2018.08.019. URL https://www.sciencedirect.com/science/article/pii/S0893608018302454.

Philipp Petersen and Jakob Zech. Mathematical theory of deep learning, 2024. URL https://arxiv.org/abs/2407.18384.

Guergana Petrova and Przemysław Wojtaszczyk. Lipschitz widths. *Constructive Approximation*, 57(2):759–805, 2023. doi: 10.1007/s00365-022-09576-3. URL https://doi.org/10.1007/s00365-022-09576-3.

Allan Pinkus. *n-Widths in Approximation Theory*, volume 7 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer-Verlag, Berlin and Heidelberg, 1985. ISBN 978-3-642-69894-1. doi: 10.1007/978-3-642-69894-1. URL https://doi.org/10.1007/978-3-642-69894-1.

Muhammad Asif Rana, Anqi Li, Dieter Fox, Byron Boots, Fabio Ramos, and Nathan Ratliff. Euclideanizing flows: Diffeomorphic reduction for learning stable dynamical systems. In Alexandre M. Bayen, Ali Jadbabaie, George Pappas, Pablo A. Parrilo, Benjamin Recht, Claire Tomlin, and Melanie Zeilinger (eds.), *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, volume 120 of *Proceedings of Machine Learning Research*, pp. 630–639. PMLR, 10–11 Jun 2020. URL https://proceedings.mlr.press/v120/rana20a.html.

Alessandro Scagliotti. Deep learning approximation of diffeomorphisms via linear-control systems, 2023a. ISSN 2156-8472. URL https://www.aimsciences.org/article/id/6306d8b94cedfd00075d1914.

Alessandro Scagliotti. Deep learning approximation of diffeomorphisms via linear-control systems, 2023b. ISSN 2156-8472. URL https://www.aimsciences.org/article/id/6306d8b94cedfd00075d1914.

Zuowei Shen, Haizhao Yang, and Shijun Zhang. Optimal approximation rate of relu networks in terms of width and depth. *Journal de Mathématiques Pures et Appliquées*, 157:101–135, 2022. ISSN 0021-7824. doi: https://doi.org/10.1016/j.matpur.2021.07.009. URL https://www.sciencedirect.com/science/article/pii/S0021782421001124.

Jonathan W. Siegel. Optimal approximation rates for deep relu neural networks on sobolev and besov spaces. *Journal of Machine Learning Research*, 24(357):1–52, 2023. URL http://jmlr.org/papers/v24/23-0025.html.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds, 2018. URL https://arxiv.org/abs/1802.08219.

Alphin J. Thottupattu, Jayanthi Sivaswamy, and Venkateswaran P. Krishnan. A method for image registration via broken geodesics. In Alessa Hering, Julia Schnabel, Miaomiao Zhang, Enzo Ferrante, Mattias Heinrich, and Daniel Rueckert (eds.), *Biomedical Image Registration*, pp. 47–56, Cham, 2022. Springer International Publishing. ISBN 978-3-031-11203-4. URL https://link.springer.com/chapter/10.1007/978-3-031-11203-4_6.

William Thurston. Foliations and groups of diffeomorphisms. *Bulletin of the American Mathematical Society*, 80(2):304–307, 1974. doi: 10.1090/S0002-9904-1974-13475-0. URL https://doi.org/10.1090/S0002-9904-1974-13475-0.

David Vigouroux, Lucas Drumetz, Ronan Fablet, and François Rousseau. Discovering data manifold geometry through geometric properties, 2026. URL https://arxiv.org/abs/2602.02611.

Benjamin Walker, Andrew Donald Mcleod, Tiexin Qin, Yichuan Cheng, Haoliang Li, and Terry Lyons. Log neural controlled differential equations: The lie brackets make a difference. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*,

volume 235 of *Proceedings of Machine Learning Research*, pp. 49822–49844. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/walker24a.html.

Haixia Wang, Shuhan Shen, and Xiao Lu. A screw axis identification method for serial robot calibration based on the POE model. *Industrial Robot*, 39(2):146–153, 2012. doi: 10.1108/01439911211201609. URL https://doi.org/10.1108/01439911211201609.

Stephen Willard. *General Topology*. Dover Books on Mathematics. Dover Publications, Mineola, NY, 2004. ISBN 978-0-486-43479-7. URL https://store.doverpublications.com/products/9780486434797. Reprint of the 1970 edition.

Dmitry Yarotsky. Error bounds for approximations with deep relu networks. *Neural Networks*, 94:103–114, 2017. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2017.07.002. URL https://www.sciencedirect.com/science/article/pii/S0893608017301545.

Dmitry Yarotsky. Optimal approximation of continuous functions by very deep relu networks. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet (eds.), *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pp. 639–649. PMLR, 06–09 Jul 2018. URL https://proceedings.mlr.press/v75/yarotsky18a.html.

Han Zhang, Xi Gao, Jacob Unterman, and Tom Arodz. Approximation capabilities of neural ODEs and invertible residual networks. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 11086–11095. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/zhang20h.html.

Shijun Zhang, Zuowei Shen, and Haizhao Yang. Deep network approximation: Achieving arbitrary accuracy with fixed number of neurons. *Journal of Machine Learning Research*, 23(276):1–60, 2022. URL http://jmlr.org/papers/v23/21-1404.html.

A Appendix

A.1 Negative Result Propositions and Lemmas for Compactly Supported$_{\mathrm{id}}$ Case

Proposition A.1 (Existence of a recurrent point)**.** *For every compact set $M \subset \mathbb{R}^d$ and any $f \in \mathcal{H}\left(\mathbb{R}^d, M\right)_0$ there exists a recurrent point in M.*

Proof. Take $f \in \mathcal{H}\left(M, M\right)_0$. Since M is compact, the set:

$$\mathcal{A} = \{S \subseteq M : S \neq \varnothing, S \text{ closed}, f(S) = S\},$$

is non-empty (e.g. $M \in \mathcal{A}$). Order $\mathcal{A}$ by inclusion. Any totally ordered chain $\{S_i\}$ in $\mathcal{A}$ has:

$$\bigcap_i S_i \;\neq\; \varnothing,$$

by compactness, and $\bigcap_i S_i$ is closed and f-invariant. By the dual form of Zorn's Lemma there is a minimal element $S_{\min} \in \mathcal{A}$, i.e. a nonempty closed invariant set containing no proper nonempty closed invariant subset. Take any $x \in S_{\min}$. If x were not recurrent then there would exist a neighborhood $x \in U$ and an integer N such that:

$$f^n(x) \notin U, \quad \forall\, n \geqslant N.$$

Set:

$$\omega(x) \;=\; \bigcap_{N \geqslant 0} \overline{\{\, f^n(x) : n \geqslant N\}}.$$

Then $\omega(x)$ is nonempty, compact, and f-invariant, but $x \notin \omega(x)$, so $\omega(x) \subsetneq S_{\min}$, contradicting minimality. Thus x must be recurrent. Since x was arbitrary in $S_{\min}$, every point of $S_{\min}$ is recurrent. □

Let (M, g) be a Riemannian manifold of dimension at least 2, with distance d_g coming from a Riemannian metric and let ϕ be a flow on M and:

$$\mathcal{Z}(\phi) \stackrel{\text{def.}}{=} \{x : \dot{\phi}(x) = 0\}.$$

Given a non-zero vector $Y \in T_xM$, where $x \notin \mathcal{Z}(\phi)$, we define the *inclination* of Y (relative to ϕ) to be the length of the normalized difference:

$$\sigma(Y) \stackrel{\text{def.}}{=} \left\| \frac{1}{\|Y\|_g} Y - \frac{1}{\|\dot{\phi}(x)\|_g} \dot{\phi}(x) \right\|_g .$$

If the angle between Y and $\dot{\phi}(x)$ is θ, $0 \leqslant \theta \leqslant \pi$, then:

$$\sigma(Y) = 2 \sin\left(\frac{\theta}{2}\right).$$

Note that $0 \leqslant \sigma \leqslant 2$, that $\sigma = 0$ if and only if Y points in the same direction as $\dot{\phi}$, and that:

$$\widetilde{Y} = \left(\frac{\|\dot{\phi}\|_g}{\|Y\|_g} \right) Y,$$

is a vector parallel to Y with:

$$\|\widetilde{Y} - \dot{\phi}\|_g = \sigma(Y) \|\dot{\phi}\|_g.$$

We define the inclination of a curve γ to be the inclination of its velocity vector $\dot{\gamma}$.

Lemma A.2. *Let (M, g) be a smooth Riemannian manifold of dimension at least 2 with metric g and distance d_g coming from a Riemannian metric. Given $\varepsilon > 0$ and a global flow ϕ, suppose γ is a C^1 curve in M (an embedded closed interval or circle) such that at each point x in the image of γ one of the following conditions holds:*

(i) $\|\dot{\phi}(x)\|_g < \varepsilon/2$, or

(ii) $x \notin \mathcal{Z}(\phi)$, and γ has inclination $\sigma < \varepsilon / \|\dot{\phi}\|_g$ at x.

Then, given any neighbourhood U of the image of γ, there exists a flow ψ on M satisfying:

(a) $\dot{\psi} = \dot{\phi}$ off U,

(b) $\|\dot{\psi} - \dot{\phi}\|_g < \varepsilon$ on M,

(c) γ is (a segment of an) integral curve of ψ.

Proof. For the proof, see Nitecki & Shub (1975, Lemma 9). Note that in the paper Nitecki & Shub (1975) the manifold is assumed to be compact but the proof of this lemma also works on non-compact smooth manifolds, provided ϕ is a global flow. This is because the proof does not explicitly use compactness of the ambient manifold M and it is a local construction near γ. This can also be deduced from the same paper as in Nitecki & Shub (1975, Lemma 13) the Riemannian manifold is just assumed to be of dimension at least 2. □

Lemma A.3. *Let (M, g) be an orientable smooth manifold of dimension at least 2 with Riemannian metric g and distance d_g coming from the Riemannian metric. Suppose a finite collection $\{(p_i, q_i) \in M \times M : i = 1, \ldots, k\}$ of pairs of points of M is specified and a compact subset $P \subset M$ with path connected interior is such that $p_i, q_i \in$ int(P) (for $i = 1, \ldots, k$), together with a small positive constant $\delta > 0$ such that:*

(a) For each i, $d_g(p_i, q_i) < \delta$.

(b) If $i \neq j$, then $p_i \neq p_j$ and $q_i \neq q_j$.

Then there exists an orientation-preserving $f \in \text{Diff}^\infty(M, P)_0$ such that:

(i) $d_g\big(f(x), x\big) < 2\pi\delta$ *for every* $x \in M$.

(ii) $f(p_i) = q_i$ *for* $i = 1, \ldots, k$.

(iii) f *is compactly supported*$_{\mathrm{id}}$ *on* P.

Proof. We consider the "suspension" of the identity, by taking $M \times S^1$ with the product orientation and with the vector field:

$$X(p, \theta) = 2\pi \frac{\partial}{\partial \theta}.$$

The induced flow is, of course:

$$\phi_t(p, \theta) = (p, \theta + 2\pi t).$$

Its time-one map, ϕ_1, takes $M \times \{0\}$ to itself and induces the identity map there.

Given the points $p_i, q_i \in M$, we consider the points:

$$(p_i, \pi/2) \quad \text{and} \quad (q_i, 3\pi/2),$$

in $M \times S^1$. We take, for each i, a curve $\gamma_i(t)$ in M, $0 \leqslant t \leqslant 1$, of constant speed, joining p_i to q_i, of length $< \delta$. Then the curve g_i in $M \times S^1$ given by:

$$g_i(t) = \left(\gamma_i(t), \pi t + \frac{\pi}{2}\right), \qquad 0 \leqslant t \leqslant 1,$$

has velocity vector:

$$\dot{\gamma}_i + \frac{1}{2} X,$$

whose inclination is $< 2\pi\delta$. We can change g_i slightly so that $\dot{g}_i(t) = \frac{1}{2}X$ at $t = 0, 1$, and then use lemma A.2 to find a vector field Y, with:

$$\|Y - X\|_g < 2\pi\delta,$$

for which the curves g_i are segments of integral curves. But then the flow of Y takes $(p_i, \pi/2)$ to $(q_i, 3\pi/2)$, so that $(p_i, 0)$ and $(q_i, 2\pi) = (q_i, 0)$ are joined by an integral curve of Y. Hence, the time-one map of the flow of Y is a diffeomorphism of $M \times \{0\}$, within distance:

$$\int_0^1 \|Y - X\|_g \, dt < 2\pi\delta,$$

of the identity, and taking p_i to q_i.

The above proof is reproduced verbatim from Nitecki & Shub (1975, Lemma 13). We add that property (a) in lemma A.2 makes the vector field $Y - X$ compactly supported on a finite number of neighborhoods $\bigcup U_i$ of the curves g_i joining p_i to q_i, which also makes its time-one flow compactly supported$_{\mathrm{id}}$ on M; as U_i were arbitrary neighborhoods around the curves, one can choose them sufficiently small that $\bigcup U_i \subset \mathrm{int}(P)$; which makes the time-one flow of Y to be compactly supported on P. By the suspension construction, the ambient flow preserves the positive S^1-normal direction. Consequently, its induced return map $h : M \to M$ is orientation-preserving. □

Lemma A.4 (C^0-closing lemma for compactly supported functions)**.** *Let* $f \colon M \to M$ *be a compactly supported*$_{\mathrm{id}}$ *homeomorphism on the compact set* $K \subset M$ *with path connected interior, where* (M, g) *is a* C^∞ *orientable manifold of dimension at least* 2 *with Riemannian metric* g*, with distance* d_g *coming from the Riemannian metric, and let* $x_0 \in M$ *be a non-wandering point. Then for every* $\varepsilon > 0$ *and compact set* P *with path connected interior such that* $K \Subset \mathrm{int}(P)$ *(*K *is compactly contained in* $\mathrm{int}(P)$*) there exists a compactly supported*$_{\mathrm{id}}$ *homeomorphism* $g \colon M \to M$ *on* P *such that* x_0 *is a periodic point of* g *and*

$$d_1(f, g) < \varepsilon.$$

Moreover, if f *is orientation-preserving, then* g *is also orientation-preserving.*

Proof. The Lemma is trivial when x_0 is a fixed point or a periodic point. Assume otherwise.
If M is compact and $K = M$, then, by Kwiecińska (1996, Theorem 4), the result follows. Suppose not.

As f is compactly supported$_{\mathrm{id}}$, $\mathcal{O}_f(x_0)$ is a subset of K. Note f is uniformly continuous on P and lemmas A.2 and A.3 are true on any Riemannian manifold like M.

Let ε be a small positive constant. Since f is uniformly continuous on P, there exists η, with $0 < \eta < \varepsilon$, such that

$$d_g\big(f(x_1), f(x_2)\big) < \varepsilon \qquad \text{if} \qquad d_g(x_1, x_2) < \eta.$$

Choose δ, with $0 < \delta < \eta/2$, such that

$$d_g\big(f^{-1}(x_1), f^{-1}(x_2)\big) < \frac{\eta}{2} \qquad \text{if} \qquad d_g(x_1, x_2) < \delta.$$

This is possible since f^{-1} is also uniformly continuous.
Let $z_0 \in B(x_0, \delta)$ be such that there exists $m \in \mathbb{Z}$, $m \neq 0$, satisfying

$$f^m(z_0) = z_m \in B(x_0, \delta), \qquad z_0 \neq x_0, \qquad z_m \neq x_0.$$

Without loss of generality, we may assume that $m > 0$; otherwise, we take z_m instead of z_0. We also take the smallest m satisfying the above conditions.

Suppose first that $m \geqslant 2$. Set

$$z_i = f^i(z_0), \qquad i \in \mathbb{Z},$$

and consider the finite collection

$$\big\{(p_i, q_i) \in M \times M : i = 0, \dots, m-1\big\},$$

where

$$\begin{aligned} &p_0 = x_0, \quad p_1 = z_1, \quad p_2 = z_2, \quad \dots, \quad p_{m-2} = z_{m-2}, \quad p_{m-1} = z_{m-1} = f^{-1}(z_m), \\ &q_0 = z_0, \quad q_1 = z_1, \quad q_2 = z_2, \quad \dots, \quad q_{m-2} = z_{m-2}, \quad q_{m-1} = f^{-1}(x_0). \end{aligned}$$

as $m \geqslant 2$ all of these points are in $K \subset \mathrm{int}(P)$ and the points p_i and q_i $(i = 0, \dots, m-1)$ satisfy:

(a) $p_i \neq p_j$ and $q_i \neq q_j$ for $0 \leqslant i < j \leqslant m-1$;

(b) $d_g(p_i, q_i) < \eta/2$ for $i = 0, \dots, m-1$.

According to lemma A.3, there exists an orientation-preserving diffeomorphism $h\colon M \to M$ with the following properties:

(a) $d_1(h, \mathrm{id}) < \eta$;

(b) $h(p_i) = q_i$ for $i = 0, \dots, m-1$.

(c) h is compactly supported$_{\mathrm{id}}$ on P.

Set $g = f \circ h$. Clearly, g is a compactly supported$_{\mathrm{id}}$ homeomorphism on P and x_0 is a periodic point of g. Moreover, since $d_1(h, \mathrm{id}) < \eta < \varepsilon$,

$$\begin{aligned} d_1(f, g) &= \max_{x \in M} \Big\{ d_g\big(f(x), (f \circ h)(x)\big), d_g\big(f^{-1}(x), (h^{-1} \circ f^{-1})(x)\big) \Big\} \\ &= \max_{x \in M} \Big\{ d_g\big(f(x), (f \circ h)(x)\big), d_g\big(x, h^{-1}(x)\big) \Big\} \\ &< \varepsilon. \end{aligned}$$

Suppose now that $m = 1$. Choose ϱ, with $0 < \varrho < \eta$, such that

$$d_g\big(f^{-1}(x_1), f^{-1}(x_2)\big) < \frac{\eta}{2} \qquad \text{if} \qquad d_g(x_1, x_2) < \varrho.$$

Choose θ , with $0 < \theta < \varrho/2$ and small enough, such that

- $d_g\big(f(x_1), f(x_2)\big) < \frac{\varrho}{2} \qquad \text{if} \qquad d_g(x_1, x_2) < 2\theta,$
- $B(x_0, \theta) \subset \mathrm{int}(P)$.

Let $y_0 \in B(x_0, \theta)$ be such that there exists $n > 0$ satisfying

$$f^n(y_0) = y_n \in B(x_0, \theta), \qquad y_0 \neq x_0, \qquad y_n \neq x_0.$$

We take the smallest n satisfying these conditions. If $n \geqslant 2$, the problem reduces to the previous case. If $n = 1$, consider the points

$$p_0 = x_0, \qquad p_1 = y_1 = f^{-1}(y_2), \qquad q_0 = y_0, \qquad q_1 = f^{-1}(x_0).$$

The points p_0, p_1, q_0, q_1 satisfy:

(a) $p_0 \neq p_1$ and $q_0 \neq q_1$;

(b)

$$\begin{aligned} d_g(p_0, q_0) &= d_g(x_0, y_0) < \theta < \frac{\varrho}{2} < \frac{\eta}{2}, \\ d_g(p_1, q_1) &= d_g\big(y_1, f^{-1}(x_0)\big) \\ &= d_g\big(f^{-1}(y_2), f^{-1}(x_0)\big) < \frac{\eta}{2}, \end{aligned}$$

since

$$\begin{aligned} d_g(x_0, y_2) &\leqslant d_g(x_0, y_1) + d_g(y_1, y_2) \\ &= d_g(x_0, y_1) + d_g\big(f(y_0), f(y_1)\big) \\ &< \theta + \frac{\varrho}{2} < \varrho. \end{aligned}$$

(c) $p_0, p_1, q_0, q_1 \in \operatorname{int}(P)$

Again, lemma A.3 gives an orientation-preserving diffeomorphism $h\colon M \to M$ with the following properties:

(a) $d_1(h, \mathrm{id}) < \eta$;

(b) $h(x_0) = y_0$ and $h(y_1) = f^{-1}(x_0)$.

(c) h is compactly supported$_{\mathrm{id}}$ on P.

Set $g = f \circ h$. Again, g is a compactly supported$_{\mathrm{id}}$ homeomorphism on P, x_0 is a periodic point of g, and, as before,

$$d_1(f, g) < \varepsilon.$$

If f is orientation-preserving then $g = f \circ h$ is also orientation-preserving. This is because h is orientation-preserving from lemma A.3. □

A.2 Positive Result Propositions and Lemmas for Compactly Supported$_{\mathrm{id}}$ Case

A.2.1 Qualitative Part

Lemma A.5 (Munkres, Connell, Bing). *Let $d \geqslant 5$, and let $\varphi \in \mathcal{H}\left(\mathbb{R}^d, B(0,1)\right)_0$. Then φ can be approximated uniformly by diffeomorphisms in* $\mathrm{Diff}^\infty\left(\mathbb{R}^d, B(0,1)\right)_0$.

Proof. Note that $\varphi|_{B(0,1)}$ is isotopic to the identity through an isotopy that fixes point-wise a neighborhood of the boundary of $B(0,1)$ by Alexander's Trick Alexander (1923). Apply Müller (2014, Lemma 2) on $\varphi|_{B(0,1)}$. Because the diffeomorphisms and $\varphi|_{B(0,1)}$ are identity near the boundary of $B(0,1)$ one can extend them on $\mathbb{R}^d$ by taking them equal to the identity outside of $B(0,1)$. Every compactly supported$_{\mathrm{id}}$ diffeomorphism is orientation preserving as the sign of the determinant of the diffeomorphisms outside of the compact supported is 1 and it cannot change sign. By Milnor (1997, §6. Lemma 2.) every orientation preserving diffeomorphism on $\mathbb{R}^d$ is isotopic to identity. This means the diffeomorphisms are in $\mathrm{Diff}^\infty\left(\mathbb{R}^d, B(0,1)\right)_0$. □

A.2.2 QUANTITATIVE FORMULATIONS PART

Independent Lipschitz Constant for Neural ODE, but Worse Rates:

Lemma A.6 (Optimal Regular Approximation by Sample-Interpolating ReLU MLPs)**.** *Let f be a function from $[0,1]^d$ to $\mathbb{R}$, and let*

$$\omega\colon [0,d] \to [0,\infty)$$

be a modulus of regularity of f. Then, for any $n \in \mathbb{N}_+$, there exists a ReLU MLP Φ on $[0,1]^d$ with width at most $8d(n+1)^d$, depth at most $\lceil \log_2 d \rceil + 4$, and at most $16d(n+1)^d$ parameters, which satisfies the approximation guarantee

$$\|f - \Phi\|_{L^\infty([0,1]^d)} \leqslant \omega\left(\frac{d}{2n}\right),$$

as well as the sample-interpolation guarantee

$$f(\mathbf{x}) = \Phi(\mathbf{x}) \quad \textit{for all} \quad \mathbf{x} \in \left\{\frac{0}{n}, \frac{1}{n}, \ldots, \frac{n}{n}\right\}^d.$$

Furthermore, ω is a modulus of regularity of Φ on $[0,1]^d$.

Proof. For the proof, see (Hong & Kratsios, 2025, Theorem 4.1). □

Lemma A.7 (Transfer: Universal Approximation of Vector Field to Flow)**.** *Let $n \in \mathbb{N}_+$ and $V = (V_1, \ldots, V_d) \in \mathcal{X}^0(\mathbb{R}^d, (0,1)^d)$ with modulus of regularity $\omega_j : [0,d] \to \mathbb{R}$ for $j \in [d]$. Consider its time-one flow:*

$$\begin{aligned} \varphi(x) &= x_1^x \\ x_t^x &= x + \int_0^t V(x_s^x)\, ds, \qquad 0 \leqslant t \leqslant 1, \end{aligned} \tag{18}$$

then there exists a ReLU MLP $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ supported on $(0,1)^d$ with width at most $8d^2(n+1)^d + 2d^2 + d$, depth at most $\lceil log_2 d \rceil + 7$, and at most $16d^2(n+1)^d + 6d^3 + 9d^2$ parameters such that the time-one flow (neural ODE):

$$\begin{aligned} \Psi(x) &= z_1^x, \\ z_t^x &= x + \int_0^t \Phi(z_s^x)\, ds, \qquad 0 \leqslant t \leqslant 1 \end{aligned} \tag{19}$$

satisfies the uniform estimate:

$$\|\Psi - \varphi\|_{L^\infty(\mathbb{R}^d; l^\infty)} \leqslant \|\omega(\frac{d}{2n})\|_{l^\infty} e^{L^V}, \tag{20}$$

and one gets convergence as $n \to \infty$. Moreover, $L^\Psi \leqslant e^{L^\Phi}$ and L^Ψ and L^Φ do not depend on n.

Proof. As $V_j : \mathbb{R}^d \to \mathbb{R}$ for $j \in [d]$ is Lipschitz and compactly supported on $(0,1)^d$, from lemma A.6 there exists a ReLU MLP ϕ_j defined on $[0,1]^d$ and modulus of regularity ω_j with width at most $8d(n+1)^d$, depth at most $\lceil log_2 d \rceil + 4$ and at most $16d(n+1)^d$ free parameters such that:

$$\|V_j - \phi_j\|_{L^\infty([0,1]^d)} \leqslant \omega_j(\frac{d}{2n}). \tag{21}$$

Define $(\phi)_j \overset{\text{def.}}{=} \phi_j$ and $(\omega)_j \overset{\text{def.}}{=} \omega_j$. V is supported on a compact set in $(0,1)^d$. Without loss of generality, suppose that V_j is supported on $[R_j, 1-R_j]^d$ for a $R > 0$. As ϕ is Lipschitz on $[0,1]^d$, there exists $M \in \mathbb{R}$ such that:

$$\|\phi\|_{L^\infty([0,1]^d; l^\infty)} \leqslant M.$$

Consider the ReLU bump function below for a small $\delta > 0$ such that $[R_j - \delta, 1 - R_j + \delta] \subseteq [0,1]^d$:

$$\begin{aligned} b_\delta(x) &\overset{\text{def.}}{=} \sigma\left(1 - \frac{1}{\delta}\sum_{i=1}^d \big[\sigma(R_j - x_i) + \sigma\big(x_i - (1 - R_j)\big)\big]\right) \\ &= \begin{cases} 0, & \text{if } x \notin [R_j - \delta, 1 - R_j + \delta]^d, \\ 1, & \text{if } x \in [R_j, 1 - R_j]^d. \end{cases} \end{aligned} \tag{22}$$

Implement the following network for $j \in [d]$:

$$\begin{aligned}\mathbf{x} &\Rightarrow (\phi_j, b_\delta) \\ &\Rightarrow (\phi_j, \sigma_{ReLU}(-Mb_\delta - \phi_j), b_\delta), \\ &\Rightarrow (\sigma_{ReLU}\big(Mb_\delta - (\phi_j + \sigma_{ReLU}(-Mb_\delta - \phi_j)), b_\delta\big), \\ &\Rightarrow Mb_\delta - \sigma_{ReLU}\big(Mb_\delta - (\phi_j + \sigma_{ReLU}(-Mb_\delta - \phi_j))\big) = \Phi_j.\end{aligned} \tag{23}$$

The resulting $\Phi_j = \min\{Mb_\delta, \max\{-Mb_\delta, \phi_j\}\}$ equals ϕ_j on $[R_j, 1-R_j]^d$ and equals zero outside $[R_j - \delta, 1 - R_j + \delta]^d$. Φ_j has width:

$$\begin{aligned}\mathrm{width}(\Phi_j) &\leqslant \mathrm{width}(\phi_j) + 2d + 1 \\ &\leqslant 8d(n+1)^d + 2d + 1.\end{aligned}$$

Its depth is:

$$\mathrm{depth}(\Phi_j) = \mathrm{depth}(\phi_j) + 3 = \lceil \log_2 d \rceil + 7.$$

Its parameter count is:

$$\mathrm{Par}(\Phi_j) = \mathrm{Par}(\phi_j) + 6d^2 + 9d = 16d(n+1)^d + 6d^2 + 9d.$$

If needed, one can increase the depth of the neural networks such that $\mathrm{depth}(\Phi_j) = \max_{1 \leqslant i \leqslant d}\{\mathrm{depth}(\Phi_i)\}$, then by parallelization (Petersen & Zech, 2024, Proposition 2.3.) the neural network:

$$\Phi(x) = (\Phi_1(x), \ldots, \Phi_d(x)) : \mathbb{R}^d \to \mathbb{R}^d$$

has the same depth. Its width is at most:

$$\mathrm{width}(\Phi) \leqslant \sum_{j=1}^{d} \mathrm{width}(\Phi_j) \leqslant 16d^2(n+1)^d + 6d^3 + 9d^2,$$

And parameters:

$$\mathrm{Par}(\Phi) \leqslant \sum_{i=1}^{d} \mathrm{Par}(\phi_j) + 6d^2 + 9d = 16d^2(n+1)^d + 6d^3 + 9d^2.$$

By equation 21 and the definition of Φ one has:

$$\|\Phi - V\|_{L^\infty(\mathbb{R}^d; l^\infty)} \leqslant \|\phi - V\|_{L^\infty([0,1]^d; l^\infty)} \leqslant \|\omega(\frac{d}{2n})\|_{l^\infty}, \tag{24}$$

this yields:

$$\begin{aligned}\|V(x_s^x) - \Phi(z_s^x)\|_{L^\infty(\mathbb{R}^d; l^\infty)} &\leqslant \|V(x_s^x) - V(z_s^x)\|_{L^\infty(\mathbb{R}^d; l^\infty)} + \|V(z_s^x) - \Phi(z_s^x)\|_{L^\infty(\mathbb{R}^d; l^\infty)} \\ &\leqslant L^V \|x_s^x - z_s^x\|_{L^\infty(\mathbb{R}^d; l^\infty)} + \|\omega(\frac{d}{2n})\|_{l^\infty},\end{aligned} \tag{25}$$

hence:

$$\begin{aligned}\|z_t^x - x_t^x\|_{L^\infty(\mathbb{R}^d; l^\infty)} &= \| \int_0^t V(x_s^x) - \Phi(z_s^x) ds \|_{L^\infty(\mathbb{R}^d; l^\infty)} \\ &\leqslant \int_0^t \|V(x_s^x) - \Phi(z_s^x)\|_{L^\infty(\mathbb{R}^d; l^\infty)} ds \\ &\leqslant L^V \int_0^t \|x_s^x - z_s^x\|_{L^\infty(\mathbb{R}^d; l^\infty)} ds + \left(\|\omega(\frac{d}{2n})\|_{l^\infty}\right) t.\end{aligned}$$

By applying Grönwall's inequality Pachpatte (1998, Theorem 1.3.1):

$$\|z_t^x - x_t^x\|_{L^\infty(\mathbb{R}^d; l^\infty)} \leqslant \left(\|\omega(\frac{d}{2n})\|_{l^\infty}\right) t e^{L^V t},$$

let $t = 1$:

$$\|\Psi(x) - \varphi(x)\|_{L^\infty(\mathbb{R}^d; l^\infty)} = \|z_1^x - x_1^x\|_{L^\infty(\mathbb{R}^d; l^\infty)} \leqslant \|\omega(\frac{d}{2n})\|_{l^\infty} e^{L^V}. \tag{26}$$

The right hand side goes to zero as $n \to \infty$.

Observe that the Lipschitz constant L^{Ψ} does *not* depend on the approximation parameter $n \in \mathbb{N}_+$. This can be shown as:

$$\begin{aligned} \left\|z_t^x - z_t^y\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} &\leqslant \left\|x-y\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} + \int_0^t \left\|\Phi(z_s^x) - \Phi(z_s^y)\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} ds \\ &\leqslant \left\|x-y\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} + L^{\Phi} \int_0^t \left\|z_s^x - z_s^y\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} ds. \end{aligned} \tag{27}$$

By applying Grönwall's inequality Pachpatte (1998, Theorem 1.2.2)

$$\|z_t^x - z_t^y\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \|x-y\|_{L^\infty(\mathbb{R}^d;l^\infty)} e^{tL^{\Phi}}.$$

Let $t = 1$:

$$\|\Psi(x) - \Psi(y)\|_{L^\infty(\mathbb{R}^d;l^\infty)} = \|z_1^x - z_1^y\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \|x-y\|_{L^\infty(\mathbb{R}^d;l^\infty)} e^{L^{\Phi}}. \tag{28}$$

This means that $L^{\Psi} \leqslant e^{L^{\Phi}}$. Since the Lipschitz constant of ϕ and the function b_δ do not depend on n, L^{Φ} is also independent of n. As a result, L^{Ψ} is independent of n and the result follows. □

Proposition A.8 (Universal Approximation by Deep Neural ODEs)**.** *Let $\varphi \in \mathcal{H}^I\left(\mathbb{R}^d, (0,1)^d\right)_0$ and $n, d, I \in \mathbb{N}_+$ then, there exist ReLU neural ODEs $\{\Psi_i = \mathrm{Flow}(\Phi_i)\}_{i=1}^I$ such that $\Psi \overset{\text{def.}}{=} \bigcirc_{i=1}^I \Psi_i$ satisfies the approximation guarantee:*

$$\left\|\varphi - \Psi\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \sum_{i=1}^{I} \left(\|\omega_i(\frac{d}{2n})\|_{l^\infty} \prod_{j=i}^{I} e^{L^{V_j}} \right), \tag{29}$$

where $(\omega_i)_j$ is the modulus of regularity of $(V_i)_j$ for $i \in [I]$ and $j \in [d]$. The right hand side converges to zero as $n \to \infty$.
Moreover, $\Psi \in \mathcal{H}\left(\mathbb{R}^d, (0,1)^d\right)_0$, with Lipschitz constant at most $\prod_{i=1}^I e^{L^{\Phi_i}}$ and $\Phi_1, \ldots, \Phi_I$ are ReLU MLP vector fields of depth at most $\lceil log_2 d \rceil + 7$, width at most $8d^2(n+1)^d + 2d^2 + d$, and at most $16d^2(n+1)^d + 6d^3 + 9d^2$ parameters. Furthermore, L^{Φ_i} and L^{Ψ} do not depend on the parameter n for $i \in [I]$.

Proof. Recall equation 4 as $\varphi \in \mathcal{H}^I\left(\mathbb{R}^d, [0,1]^d\right)_0$ there exist $\varphi_i = \mathrm{Flow}\left(V_i\right)$ for every $i \in [I]$ such that:

$$\varphi = \bigcirc_{i=1}^I \varphi_i(x). \tag{30}$$

We use lemma A.7 for each φ_i to find neural ODEs $\{\Psi_i\}_{i=1}^I$.

Define $\Psi \overset{\text{def.}}{=} \bigcirc_{i=1}^I \Psi_i$. We try to show that the Lipschitz constant L^{Ψ} is at most $\prod_{i=1}^I e^{L^{\Phi_i}}$ using induction. The base case is already proved in lemma A.7; for the step of the induction, consider the induction assumption below :

$$\left\| \bigcirc_{i=1}^{I-1} \Psi_i(x) - \bigcirc_{i=1}^{I-1} \Psi_i(y) \right\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \prod_{i=1}^{I-1} e^{L^{\Phi_i}} \|x-y\|_{L^\infty(\mathbb{R}^d;l^\infty)},$$

we can write:

$$\begin{aligned} \left\| \bigcirc_{i=1}^{I} \Psi_i(x) - \bigcirc_{i=1}^{I} \Psi_i(y) \right\|_{L^\infty(\mathbb{R}^d;l^\infty)} &\leqslant e^{L^{\Phi_I}} \left\| \bigcirc_{i=1}^{I-1} \Psi_i(x) - \bigcirc_{i=1}^{I-1} \Psi_i(y) \right\|_{L^\infty(\mathbb{R}^d;l^\infty)} \\ &\leqslant \prod_{i=1}^{I} e^{L^{\Phi_i}} \|x-y\|_{L^\infty(\mathbb{R}^d;l^\infty)}, \end{aligned} \tag{31}$$

So one can deduce the Lipschitz constant is at most $\prod_{i=1}^I e^{L^{\Phi_i}}$ and by lemma A.7 it doesn't depend on n.

We again use induction to show inequality equation 29. The base case is already proved in lemma A.7. For the step of the induction, consider the induction assumption below:

$$\| \bigcirc_{i=1}^{I-1} \varphi_i - \bigcirc_{i=1}^{I-1} \Psi_i \|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \sum_{i=1}^{I-1} \left(\|\omega_i(\frac{d}{2n})\|_{l^\infty} \prod_{j=i}^{I-1} e^{L^{V_j}} \right).$$

Using equation 30 and lemma A.7 one can write:

$$\begin{aligned} \|\varphi - \bigcirc_{i=1}^{I} \Psi_i\|_{L^\infty(\mathbb{R}^d;l^\infty)} &= \| \bigcirc_{i=1}^{I} \varphi_i - \bigcirc_{i=1}^{I} \Psi_i \|_{L^\infty(\mathbb{R}^d;l^\infty)} \\ &\leqslant \| \bigcirc_{i=1}^{I} \varphi_i - \varphi_I \circ \bigcirc_{i=1}^{I-1} \Psi_i \|_{L^\infty(\mathbb{R}^d;l^\infty)} \\ &\quad + \| \varphi_I \circ \bigcirc_{i=1}^{I-1} \Psi_i - \bigcirc_{i=1}^{I} \Psi_i \|_{L^\infty(\mathbb{R}^d;l^\infty)} \\ &\leqslant e^{L^{V_I}} \| \bigcirc_{i=1}^{I-1} \varphi_i - \bigcirc_{i=1}^{I-1} \Psi_i \|_{L^\infty(\mathbb{R}^d;l^\infty)} + e^{L^{V_I}} \|\omega_I(\frac{d}{2n})\|_{l^\infty}. \end{aligned}$$

Then by the induction assumption:

$$\|\varphi - \bigcirc_{i=1}^{I} \Psi_i\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \sum_{i=1}^{I} \left(\|\omega_i(\frac{d}{2n})\|_{l^\infty} \prod_{j=i}^{I} e^{L^{V_j}} \right).$$

By lemma A.7 the right hand side goes to zero as $n \to \infty$. □

Dependent Lipschitz Constant for Neural ODE, Better Rates: Naturally, one may wonder if improved rates are achievable under additional smoothness of the target homeomorphism.

Lemma A.9 (Transfer: Universal Approximation of Vector Field to Flow (Differentiable Case))**.** *Fix $k, d, L \in \mathbb{N}_+$ and $N_j \in \mathbb{N}_+$ for $j \in [d]$, let $V = (V_1, \ldots, V_d) \in \mathcal{X}^k(\mathbb{R}^d, (0,1)^d)$. Define for $j \in [d]$, $\upsilon_j \overset{\text{def.}}{=} 85(k+1)^d 8^k \|V_j\|_{C^k([0,1]^d)} (N_j L)^{-2k/d}$ and $\upsilon(N_1, \ldots, N_d, L) \overset{\text{def.}}{=} (\upsilon_1, \ldots, \upsilon_d)$; then for every time-one flow:*

$$\begin{aligned} \varphi(x) &= x_1^x \\ x_t^x &= x + \int_0^t V(x_s^x)\, ds, \qquad 0 \leqslant t \leqslant 1, \end{aligned} \tag{32}$$

there exists a ReLU MLP $\Phi : \mathbb{R}^d \to \mathbb{R}^d$, compactly supported on $(0,1)^d$ with width at most $\sum_{j=1}^{d} 1 + 2d + 17kd^{d+13}d(N_j+2)\log_2(8N_j)$ and depth at most $18k^2(L+2)\log_2(4L) + 2d + 3$ such that the neural ODE:

$$\begin{aligned} \Psi(x) &= z_1^x, \\ z_t^x &= x + \int_0^t \Phi(z_s^x)\, ds, \qquad 0 \leqslant t \leqslant 1, \end{aligned} \tag{33}$$

satisfies the uniform estimate:

$$\|\Psi - \varphi\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \|\upsilon(N_1, \ldots, N_d, L)\|_{l^\infty} e^{L^V}. \tag{34}$$

In terms of number of parameters P one can write:

$$\|\Psi - \varphi\|_{L^\infty(\mathbb{R}^d;l^\infty)} \in \mathcal{O}\big((\frac{P}{\log P})^{-2k/d}\big). \tag{35}$$

Note that $\Psi \in \mathcal{H}\left(\mathbb{R}^d, (0,1)^d\right)_0$.

Proof. For any $N_j, L \in \mathbb{N}_+$ by Lu et al. (2021, Theorem 1.1) for V_j, $j \in [d]$ there exists a function ϕ_j defined on $[0,1]^d$, implemented by a ReLU FNN with width at most $C_1(N_j+2)\log_2(8N_j)$ and depth at most $C_2(L+2)\log_2(4L) + 2d$ such that:

$$\|\phi_j - V_j\|_{L^\infty([0,1]^d)} \leqslant C_3 \|V_j\|_{C^k([0,1]^d)} (N_j L)^{-2k/d}, \tag{36}$$

where $C_1 = 17k^{d+1}3^d d$, $C_2 = 18k^2$, and $C_3 = 85(k+1)^d 8^k$. Define $(\phi)_j \overset{\text{def.}}{=} \phi_j$. V is supported on a compact set in $(0,1)^d$. Without loss of generality, suppose that V_j is supported on $[R, 1-R]^d$ for a $R > 0$. As ϕ is Lipschitz on $[0,1]^d$, there exists $M \in \mathbb{R}$ such that:

$$\|\phi\|_{L^\infty([0,1]^d;l^\infty)} \leqslant M.$$

Consider the ReLU bump function below for a small $\delta > 0$ such that $[R-\delta, 1-R+\delta] \subseteq [0,1]^d$:

$$\begin{aligned} b_{R,\delta}(x) &\overset{\text{def.}}{=} \sigma\left(1 - \frac{1}{\delta}\sum_{i=1}^{d}\left[\sigma(R-x_i) + \sigma\big(x_i - (1-R)\big)\right]\right) \\ &= \begin{cases} 0, & \text{if } x \notin [R-\delta, 1-R+\delta]^d, \\ 1, & \text{if } x \in [R, 1-R]^d. \end{cases} \end{aligned} \tag{37}$$

Implement the following network for $j \in [d]$:

$$\begin{aligned} \mathbf{x} &\Rightarrow (\phi_j, b_\delta) \\ &\Rightarrow (\phi_j, \sigma_{ReLU}(-Mb_\delta - \phi_j), b_\delta), \\ &\Rightarrow (\sigma_{ReLU}\big(Mb_\delta - (\phi_j + \sigma_{ReLU}(-Mb_\delta - \phi_j)), b_\delta\big), \\ &\Rightarrow Mb_\delta - \sigma_{ReLU}\big(Mb_\delta - (\phi_j + \sigma_{ReLU}(-Mb_\delta - \phi_j))\big) = \Phi_j. \end{aligned} \tag{38}$$

The resulting $\Phi_j = \min\{Mb_\delta, \max\{-Mb_\delta, \phi_j\}\}$ equals ϕ_j on $[R, 1-R]^d$ and equals zero outside $[R-\delta, 1-R+\delta]^d$. Φ_j has width:

$$\begin{aligned} \operatorname{width}(\Phi_j) &\leqslant \operatorname{width}(\phi_j) + 2d + 1 \\ &\leqslant C_1(N_j+2)\log_2(8N_j) + 2d + 1 \\ &\leqslant 17k^{d+1}3^d d(N_j+2)\log_2(8N_j) + 2d + 1. \end{aligned}$$

Its depth is:

$$\operatorname{depth}(\Phi_j) = \operatorname{depth}(\phi_j) + 3 = C_2(L+2)\log_2(4L) + 2d + 3.$$

If needed, one can increase the depth of the neural networks such that $\operatorname{depth}(\Phi_j) = \max_{1\leqslant i\leqslant d}\{\operatorname{depth}(\Phi_i)\}$, then by parallelization (Petersen & Zech, 2024, Proposition 2.3.) the neural network:

$$\Phi(x) = (\Phi_1(x), \ldots, \Phi_d(x)) : \mathbb{R}^d \to \mathbb{R}^d$$

has the same depth. Its width is at most:

$$\operatorname{width}(\Phi) \leqslant \sum_{j=1}^{d} \operatorname{width}(\Phi_j),$$

By equation 36 and the definition of Φ one has:

$$\|\Phi - V\|_{L^\infty(\mathbb{R}^d; l^\infty)} \leqslant \|\phi - V\|_{L^\infty([0,1]^d; l^\infty)} \leqslant \|\upsilon(N_1, \ldots, N_d, L)\|_{l^\infty}, \tag{39}$$

this yields:

$$\begin{aligned} \|V(x_s^x) - \Phi(z_s^x)\|_{L^\infty(\mathbb{R}^d; l^\infty)} &\leqslant \|V(x_s^x) - V(z_s^x)\|_{L^\infty(\mathbb{R}^d; l^\infty)} + \|V(z_s^x) - \Phi(z_s^x)\|_{L^\infty(\mathbb{R}^d; l^\infty)} \\ &\leqslant L^V \|x_s^x - z_s^x\|_{L^\infty(\mathbb{R}^d; l^\infty)} + \|\upsilon(N_1, \ldots, N_d, L)\|_{l^\infty}. \end{aligned} \tag{40}$$

Then:

$$\begin{aligned} \|x_t^x - z_t^x\|_{L^\infty(\mathbb{R}^d; l^\infty)} &= \| \int_0^t V(x_s^x) - \Phi(z_s^x) ds\|_{L^\infty(\mathbb{R}^d; l^\infty)} \\ &\leqslant \int_0^t \|V(x_s^x) - \Phi(z_s^x)\|_{L^\infty(\mathbb{R}^d; l^\infty)} ds \\ &\leqslant L^V \int_0^t \|x_s^x - z_s^x\|_{L^\infty(\mathbb{R}^d; l^\infty)} ds + \left(\|\upsilon(N_1, \ldots, N_d, L)\|_{l^\infty}\right) t. \end{aligned} \tag{41}$$

By applying Grönwall's inequality Pachpatte (1998, Theorem 1.3.1)

$$\|z_t^x - x_t^x\|_{L^\infty(\mathbb{R}^d; l^\infty)} \leqslant \|\upsilon(N_1, \ldots, N_d, L)\|_{l^\infty} t e^{L^V t}.$$

Let $t = 1$:

$$\|\Psi(x) - \varphi(x)\|_{L^\infty(\mathbb{R}^d; l^\infty)} = \|z_1^x - x_1^x\|_{L^\infty(\mathbb{R}^d; l^\infty)} \leqslant \|\upsilon(N_1, \ldots, N_d, L)\|_{l^\infty} e^{L^V}.$$

As the vector field is C^k on a bounded domain then $\|V\|_{C^k(0,1)^d}$ is bounded and by Lu et al. (2021, Corollary 1.3) one can express the rate in terms of the parameter count P as:

$$\|\Psi - \varphi\|_{L^\infty(\mathbb{R}^d;l^\infty)} \in \mathcal{O}\big((\frac{P}{\log P})^{-2k/d}\big). \tag{42}$$

To achieve the Lipschitz constant, as Φ is Lipschitz one can write:

$$\begin{aligned}\left\|z_t^x - z_t^y\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} &\leqslant \left\|x-y\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} + \int_0^t \left\|\Phi(z_s^x) - \Phi(z_s^y)\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} ds \\ &\leqslant \left\|x-y\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} + L^\Phi \int_0^t \left\|z_s^x - z_s^y\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} ds.\end{aligned} \tag{43}$$

By applying Grönwall's inequality Pachpatte (1998, Theorem 1.2.2):

$$\|z_t^x - z_t^y\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \left(\|x-y\|_{L^\infty(\mathbb{R}^d;l^\infty)}\right) e^{L^\Phi t}.$$

Set $t = 1$:

$$\left\|\Psi(x) - \Psi(y)\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} = \left\|z_1^x - z_1^y\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \left(\|x-y\|_{L^\infty(\mathbb{R}^d;l^\infty)}\right) e^{L^\Phi}. \tag{44}$$

The result follows. □

Proposition A.10 (Universal Approximation by Deep Neural ODEs (Differentiable Case))**.** *Consider $k, d, I, L \in \mathbb{N}_+$ and $N_j \in \mathbb{N}_+$ for $j \in [d]$ and $\varphi \in \mathrm{Diff}^{k,I}\left(\mathbb{R}^d, (0,1)^d\right)_0$; then, there exist ReLU neural ODEs $\{\Psi_i = \mathrm{Flow}(\Phi_i)\}_{i=1}^I$ such that $\Psi \overset{\text{def.}}{=} \bigcirc_{i=1}^I \Psi_i$ satisfies the approximation guarantee:*

$$\left\|\varphi - \Psi\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \sum_{i=1}^I \left(\|\upsilon_i(N_1, \ldots, N_d, L)\|_{l^\infty} \prod_{j=i}^I e^{L^{V_j}} \right). \tag{45}$$

Where $(\upsilon_i)_j = 85(k+1)^d 8^k \|(V_i)_j\|_{C^k([0,1]^d)} (N_j L)^{-2k/d}$ for $i \in [I]$ and $j \in [d]$. The right hand side goes to zero as $L \to \infty$ and $N_j \to \infty$ for each $j \in [d]$.
Moreover, $\Psi \in \mathcal{H}\left(\mathbb{R}^d, (0,1)^d\right)_0$, with Lipschitz constant at most $\prod_{i=1}^I e^{L^{\Phi_i}}$ where Φ_i are ReLU MLPs of width at most $\sum_{j=1}^d 1 + 2d + 17k^{d+1}3^d d(N_j+2)\log_2(8N_j)$ and depth at most $18k^2(L+2)\log_2(4L) + 2d + 3$ and $N_j, L \in \mathbb{N}_+$ for each $j \in [d]$, $i \in [I]$. Note that L^Ψ depends on the parameters $N_1, \ldots, N_d$. Furthermore, if L is not fixed, then in terms of the number of parameters P one can write:

$$\|\Psi - \varphi\|_{L^\infty(\mathbb{R}^d;l^\infty)} \in \mathcal{O}\big((\frac{P}{\log P})^{-2k/d}\big). \tag{46}$$

Proof. Recall as in equation 5, $\varphi \in \mathrm{Diff}^{k,I}\left(\mathbb{R}^d, (0,1)^d\right)_0$ means $\varphi_i = \mathrm{Flow}\left(V_i\right)$ for every $i \in [I]$ and :

$$\varphi = \bigcirc_{i=1}^I \varphi_i(x). \tag{47}$$

We use lemma A.9 for each φ_i to find neural ODEs $\{\Psi_i = \mathrm{Flow}(\Phi_i)\}_{i=1}^I$.

Define $\Psi \overset{\text{def.}}{=} \bigcirc_{i=1}^I \Psi_i$. We try to show that the Lipschitz constant L^Ψ is at most $\prod_{i=1}^I e^{L^{\Phi_i}}$ using induction. The base case is already proved in lemma A.9; for the step of the induction, consider the induction assumption below :

$$\left\| \bigcirc_{i=1}^{I-1} \Psi_i(x) - \bigcirc_{i=1}^{I-1} \Psi_i(y) \right\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \prod_{i=1}^{I-1} e^{L^{\Phi_i}} \|x-y\|_{L^\infty(\mathbb{R}^d;l^\infty)}.$$

Note:

$$\begin{aligned}\left\| \bigcirc_{i=1}^{I} \Psi_i(x) - \bigcirc_{i=1}^{I} \Psi_i(y) \right\|_{L^\infty(\mathbb{R}^d;l^\infty)} &\leqslant e^{L^{\Phi_I}} \left\| \bigcirc_{i=1}^{I-1} \Psi_i(x) - \bigcirc_{i=1}^{I-1} \Psi_i(y) \right\|_{L^\infty(\mathbb{R}^d;l^\infty)} \\ &\leqslant \prod_{i=1}^{I} e^{L^{\Phi_i}} \|x-y\|_{L^\infty(\mathbb{R}^d;l^\infty)}.\end{aligned} \tag{48}$$

So one can deduce the Lipschitz constant is at most $\prod_{i=1}^{I} e^{L^{\Phi_i}}$.
We again use induction to show inequality equation 45. The base case is already proved in lemma A.9. For the step of the induction consider the induction assumption below:

$$\| \bigcirc_{i=1}^{I-1} \varphi_i - \bigcirc_{i=1}^{I-1} \Psi_i \|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \sum_{i=1}^{I-1} \left(\|v_i(N_1,\dots,N_d,L)\|_{l^\infty} \prod_{j=i}^{I-1} e^{L^{V_j}} \right).$$

Using equation 47 one can write:

$$\begin{aligned}
\|\varphi - \bigcirc_{i=1}^{I} \Psi_i\|_{L^\infty(\mathbb{R}^d;l^\infty)} &= \| \bigcirc_{i=1}^{I} \varphi_i - \bigcirc_{i=1}^{I} \Psi_i \|_{L^\infty(\mathbb{R}^d;l^\infty)} \\
&\leqslant \| \bigcirc_{i=1}^{I} \varphi_i - \varphi_I \circ \bigcirc_{i=1}^{I-1} \Psi_i \|_{L^\infty(\mathbb{R}^d;l^\infty)} \\
&\quad + \|\varphi_I \circ \bigcirc_{i=1}^{I-1} \Psi_i - \bigcirc_{i=1}^{I} \Psi_i\|_{L^\infty(\mathbb{R}^d;l^\infty)} \\
&\leqslant \| \bigcirc_{i=1}^{I-1} \varphi_i - \bigcirc_{i=1}^{I-1} \Psi_i \|_{L^\infty(\mathbb{R}^d;l^\infty)} e^{L^{V_I}} \\
&\quad + \|v_I(N_1,\dots,N_d,L)\|_{l^\infty} e^{L^{V_I}}.
\end{aligned}$$

Then, by the induction assumption:

$$\|\varphi - \bigcirc_{i=1}^{I} \Psi_i\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \sum_{i=1}^{I} \left(\|v_i(N_1,\dots,N_d,L)\|_{l^\infty} \prod_{j=i}^{I} e^{L^{V_j}} \right).$$

By lemma A.9 the right hand side goes to zero as $L \to \infty$ and $N_j \to \infty$ for each $j \in [d]$. In terms of the number of parameters P, one can write from lemma A.9:

$$\|\varphi - \bigcirc_{i=1}^{I} \Psi_i\|_{L^\infty(\mathbb{R}^d;l^\infty)} \in \mathcal{O}\big((\frac{P}{\log P})^{-2k/d}\big). \tag{49}$$

We have thus concluded the proof. □

Proposition A.11 (Universal Approximation by Deep Neural ODEs (Differentiable Case)). *Let $\varphi \in \mathrm{Diff}^{k,I}\big(\mathbb{R}^d,(0,1)^d\big)_0$ and $N,\ L,\ k \in \mathbb{N}_+$ then, there exist ReLU neural ODEs $\{\Psi_i = \mathrm{Flow}(\Phi_i)\}_{i=1}^{I}$ such that $\Psi \overset{\text{def.}}{=} \bigcirc_{i=1}^{I} \Psi_i$ satisfies the approximation guarantee:*

$$\|\varphi - \Psi\|_{L^\infty(\mathbb{R}^d;l^\infty)} \leqslant \sum_{i=1}^{I} \left(\|v_i(N,L)\|_{l^\infty} \prod_{j=i}^{I} e^{L^{V_j}} \right). \tag{50}$$

Where $(v_i)_j = 85(k+1)^d 8^k \|(V_i)_j\|_{C^k([0,1]^d)} (NL)^{-2k/d}$ and the right hand side goes to zero as $L \to \infty$ and $N \to \infty$.
Moreover, $\Psi \in \mathcal{H}\left(\mathbb{R}^d,(0,1)^d\right)_0$ and its Lipschitz constant is at most $\prod_{i=1}^{I} e^{L^{\Phi_i}}$. $\Phi_1,\dots,\Phi_i$ are ReLU MLPs of width at most $d + 2d^2 + 17k^{d+1}3^d d^2(N+2)\log_2(8N)$ and depth at most $18k^2(L+2)\log_2(4L) + 2d + 3$. Note that L^Ψ depends on the parameter N. Furthermore, in terms of the number of parameters P, one can write:

$$\|\Psi - \varphi\|_{L^\infty(\mathbb{R}^d;l^\infty)} \in \mathcal{O}\big((\frac{P}{\log P})^{-2k/d}\big). \tag{51}$$

Proof. Set $N = N_1 = N_2 = \cdots = N_d$ then the result follows from proposition A.10. □

A.2.3 Decomposition

Lemma A.12 (Normality). *For an open set $K \subseteq M$, the set $F(K)$ generates a normal subgroup (denoted as) $\langle F(K)\rangle$ of $\mathrm{Diff}^\infty(K)_0$.*

Proof. Let $F \in \mathrm{Diff}^\infty(K)_0$ and $\varphi \in F(K)$ so there exists $V \in \mathcal{X}^\infty(K)$ such that $\varphi = \mathrm{Flow}(V)$, then by Lee (2013, Corollary 9.14) the time-one flow of the (pushforward) vector field $F_*V \in \mathcal{X}^\infty(K)$ is $\eta = F \circ \varphi \circ F^{-1}$, in other words $\eta = \mathrm{Flow}(F_*V)$ which means $\eta \in F(K)$.

Table 1: Parametric complexity of the ReLU MLP vector fields used in proposition A.8 and proposition A.11.

Complexity of MLP	**Lipschitz Case**	C^k**-Differentiable Case** ($k \geqslant 1$)
Depth	$\lceil \log_2(d) \rceil + 7$	$18k^2(L+2)\log_2(4L) + 2d + 3$
Width	$d + 2d^2 + 8d^2(n+1)^d$	$d + 2d^2 + 17k^{d+1}3^d d^2(N+2)\log_2(8N)$
Nonzero Parameters	$16d^2(n+1)^d + 6d^3 + 9d^2$	Explicit count not available

Now let $\varphi \in \langle F(K) \rangle$. This means there exists $n \in \mathbb{N}_+$ such that one can write φ as a composition of n time-one flows:

$$\varphi = \varphi_n \circ \cdots \circ \varphi_1.$$

Now if $F \in \mathrm{Diff}^\infty(K)_0$, one can write:

$$F \circ \varphi \circ F^{-1} = (F \circ \varphi_n \circ F^{-1}) \circ (F \circ \varphi_{n-1} \circ F^{-1}) \circ \cdots \circ (F \circ \varphi_1 \circ F^{-1}) = \eta_n \circ \cdots \circ \eta_1 \in \langle F(K) \rangle.$$

□

Lemma A.13 (Decomposition)**.** *Let $K \subseteq M$ where K is open and connected. Then for every diffeomorphism $\varphi \in \mathrm{Diff}^\infty(M, K)_0$, there exist $I \in \mathbb{N}_+$ and flows $\varphi_1, \ldots, \varphi_i \in F(M, K)$ such that:*

$$\varphi = \bigcirc_{i=1}^{I} \varphi_i \overset{\text{def.}}{=} \varphi_I \circ \cdots \circ \varphi_1. \tag{52}$$

Proof. By Thurston's Theorem, see e.g. Banyaga (1997, Theorem 2.1.1), the group $\mathrm{Diff}^\infty(K, K)_0$ is simple; i.e. it has no *proper* normal subgroups besides the trivial group consisting only of the identity. Since $\langle F(K, K) \rangle \neq \{\mathrm{id}_M\}$ and by Lemma equation A.12, $\langle F(K, K) \rangle$ is a normal subgroup of $\mathrm{Diff}^\infty(K, K)_0$, which means $F(K, K)$ generates the entire group $\mathrm{Diff}^\infty(K, K)_0$. Now for any $\varphi \in \mathrm{Diff}^\infty(M, K)_0$, one has $\varphi_{|K} \in \mathrm{Diff}^\infty(K, K)_0$ which means:

$$\varphi_{|K} = \bigcirc_{i=1}^{I} \varphi_i, \qquad \varphi_i = \mathrm{Flow}\,(V_i). \tag{53}$$

As V_i are compactly supported on the open set K one can extend each V_i by zero outside K. Note that $\varphi = \mathrm{id}$ outside K, so considering the composition of the time-one flows of the extended vector fields finishes the proof. □

A.2.4 The uniform bound

Conjugation-invariant norm: A conjugation-invariant norm $\nu : G \to [0; +\infty)$ on a group G is a function which satisfies the following axioms:

(i) $\nu(1) = 0$,
(ii) $\nu(f) = \nu(f^{-1}), \quad \forall f \in G$;
(iii) $\nu(fg) \leqslant \nu(f) + \nu(g), \quad \forall f, g \in G$;
(iv) $\nu(f) = \nu(gfg^{-1}), \quad \forall f, g \in G$;
(v) $\nu(f) > 0$, for all $f \neq 1$

BIP-meagre groups: A group G is called BIP-meagre if every conjugation-invariant norm on G is bounded and discrete.

Portable manifold: We say that a smooth connected open manifold M is portable if it admits a complete vector field X and a compact subset M_0 with the following properties:

- M_0 is an attractor of the flow x_t generated by X: for every compact subset $K \subset M$, there exists $T > 0$ such that
$$x_T(K) \subset M_0.$$
- There exists a diffeomorphism $f \in \mathrm{Diff}(M, M)_0$ such that
$$f(M_0) \cap M_0 = \varnothing.$$

The set M_0 is called the core of a portable manifold M.

Lemma A.14. *The group* $\mathrm{Diff}^\infty(M,M)_0$ *is BIP-meagre provided that the smooth connected open manifold* M *is portable.*

Proof. For the proof, see Burago et al. (2008, Theorem 1.17). □

Lemma A.15. *There exists* $I_d \in \mathbb{N}_+$ *such that for any integer* $I > I_d$, $\mathrm{Diff}^{\infty,I}\big(\mathbb{R}^d,(0,1)^d\big)_0$ *is empty.*

Proof. Define $S \stackrel{\text{def.}}{=} \{\,\mathrm{Flow}\,(V) \mid V \in \mathcal{X}^\infty((0,1)^d,(0,1)^d)\,\}$ and the norm:

$$\|f\|_{\mathrm{frag}} = \begin{cases} 0, & f = \mathrm{id}, \\ \min\{\, m \in \mathbb{N} \mid f = h_1 \cdots h_m,\ h_i \in S \text{ for } i \in [m]\,\}, & f \neq \mathrm{id}. \end{cases}$$

One can check that this norm is conjugation-invariant. As $(0,1)^d$ is portable, (it is diffeomorphic to an open ball), by lemma A.14 the autonomous norm on the group $\mathrm{Diff}^\infty\big((0,1)^d,(0,1)^d\big)_0$ is bounded by a constant I_d. As for any $\varphi \in \mathrm{Diff}^\infty\big(\mathbb{R}^d,(0,1)^d\big)_0$ one has $\varphi_{|(0,1)^d} \in \mathrm{Diff}^\infty\big((0,1)^d,(0,1)^d\big)_0$, this means any $\varphi \in \mathrm{Diff}^\infty\big(\mathbb{R}^d,(0,1)^d\big)_0$ has a minimal representation less than or equal to I_d. □

A.3 Positive Result Propositions and Lemmas ($[0,1]^d$ Case)

A.3.1 Independent Lipschitz Constant for Neural ODE, but Worse Rates

Lemma A.16 (Transfer: Universal Approximation of Vector Field to Flow)**.** *Let* $n \in \mathbb{N}_+$ *and* $V = (V_1,\ldots,V_d) \in \mathcal{X}^0([0,1]^d)$ *with modulus of regularity* $\omega_j : [0,d] \to \mathbb{R}$ *for* $j \in [d]$. *Consider its time-one flow:*

$$\begin{aligned} \varphi(x) &= x_1^x \\ x_t^x &= x + \int_0^t V(x_s^x)\,ds, \qquad 0 \leqslant t \leqslant 1, \end{aligned} \tag{54}$$

then there exists a ReLU MLP $\Phi : [0,1]^d \to \mathbb{R}^d$ *tangent to the faces of the cube with width at most* $8d^2(n+1)^d$, *depth at most* $\lceil log_2 d \rceil + 4$ *and at most* $16d^2(n+1)^d$ *free parameters such that the time-one flow (neural ODE):*

$$\begin{aligned} \Psi(x) &= z_1^x, \\ z_t^x &= x + \int_0^t \Phi(z_s^x)\,ds, \qquad 0 \leqslant t \leqslant 1 \end{aligned} \tag{55}$$

satisfies the uniform estimate:

$$\|\Psi - \varphi\|_{L^\infty([0,1]^d;l^\infty)} \leqslant \|\omega(\frac{d}{2n})\|_{l^\infty} e^{L^V}, \tag{56}$$

and one gets convergence as $n \to \infty$. *Moreover,* $L^\Psi \leqslant e^{L^\Phi}$ *and* L^Ψ *and* L^Φ *do not depend on* n.

Proof. As $V_j : [0,1]^d \to \mathbb{R}$ for $j \in [d]$ is Lipschitz on $[0,1]^d$, from lemma A.6 there exists a ReLU MLP $\Phi_j : [0,1]^d \to \mathbb{R}$ tangent to the faces of the cube and modulus of regularity ω_j with width at most $8d(n+1)^d$, depth at most $\lceil log_2 d \rceil + 4$ and at most $16d(n+1)^d$ free parameters such that:

$$\|V_j - \Phi_j\|_{L^\infty([0,1]^d)} \leqslant \omega_j(\frac{d}{2n}). \tag{57}$$

If needed, one can increase the depth of the neural networks such that $\mathrm{depth}(\Phi_j) = \max_{1\leqslant i\leqslant d}\{\mathrm{depth}(\Phi_i)\}$, then by parallelization (Petersen & Zech, 2024, Proposition 2.3.) the neural network:

$$\Phi(x) = (\Phi_1(x),\ldots,\Phi_d(x)) : \mathbb{R}^d \to \mathbb{R}^d$$

has the same depth. Its width is at most:

$$\text{width}(\Phi) \leqslant \sum_{j=1}^{d} \text{width}(\Phi_j) = 8d^2(n+1)^d,$$

and free parameters at most $16d^2(n+1)^d$. Define $(\omega)_j \stackrel{\text{def.}}{=} \omega_j$ then:

$$\|\Phi - V\|_{L^\infty([0,1]^d;l^\infty)} \leqslant \|\omega(\frac{d}{2n})\|_{l^\infty}, \tag{58}$$

Note that if V is such that it is tangent on the faces of the cube, then the MLP has the same value of the vector field on the vertices of boundary simplices. Also note that the proof uses hat functions and triangulation which makes the MLP tangent to the faces. This means there won't be any outward normal vector on the boundary and in particular, there will be no flow-out of $\text{Flow}(\Phi)$. this yields:

$$\begin{aligned}\|V(x_s^x) - \Phi(z_s^x)\|_{L^\infty([0,1]^d;l^\infty)} &\leqslant \|V(x_s^x) - V(z_s^x)\|_{L^\infty([0,1]^d;l^\infty)} + \|V(z_s^x) - \Phi(z_s^x)\|_{L^\infty([0,1]^d;l^\infty)} \\ &\leqslant L^V \|x_s^x - z_s^x\|_{L^\infty([0,1]^d;l^\infty)} + \|\omega(\frac{d}{2n})\|_{l^\infty},\end{aligned} \tag{59}$$

hence:

$$\begin{aligned}\|z_t^x - x_t^x\|_{L^\infty([0,1]^d;l^\infty)} &= \|\int_0^t V(x_s^x) - \Phi(z_s^x)ds\|_{L^\infty([0,1]^d;l^\infty)} \\ &\leqslant \int_0^t \|V(x_s^x) - \Phi(z_s^x)\|_{L^\infty([0,1]^d;l^\infty)}ds \\ &\leqslant L^V \int_0^t \|x_s^x - z_s^x\|_{L^\infty([0,1]^d;l^\infty)}ds + \left(\|\omega(\frac{d}{2n})\|_{l^\infty}\right)t.\end{aligned}$$

By applying Grönwall's inequality Pachpatte (1998, Theorem 1.3.1):

$$\|z_t^x - x_t^x\|_{L^\infty([0,1]^d;l^\infty)} \leqslant \left(\|\omega(\frac{d}{2n})\|_{l^\infty}\right) te^{L^V t},$$

let $t = 1$:

$$\|\Psi(x) - \varphi(x)\|_{L^\infty([0,1]^d;l^\infty)} = \|z_1^x - x_1^x\|_{L^\infty([0,1]^d;l^\infty)} \leqslant \|\omega(\frac{d}{2n})\|_{l^\infty} e^{L^V}. \tag{60}$$

The right hand side goes to zero as $n \to \infty$.

Observe that the Lipschitz constant L^Ψ does *not* depend on the approximation parameter $n \in \mathbb{N}_+$. This can be shown as:

$$\begin{aligned}\left\|z_t^x - z_t^y\right\|_{L^\infty([0,1]^d;l^\infty)} &\leqslant \left\|x - y\right\|_{L^\infty([0,1]^d;l^\infty)} + \int_0^t \left\|\Phi(z_s^x) - \Phi(z_s^y)\right\|_{L^\infty([0,1]^d;l^\infty)}ds \\ &\leqslant \left\|x - y\right\|_{L^\infty([0,1]^d;l^\infty)} + L^\Phi \int_0^t \left\|z_s^x - z_s^y\right\|_{L^\infty([0,1]^d;l^\infty)}ds.\end{aligned} \tag{61}$$

By applying Grönwall's inequality Pachpatte (1998, Theorem 1.2.2)

$$\|z_t^x - z_t^y\|_{L^\infty([0,1]^d;l^\infty)} \leqslant \|x - y\|_{L^\infty([0,1]^d;l^\infty)} e^{tL^\Phi}.$$

Let $t = 1$:

$$\|\Psi(x) - \Psi(y)\|_{L^\infty([0,1]^d;l^\infty)} = \|z_1^x - z_1^y\|_{L^\infty([0,1]^d;l^\infty)} \leqslant \|x - y\|_{L^\infty([0,1]^d;l^\infty)} e^{L^\Phi}. \tag{62}$$

This means that $L^\Psi \leqslant e^{L^\Phi}$. Since L^Φ is independent of n, L^Ψ is also independent of n and the result follows. □

Proposition A.17 (Universal Approximation by Deep Neural ODEs). *Let $\varphi \in \mathcal{H}^I\left([0,1]^d\right)_0$ and $n, d, I \in \mathbb{N}_+$ then, there exist ReLU neural ODEs $\{\Psi_i = \mathrm{Flow}(\Phi_i)\}_{i=1}^I$ such that $\Psi \overset{\text{def.}}{=} \bigcirc_{i=1}^I \Psi_i$ satisfies the approximation guarantee:*

$$\left\|\varphi - \Psi\right\|_{L^\infty([0,1]^d; l^\infty)} \leqslant \sum_{i=1}^{I} \left(\|\omega_i(\frac{d}{2n})\|_{l^\infty} \prod_{j=i}^{I} e^{L^{V_j}} \right), \tag{63}$$

where $(\omega_i)_j$ is the modulus of regularity of $(V_i)_j$ for $i \in [I]$ and $j \in [d]$. The right hand side converges to zero as $n \to \infty$.
Moreover, $\Psi \in \mathcal{H}\left([0,1]^d\right)_0$, with Lipschitz constant at most $\prod_{i=1}^I e^{L^{\Phi_i}}$ and $\Phi_1, \ldots, \Phi_I$ are ReLU MLP vector fields with width at most $8d^2(n+1)^d$, depth at most $\lceil log_2 d \rceil + 4$ and at most $16d^2(n+1)^d$ free parameters. Furthermore, L^{Φ_i} and L^Ψ do not depend on the parameter n for $i \in [I]$.

Proof. The proof is similar to proposition A.8 but instead use lemma A.16. □

A.4 Proofs of Theorems

Theorem 3.1. : Choose an arbitrary flowable $f \in \mathcal{H}\left(\mathbb{R}^d, K\right)_0$. By proposition A.1 (for $M = K$) there exists $x \in R(f)$.

First, consider the case $R(f) \backslash \mathrm{Fix}(f) \neq \varnothing$ and take $x \in R(f) \backslash \mathrm{Fix}(f)$. As calculation in d_1 distance results in d_{C^0} distance, by lemma A.4, there exists an $\frac{\varepsilon}{2}$-C^0-perturbation $f_1 = f \circ h$ such that $f_1^k(x) = x$ for some $k > 1$ and $f_1 \in \mathcal{H}\left(\mathbb{R}^d, \tilde{K}\right)_0$ where $K \subset \tilde{K}$ is compact. Choose r small enough such that $\mathcal{B} = \{B(f_1^i(x), r) : 1 \leqslant i \leqslant k\}$ forms a pairwise disjoint collection of balls inside Q.

Set $p_i \overset{\text{def.}}{=} f_1^i(x)$. Choose small balls U_i and compact closed balls V_i such that $p_i \in V_i \subset U_i \subset \overline{B(p_i, r)}$ and $f_1(V_i) \subset B(p_{i+1}, r)$; and corresponding orientation-preserving coordinate charts $\psi_i : U_i \to \mathbb{R}^d$ such that $\psi(p_i) = 0$. Define $g_i : U_i \to U_{i+1}$, $g_i \overset{\text{def.}}{=} \psi_{i+1}^{-1}(\lambda \psi_i(y))$ for $0 < \lambda < 1$ on U_i, thus $g_i(p_i) = p_{i+1}$ and in coordinates $\|\psi_{i+1}(g_i(y)) - \psi_{i+1}(g_i(z))\|_2 = \lambda \|\psi_i(y) - \psi_i(z)\|_2$ so g_i is smooth and contracting. Set $c_i : f_1(U_i) \to U_{i+1}$, $c_i \overset{\text{def.}}{=} g_i \circ f_1^{-1}$ which is a homeomorphism from $f_1(V_i)$ to the range of g_i; Define $\tilde{c}_i : f_1(V_i) \to c_i(f_1(V_i))$, $\tilde{c}_i \overset{\text{def.}}{=} c_i|_{f_1(V_i)}$. The closed sets $f_1(V_i)$ and $c_i(f_1(V_i))$ are flat topological closed balls. This is because one can consider a homeomorphism $\varphi : B(0,2) \to U_i$ such that $\varphi|_{\overline{B(0,1)}}$ is a homeomorphism from $\overline{B(0,1)}$ to V_i. Then by the way the functions φ, f_1 and c_i are constructed, the homeomorphism $\eta \overset{\text{def.}}{=} c_i \circ f_1 \circ \varphi|_{\overline{B(0,1)}}$:

$$\overline{B(0,1)} \xrightarrow{\varphi|_{\overline{B(0,1)}}} V_i \xrightarrow{f_1} f_1(V_i) \xrightarrow{c_i} c_i(f_1(V_i)),$$

can be extended to a homeomorphism from the neighborhood $B(0,2)$ of $\overline{B(0,1)}$:

$$B(0,2) \xrightarrow{\varphi} U_i \xrightarrow{f_1} f_1(U_i) \xrightarrow{c_i} U_{i+1}.$$

This proves that $c_i(f_1(V_i))$ is a flat topological ball. Similarly, $f_1(V_i)$ is a flat topological ball (without composing c_i, i.e. $\tilde{\eta} \overset{\text{def.}}{=} f_1 \circ \varphi|_{\overline{B(0,1)}}$).

Then by Goldstein et al. (2025, Theorem 1.9.) one can obtain homeomorphisms H_i that are the identity outside of $B(p_{i+1}, r)$ and are equal to $\tilde{c}_i$ on $f_1(V_i)$. Now note that $H_i \circ f_1 = g_i$ on V_i and $H_i \circ f_1 = f_1$ outside of $B(p_{i+1}, r)$. As the balls in $\mathcal{B}$ are disjoint, $f_2 \overset{\text{def.}}{=} \bigcirc_{i=1}^k (H_i) \circ f_1$ is a homeomorphism which is equal to f_1 outside of $\mathcal{B}$, $f_2^i(x)$ is a periodic attractor for f_2^k and $f_2^k(p_i)$ are isolated fixed points of f_2^k for each $i \in \{1, \ldots, k\}$. Having isolated attracting points yields a homeomorphism which is not flowable and by choosing r sufficiently small, one gets $d_{C^0}(f_2, f_1) = \|f_2 - f_1\|_{C^0} < \frac{\varepsilon}{2}$.

This shows that such homeomorphisms $\{f_2\}$ form a dense set of homeomorphisms compactly supported$_{\mathrm{id}}$ on Q which are not flowable and they have (non-fixed) periodic points of period $k > 1$.

Note that for each homeomorphism f_2, there exists a sufficiently small neighborhood U of x such that $f_{2\,|U}^k$ has no fixed point on ∂U. For every C^0-Perturbation g of f_2 in $\mathcal{H}\left(\mathbb{R}^d, Q\right)_0$, one has that $0 \neq \deg(f_2^k(x) - x, U, 0) = \deg(g^k(x) - x, U, 0)$, which means, there exists a point z such that $g^k(z) - z = 0$, or in other words, g^k has a fixed point in U and no fixed point on the boundary of U. Thus, g has a periodic point of period $k > 1$ that is not contained in a continuum subset of the set of k-periodic points. It means f_2 has a C^0-open neighbourhood. This proves the openness.

Now consider $x \in R(f) \subseteq \mathrm{Fix}(f)$. Take a point $x \in \mathrm{int}(Q \backslash K)$ and $0 < r < \varepsilon$ small enough such that $B(x, r) \subseteq \mathrm{int}(Q \backslash K)$, then consider ϕ the time-one flow of the vector field $\rho \cdot R$, where R is the vector field whose time-one map gives the π rotation around the point x and ρ is a smooth bump function equal to 1 on $B(x, \frac{r}{2})$ and zero outside $B(x, r)$. The composition $f_1 \overset{\text{def.}}{=} f \circ \phi \in \mathcal{H}\left(\mathbb{R}^d, Q\right)_0$ satisfies $d_{C^0}(f_1, f) = \|f_1 - f\|_{C^0} < \varepsilon$ having a period-two point (because f is the identity outside K), which is the condition of having a non-fixed periodic point. This finishes the proof. □

Theorem 3.2. : Let $K_N \overset{\text{def.}}{=} [\frac{1}{N}, 1 - \frac{1}{N}]^d$. By definition for $N \geqslant 3$:

$$\mathcal{H}\left(\mathbb{R}^d, (0,1)^d\right)_0 = \bigcup_{K_N} \mathcal{H}\left(\mathbb{R}^d, K_N\right)_0,$$

and by theorem 3.1, for any compact set $K_N \subseteq K_{N+1}$, each flowable $\mathcal{H}\left(\mathbb{R}^d, K_N\right)_0$ is nowhere dense in $\mathcal{H}\left(\mathbb{R}^d, K_{N+1}\right)_0 \subseteq \mathcal{H}\left(\mathbb{R}^d, (0,1)^d\right)_0$, so the set of flowable $\mathcal{H}(\mathbb{R}^d, (0,1)^d)$ is meagre in $\mathcal{H}\left(\mathbb{R}^d, (0,1)^d\right)_0$. □

Theorem 3.3. : By lemma A.13, $\varphi \in \mathrm{Diff}^\infty\left(\mathbb{R}^d, (0,1)^d\right)_0$ has a representation. By lemma A.15 we conclude that among all the representations, there is one with I autonomous flows where $I \leqslant I_d$. This means one can write any diffeomorphism as a composition of at most I_d flows $\varphi_i \in F(\mathbb{R}^d, (0,1)^d)$. □

Theorem 3.4. : As $\varphi \in \mathrm{Diff}^\infty\left(\mathbb{R}^d, (0,1)^d\right)_0$ by theorem 3.3 there exists $I \leqslant I_d$ such that $\varphi \in \mathrm{Diff}^{k,I}\left(\mathbb{R}^d, (0,1)^d\right)_0$ for any $k \in N_+$. Then by proposition A.11 there exists Ψ such that:

$$\begin{aligned} \left\|\varphi - \Psi\right\|_{L^\infty(\mathbb{R}^d; l^\infty)} &\leqslant \sum_{i=1}^{I} \left(\|v_i(n, L)\|_{l^\infty} \prod_{j=i}^{I} e^{L^{V_j}} \right) \\ &\leqslant 85(k+1)^d 8^k \max_{i,j} \{\|(V_i)_j\|_{C^k([0,1]^d)}\} (NL)^{-2k/d} e^{I_d \max_i \{L^{V_i}\}} I_d \\ &\in \mathcal{O}\big((NL)^{-2k/d}\big), \end{aligned} \tag{64}$$

as $(v_i)_j = 85(k+1)^d 8^k \|(V_i)_j\|_{C^k([0,1]^d)} (NL)^{-2k/d}$. equation 8 is a direct result of proposition A.11. □

Theorem 3.6. : By lemma A.5, one has for every $\varepsilon > 0$, there exists a diffeomorphism $\varphi_\varepsilon \in \mathrm{Diff}^\infty\left(\mathbb{R}^d, (0,1)^d\right)_0$ satisfying the uniform approximation guarantee:[4]

$$\|\varphi_\varepsilon - \varphi\|_{L^\infty(\mathbb{R}^d; l^\infty)} \leqslant \tfrac{\varepsilon}{2}. \tag{65}$$

By applying theorem 3.3 and proposition A.8, there exist ReLU neural ODEs $\{\Psi_i = \mathrm{Flow}(\Phi_i)\}_{i=1}^I$ for $I \leqslant I_d$ such that the homeomorphism $\Psi \overset{\text{def.}}{=} \bigcirc_{i=1}^I \Psi_i$ is Lipschitz and compactly supported$_{\mathrm{id}}$ on $(0,1)^d$ and satisfies the approximation guarantee:

$$\left\|\varphi_\varepsilon - \Psi\right\|_{L^\infty(\mathbb{R}^d; l^\infty)} \leqslant \tfrac{\varepsilon}{2}. \tag{66}$$

Inequalities equation 65 and equation 66 yields the theorem. □

Theorem 4.1. : By definition, $\mathrm{NODE}_\sigma([0,1]^d)$ is a subset of flowable $\mathcal{H}\left([0,1]^d\right)_0$. From Bonomo & Varandas (2020, Theorem A) flowable $\mathcal{H}\left([0,1]^d\right)_0$ is nowhere dense in $\mathcal{H}\left([0,1]^d\right)_0$. This proves the theorem. □

[4] Recall the norm in lemma A.5 is the uniform norm

Theorem 4.2. : Define $h \in \mathcal{H}\left([0,1]^2\right)_0$ such that it rotates $(0,1)^2$ around $(\frac{1}{2},\frac{1}{2})$ and shifts the boundary by $\frac{1}{2}$. In particular, let $Q=[0,1]^2$. Parametrize its boundary counterclockwise by:

$$\gamma\colon \mathbb{R}/4\mathbb{Z} \longrightarrow \partial Q, \qquad \gamma(s) \overset{\text{def.}}{=} \begin{cases} (s,0), & 0 \leqslant s \leqslant 1, \\ (1,s-1), & 1 \leqslant s \leqslant 2, \\ (3-s,1), & 2 \leqslant s \leqslant 3, \\ (0,4-s), & 3 \leqslant s \leqslant 4. \end{cases}$$

Define a boundary homeomorphism $b\colon \partial Q \to \partial Q$ by $b(\gamma(s)) \overset{\text{def.}}{=} \gamma\left(s+\frac{1}{2} \pmod 4\right)$. In particular, $b(0,0) = b(\gamma(0)) = \gamma\left(\frac{1}{2}\right) = \left(\frac{1}{2},0\right)$. Set $c \overset{\text{def.}}{=} (1/2,1/2)$. Every $x \in Q\backslash\{c\}$ admits a unique representation:

$$x = c + r(p-c), \qquad p \in \partial Q, \quad 0 < r \leqslant 1,$$

where $r = 2\|x-c\|_\infty, p = c + \frac{x-c}{r}$. Define $h\colon Q \to Q$ by:

$$h(c) = c, \qquad h\big(c+r(p-c)\big) = c + r\big(b(p)-c\big).$$

Then h is a homeomorphism of Q, with inverse obtained by replacing b with $b^{-1}(\gamma(s)) = \gamma\left(s-\frac{1}{2} \pmod 4\right)$. Moreover, $h(0,0) = \left(\frac{1}{2},0\right)$.

By (Jakob, 2026, Lemma 4.2.), a C^1-diffeomorphism on the cube maps faces to faces and vertices to vertices. Define $f \in \mathcal{H}\left([0,1]^d\right)_0$, $f \overset{\text{def.}}{=} h \times \mathrm{id}_{[0,1]^{d-2}}$. f has uniform distance at least $\frac{1}{2}$ from any diffeomorphism $g \in \mathrm{Diff}^1\left([0,1]^d\right)_0$. In particular,

$$\|g-f\|_{L^\infty([0,1]^d;l^\infty)} \geqslant \|g(0)-f(0)\|_{l^\infty} \geqslant \frac{1}{2}.$$

Thus h cannot be uniformly approximated by diffeomorphisms of the square. □

Theorem 4.3. : According to Glöckner et al. (2026) the cube is a simple polytope. By Glöckner et al. (2026, Corollary 1.6.) the identity group $\mathrm{Diff}^\infty\left([0,1]^d\right)_0$ can be written as a composition of time-one autonomous flows. □

Theorem 4.4. : As $\varphi \in \mathrm{Diff}^\infty\left([0,1]^d\right)_0$ by theorem 4.3 there exists an I such that $\varphi \in \mathrm{Diff}^{k,I}\left([0,1]^d\right)_0$ for any $k \in \mathbb{N}_+$. Then by proposition A.17 there exists Ψ such that:

$$\begin{aligned} \left\|\varphi-\Psi\right\|_{L^\infty([0,1]^d;l^\infty)} &\leqslant \sum_{i=1}^{I}\left(\|\omega_i(\frac{d}{2n})\|_{l^\infty}\prod_{j=i}^{I}e^{L^{V_j}}\right), \\ &\leqslant \max_i\{\|\omega_i(\frac{d}{2n})\|\}e^{I_d\max_i\{L^{V_i}\}}I_d \\ &\in \mathcal{O}\big(\frac{d}{n}\big), \end{aligned} \tag{67}$$

as $(v_i)_j = 85(k+1)^d 8^k \|(V_i)_j\|_{C^k([0,1]^d)}(NL)^{-2k/d}$. as the number of parameters $P \in \mathcal{O}\left(d^2(n+1)^d\right)$ then one can write the upper bound in $O\big(d(\frac{d}{P})^{1/d}\big)$ which means for fixed d:

$$\left\|\varphi-\Psi\right\|_{L^\infty(\mathbb{R}^d;l^\infty)} \in \mathcal{O}\big(P^{-1/d}\big). \tag{68}$$

□

A.5 Proof of Corollary

Corollary 3.7. : Recall in equation 13, V_{f_i} is compactly supported for each $i \in [D]$, by proposition A.8 there exist Lipschitz ReLU neural ODEs $\{\Psi_i = \mathrm{Flow}\,(\Phi_i)\}_{i=1}^D$ such that for $\varepsilon \in O\big(\frac{d+1}{n}\big)$ and each $i \in [D]$ one has:

$$\|\,\mathrm{Flow}\,(V_{f_i}) - \Psi_i\|_{L^\infty(\mathbb{R}^{d+1};l^\infty)} < \epsilon.$$

By restricting the domain and only considering the last coordinate of the output, one has:

$$\|\pi_1^{d+1} \circ \mathrm{Flow}\,(V_{f_i}) \circ \iota_d^{d+1} - \pi_1^{d+1} \circ \Psi_i \circ \iota_d^{d+1}\|_{L^\infty(\mathbb{R}^d)} < \epsilon.$$

Recall from equation 13 that $f_i = \pi_1^{d+1} \circ \mathrm{Flow}\,(V_{f_i}) \circ \iota_d^{d+1}$. By enlarging the depth and width of each Φ_i to be equal to $\max_{i\in[D]}\{\mathrm{depth}(\Phi_i)\}$ and $\max_{i\in[D]}\{\mathrm{width}(\Phi_i)\}$ if one needs to have the same width and depth, one can use concatenation (Petersen & Zech, 2024, Proposition 2.3.):

$$\Psi \overset{\text{def.}}{=} \bigoplus_{i=1}^{D} (\pi_1^{d+1} \circ \Psi_i \circ \iota_d^{d+1}),$$

where Ψ_i has width and parameters at most $\mathcal{O}(D(d+1)^2(n+1)^{d+1})$ and depth at most $\mathcal{O}(\lceil \log_2(d+1)\rceil)$ for each $i \in [D]$. This yields:

$$\|f - \Psi\|_{L^\infty(\mathbb{R}^d;l^\infty)} < \epsilon.$$

Note that by lifting, projecting and concatenation, the function remains Lipschitz. Finally, recall by proposition A.8 the Lipschitz constant is independent of n. □

A.6 Additional Notations and Definitions:

- Compositional Notation: The structure of a ReLU MLP Φ is represented in the following way: suppose $\Phi = \mathcal{L}_m \circ (\sigma \circ \mathcal{L}_{m-1}) \circ \cdots \circ (\sigma \circ \mathcal{L}_2) \circ (\sigma \circ \mathcal{L}_1)$ where the $\mathcal{L}_i$'s are affine transformations. We express this notationally as:

$$\begin{aligned}\mathbf{x} &\Longrightarrow (\sigma\circ\mathcal{L}_1)(\mathbf{x}) \Longrightarrow (\sigma\circ\mathcal{L}_2)\circ(\sigma\circ\mathcal{L}_1)(\mathbf{x}),\\ &\Longrightarrow (\sigma\circ\mathcal{L}_{m-1})\circ\cdots\circ(\sigma\circ\mathcal{L}_2)\circ(\sigma\circ\mathcal{L}_1)(\mathbf{x}),\\ &\Longrightarrow \mathcal{L}_m\circ(\sigma\circ\mathcal{L}_{m-1})\circ\cdots\circ(\sigma\circ\mathcal{L}_2)\circ(\sigma\circ\mathcal{L}_1)(\mathbf{x}),\\ &= \Phi(\mathbf{x})\,.\end{aligned}$$

  In other words, if $x_1, x_2, \ldots, x_{m-1}$ are the $1, 2, \ldots, (m-1)$-th hidden layers of Φ and x_m is the output layer, then the structure of Φ is expressed as:

$$x \implies x_1 \implies x_2 \implies \cdots \implies x_{m-1} \implies x_m = \Phi(x)\,.$$

- For a function $f : \mathbb{R}^d \to \mathbb{R}^d$ let $\deg(f, U, x)$ be the Brouwer degree of f at the point $x \in \mathbb{R}^d$, assuming $x \notin f(\partial U)$ and U open and bounded and f is continuous on $\overline{U}$.
- A continuum is a compact and connected metric space that contains at least two points.
- Recurrent point: Let X be a topological space and let $f : X \to X$ be a continuous map. A point $x \in X$ is called a *recurrent point* of f if there exists a sequence of integers (n_k) with $n_k \to \infty$ such that:

$$f^{n_k}(x) \longrightarrow x \quad \text{as } k \to \infty.$$

  Equivalently, x is recurrent if it belongs to its own ω-limit set:

$$R(f) \overset{\text{def.}}{=} \{y \in X : f^{n_k}(x) \to y \text{ for some sequence } n_k \to \infty\}.$$

- Orbit, fixed point and periodic point: Let X be a set and $f : X \to X$ a map. For $x \in X$, the *forward orbit* of x under f is the set:

$$\mathcal{O}^+(x) \overset{\text{def.}}{=} \{f^n(x) : n \in \mathbb{N}_0\},$$

  where $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and $f^0 = \mathrm{id}_X$.
  If f is invertible, the *(full) orbit* of x is:

$$\mathcal{O}(x) \overset{\text{def.}}{=} \{f^n(x) : n \in \mathbb{Z}\}.$$

  A point $x \in X$ is called a fixed point of f if:

$$f(x) = x.$$

  A point $x \in X$ is called a periodic point of period $k \geqslant 1$ if:

$$f^k(x) = x,$$

  and k is the smallest positive integer with this property.
  The set of all periodic points of exact period k is denoted:

$$\mathrm{Per}_k(f) \overset{\text{def.}}{=} \{\, x \in X : f^k(x) = x \text{ and } f^j(x) \neq x \text{ for all } 0 < j < k \,\}.$$

Given a subset $P \subseteq M$ of a manifold M, for $k \in \mathbb{N}_+ \cup \{\infty\}$:

- We say that a closed set $D \subset M$ is a flat topological closed ball if there is a surjective homeomorphism $F : \overline{B(0,1)} \to D$ that can be extended to a homeomorphism on a neighborhood of $\overline{B(0,1)}$.
- Define $F(M,K) \stackrel{\text{def.}}{=} \{\mathrm{Flow}(V) | V \in \mathcal{X}^\infty(M,K)\}$. For simplicity, when there is no confusion, we might write $F(K,K)$ as $F(K)$.

Let (M,g) be a connected Riemannian manifold:

- The length of a piecewise C^1 curve $\gamma\colon [a,b] \to M$ is defined by:
$$L_g(\gamma) \stackrel{\text{def.}}{=} \int_a^b \sqrt{g_{\gamma(t)}\big(\dot\gamma(t),\dot\gamma(t)\big)}\,dt.$$
- The distance induced by the Riemannian metric g is:
$$d_g(p,q) \stackrel{\text{def.}}{=} \inf_{\substack{\gamma\colon [a,b]\to M\\ \gamma(a)=p,\ \gamma(b)=q}} L_g(\gamma), \qquad p,q \in M,$$
where the infimum is taken over all piecewise C^1 curves joining p to q. For the Euclidean metric on $M = \mathbb{R}^d$: $d_g = \|p-q\|_2$.
- The norm induced by metric g is:
$$\|v\|_g \stackrel{\text{def.}}{=} \sqrt{g_p(v,v)}, \qquad v \in T_pM.$$
- We denote the ball in M with radius r around point p as $B_M(p,r) \stackrel{\text{def.}}{=} \{x \in M |\ d_g(x,p) < r\}$.
- For two homeomorphisms f, g on M define:
$$d_1(f,g) \stackrel{\text{def.}}{=} \sup_{x\in M}\Big\{d_g\big(f(x),g(x)\big), d_g\big(f^{-1}(x),g^{-1}(x)\big)\Big\}.$$