UNIVERSAL APPROXIMATION POWER OF DEEP RESIDUAL NEURAL NETWORKS VIA NONLINEAR CONTROL THEORY

Anonymous authors
Paper under double-blind review

ABSTRACT

In this paper, we explain the universal approximation capabilities of deep residual neural networks through geometric nonlinear control. Inspired by recent work establishing links between residual networks and control systems, we provide a general sufficient condition for a residual network to have the power of universal approximation by asking the activation function, or one of its derivatives, to satisfy a quadratic differential equation. Many activation functions used in practice satisfy this assumption, exactly or approximately, and we show this property to be sufficient for an adequately deep neural network with $2n$ states to approximate arbitrarily well, on a compact set and with respect to the supremum norm, any continuous function from $\mathbb{R}^n$ to $\mathbb{R}^n$. We further show this result to hold for very simple architectures for which the weights only need to assume two values. The first key technical contribution consists of relating the universal approximation problem to controllability of an ensemble of control systems corresponding to a residual network and to leverage classical Lie algebraic techniques to characterize controllability. The second technical contribution is to identify monotonicity as the bridge between controllability of finite ensembles and uniform approximability on compact sets.

1 INTRODUCTION

In the past few years, we have witnessed a resurgence in the use of techniques from dynamical and control systems for the analysis of neural networks. This recent development was sparked by the papers (Weinan, 2017; Haber & Ruthotto, 2017; Lu et al., 2018) establishing a connection between certain classes of neural networks, such as residual networks (He et al., 2016), and control systems. However, the use of dynamical and control systems to describe and analyze neural networks goes back at least to the 70's. For example, Wilson-Cowan's equations (Wilson & Cowan, 1972) are differential equations and so is the model proposed by Hopfield in (Hopfield, 1984). These techniques have been used to study several problems such as weight identifiability from data (Albertini & Sontag, 1993; Albertini et al., 1993), controllability (Sontag & Qiao, 1999; Sontag & Sussmann, 1997), and stability (Michel et al., 1989; Hirsch, 1989).

The objective of this paper is to shed new light into the approximation power of deep neural networks and, in particular, of residual deep neural networks (He et al., 2016). It has been empirically observed that deep networks have better approximation capabilities than their shallow counterparts and are easier to train (Ba & Caruana, 2014; Urban et al., 2017). An intuitive explanation for this fact is based on the different ways in which these types of networks perform function approximation. While shallow networks prioritize parallel compositions of simple functions (the number of neurons per layer is a measure of parallelism), deep networks prioritize sequential compositions of simple functions (the number of layers is a measure sequentiality). It is therefore natural to seek insights using control theory where the problem of producing interesting behavior by manipulating a few inputs over time, i.e., by sequentially composing them, has been extensively studied. Even though control-theoretic techniques have been utilized in the literature to showcase the controllability properties of neural networks, to best of our knowledge, this paper is the first to use tools from geometric control theory to establish universal approximation properties with respect to the infinity norm.

1.1 CONTRIBUTIONS

In this paper we focus on residual networks (He et al., 2016). This being said, as explained in (Lu et al., 2018), similar techniques can be exploited to analyze other classes of networks. It is known that deep residual networks have the power of universal approximation. What is less understood is where this power comes from. We show in this paper that it stems from the activation functions in the sense that when using a sufficiently rich activation function, even networks with very simple architectures and weights taking only two values suffice for universal approximation. It is the power of sequential composition, analyzed in this paper via geometric control theory, that unpacks the richness of the activation function into universal approximability. Surprisingly, the level of richness required from an activation function also has a very simple characterization; it suffices for activation functions (or a suitable derivative) to satisfy a quadratic differential equation. Most activation functions in the literature either satisfy this condition or can be suitably approximated by functions satisfying it.

More specifically, given a finite ensemble of data points, we cast the problem of designing weights for training a deep residual network as the problem of driving the state of a finite ensemble of initial points with a single open-loop control input to the finite ensemble of target points produced by the function to be learned when evaluated at the initial points. In spite of the fact that we only have access to a single open-loop control input, we prove that the corresponding ensemble of control systems is controllable. This result can also be understood in terms of the memorization capacity of deep networks, almost any finite set of samples can be memorized, see (Yun et al., 2019; Vershynin, 2020) for some recent work on this problem. We then utilize this controllability property to obtain universal approximability results for continuous functions in a uniform sense, i.e., with respect to the supremum norm. This is achieved by using the notion of monotonicity that lets us conclude uniform approximability on compact sets from controllability of finite ensembles.

1.2 RELATED WORK

Several papers have studied and established that residual networks have the power of universal approximation. This was done in (Lin & Jegelka, 2018) by focusing on the particular case of residual networks with the ReLU activation function. It was shown that any such network with n states and one neuron per layer can approximate an arbitrary Lebesgue integrable function $f : \mathbb{R}^n \to \mathbb{R}$ with respect to the L^1 norm. The paper (Zhang et al., 2019) shows that the functions described by deep networks with n states per layer, when these networks are modeled as control systems, are restricted to be homeomorphisms. The authors then show that increasing the number of states per layer to $2n$ suffices to approximate arbitrary homeomorphisms $f : \mathbb{R}^n \to \mathbb{R}^n$ under the assumption the underlying network already has the power of universal approximation. Note that the results in (Lin & Jegelka, 2018) do not model deep networks as control systems and, for this reason, bypass the homeomorphism restriction. There is also an important distinction to be made between requiring a network to exactly implement a function and to approximate it. The homeomorphism restriction does not prevent a network from approximating arbitrary functions; it just restricts the functions that can be implemented as a network. Closer to this paper are the results in (Li et al., 2019) establishing universal approximation, with respect to the L^p norm, $1 \leq p < \infty$, based on a general sufficient condition satisfied by several examples of activation functions. These results are a major step forward in identifying what is needed for universal approximability, as they are not tied to specific architectures or activation functions. In this paper we establish universal approximation in the stronger sense of the infinity norm L^∞ which implies, as a special case, universal approximation with respect to the L^p norm for $1 \leq p < \infty$.

At the technical level, our results build upon the controllability properties of deep residual networks. Earlier work on controllability of differential equation models for neural networks, e.g., (Sontag & Qiao, 1999), assumed the weights to be constant and that an exogenous control signal was fed into the neurons. In contrast, we regard the weights as control inputs and that no additional control inputs are present. These two different interpretations of the model lead to two very different technical problems. More recent work in the control community includes (Agrachev & Caponigro, 2009), where it is shown that any orientation preserving diffeomorphism on a compact manifold, can be obtained as the flow of a control system when using a time-varying feedback controller. In the context of this paper those results can be understood as: residual networks can represent any orientation preserving diffeomorphism provided that we can make the weights depend on the state.

Although quite insightful, such results are not applicable to the standard neural network models where the weights are not allowed to depend on the state. Another relevant topic is ensemble control. Most of the work on the control of ensembles, see for instance (Li & Khaneja, 2006; Helmke & Schönlein, 2014; Brockett, 2007), considers parametrized ensembles of vector fields. In other words, the individual systems that drive the state of the whole ensemble are different, whereas in our setting the ensemble consists of exact copies of the same system, albeit initialized differently. In this sense, our work is most closely related to the setting of (Agrachev & Sarychev, 2020a;b) where controllability results for ensembles of infinitely many control systems are provided. In this paper, in contrast, we use Lie algebraic techniques to study controllability of finite ensembles and obtain approximation results for infinite ensembles by using the notion of monotonicity rather than Lie algebraic techniques as is done in (Agrachev & Sarychev, 2020a;b). Moreover, by focusing on the specific control systems arising from deep residual networks we are able to provide easier to verify controllability conditions than those provided in (Agrachev & Sarychev, 2020a;b) for more general control systems. Controllability of finite ensembles of control systems motivated by neural network applications was investigated in (Cuchiero et al., 2019) where it is shown that controllability is a generic property and that, for control systems that are linear in the inputs, 5 inputs suffice. These results are insightful but they do not apply to specific control systems such as those describing residual networks and studied in this paper. Moreover the results in (Cuchiero et al., 2019) do not address the problem of universal approximation in the infinity norm.

To conclude the review of related work, we note that universal approximation with respect to the infinity norm for non-residual deep networks, allowing for general classes of activation functions, was recently established in (Kidger & Lyons, 2020). In particular, it is shown in (Kidger & Lyons, 2020) that under very mild conditions on the activation functions any continuous function $f : K \to \mathbb{R}^m$, where $K \subset \mathbb{R}^n$ is compact, can be approximated in the infinity norm using a deep neural network of width $n + m + 2$. These results do not directly carry over to residual networks. On the one hand, residual networks have skip connections that are not directly allowed by the formulation in (Kidger & Lyons, 2020). On the other hand, simulating the effect of skip connections with the feedforward networks used in (Kidger & Lyons, 2020) would lead to an increase in width rendering the bounds reported in (Kidger & Lyons, 2020) not applicable. Moreover, even if the bounds in (Kidger & Lyons, 2020) would apply to residual networks, the bounds proposed in this paper are tighter: when $n = m$ one of our main results, Corollary 4.5, asserts that a width of $2n$ is sufficient for universal approximation.

2 Control-theoretic view of residual networks

2.1 From residual networks to control systems and back

We start by providing a control system perspective on residual neural networks. We mostly follow the treatment proposed in (Weinan, 2017; Haber & Ruthotto, 2017; Lu et al., 2018), where it was suggested that residual neural networks with an update equation of the form:

$$x(k+1) = x(k) + S(k)\Sigma(W(k)x(k) + b(k)), \tag{2.1}$$

where $k \in \mathbb{N}_0$ indexes each layer, $x(k) \in \mathbb{R}^n$, and $(S(k), W(k), b(k)) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$, can be interpreted as a control system when k is viewed as indexing time. In (2.1), S, W, and b are the weights functions assigning weights to each time instant k, and $\Sigma : \mathbb{R}^n \to \mathbb{R}^n$ is of the form $\Sigma(x) = (\sigma(x_1), \sigma(x_2), \ldots, \sigma(x_n))$, where $\sigma : \mathbb{R}^n \to \mathbb{R}^n$ is an *activation function*. By drawing an analogy between (2.1) and Euler's forward method to discretize differential equations, one can interpret (2.1) as the time discretization of the continuous-time control system:

$$\dot{x}(t) = S(t)\Sigma(W(t)x(t) + b(t)), \tag{2.2}$$

where $x(t) \in \mathbb{R}^n$ and $(S(t), W(t), b(t)) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$; in what follows, and in order to make the presentation simpler, we sometimes drop the dependency on time. To make the connection between the discretization and (2.2) precise, let $x : [0, \tau] \to \mathbb{R}^n$ be a solution of the control system (2.2) for the control input $(S, W, b) : [0, \tau] \to \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$, where $\tau \in \mathbb{R}^+$. Then, given any desired accuracy $\varepsilon \in \mathbb{R}^+$ and any norm $|\cdot|$ in $\mathbb{R}^n$, there exists a sufficiently small time step $T \in \mathbb{R}^+$ so that the function $z : \{0, 1, \ldots, \lfloor \tau/T \rfloor\} \to \mathbb{R}^n$ defined by:

$$z(0) = x(0), \quad z(k+1) = z(k) + TS(kT)\Sigma(W(kT)z(k) + b(kT)),$$

approximates the sequence $\{x(kT)\}_{k=0,\dots,\lfloor\tau/T\rfloor}$ with error ε, i.e.:

$$|z(k) - x(kT)| \leq \varepsilon,$$

for all $k \in \{0, 1, \dots, \lfloor\tau/T\rfloor\}$. Intuitively, any statement about the solutions of (2.2) holds for the solutions of (2.1) with arbitrarily small error ε, provided that we can choose the depth to be arbitrarily large since by making T small we increase the depth, given by $1 + \lfloor\tau/T\rfloor$.

2.2 Neural network training and controllability

Given a function $f : \mathbb{R}^n \to \mathbb{R}^n$ and a finite set of samples $E_{\text{samples}} \subset \mathbb{R}^n$, the problem of training a residual network so that it maps $x \in E_{\text{samples}}$ to $f(x)$ can be phrased as the problem of constructing an open-loop control input $(S, W, b) : [0, \tau] \to \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$ so that the resulting solution of (2.2) takes the states $x \in E_{\text{samples}}$ to the states $f(x)$. It should then come as no surprise that the ability to approximate a function f is tightly connected with the control-theoretic problem of controllability: given, one initial state $x^{\text{init}} \in \mathbb{R}^n$ and one final state $x^{\text{fin}} \in \mathbb{R}^n$, when does there exist a finite time $\tau \in \mathbb{R}^+$ and a control input $(S, W, b) : [0, \tau] \to \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$ so that the solution of (2.2) starting at x^{init} at time 0 ends at x^{fin} at time τ?

To make the connection between controllability and the problem of mapping every $x \in E_{\text{samples}}$ to $f(x)$ clear, it is convenient to consider the ensemble of $d = |E_{\text{samples}}|$ copies of (2.2) given by the matrix differential equation:

$$\dot{X}(t) = [S(t)\Sigma(W(t)X_{\bullet 1}(t) + b(t))|S(t)\Sigma(W(t)X_{\bullet 2}(t) + b(t))|\dots|S(t)\Sigma(WX_{\bullet d}(t) + b(t)))], \tag{2.3}$$

where for time $t \in \mathbb{R}_0^+$ the ith column of the matrix $X(t) \in \mathbb{R}^{n\times d}$, denoted by $X_{\bullet i}(t)$, is the solution of the ith copy of (2.2) in the ensemble. If we now index the elements of E_{samples} as $\{x^1, \dots, x^d\}$, where d is the cardinality of E_{samples}, and consider the matrices $X^{\text{init}} = [x^1|x^2|\dots|x^n]$ and $X^{\text{fin}} = [f(x^1)|f(x^2)|\dots|f(x^n)]$, we see that the existence of a control input resulting in a solution of (2.3) starting at X^{init} and ending at X^{fin}, i.e., controllability of (2.3), is equivalent to existence of an input for (2.2) so that the resulting solution starting at $x^i \in E_{\text{samples}}$ ends at $f(x^i)$, for all $i \in \{1, \dots, d\}$.

Note that achieving controllability of (2.3) is especially difficult, since all the copies of (2.2) in (2.3) are *identical* and they all use the *same input*. Therefore, to achieve controllability, we must have sufficient diversity in the initial conditions to overcome the symmetries present in (2.3), see (Aguilar & Gharesifard, 2014). Our controllability result, Theorem 4.2, describes precisely such diversity. As mentioned in the introduction, this observation also distinguishes the problem under study here from the classical setting of ensemble control (Li & Khaneja, 2006; Helmke & Schönlein, 2014), with the exception of the recent work (Cuchiero et al., 2019; Agrachev & Sarychev, 2020a;b), where a collection of systems with *different* dynamics are driven by the same control input.

3 Problem formulation

Our starting point is the control system:

$$\dot{x}(t) = s(t)\Sigma(W(t)x(t) + b(t)), \tag{3.1}$$

a slightly simplified version of (2.2), where $x(t) \in \mathbb{R}^n$, $(s(t), W(t), b(t)) \in \mathbb{R} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$, and the input S in (2.2) is now the scalar-valued function s; as we will prove in what follows, this model is enough for universal approximation. In fact, we will later see[1] that it suffices to let s assume two arbitrary values only (one positive and one negative). Moreover, for certain activation functions, we can dispense with s altogether.

We make the following assumptions regarding the model (3.1):

- The function Σ is defined as $\Sigma : x \mapsto (\sigma(x_1), \sigma(x_2), \dots, \sigma(x_n))$, where the activation function $\sigma : \mathbb{R} \to \mathbb{R}$, or a suitable derivative of it, satisfies a quadratic differential equation, i.e., $D\xi = a_0 + a_1\xi + a_2\xi^2$ with $a_1, a_2, a_3 \in \mathbb{R}$, $a_2 \neq 0$, and $\xi = D^j\sigma$ for some $j \in \mathbb{N}_0$. Here, $D^j\sigma$ denotes the derivative of σ of order j and $D^0\sigma = \sigma$.
- The activation function $\sigma : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous, $D\sigma \geq 0$, and $\xi = D^j\sigma$ defined above is injective.

Table 1: Activation functions and the differential equations they satisfy.

Function name	**Definition**	**Satisfied differential equation**
Logistic function	$\sigma(x) = \frac{1}{1+e^{-x}}$	$D\sigma - \sigma + \sigma^2 = 0$
Hyperbolic tangent	$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$D\sigma - 1 + \sigma^2 = 0$
Soft plus	$\sigma(x) = \frac{1}{r}\log(1 + e^{rx})$	$D^2\sigma - rD\sigma + r(D\sigma)^2 = 0$

Several activation functions used in the literature are solutions of quadratic differential equations as can be seen in Table 1. Moreover, activation functions that are not differentiable can also be handled via approximation. For example, the ReLU function defined by $\max\{0, x\}$ can be approximated by $\sigma(x) = \log(1 + e^{rx})/r$, as $r \to \infty$, which satisfies the quadratic differential equation given in Table 1.

The Lipschitz continuity assumption is made to simplify the presentation and can be replaced with local Lipschitz continuity, which then does not need to be assumed, since σ is analytic in virtue of being the solution of an analytic (quadratic) differential equation. Moreover, all the activation functions in Table 1 are Lipschitz continuous, have positive derivative and are thus injective.

To formally state the problem under study in this paper, we need to discuss a different point of view on the solutions of the control system (3.1) given by *flows*. A continuously differentiable curve $x : [0, \tau] \to \mathbb{R}^n$ is said to be a solution of (3.1) under the piecewise continuous input $(s, W, b) : [0, \tau] \to \mathbb{R} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$ if it satisfies (3.1). Under the stated assumptions on σ, given a piecewise continuous input and a state $x^{\text{init}} \in \mathbb{R}^n$, there is one and at most one solution $x(t)$ of (3.1) satisfying $x(0) = x^{\text{init}}$. Moreover, solutions are defined for all $\tau \in \mathbb{R}_0^+$. We can thus define the flow of (3.1) under the input (s, W, b) as the map $\phi^\tau : \mathbb{R}^n \to \mathbb{R}^n$ given by the assignment $x^{\text{init}} \mapsto x(\tau)$. In other words, $\phi^\tau(x^{\text{init}})$ is the point reached at time τ by the unique solution starting at x^{init} at time 0. When the time τ is clear from context, we denote a flow simply by ϕ. It will also be convenient to denote the flow ϕ^τ by Z^τ when ϕ is defined by the solution of the differential equation $\dot{x} = Z(x)$ for some vector field $Z : \mathbb{R}^n \to \mathbb{R}^n$.

We will use flows to approximate arbitrary continuous functions $f : \mathbb{R}^n \to \mathbb{R}^m$. Since flows have the same domain and co-domain, and $f : \mathbb{R}^n \to \mathbb{R}^m$ may not, we first lift f to a map $\tilde{f} : \mathbb{R}^k \to \mathbb{R}^k$. When $n > m$, we lift f to $\tilde{f} = \imath \circ f : \mathbb{R}^n \to \mathbb{R}^n$, where $\imath : \mathbb{R}^m \to \mathbb{R}^n$ is the injection given by $\imath(x) = (x_1, \ldots, x_n, 0, \ldots, 0)$. In this case $k = n$. When $n < m$, we lift f to $\tilde{f} = f \circ \pi : \mathbb{R}^m \to \mathbb{R}^m$, where $\pi : \mathbb{R}^m \to \mathbb{R}^n$ is the projection $\pi(x_1, \ldots, x_n, x_{n+1}, \ldots, x_m) = (x_1, \ldots, x_n)$. In this case $k = m$. Although we could consider factoring f through a map $g : \mathbb{R}^n \to \mathbb{R}^m$, i.e., to construct $\tilde{f} : \mathbb{R}^n \to \mathbb{R}^n$ so that $f = g \circ \tilde{f}$ as done in, e.g., (Li et al., 2019), the construction of g requires a deep understanding of f, since a necessary condition for this factorization is $f(\mathbb{R}^n) \subseteq g(\mathbb{R}^n)$. Constructing g so as to contain $f(\mathbb{R}^n)$ on its image requires understanding what $f(\mathbb{R}^n)$ is and this information is not available in learning problems. Given this discussion, in the remainder of this paper we directly assume we seek to approximate a map $f : \mathbb{R}^n \to \mathbb{R}^n$.

The final ingredient we need before stating the problem solved in this paper is the precise notion of approximation. Throughout the paper, we will investigate approximation in the sense of the L^∞ (supremum) norm, i.e.:

$$\|f\|_{L^\infty(E)} = \sup_{x\in E} |f(x)|_\infty,$$

where $E \subset \mathbb{R}^n$ is the compact set over which is the approximation is going to be conducted and $|f(x)|_\infty = \max_{i\in\{1,\ldots,n\}} |f_i(x)|$. Some approximation results will be stated for networks modeled by a control system (3.1) with state space $\mathbb{R}^n$. In such cases the approximation quality is measured by $\|f - \phi\|_{L^\infty(E)}$ where ϕ is the flow of (3.1). Other results will require networks with state space $\mathbb{R}^{2n}$. For those cases the approximation quality is measured by $\|f - \beta \circ \phi \circ \alpha\|_{L^\infty(E)}$ where $\alpha : \mathbb{R}^n \to \mathbb{R}^{2n}$ is an injection and $\beta : \mathbb{R}^{2n} \to \mathbb{R}^n$ is a projection. These maps will be linear and can be implemented as the first and last layers of a residual network.

[1] See the discussion after the proof of Theorem 4.2

We are now ready to state the two problems we study in this paper.

Problem 3.1. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous function, $E_{\text{samples}} \subset \mathbb{R}^n$ be a finite set, and $\varepsilon \in \mathbb{R}_0^+$ be the desired approximation accuracy. Under what conditions on the activation function of control system* (3.1) *does there exist a time $\tau \in \mathbb{R}^+$ and an input $(s, W, b) : [0, \tau] \to \mathbb{R} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$ so that the flow $\phi^\tau : \mathbb{R}^n \to \mathbb{R}^n$ defined by the solution of* (3.1) *with state space $\mathbb{R}^n$ under the said input satisfies:*

$$\|f - \phi^\tau\|_{L^\infty(E_{\text{samples}})} \leq \varepsilon.$$

Note that we allow ε to be zero in which case the flow ϕ^τ matches f exactly on E_{samples}, i.e., $f(x) = \phi^\tau(x)$ for every $x \in E_{\text{samples}}$.

The next problem considers the more challenging case of approximation on compact sets and allows for residual networks with $2n$ neurons per layer when approximating functions on $\mathbb{R}^n$.

Problem 3.2. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous function, $E \subset \mathbb{R}^n$ be a compact set, and $\varepsilon \in \mathbb{R}^+$ be the desired approximation accuracy. Under what conditions on the activation function of control system* (3.1) *does there exist a time $\tau \in \mathbb{R}^+$, an injection $\alpha : \mathbb{R}^n \to \mathbb{R}^{2n}$, a projection $\beta : \mathbb{R}^{2n} \to \mathbb{R}^n$, and an input $(s, W, b) : [0, \tau] \to \mathbb{R} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$ so that the flow $\phi^\tau : \mathbb{R}^n \to \mathbb{R}^n$ defined by the solution of* (3.1) *with state space $\mathbb{R}^{2n}$ under the said input satisfies:*

$$\|f - \beta \circ \phi^\tau \circ \alpha\|_{L^\infty(E)} \leq \varepsilon.$$

In the next section, we will show the answer to these problems is remarkably simple. The first problem is solved under an assumption on the activation function: σ satisfies a quadratic differential equation. As we argued in the previous section, several activation functions satisfy this assumption exactly or approximately. The second problem is solved based on the additional assumption of monotonicity which is satisfied by construction when we allow the network to have $2n$ neurons per layer.

4 Main results

The proofs of all the results in this section are provided in the Appendix.

We first discuss the problem of constructing an input for (3.1) so that the resulting flow ϕ satisfies $\phi(x) = f(x)$ for all the points x in a given finite set $E_{\text{samples}} \subset \mathbb{R}^n$. We explained in Section 2.2 that this is equivalent to determining if the ensemble control system (2.3) is controllable. It is simple to see that controllability of (2.3) cannot hold on all of $\mathbb{R}^{n\times d}$, since if the initial state $X(0)$ satisfies $X_{\bullet i}(0) = X_{\bullet j}(0)$ for some $i \neq j$, we must have $X_{\bullet i}(t) = X_{\bullet j}(t)$ for all $t \in [0, \tau]$ by uniqueness of solutions of differential equations.

Our first result establishes that the controllability property holds for the ensemble control system (2.3) on a dense and connected submanifold of $\mathbb{R}^{n\times d}$, independently of the (finite) number of copies d, as long the activation function satisfies a quadratic differential equation. Before stating this result, we recall the formal definition of controllability.

Definition 4.1. *A point $X^{\text{fin}} \in \mathbb{R}^{n\times d}$ is said to be reachable from a point $X^{\text{init}} \in \mathbb{R}^{n\times d}$ for the control system* (2.3) *if there exist $\tau \in \mathbb{R}^+$ and a control input $(s, W, b) : [0, \tau] \to \mathbb{R} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$ so that the solution X of* (2.3) *under said input satisfies $X(0) = X^{\text{init}}$ and $X(\tau) = X^{\text{fin}}$. Control system* (2.3) *is said to be controllable on a submanifold M of $\mathbb{R}^{n\times d}$ if any point in M is reachable from any point in M.*

Theorem 4.2. *Let $N \subset \mathbb{R}^{n\times d}$ be the set defined by:*

$$N = \left\{ A \in \mathbb{R}^{n\times d} \mid \prod_{1\leq i<j\leq d} (A_{\ell i} - A_{\ell j}) = 0,\ \ell \in \{1, \ldots, n\} \right\}.$$

Suppose that σ is injective and satisfies the quadratic differential equation $D\sigma = a_0 + a_1\sigma + a_2\sigma^2$ with $a_2 \neq 0$. If $n > 1$, then the ensemble control system (2.3) *is controllable on the submanifold $M = \mathbb{R}^{n\times d} \backslash N$.*

It is worth mentioning that the assumption of $n \neq 1$ ensures connectedness of the submanifold M, which we rely on to obtain controllability. The following corollary of Theorem 4.2 weakens controllability to reachability but applies to a larger set.

Corollary 4.3. *Let $M \subset \mathbb{R}^{n\times d}$ be the submanifold defined in Theorem 4.2. Under assumptions of Theorem 4.2, any point in M is reachable from a point $A \in \mathbb{R}^{n\times d}$ for which:*

$$A_{\bullet i} \neq A_{\bullet j},$$

holds for all $i \neq j$, where $i, j \in \{1, \dots, d\}$.

The assumption $A_{\bullet i} \neq A_{\bullet j}$ in Corollary 4.3 requires all the columns of A to be different and is always satisfied when $A = \left[x^1|x^2|\dots|x^d\right]$, $x^i \in E_{\text{samples}}$. Hence, for any finite set E_{samples} there exists a flow ϕ of (3.1) satisfying $f(x) = \phi(x)$ for all $x \in E_{\text{samples}}$ provided that $f(E_{\text{samples}}) \subset M$, i.e., Problem 3.1 is solved with $\varepsilon = 0$. Moreover, since M is dense in $\mathbb{R}^{n\times d}$, when $f(E_{\text{samples}}) \subset M$ fails, there still exists a flow ϕ of (3.1) taking $\phi(x)$ arbitrarily close to $f(x)$ for all $x \in E_{\text{samples}}$, i.e., Problem 3.1 is solved for any $\varepsilon > 0$. This result also sheds light on the memorization capacity of residual networks as it states that almost any finite set of samples can be memorized, independently of its cardinality. See, e.g., Yun et al. (2019); Vershynin (2020), for recent results on this problem that do not rely on differential equation models.

Some further remarks are in order. The assumptions above on σ can be relaxed; in particular, it is enough for $D^j\sigma$ to be injective and to satisfy the mentioned quadratic differential equation for some $j \in \mathbb{N}_0$. Moreover, Theorem 4.2 and Corollary 4.3 do not directly apply to the ReLU activation function, defined by $\max\{0, x\}$, since this function is not differentiable. However, the ReLU is approximated by the activation function:

$$\frac{1}{r}\log(1 + e^{rx}),$$

as $r \to \infty$. In particular, as $r \to \infty$ the ensemble control system (2.3) with $\sigma(x) = \log(1 + e^{rx})/r$ converges to the ensemble control system (2.3) with $\sigma(x) = \max\{0, x\}$ and thus the solutions of the latter are arbitrarily close to the solutions of the former whenever r is large enough. Moreover, $\xi = D\sigma$ satisfies $D\xi = r\xi - r\xi^2$ and $D\xi = re^{rx}/(1+e^{rx})^2 > 0$ for $x \in \mathbb{R}$ and $r > 0$ thus showing that ξ is an increasing function and, consequently, injective.

The conclusions of Theorem 4.2 and Corollary 4.3 also hold if we weaken the assumptions on the inputs of (3.1). It suffices for the entries of W and b to take values on a set with two elements, see the discussion after the proof of Theorem 4.2 for details. Moreover, when the activation function is and odd function, i.e., $\sigma(-x) = -\sigma(x)$, as is the case for the hyperbolic tangent, the conclusions of Theorem 4.2 hold for the simpler version of (3.1), where we fix s to be 1.

In order to extend the approximation guarantees from a finite set $E_{\text{samples}} \subset \mathbb{R}^n$ to an arbitrary compact set $E \subset \mathbb{R}^n$, we rely on the notion of monotonicity. On $\mathbb{R}^n$ we consider the ordering relation $x \preceq x'$ defined by $x_i \leq x'_i$ for all $i \in \{1, \dots, n\}$ and $x, x' \in \mathbb{R}^n$. A map $f : \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotone when it respects this ordering relation, i.e., when $x \preceq x'$ implies $f(x) \preceq f(x')$. A vector field $Z : \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotone when its flow $\phi^\tau : \mathbb{R}^n \to \mathbb{R}^n$ is a monotone map. Monotone vector fields admit a simple characterization (Smith, 2008):

$$\frac{\partial Z_i}{\partial x_j} \geq 0, \quad \forall i, j \in \{1, \dots, n\}, i \neq j. \tag{4.1}$$

When the function $f : \mathbb{R}^n \to \mathbb{R}^n$ to be approximated is representable as the flow ϕ of an analytic monotone vector field $Z : \mathbb{R}^n \to \mathbb{R}^n$, i.e., $f = Z^1$, f can be uniformly approximated to any desired accuracy by a residual network with n neurons per layer.

Theorem 4.4. *Let $n > 1$, assume $D\sigma \geq 0$ and the existence of $k \in \mathbb{N}_0$ so that $\xi = D^k\sigma$ is injective and satisfies a quadratic differential equation $D\xi = a_0 + a_1\xi + a_2\xi^2$ with $a_2 \neq 0$. Then, for every continuous function $f : \mathbb{R}^n \to \mathbb{R}^n$ satisfying $f = Z^1$ for an analytic monotone vector field $Z : \mathbb{R}^n \to \mathbb{R}^n$, for every compact set $E \subset \mathbb{R}^n$, and for every $\varepsilon \in \mathbb{R}^+$ there exist a time $\tau \in \mathbb{R}^+$ and an input $(s, W, b) : [0, \tau] \to \mathbb{R} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$ so that the flow $\phi^\tau : \mathbb{R}^n \to \mathbb{R}^n$ defined by the solution of* (3.1) *with state space $\mathbb{R}^n$ under the said input satisfies:*

$$\|f - \phi^\tau\|_{L^\infty(E)} \leq \varepsilon. \tag{4.2}$$

Not every function can be represented as the flow of a vector field, much less an analytic monotone one (Fort, 1955; Utz, 1981). Yet, the following corollary is based on a simple construction that

embeds a continuous function $f : \mathbb{R}^n \to \mathbb{R}^n$ into the flow of a monotone vector field on $\mathbb{R}^{2n}$; a similar approach is used in (Zhang et al., 2020). As a direct consequence, any continuous function $f : \mathbb{R}^n \to \mathbb{R}^n$ can be uniformly approximated to any desired accuracy by a residual network with $2n$ neurons per layer.

Corollary 4.5. *Let $n > 1$, assume $D\sigma \geq 0$ and the existence of $k \in \mathbb{N}_0$ so that $\xi = D^k\sigma$ is injective and satisfies a quadratic differential equation $D\xi = a_0 + a_1\xi + a_2\xi^2$ with $a_2 \neq 0$. Then, for every continuous function $f : \mathbb{R}^n \to \mathbb{R}^n$, for every compact set $E \subset \mathbb{R}^n$, and for every $\varepsilon \in \mathbb{R}^+$ there exist a time $\tau \in \mathbb{R}^+$, an injection $\alpha : \mathbb{R}^n \to \mathbb{R}^{2n}$, a projection $\beta : \mathbb{R}^{2n} \to \mathbb{R}^n$, and an input $(s, W, b) : [0, \tau] \to \mathbb{R} \times \mathbb{R}^{2n\times 2n} \times \mathbb{R}^{2n}$ so that the flow $\phi^\tau : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ defined by the solution of* (3.1) *with state space $\mathbb{R}^{2n}$ under the said input satisfies:*

$$\|f - \beta \circ \phi^\tau \circ \alpha\|_{L^\infty(E)} \leq \varepsilon.$$

It is worth pointing out that, contrary to Theorem 4.4, no requirements are placed on f in addition to continuity. In (Agrachev & Sarychev, 2020a;b), sufficient conditions for the existence of a flow ϕ^τ satisfying (4.2) are given for a more general class of control systems. The assumptions used in Theorem (4.4) are not easy to compare with the assumptions in Theorem 5.1 of (Agrachev & Sarychev, 2020b). Checking the existence of an analytic monotone vector field Z satisfying $Z^1 = f$ is a non-trivial task. However, by employing a deep network of width $2n$, i.e., by using Corollary 4.5, we can rely on much simpler assumptions on the activation functions which are satisfied by the networks used in practice. In contrast, (Agrachev & Sarychev, 2020b, Theorem 5.1) requires a strong Lie algebra approximation property to be satisfied by the ensemble control system that does not appear to be easy to verify.

REFERENCES

A. Agrachev and A. Sarychev. Control in the spaces of ensembles of points. *SIAM Journal on Control and Optimization*, 58(3):1579–1596, 2020a.

A. Agrachev and A. Sarychev. Control on the manifolds of mappings as a setting for deep learning. *arXiv preprint arXiv:2008.12702*, 2020b.

A.A. Agrachev and M. Caponigro. Controllability on the group of diffeomorphisms. *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*, 26(6):2503 – 2509, 2009.

C. Aguilar and B. Gharesifard. Necessary conditions for controllability of nonlinear networked control systems. In *American Control Conference*, pp. 5379–5383, Portland, OR, 2014.

D. Albertini, E. D. Sontag, and V. Maillot. Uniqueness of weights for neural networks. *Artificial Neural Networks for Speech and Vision*, pp. 115–125, 1993.

F. Albertini and E. D. Sontag. For neural networks, function determines form. *Neural Networks*, 6 (7):975 – 990, 1993.

L. J. Ba and R. Caruana. Do deep nets really need to be deep? In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pp. 2654–2662, Cambridge, MA, USA, 2014. MIT Press.

R. W. Brockett. Optimal control of the Liouville equation. *AMS IP Studies in Advanced Mathematics*, 39:23, 2007.

C. Cuchiero, M. Larsson, and J. Teichmann. Deep neural networks, generic universal interpolation, and controlled ODEs. *https://arxiv.org/abs/1908.07838*, 2019.

M. K. Fort. The embedding of homeomorphisms in flows. *Proceedings of the American Mathematical Society*, 6(6):960–967, 1955.

E. Haber and L. Ruthotto. Stable architectures for deep neural networks. *Inverse Problems*, 34(1), 2017.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

U. Helmke and M. Schönlein. Uniform ensemble controllability for one-parameter families of time-invariant linear systems. *Systems & Control Letters*, 71:69–77, 2014.

Morris W. Hirsch. Convergent activation dynamics in continuous time networks. *Neural Networks*, 2(5):331 – 349, 1989.

J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10):3088–3092, 1984.

V. Jurdjevic. *Geometric Control Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1996.

P. Kidger and T. Lyons. Universal approximation with deep narrow networks. *arXiv preprint arXiv:1905.08539*, 2020.

C. Krattenthaler. Advanced determinant calculus. In *The Andrews Festschrift*, pp. 349–426. Springer, 2001.

John M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2013.

J-S. Li and N. Khaneja. Control of inhomogeneous quantum ensembles. *Physical Review A*, 73(3): 030302, 2006.

Q. Li, T. Lin, and Z. Shen. Deep learning via dynamical systems: An approximation perspective. *arXiv preprint arXiv:1912.10382*, 2019.

H. Lin and S. Jegelka. ResNet with one-neuron hidden layers is a universal approximator. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 6172–6181, Red Hook, NY, USA, 2018. Curran Associates Inc.

Y. Lu, A. Zhong, Q. Li, and B. Dong. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. In *International Conference on Machine Learning*, pp. 3276–3285, 2018.

A. N. Michel, J. A. Farrell, and W. Porod. Qualitative analysis of neural networks. *IEEE Transactions on Circuits and Systems*, 36(2):229–243, 1989.

H. L. Smith. *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems*. Mathematical Surveys and Monographs. American Mathematical Society, 2008.

E. D. Sontag. *Mathematical Control Theory. Deterministic Finite-Dimensional Systems*, volume 6 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 1998.

E. D. Sontag and Y. Qiao. Further results on controllability of recurrent neural networks. *Systems & Control Letters*, 36(2):121 – 129, 1999.

E. D. Sontag and H. Sussmann. Complete controllability of continuous-time recurrent neural networks. *Systems & Control Letters*, 30(4):177–183, 1997.

G. Urban, K. J. Geras, S. Ebrahimi Kahou, Ö. Aslan, S. Wang, A. Mohamed, M. Philipose, M. Richardson, and R. Caruana. Do deep convolutional nets really need to be deep and convolutional? In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

W. R. Utz. The embedding of homeomorphisms in continuous flows. In *Topology Proc.*, volume 6 (1), pp. 159–177, 1981.

Roman Vershynin. Memory capacity of neural networks with threshold and ReLU activations. *arXiv preprint arXiv:2001.06938*, 2020.

E. Weinan. A proposal on machine learning via dynamical systems. *Communications in Mathematics and Statistics*, 5, 2017.

H.R. Wilson and J. D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12(1):1–24, 1972.

Chulhee Yun, Suvrit Sra, and Ali Jadbabaie. Small ReLU networks are powerful memorizers: a tight analysis of memorization capacity. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pp. 15532–15543, 2019.

H. Zhang, X. Gao, J. Unterman, and T. Arodz. Approximation capabilities of neural ODEs and invertible residual networks. *arXiv preprint arXiv:1907.12998*, 2019.

H. Zhang, X. Gao, J. Unterman, and T. Arodz. Approximation capabilities of neural ODEs and invertible residual networks. *arXiv preprint arXiv:1907.12998*, 2020.

A APPENDIX

The proof of Theorem 4.2 is based on two technical results. The first characterizes the rank of a certain matrix that will be required for our controllability result. In essence, the proof of this result follows from (Krattenthaler, 2001, Proposition 1), however, we provide a proof for completeness.

Lemma A.1. *Let $\xi : \mathbb{R} \to \mathbb{R}$ be a function that satisfies the quadratic differential equation:*

$$D\xi(x) = a_0 + a_1\xi(x) + a_2\xi^2(x),$$

where $a_0, a_1, a_2 \in \mathbb{R}$. Suppose that derivatives of ξ of up to order $(\ell - 2)$ exist at ℓ points $x_1, \ldots, x_\ell \in \mathbb{R}$. Then, the determinant of the matrix:

$$L(x_1, x_2, \ldots, x_\ell) = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \xi(x_1) & \xi(x_2) & \ldots & \xi(x_\ell) \\ D\xi(x_1) & D\xi(x_2) & \ldots & D\xi(x_\ell) \\ \vdots & \vdots & \ddots & \vdots \\ D^{\ell-2}\xi(x_1) & D^{\ell-2}\xi(x_2) & \ldots & D^{\ell-2}\xi(x_\ell) \end{bmatrix}, \tag{A.1}$$

is given by:

$$\det L(x_1, x_2, \ldots, x_\ell) = \prod_{i=1}^{\ell-2} i! a_2^i \prod_{1 \le i < j \le \ell} (\xi(x_i) - \xi(x_j)). \tag{A.2}$$

Proof. We assume that the elements of the set $\{x_1, x_2, \ldots, x_\ell\}$ are distinct, as otherwise, the determinant is clearly zero. We also assume that $\ell \geq 3$ to exclude the trivial case. First, by the Vandermonde determinant formula, we have that:

$$V_0(x_1, x_2, \ldots, x_\ell) := \begin{vmatrix} 1 & 1 & \ldots & 1 \\ \xi(x_1) & \xi(x_2) & \ldots & \xi(x_\ell) \\ \xi^2(x_1) & \xi^2(x_2) & \ldots & \xi^2(x_\ell) \\ \vdots & \vdots & \ddots & \vdots \\ \xi^{\ell-1}(x_1) & \xi^{\ell-1}(x_2) & \ldots & \xi^{\ell-1}(x_\ell) \end{vmatrix} = \prod_{1 \le i < j \le \ell} (\xi(x_i) - \xi(x_j)). \tag{A.3}$$

Our proof technique is to use elementary row operations to construct the determinant of $L(x_1, x_2, \ldots, x_\ell)$ from (A.3). To illustrate the idea, let us use (A.3) to show that:

$$V_1(x_1, x_2, \ldots, x_\ell) := \begin{vmatrix} 1 & 1 & \ldots & 1 \\ \xi(x_1) & \xi(x_2) & \ldots & \xi(x_\ell) \\ D\xi(x_1) & D\xi(x_2) & \ldots & D\xi(x_\ell) \\ \xi^3(x_1) & \xi^3(x_2) & \ldots & \xi^3(x_\ell) \\ \vdots & \vdots & \ddots & \vdots \\ \xi^{\ell-1}(x_1) & \xi^{\ell-1}(x_2) & \ldots & \xi^{\ell-1}(x_\ell) \end{vmatrix} = a_2 \prod_{1 \le i < j \le \ell} (\xi(x_i) - \xi(x_j)).$$

For later use, we denote by $V_i(x_1, x_2, \ldots, x_\ell)$ the determinant of the matrix constructed by substituting rows 3 to i in $V_0(x_1, x_2, \ldots, x_\ell)$ by derivatives of order 1 to $i-2$, respectively. First, note that multiplying the third row of $L(x_1, x_2, \ldots, x_\ell)$ by a_2 leads to:

$$\begin{vmatrix} 1 & 1 & \ldots & 1 \\ \xi(x_1) & \xi(x_2) & \ldots & \xi(x_\ell) \\ a_2\xi^2(x_1) & a_2\xi^2(x_2) & \ldots & a_2\xi^2(x_\ell) \\ \xi^3(x_1) & \xi^3(x_2) & \ldots & \xi^3(x_\ell) \\ \vdots & \vdots & \ddots & \vdots \\ \xi^{\ell-1}(x_1) & \xi^{\ell-2}(x_1) & \ldots & \xi^{\ell-1}(x_\ell) \end{vmatrix} = a_2 V_0(x_1, x_2, \ldots, x_\ell).$$

Moreover, by the fact that the determinant is unchanged by adding a constant multiple of a row to another row, using rows one and two for this purpose, we have that:

$$\begin{vmatrix} 1 & 1 & \ldots & 1 \\ \xi(x_1) & \xi(x_2) & \ldots & \xi(x_\ell) \\ a_0 + a_1\xi(x_1) + a_2\xi^2(x_1) & a_0 + a_1\xi(x_2) + a_2\xi^2(x_2) & \ldots & a_0 + a_1\xi(x_\ell) + a_2\xi^2(x_\ell) \\ \xi^3(x_1) & \xi^3(x_2) & \ldots & \xi^3(x_\ell) \\ \vdots & \vdots & \ddots & \vdots \\ \xi^{\ell-1}(x_1) & \xi^{\ell-1}(x_2) & \ldots & \xi^{\ell-1}(x_\ell) \end{vmatrix},$$

equals $a_2 V_0(x_1, x_2, \ldots, x_\ell)$ and yelds:

$$V_1(x_1, x_2, \ldots, x_\ell) = a_2 V_0(x_1, x_2, \ldots, x_\ell),$$

proving the claim. The idea of the proof is to use this same procedure, row by row, to construct $D^i\xi(x_j)$ in the entry $(i+2) \times j$ of the matrix. In order to proceed, however, we need to find a formula for $D^i\xi(x)$, where $x \in \mathbb{R}$. Note that, for $i \geq 2$, we have that:

$$\begin{aligned} D^i\xi(x) &= a_1 D^{i-1}\xi(x) + 2a_2 \frac{d}{dx^{i-2}}(\xi(x)D\xi(x)) \\ &= a_1 D^{i-1}\xi(x) + 2a_2 \sum_{k=0}^{i-2} \binom{i-1}{k} D^{i-k-2}\xi(x) D^{k+1}\xi(x), \end{aligned}$$

and $D^i\xi(x)$, as a polynomial in $\xi(x)$, is of degree $(i+1)$. We now make an observation that finishes the proof. In particular, in the computation of $V_1(x_1, x_2, \ldots, x_\ell)$ and in order to construct $D\xi(x)$ in the third row, we only needed to know the coefficient of the highest degree monomial, in terms of $\xi(x)$, that constitutes $D\xi(x)$. In other words, the lower degree terms do not contribute to the determinant, as they can be constructed, without changing the determinant, from previous rows. Using this observation, the term $a_1 D^{i-1}\xi(x)$ in the expansion of $D^i\xi(x)$ does not contributed to $V_i(x_1, x_2, \ldots, x_\ell)$, as it can be added from the previously constructed rows. Using this reasoning for all i, we conclude that the determinant of $L(x_1, \ldots, x_\ell)$ is independent of a_1, and a_0. Substituting $a_0 = 0$ and $a_1 = 0$, since $D^i\xi(x) = i! a_2^i \xi^{i+1}$, we have that:

$$\begin{aligned} &\det L(x_1, \ldots, x_\ell) \\ &= \begin{vmatrix} 1 & 1 & \ldots & 1 \\ \xi(x_1) & \xi(x_2) & \ldots & \xi(x_\ell) \\ a_2\xi^2(x_1) & a_2\xi^2(x_1) & \ldots & a_2\xi^2(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ (\ell-2)! a_2^{(\ell-2)} \xi^{\ell-1}(x_1) & (\ell-2)! a_2^{(\ell-2)} \xi^{\ell-1}(x_2) & \ldots & (\ell-2)! a_2^{(\ell-2)} \xi^{\ell-1}(x_\ell) \end{vmatrix} \\ &= \prod_{i=1}^{\ell-2} i! a_2^i V_0(x_1, x_2, \ldots, x_\ell) \\ &= \prod_{i=1}^{\ell-2} i! a_2^i \prod_{1 \leq i < j \leq \ell} (\xi(x_i) - \xi(x_j)), \end{aligned}$$

as claimed. □

Our second technical result is stated next, for which we provide an elementary proof to keep the manuscript self-contained.

Proposition A.2. *Let $N \subset \mathbb{R}^{n\times d}$ be the set defined by:*

$$N = \left\{ A \in \mathbb{R}^{n\times d} \mid \prod_{1\leq i<j\leq d} (A_{\ell i} - A_{\ell j}) = 0,\ \ell \in \{1,\ldots,n\} \right\}.$$

The set $M = \mathbb{R}^{n\times d}\backslash N$ is an open and dense submanifold of $\mathbb{R}^{n\times d}$ which is connected when $n > 1$.

Proof. Note that N is a finite union of vector subspaces of $\mathbb{R}^{n\times d}$, hence topologically closed. Therefore, $\mathbb{R}^{n\times d}\backslash N$ is an open and dense subset of $\mathbb{R}^{n\times d}$ and thus a submanifold of dimension nd. It remains to show that M is connected.

Let $A^{\text{init}}, A^{\text{fin}} \in M$, and assume that $n > 1$. We prove that there exists a continuous curve $\gamma : [0,n] \to M$ connecting A^{init} to A^{fin}, i.e., $\gamma(0) = A^{\text{init}}$ and $\gamma(n) = A^{\text{fin}}$. Since $A^{\text{init}} \in M$ there exists $\ell^{\text{init}} \in \{1,\ldots,n\}$ so that $\prod_{1\leq i<j\leq d}(A_{\ell^{\text{init}}i} - A_{\ell^{\text{init}}j}) \neq 0$. Similarly, since $A^{\text{fin}} \in M$ there exists $\ell^{\text{fin}} \in \{1,\ldots,n\}$ so that $\prod_{1\leq i<j\leq d}(A_{\ell^{\text{fin}}i} - A_{\ell^{\text{fin}}j}) \neq 0$. We first consider the case where $\ell^{\text{init}} \neq \ell^{\text{fin}}$ (which is possible since $n > 1$). Without loss of generality assume that $\ell^{\text{init}} = n$ and $\ell^{\text{fin}} = 1$ and let $\gamma^k : \mathbb{R}^{n\times d} \times [k-1,k] \to \mathbb{R}^{n\times d}$ be defined as:

$$\gamma^k_\lambda(A) = \gamma^k(A,\lambda) = \begin{bmatrix} A_{1\bullet} \\ \vdots \\ A_{k-1\bullet} \\ A_{k\bullet} + (\lambda - (k-1))A^{\text{fin}}_{k\bullet} \\ A_{k+1\bullet} \\ \vdots \\ A_{n\bullet} \end{bmatrix}, \quad k \in \{1,\ldots,n\},$$

where $A_{k\bullet}$ denotes the kth row of A. We now define the curve $\gamma : [0,n] \to \mathbb{R}^{n\times d}$ by:

$$\gamma(\lambda) = \gamma^k_\lambda \circ \gamma^{k-1}_{k-1} \circ \ldots \circ \gamma^2_2 \circ \gamma^1_1(A^{\text{init}}), \quad \lambda \in [k-1,k],$$

and note that $\gamma(\lambda) \in M$ for all $\lambda \in [0,n]$. This is because, by definition, there exists at least one index $\ell \in \{1,\ldots,n\}$ such that $\prod_{1\leq i<j\leq d}(\gamma_{\ell i}(\lambda) - \gamma_{\ell j}(\lambda)) \neq 0$. When $\lambda \leq n-1$, we can choose ℓ to be ℓ^{init} because $\gamma_{\ell^{\text{init}}\bullet}(\lambda) = \gamma_{n\bullet}(\lambda) = A^{\text{init}}_{n\bullet}$. When $\lambda \geq n-1$, we can choose ℓ to be ℓ^{fin} because $\gamma_{\ell^{\text{fin}}\bullet}(\lambda) = \gamma_{1\bullet}(\lambda) = A^{\text{fin}}_{1\bullet}$. Since γ is the composition of continuous functions, it is continuous. Moreover, by construction, $\gamma(0) = A^{\text{init}}$ and $\gamma(n) = A^{\text{fin}}$.

We now consider the case where $\ell^{\text{init}} = \ell^{\text{fin}}$. Since $n > 1$, we can choose $A \in M$ so that $\prod_{1\leq i<j\leq d}(A_{\ell i} - A_{\ell j}) \neq 0$ with $\ell \neq \ell^{\text{fin}}$ and $\ell \neq \ell^{\text{init}}$. By the previous argument, there is a continuous curve connecting A^{init} to A without leaving M and there is also a continuous curve connecting A to A^{fin} without leaving M. Therefore, their concatenation produces the desired continuous curve γ connecting A^{init} to A^{fin} and the proof is finished. □

The proof of Theorem 4.2 uses several key ideas from geometric control that we now review. A collection of vector fields $\mathcal{F} = \{Z_1,\ldots,Z_k\}$ on a manifold M is said to be controllable if given $x^{\text{init}}, x^{\text{fin}} \in M$, there exists a finite sequence of times $0 < t_1 < t_1 + t_2 < \ldots < t_1 + \ldots + t_q$ so that:

$$Z^{t_q}_\ell \circ \ldots \circ Z^{t_2}_2 \circ Z^{t_1}_1(x^{\text{init}}) = x^{\text{fin}},$$

where $Z_i \in \mathcal{F}$ and Z^t_i is the flow of Z_i. When the vector fields Z_i are smooth, M is smooth and connected, and the collection $\mathcal{F}$ satisfies:

$$Z \in \mathcal{F} \implies \alpha Z \in \mathcal{F} \text{ for some } \alpha < 0,$$

then $\mathcal{F}$ is controllable provided the evaluation of the Lie algebra generated by $\mathcal{F}$ at every point $x \in M$ has the same dimension as M, see, e.g., (Jurdjevic, 1996)[2]. Recall that the Lie algebra generated

[2]In this footnote we provide additional details relating controllability of a family of vector fields to the Lie algebra rank condition. Let us denote by $\mathcal{A}_\mathcal{F}(x)$ the reachable set of the family of smooth vector fields $\mathcal{F}$ from

by $\mathcal{F}$, and denoted by $Lie(\mathcal{F})$, is the smallest vector space of vector fields on M containing $\mathcal{F}$ and closed under the Lie bracket. By evaluation of $Lie(\mathcal{F})$ at $x \in M$, we mean the finite-dimensional vector subspace of the tangent space of M at x that is obtained by evaluating every vector field in $Lie(\mathcal{F})$ at x. The proof consists in establishing controllability by determining the points at which $Lie(\mathcal{F})$ has the right dimension for a collection of vector fields $\mathcal{F}$ induced by the ensemble control system (2.3).

We are now in position to prove Theorem 4.2.

Proof of Theorem 4.2. Consider the control system given in (2.3). We prove that under the mentioned assumptions, there is a choice of the control inputs (s, W, b) that renders (2.3) controllable in M.

It will be sufficient to work with inputs that are piecewise constant, and we can further simplify the analysis by choosing the family of inputs (s, W, b) given by (A.4) and (A.5), where:

- the first class of inputs is given by:

$$(\pm 1, 0, ce_j), \tag{A.4}$$

  where $j \in \{1, 2, \ldots, n\}$ and $c \in \mathbb{R}$ is any value such that $\sigma(c) \neq 0$ and $e_j \in \mathbb{R}^n$ has zeros in all its entries except for a 1 on its jth entry;

- the second class of inputs is given by:

$$(\pm 1, E_{jk}, 0), \tag{A.5}$$

  where $j, k \in \{1, 2, \ldots, n\}$ and E_{ij} is the $n \times n$ matrix that has zeros in all its entries except for a 1 in its jth row and kth column.

Once we substitute these inputs into the right hand side of the ensemble control system (2.3), we obtain a family of vector fields on $\mathbb{R}^{n \times d}$. More specifically, the vector fields arising from the inputs (A.4), denoted by $\{X_j^{\pm}\}_{j \in \{1,\ldots,n\}}$, are given by:

$$X_j^+ = \sigma(c) \sum_{i=1}^{d} \frac{\partial}{\partial A_{ji}} \quad \text{and} \quad X_j^- = -X_j^+. \tag{A.6}$$

Similarly, the vector fields arising from the inputs (A.5), denoted by $\{Y_{j,k}^{\pm}\}_{j,k \in \{1,\ldots,n\}}$, are given by:

$$Y_{jk}^+ = \sum_{i=1}^{d} \sigma\left(A_{ki}\right) \frac{\partial}{\partial A_{ji}} \quad \text{and} \quad Y_{jk}^- = -Y_{jk}^+. \tag{A.7}$$

$x \in M$, i.e., the set of all points $x^{\text{fin}} \in \mathbb{R}^n$ of the form:

$$x^{\text{fin}} = Z_\ell^{t_q} \circ \ldots \circ Z_2^{t_2} \circ Z_1^{t_1}(x),$$

for $Z_i \in \mathcal{F}$ and $0 < t_1 < t_1 + t_2 < \ldots < t_1 + \ldots + t_q$,, and denote by $Lie_x(\mathcal{F})$ the evaluation of the Lie algebra generated by $\mathcal{F}$ at $x \in M$. By $\mathcal{F}'$ we denote the family of vector fields of the form $\sum_i \lambda_i X_i$ with $X_i \in \mathcal{F}$ and $\lambda_i \geq 0$. Since $\mathcal{F} \subseteq \mathcal{F}'$ we have $\mathcal{A}_{\mathcal{F}}(x) \subseteq \mathcal{A}_{\mathcal{F}'}(x)$. By Theorem 8 in Chapter 3 of (Jurdjevic, 1996) we have that:

$$\mathcal{A}_{\mathcal{F}}(x) \subseteq \mathcal{A}_{\mathcal{F}'}(x) \subseteq \text{cl}(\mathcal{A}_{\mathcal{F}}(x)),$$

where cl denotes topological closure. Moreover, by Theorem 2 in Chapter 3 of (Jurdjevic, 1996), if $Lie_x(\mathcal{F}) = T_x M$ for every $x \in M$, then $\text{int}(\text{cl}(\mathcal{A}_{\mathcal{F}}(x))) = \text{int}(\mathcal{A}_{\mathcal{F}}(x))$. We thus obtain:

$$\text{int}(\mathcal{A}_{\mathcal{F}}(x)) \subseteq \text{int}(\mathcal{A}_{\mathcal{F}'}(x)) \subseteq \text{int}(\text{cl}(\mathcal{A}_{\mathcal{F}}(x))) = \text{int}(\mathcal{A}_{\mathcal{F}}(x)).$$

But if $\mathcal{F}'$ is controllable, $\text{int}(\mathcal{A}'_{\mathcal{F}}(x)) = M$ and thus $\mathcal{F}$ is also controllable. Therefore, we now focus on determining if $\mathcal{F}'$ is controllable. Provided that for each $X \in \mathcal{F}$ there exists $X' \in \mathcal{F}$ satisfying $X = \sigma X'$ with $\sigma < 0$ (this is weaker than symmetry, symmetry is this property for $\sigma = -1$), $\mathcal{F}'$ is simply the vector space spanned by $\mathcal{F}$. Moreover, since the control system $\dot{x} = \sum_i X_i u_i$ with $X_i \in \mathcal{F}$ and $u_i \in \mathbb{R}$ generates the same family of vector fields as $\mathcal{F}'$, we conclude that we can instead study the reachable set of $\dot{x} = \sum_i X_i u_i$ with $X_i \in \mathcal{F}$ which is driftless. By Theorem 2 in Chapter 4, in (Jurdjevic, 1996) the control system $\dot{x} = \sum_i X_i u_i$ is controllable provided that $Lie_x(\mathcal{F}') = T_x M$ for every $x \in M$.

This definition abuses notation, since defining a vector field on $\mathbb{R}^{n\times d}$ requires one summation over i and one over j. However, summation over j, i.e., summation over rows, only produces non-zero terms for one row, that we decided to index by j.

We make the observation that, since $\sigma(c) \neq 0$, we can simplify the vector fields $X_j^{\pm}$ to:

$$X_j^+ = \sum_{i=1}^{d} \frac{\partial}{\partial A_{ji}} \quad \text{and} \quad X_j^- = -X_j^+,$$

without altering controllability. This follows from the observation that for any vector field X with flow X^t we have $X^{\alpha\tau} = (\alpha X)^\tau$ for any $\alpha \in \mathbb{R}$.

By Proposition A.2, M is a connected smooth submanifold of $\mathbb{R}^{n\times d}$. The remainder of the proof consists of showing that the family of vector fields $\mathcal{F} = \{X_j^{\pm}, Y_{jk}^{\pm}\}_{j,k\in\{1,\dots,n\}}$, restricted to M, is controllable on M. As discussed prior to this proof, since these vector fields in $\mathcal{F}$ are smooth and satisfy $Z \in \mathcal{F} \implies -Z \in \mathcal{F}$, it suffices to establish that $\dim(Lie_A(\mathcal{F})) = \dim(M) = nd$ for every $A \in M$ and where $Lie_A(\mathcal{F})$ denotes the evaluation at A of the Lie algebra generated by $\mathcal{F}$.

We generate $Lie(\mathcal{F})$ by iteratively computing Lie brackets. For two vector fields X and Y on $\mathbb{R}^{n\times d}$, we use the notation $\text{ad}_X Y = [X, Y]$ and $\text{ad}_X^{\ell+1} Y = [X, \text{ad}_X^{\ell} Y]$ where $[X, Y]$ denotes the Lie bracket between X and Y. For our purpose, it is enough to compute $\text{ad}_{X_k^{\pm}}^{\ell} Y_{jk}^{\pm}$ and, given the implication $Z \in \mathcal{F} \implies -Z \in \mathcal{F}$, it suffices to compute:

$$(\text{ad}_{X_k^+}^{\ell} Y_{jk}^+)(A) = \sum_{i=1}^{d} D^\ell \sigma(A_{ki}) \frac{\partial}{\partial A_{ji}}. \tag{A.8}$$

In order to show that $\dim(Lie_A(\mathcal{F})) = \dim(M)$ at every $A \in M$, we find it convenient to work with the vectorization of elements of $\mathbb{R}^{n\times d}$. In particular, we associate the vector $\text{vec}(A) \in \mathbb{R}^{nd}$ to each matrix $A \in \mathbb{R}^{n\times d}$ where the entry (i, j) of A is identified with the entry $d^{i-1} + j$ of $\text{vec}(A)$. For a collection of matrices $\{A_1, \dots, A_k\}$, we denote by $\text{vec}\{A_1, \dots, A_k\}$ the collection of vectors $\text{vec}\{A_1, \dots, A_k\} = \{\text{vec}(A_1), \dots, \text{vec}(A_k)\}$.

Consider now the indexed collection of vector fields $\mathcal{S} = \{Z_\ell\}_{\ell\in\{1,\dots,n^2(d-1)\}}$ where:

$$Z_{1+(j-1)(n^2+1)} = \text{vec}(X_j), \qquad Z_{1+i+kn+(j-1)(n^2+1)} = \text{vec}(\text{ad}_{X_j}^k Y_{ji}).$$

We note that every $Z \in \mathcal{S}$ belongs to $Lie(\mathcal{F})$ since the vector fields in $\mathcal{S}$ either belong to $\mathcal{F}$ or are obtained by computing Lie brackets between elements of $\mathcal{F}$ and elements of $\mathcal{S}$. Moreover, we claim the evaluation of the vector fields in $\mathcal{S}$ at every $A \in M$ results in nd linearly independent vectors. To establish this claim, we form the matrix:

$$G(\text{vec}(A)) = \left[Z_1(\text{vec}(A)) | Z_2(\text{vec}(A)) | \dots | Z_{n^2(d-1)}(\text{vec}(A))\right],$$

and note that a simple but tedious computation, using (A.8), shows that G is a block diagonal matrix with d blocks, all of which being equal to:

$$G_{\text{blk}}(\text{vec}(A)) =$$
$$\begin{bmatrix} 1 & \sigma(A_{11}) & \cdots & \sigma(A_{1n}) & D\sigma(A_{11}) & \cdots & D\sigma(A_{1n}) & D^{d-2}\sigma(A_{11}) & \cdots & D^{d-2}(A_{1n}) \\ 1 & \sigma(A_{21}) & \cdots & \sigma(A_{2n}) & D\sigma(A_{21}) & \cdots & D\sigma(A_{2n}) & D^{d-2}\sigma(A_{21}) & \cdots & D^{d-2}(A_{2n}) \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & \sigma(A_{n1}) & \cdots & \sigma(A_{nn}) & D\sigma(A_{n1}) & \cdots & D\sigma(A_{nn}) & D^{d-2}\sigma(A_{n1}) & \cdots & D^{d-2}(A_{nn}) \end{bmatrix}.$$

To finish the proof, it suffices to show that G_{blk} has rank n (since it has n rows) and this is accomplished by showing there is a choice of n columns that are linearly independent. Since $A \in M$ implies $A \notin N$, by definition, there exists $\ell \in \{1, \dots, n\}$ such that:

$$\prod_{1\le i<j\le d} (A_{\ell i} - A_{\ell j}) \neq 0.$$

Moreover, by our assumption on injectivity of σ, we conclude that:

$$\prod_{1\le i<j\le d} (\sigma(A_{\ell i}) - \sigma(A_{\ell j})) \neq 0,$$

and it follows from Lemma A.1 that the matrix:

$$\begin{bmatrix} 1 & \sigma(A_{1\ell}) & D\sigma(A_{1\ell}) & \cdots & D^{d-2}\sigma(A_{1\ell}) \\ 1 & \sigma(A_{2\ell}) & D\sigma(A_{2\ell}) & \cdots & D^{d-2}\sigma(A_{2\ell}) \\ \vdots & \vdots & & \vdots & \\ 1 & \sigma(A_{n\ell}) & D\sigma(A_{n\ell}) & \cdots & D^{d-2}\sigma(A_{n\ell}) \end{bmatrix}, \tag{A.9}$$

has rank n, i.e., for every $A \in M$ there exists n columns of $G_{\text{blk}}(\text{vec}(A))$ that are linearly independent. The proof is then complete by noting that for $n > 1$, M is connected, as asserted by Proposition A.2. □

The preceding proof used the controllability properties of the vector fields (A.6) and (A.7); upon a closer look, the reader can observe that it suffices for s to take values in the set $\{-1, 1\}$ (or any set with two elements, one being positive and one being negative), for W to take values on $\{1, 0\}$ (or any other set $\{0, c\}$ with $c \neq 0$) and for b to take values on $\{0, d\}$ for some $d \in \mathbb{R}$ such that $\sigma(d) \neq 0$. Taking this observation one step further, one can establish controllability of an alternative network architecture defined by:

$$\dot{x} = S\Sigma(x) + b,$$

where the $n \times n$ matrix S and the n vector b only need to assume values in a set of the form $\{c^-, 0, c^+\}$ where $c^- \in \mathbb{R}^-$ and $c^+ \in \mathbb{R}^+$.

Proof of Corollary 4.3. The result follows from Theorem 4.2 once we establish the existence of a solution of (2.3) taking X^{init} to some point $X^{\text{fin}} \in M$. This is because Theorem 4.2 states that any other point in M will then be reachable. We proceed by showing the existence of a solution taking X^{init} to a point X^{fin} satisfying $X^{\text{fin}}_{1i} \neq X^{\text{fin}}_{1j}$ for all $i \neq j$, $i, j \in \{1, \ldots, d\}$. Clearly, $X^{\text{fin}} \in M$.

Assume, without loss of generality, that $X^{\text{init}}_{11} = X^{\text{init}}_{12}$. We will design an input, for a duration $\tau > 0$, that will result in a solution $X(t)$ with $X_{11}(\tau) \neq X_{12}(\tau)$, while ensuring that if X^{init}_{1i} is different from X^{init}_{1j} then $X_{1i}(\tau)$ is different from $X_{1j}(\tau)$.

By assumption, $X^{\text{init}}_{\bullet 1} \neq X^{\text{init}}_{\bullet 2}$. Hence, there must exist $k \in \{1, \ldots, n\}$ so that $X^{\text{init}}_{k1} \neq X^{\text{init}}_{k2}$. We use k to define the input $s = 1$, $b = 0$, and the matrix W all of whose entries are zero except for W_{1k} that is equal to 1. This choice of input results in the solution:

$$X(t) = X^{\text{init}} + t \begin{bmatrix} \sigma(X^{\text{init}}_{k1}) & \sigma(X^{\text{init}}_{k2}) & \ldots & \sigma(X^{\text{init}}_{kd}) \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}.$$

We note that $\frac{d}{dt}\big|_{t=0}(X_{11}(t) - X_{12}(t)) = \sigma(X^{\text{init}}_{k1}) - \sigma(X^{\text{init}}_{k2}) \neq 0$ since σ is injective. Therefore, there exists $\tau_1 \in \mathbb{R}^+$ such that $X_{11}(t) - X_{12}(t) \neq 0$ for all $t \in]0, \tau_1]$, i.e., $X_{11}(t) \neq X_{12}(t)$ for all $t \in]0, \tau_1]$. Moreover, we now show existence of τ_2 so that for all $t \in [0, \tau_2]$ we have $X_{1i}(t) \neq X_{2j}(t)$ whenever $X_{1i}(0) = X^{\text{init}}_{1i} \neq X^{\text{init}}_{2j} = X_{2j}(0)$. For a particular pair (X_{1i}, X_{2j}) for which $X^{\text{init}}_{1i} \neq X^{\text{init}}_{2j}$, the equality $X^{\text{init}}_{1i} + t\sigma(X^{\text{init}}_{ki}(0)) = X^{\text{init}}_{1j} + t\sigma(X^{\text{init}}_{kj}(0))$ defines the intersection of two lines. If they intersect for positive t, say t_2, it suffices to choose τ_2 smaller t_2. Moreover, by choosing τ_2 to be smaller than the positive intersection points for all pairs of lines corresponding to all pairs (X_{1i}, X_{2j}) for which $X^{\text{init}}_{1i} \neq X^{\text{init}}_{2j}$, we conclude that for all $t \in [0, \tau_2]$, $X_{1i}(0) = X^{\text{init}}_{1i} \neq X^{\text{init}}_{2j} = X_{2j}(0)$ implies $X_{1i}(t) \neq X_{2j}(t)$. Let now $\tau = \min\{\tau_1, \tau_2\}$. The point $X(\tau)$ satisfies the two properties we set to achive: 1) $X_{11}(\tau) \neq X_{12}(\tau)$; and 2) $X_{1i}(\tau) \neq X_{1j}(\tau)$ if $X^{\text{init}}_{1i} \neq X^{\text{init}}_{1j}$.

By noticing that $X^{\text{init}}_{ij} = X_{ij}(\tau)$ for $i > 1$ and any $j \in \{1, \ldots, d\}$, we can repeat this process iteratively to force all the entires of the first row of X to become different, the same way we forced the first two. □

Next, we state and prove a technical lemma that identifies monotonicity as a key property to establish function approximability in an L^∞ sense.

Lemma A.3. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous map and $E \subset \mathbb{R}^n$ a compact set. Suppose $E_{\text{samples}} \subset \mathbb{R}^n$ is a finite set satisfying:*

$$\forall x \in E \quad \exists\, \underline{x}, \overline{x} \in E_{\text{samples}}, \qquad |\underline{x} - \overline{x}|_\infty \leq \delta \quad \wedge \quad \underline{x}_i \leq x_i \leq \overline{x}_i \quad \forall i \in \{1, \dots, n\}, \tag{A.10}$$

with $\delta \in \mathbb{R}^+$, and $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is a monotone map satisfying:

$$\|f - \phi\|_{L^\infty(E_{\text{samples}})} \leq \zeta, \tag{A.11}$$

with $\zeta \in \mathbb{R}^+$. Then, we have that:

$$\|f - \phi\|_{L^\infty(E)} \leq 2\omega_f(\delta) + 3\zeta,$$

where ω_f is the modulus[3] of continuity of f.

Proof. The result is established by direct computation:

$$\begin{aligned}
|f(x) - \phi(x)|_\infty &\leq |f(x) - \phi(\underline{x})|_\infty + |\phi(\underline{x}) - \phi(x)|_\infty \\
&\leq |f(x) - f(\underline{x})|_\infty + |f(\underline{x}) - \phi(\underline{x})|_\infty + |\phi(\underline{x}) - \phi(x)|_\infty \\
&\leq \omega_f(|x - \underline{x}|_\infty) + \zeta + |\phi(\underline{x}) - \phi(x)|_\infty \\
&\leq \omega_f(|x - \underline{x}|_\infty) + \zeta + |\phi(\underline{x}) - \phi(\overline{x})|_\infty \\
&\leq \omega_f(|x - \underline{x}|_\infty) + \zeta + |f(\underline{x}) - f(\overline{x})|_\infty + |\phi(\underline{x}) - f(\underline{x})|_\infty + |f(\overline{x}) - \phi(\overline{x})|_\infty \\
&\leq \omega_f(|x - \underline{x}|_\infty) + |f(\underline{x}) - f(\overline{x})|_\infty + 3\zeta \\
&\leq \omega_f(|\overline{x} - \underline{x}|_\infty) + \omega_f(|\overline{x} - \underline{x}|_\infty) + 3\zeta \leq 2\omega_f(\delta) + 3\zeta,
\end{aligned}$$

where we used (A.11) to obtain the third and sixth inequalities. The fourth inequality was obtained by using monotonicity of ϕ to conclude $\phi(\underline{x}) \preceq \phi(x) \preceq \phi(\overline{x})$ from $\underline{x} \preceq x \preceq \overline{x}$. □

The next result shows that by restricting the input function W to assume values on the set of diagonal matrices leads to controllability being restricted to a smaller set but with the benefit of the resulting flows being monotone.

Proposition A.4. *Assume there exists $k \in \mathbb{N}_0$ so that $\xi = D^k\sigma$ is injective and satisfies a quadratic differential equation $D\xi = a_0 + a_1\xi + a_2\xi^2$ with $a_2 \neq 0$. Then, the ensemble control system* (2.3)*, with the image of W restricted to the class of diagonal matrices, is controllable on any connected component of the manifold:*

$$M = \left\{ A \in \mathbb{R}^{n \times d} \mid \prod_{1 \leq i < j \leq d} (A_{\ell i} - A_{\ell j}) \neq 0, \quad \ell \in \{1, \dots, n\} \right\}.$$

Moreover, the flow of (2.3) *joining two states in the same connected component of M is monotone.*

Proof. Since the proof of this result is analogous to the proof of Theorem 4.2 we discuss only where it differs. The restriction to the set of diagonal matrices restricts the family of vector fields $\mathcal{F}$ in the proof of Theorem 4.2 to $\{X_j^\pm, Y_{jj}^\pm\}_{j \in \{1,\dots,n\}}$. Computing the matrix $G(\text{vec}(A))$, we still obtain a block diagonal matrix but its blocks are now distinct and given by:

$$G_\ell(\text{vec}(A)) = \begin{bmatrix} 1 & \sigma(A_{1\ell}) & D\sigma(A_{1\ell}) & \dots & D^{d-2}\sigma(A_{1\ell}) \\ 1 & \sigma(A_{2\ell}) & D\sigma(A_{2\ell}) & \dots & D^{d-2}\sigma(A_{2\ell}) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \sigma(A_{n\ell}) & D\sigma(A_{n\ell}) & \dots & D^{d-2}\sigma(A_{n\ell}) \end{bmatrix}, \quad \ell \in \{1, \dots, n\}.$$

It now follows from injectivity of σ, Lemma A.1, and the definition of M that all these matrices are of full rank and we conclude controllability. Moreover, since the employed vector fields satisfy (4.1), they are monotone. Hence, the resulting flow is also monotone. □

[3]Note that f, being continuous, is uniformly continuous on any compact set.

Proof of Theorem 4.4. The result to be proved will follow at once from Lemma A.3 when we show existence of a finite set E_{samples} and a flow ϕ of (3.1) satisfying the assumptions. Existence of ϕ will be established by constructing an input $(s, W, b) : [0, \tau] \to \mathbb{R} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$ that is piecewise constant. While the input is held constant, the righthand side of (3.1) is a vector field, which we prove to be monotone. Since the composition of monotone flows is a monotone flow, the desired monotonicity of ϕ ensues.

Let $E_{\text{samples}} = \{x^1, x^2, \ldots, x^d\} \subset \mathbb{R}^n$ satisfy (A.10) for a constant $\delta \in \mathbb{R}^+$ to be later specified. The solution $Y(t)$ of the differential equation defined by Z, having every point in E_{samples} as its initial condition, can be described by the ensemble dynamical system:

$$\dot{Y}(t) = [Z(Y_{\bullet 1}(t))|Z(Y_{\bullet 2}(t))|\ldots|Z(Y_{\bullet d}(t))], \quad Y(0) = \left[x^1|x^2|\ldots|x^d\right]. \tag{A.12}$$

Since $Z^1 = f$, we will show that for every $\zeta \in \mathbb{R}^+$ there exist $\tau \in \mathbb{R}_0^+$ and an input $(s, W, b) : [0, \tau] \to \mathbb{R} \times \mathbb{R}^{n\times n} \times \mathbb{R}^n$ for the ensemble control system (2.3) so that its solution $X(t)$ starting at $Y(0)$ satisfies $|X_{\bullet j}(\tau) - Y_{\bullet j}(1)|_\infty \leq \zeta$ for $j \in \{1, \ldots, d\}$ which is a restatement of $\|\phi - f\|_{L^\infty(E_{\text{samples}})} \leq \zeta$. In particular, the flow ϕ will be defined by the solution $X(t)$.

To simplify the proof we make two claims whose proofs are postponed to after the conclusion of the main argument.

Claim 1: Along the flow of (A.12), the ordering of the entries of multiple rows of $Y(t)$ does not change at the same time instant. More precisely, for every $t \in [0, \tau]$, there exists a sufficiently small $\rho \in \mathbb{R}^+$ so that there exists at most one $i \in \{1, \ldots, n\}$ and at most one pair $(j, k) \in \{1, \ldots, d\}^2$ so that $Y_{ij}(t_1) - Y_{ik}(t_1) > 0$ for all $t_1 \in [t - \rho, t[$ and $Y_{ij}(t_1) - Y_{ik}(t_1) < 0$ for all $t_1 \in]t, t + \rho]$.

Claim 2: The interval $[0, \tau]$ can be divided into finitely many intervals:

$$]0 = t_0, t_1[\cup]t_1, t_2[\cup \ldots \cup]t_{Q-1}, t_Q = \tau[,$$

where Q is a positive integer, so that the ordering of the elements in the rows of Y does not change in these intervals.

We now proceed with the main argument. We assume that:

$$\prod_{1\leq i<j\leq d} (A_{\ell i} - A_{\ell j}) \neq 0, \quad \forall \ell \in \{1, \ldots, n\}, \tag{A.13}$$

where A is the matrix whose columns are the d elements of E_{samples}. Since the set of points violating (A.13) is a zero measure set, we can always perturb E_{samples} to ensure this assumption is satisfied. Note that (A.13) is violated at the time instants $t_1, \ldots, t_{Q-1}$ and possibly also at $t_Q = \tau$.

Recall that by Claim 2, no changes in the ordering of the entries of the rows of $Y(t)$ occur in the intervals $]t_q, t_{q+1}[$, $q \in \{0, \ldots, Q - 1\}$. Hence, we denote by S_q the set of matrices in $\mathbb{R}^{n\times d}$ that have the same ordering as $Y(t)$ in the interval $]t_q, t_{q+1}[$. Note that the sequence of visited sets S_q is uniquely determined by $Y(t)$, and hence this dependence is implicit in our chosen notation. Moreover, by (A.13) we have $Y(0) \in S_0$. The control input will be constructed so that the sequence of sets S_q visited by $X(t)$ as t ranges from 0 to τ will be the same as the sequence of sets S_q visited by $Y(t)$ as t ranges from 0 to 1. However, the time instants at which the switch from S_q to S_{q+1} occurs along the solution $X(t)$ are different from those along the solution $Y(t)$, which are given by t_q. The ability to design an input ensuring that a solution of (2.3) starting at an arbitrary point in S_q, for any given q, can reach an arbitrary point of S_q is ensured by Proposition A.4. Moreover, such input results in a flow that is monotone. Therefore, in the remainder of the proof we only need to establish that the solution of (2.3) can move from S_q to S_{q+1} along a monotone flow. Once this is established, we can compose the intermediate flows specifying how to select the inputs for the part of the flow that is in S_q, as well as the part that corresponds to exiting S_q and entering S_{q+1}. This allows us to obtain a monotone flow ϕ taking $Y(0)$ to $Y(1)$, if $Y(1)$ belongs to the interior of S_{Q-1}. If $Y(1)$ belongs to the boundary of S_{Q-1}, we can design the flow ϕ to take $Y(0)$ to any point in the interior of S_{Q-1} and, in particular, to a point that is arbitrarily close to $Y(1)$ since Proposition A.4 asserts controllability on the interior of S_{Q-1}. This will establish the desired claim that $\|\phi - f\|_{L^\infty(E_{\text{samples}})} \leq \zeta$ and any desired $\zeta \in \mathbb{R}+$. If we then choose δ and ζ so as to satisfy $2\omega_f(\delta) + 3\zeta \leq \varepsilon$, we can invoke Lemma A.3 to conclude the proof.

It only remains to show that the solution of (2.3) can move from S_q to S_{q+1} along a monotone flow. There are two situations to consider: $Y_{ij}(t_q - \rho) > Y_{ik}(t_q - \rho)$ changes to $Y_{ij}(t_q + \rho) < Y_{ik}(t_q + \rho)$

or $Y_{ij}(t_q - \rho) < Y_{ik}(t_q - \rho)$ changes to $Y_{ij}(t_q + \rho) > Y_{ik}(t_q + \rho)$, for some i, j, and $k > j$. It is clearly enough to consider one of these cases, and we assume the latter in what follows. In addition to this, from now on, we fix the indices i, j, and k.

The vectors $Y_{\bullet j}(t_q)$ and $Y_{\bullet k}(t_q)$ cannot satisfy $Y_{\bullet j}(t_q) \preceq Y_{\bullet k}(t_q)$, since monotonicity of the flow Z^t would imply the order is maintained for all future times, i.e., $Y_{\bullet j}(t) \preceq Y_{\bullet k}(t)$ for $t \geq t_q$. Since $Y_{\bullet j}(t_q) \preceq Y_{\bullet k}(t_q)$ does not hold there must exist $r \in \{1, \ldots, n\}$ such that $Y_{rj}(t_q) > Y_{rk}(t_q)$. We claim the input defined by $s = 1$, $b = 0$, and W being the matrix whose only non-zero entry is $W_{ir} = 1$, can be used to drive a suitably[4] chosen state $X^{\text{init}} \in S_q$ at time t^{init} to the some state $X^{\text{fin}} \in S_{q+1}$ at time t^{fin}. To establish this claim we need to specify the states X^{init} and X^{fin} as well the time instants t^{init} and t^{fin}. First, however, we observe that when using this input, the control system (3.1) becomes the vector field:

$$\sigma(x_r)\frac{\partial}{\partial x_i}. \tag{A.14}$$

Since by our assumption $D\sigma \geq 0$, we conclude that this vector field is monotone. Moreover, if we integrate the ensemble differential equation defined by the vector field (A.14) we obtain:

$$X_{i'j}(t^{\text{init}} + t) = X_{i'j}(t^{\text{init}}), \quad t \in [0, t^{\text{fin}} - t^{\text{init}}],$$

for all $i' \in \{1, \ldots, n\}$ with $i' \neq i$, and:

$$X_{ij}(t^{\text{init}} + t) = X_{ij}(t^{\text{init}}) + t\sigma(X_{rj}(t^{\text{init}})), \quad t \in [0, t^{\text{fin}} - t^{\text{init}}]. \tag{A.15}$$

We now assume, without loss of generality, that $X^{\text{init}}_{i\bullet}$ is ordered as follows: $X^{\text{init}}_{i1} < X^{\text{init}}_{i2} < \ldots < X^{\text{init}}_{id}$. Recall that j and $k > j$ were indices where the order of entries of $Y_{i\bullet}$ are swapped, at time t_q. We claim that $k = j + 1$; suppose on the contrary that there is an index k' such that $X^{\text{init}}_{ij} < X^{\text{init}}_{ik'} < X^{\text{init}}_{ik}$. This would violate the existence of a continuous path from $Y(t_q - \rho)$ to $Y(t_q + \rho)$ for which claim 1 holds. We already established that there exists $r \in \{1, \ldots, n\}$ such that $Y_{rj}(t_q) > Y_{r(j+1)}(t_q)$. By continuity of Y, we have $Y_{rj}(t_q - \theta) > Y_{r(j+1)}(t_q - \theta)$ for sufficiently small $\theta \in \mathbb{R}^+$. This shows that elements $A \in S_q$ satisfy $A_{rj} > A_{r(j+1)}$. As $X(t^{\text{init}}) \in S_q$, we also have $X_{rj}(t^{\text{init}}) > X_{r(j+1)}(t^{\text{init}})$. Moreover, σ being an increasing function (recall the assumption $D\sigma \geq 0$) implies $\sigma(X_{rj}(t^{\text{init}})) > \sigma(X_{r(j+1)}(t^{\text{init}}))$. Hence, and for any $t^*_j \in \mathbb{R}^+$ satisfying:

$$t^*_j > \frac{X_{i(j+1)}(t^{\text{init}}) - X_{ij}(t^{\text{init}})}{\sigma(X_{rj}(t^{\text{init}})) - \sigma(X_{r(j+1)}(t^{\text{init}}))}, \tag{A.16}$$

it follows from (A.15) that $X_{ij}(t^{\text{init}} + t^*_j) > X_{i(j+1)}(t^{\text{init}} + t^*_j)$. For any other entries $X_{ij'}(t^{\text{init}})$ and $X_{i(j'+1)}(t^{\text{init}})$ with $j' \neq j$, we will have $X_{ij'}(t^{\text{init}} + t^*_{j'}) = X_{i(j'+1)}(t^{\text{init}} + t^*_{j'})$ at time:

$$t^*_{j'} = \frac{X_{i(j'+1)}(t^{\text{init}}) - X_{ij'}(t^{\text{init}})}{\sigma(X_{rj'}(t^{\text{init}})) - \sigma(X_{r(j'+1)}(t^{\text{init}}))}. \tag{A.17}$$

Noting that $t^*_{j'}$ is an increasing function of $X_{i(j'+1)}(t^{\text{init}}) - X_{ij'}(t^{\text{init}})$, we conclude that if $X_{i(j'+1)}(t^{\text{init}}) - X_{ij'}(t^{\text{init}})$ is sufficiently large, we have $t^*_{j'} > t^*_j$. Hence, for any t^{init}, if we choose $X^{\text{init}} = X(t^{\text{init}}) \in S_q$ such that $\min_{j' \neq j} t^*_{j'} > t^*_j$, and choose $t^{\text{fin}} = t^*_j$ and $X^{\text{fin}} = X(t^{\text{init}} + t^*_j)$, we have $X^{\text{init}} \in S_q$ and $X^{\text{fin}} \in S_{q+1}$ as desired.

Proof of Claim 1: We argue that if the statement does not hold for the chosen set E_{samples}, it can always be enforced by an arbitrarily small change to the elements of E_{samples}. Let us fix $i \in \{1, \ldots, n\}$ and $j, k, l \in \{1, \ldots, d\}$, and suppose we want to avoid $Y_{ij}(t) = Y_{ik}(t) = Y_{il}(t)$ for any $t \in [0, \tau]$. The set of initial conditions to be avoided is thus:

$$B = \bigcup_{t \in [0,\tau]} \left\{ A \in \mathbb{R}^{n \times d} \mid Z^t_i(A_{\bullet j}) = Z^t_i(A_{\bullet k}) \wedge Z^t_i(A_{\bullet k}) = Z^t_i(A_{\bullet l}) \right\}.$$

Here $Z^t_i(A_{\bullet j})$ is the ij entry of the solution $Y(t)$ of (A.12) satisfying $Y(0) = A$, i.e., $Z^t_i(A_{\bullet j}) = Y_{ij}(t)$. It is convenient to define this set by the image of the smooth map $F : [0, \tau] \times \mathbb{R}^{nd-2} \to \mathbb{R}^{n \times d}$. To define F, note that the set:

$$N = \{A \in \mathbb{R}^{n \times d} \mid A_{ij} = A_{ik} = A_{il}\},$$

[4]Since Proposition A.4 asserts controllability in the set S_q, we are free to choose the state X^{init}.

is an affine subspace of $\mathbb{R}^{n\times d}$ and thus a submanifold of dimension $nd-2$. Let $W_1, \dots W_{nd-2}$ be a collection of vector fields on $\mathbb{R}^{n\times d}$ spanning[5] the tangent space to N. Using these vector fields, we define the map F as:

$$F(t, r_1, \dots, r_{nd-2}) = Z^{-t} \circ W_1^{r_1} \circ \dots \circ W_{nd-2}^{r_{nd-2}}(0).$$

We can observe that:

$$\bigcup_{(r_1,\dots,r_{nd-2})\in\mathbb{R}^{nd-2}} W_1^{r_1} \circ \dots \circ W_{nd-2}^{r_{nd-2}}(0) = N,$$

and thus:

$$\bigcup_{(t,r_1,\dots,r_{nd-2})\in[0,\tau]\times\mathbb{R}^{nd-2}} Z^{-t} \circ W_1^{r_1} \circ \dots \circ W_{nd-2}^{r_{nd-2}}(0) = B.$$

Also note that F is a smooth map, as it is a composition of smooth flows. Moreover, its domain is a manifold with boundary of dimension smaller than the dimension of its co-domain. Hence, it follows from Corollary 6.11 in (Lee, 2013) that the image of F has zero measure in $\mathbb{R}^{n\times d}$. We can similarly show that all the other ordering changes to be avoided result in zero measure sets. Since there are finitely many of these sets to be avoided, and a finite union of zero measure sets still has zero measure, we conclude that Claim 1 can always be enforced by suitably perturbing the elements of E_{samples} if necessary.

Proof of Claim 2: To show this claim is satisfied, let $\gamma_{ijk} : \mathbb{R} \to \mathbb{R}$ be the function defined by $\gamma_{ijk}(t) = Y_{ij}(t) - Y_{ik}(t)$. The instants $t_q \in \{0, 1, \dots, Q\}$, correspond to the zeros of γ_{ijk}, i.e., $\gamma_{ijk}(t_i) = 0$. Since Y is an analytic vector field, by (Sontag, 1998, Proposition C.3.12), the function γ_{ijk} is also analytic and its zeros are isolated. Therefore, the function γ_{ijk} restricted to the compact set $[0, \tau]$ only has finitely many zeros. Since there are finitely many functions γ_{ijk} as (i, j, k) ranges on $\{1, \dots, d\}^3$, there are only finitely many instants t_q. □

Proof of Corollary 4.5. Since the map f is continuous and defined on a compact set, it follows from the Stone-Weierstass theorem that there exists a polynomial $\tilde{f} : E \to \mathbb{R}^n$ satisfying $\|f-\tilde{f}\|_{L^\infty(E)} \leq \frac{\varepsilon}{2}$. We now construct an analytic vector field $Z : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, an injection $\alpha : \mathbb{R}^n \to \mathbb{R}^{2n}$, and a projection $\beta : \mathbb{R}^{2n} \to \mathbb{R}^n$ satisfying:

$$\left\|f - \beta \circ Z^1 \circ \alpha\right\|_{L^\infty(E)} \leq \frac{\varepsilon}{2}. \tag{A.18}$$

The vector field Z is given by $Z(x, y) = (\tilde{f}(y) - y + Ky)\frac{\partial}{\partial x} + 0\frac{\partial}{\partial y}$, where $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ and the matrix K satisfies:

$$\frac{\partial \tilde{f}}{\partial y} - I + K \geq 0, \tag{A.19}$$

for all $y \in E$ and where I is the identity matrix. Note that K exists since $\partial\tilde{f}/\partial y$ is continuous and E compact. Vector field Z is analytic, since $\tilde{f}$ is so, and is also monotone, since its mixed partial derivatives are given by (A.19), and thus non-negative, see (4.1). The injection α is given by $\alpha(x) = (x, x)$, and the projection β is given by $\beta(x, y) = x - Ky$. By noting that the flow of Z is $Z^t(x, y) = (x + t(\tilde{f}(y) - y + Ky), y)$, we compute:

$$\beta \circ Z^1 \circ \alpha(x) = \beta \circ Z^1(x, x) = \beta\left(\tilde{f}(x) + Kx, x\right) = \tilde{f}(x),$$

which shows that (A.18) holds. By Theorem 4.4, there exists an input $(s, W, b) : [0, \tau] \to \mathbb{R} \times \mathbb{R}^{2n\times 2n} \times \mathbb{R}^{2n}$ so that the flow ϕ^τ of (3.1) satisfies:

$$\|Z^1 - \phi^\tau\|_{L^\infty(E)} \leq \frac{\varepsilon}{2(1 + \|K\|)}. \tag{A.20}$$

We therefore have:

$$\begin{aligned}
\left\|\tilde{f} - \beta \circ \phi^\tau \circ \alpha\right\|_{L^\infty(E)} &= \left\|\beta \circ Z^1 \circ \alpha - \beta \circ \phi^\tau \circ \alpha\right\|_{L^\infty(E)} \\
&\leq (1 + \|K\|)\left\|Z^1 \circ \alpha - \phi^\tau \circ \alpha\right\|_{L^\infty(E)} \\
&\leq \frac{\varepsilon}{2},
\end{aligned}$$

[5]A globally defined basis for the tangent space of N exists since N is an affine manifold.

where the first inequality follows from $(1 + \|K\|)$ being the Lipschitz constant of β and the second from (A.20). Finally, we use the preceding inequality to establish:

$$\|f - \beta \circ \phi^\tau \circ \alpha\|_{L^\infty(E)} \leq \left\|f - \tilde{f}\right\|_{L^\infty(E)} + \left\|\tilde{f} - \beta \circ \phi^\tau \circ \alpha\right\|_{L^\infty(E)} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

□

We finish by making the remark that the choice of the vector field Z in this proof is certainly not unique, and any other choice that still results in the required monotonicity can be used.