Journal of Machine Learning Research 26 (2025) 1-33 Submitted 11/24; Revised 6/25; Published 11/25

Equivariant Manifold Neural ODEs and Differential Invariants

Emma Andersdotter EMMA.ANDERSDOTTER@UMU.SE
Department of Mathematics and Mathematical Statistics
Umeå University
Umeå, SE-901 87, Sweden

Daniel Persson DANIEL.PERSSON@CHALMERS.SE
Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
Gothenburg, SE-412 96, Sweden

Fredrik Ohlsson FREDRIK.OHLSSON@UMU.SE
Department of Mathematics and Mathematical Statistics
Umeå University
Umeå, SE-901 87, Sweden

Editor: Joan Bruna

Abstract

In this paper we develop a geometric framework for equivariant manifold neural ordinary differential equations (NODEs), and use it to analyse their modelling capabilities for symmetric data. First, we consider the action of a Lie group G on a smooth manifold M and establish the equivalence between equivariance of vector fields, symmetries of the corresponding Cauchy problems, and equivariance of the associated NODEs. We also propose a novel formulation of the equivariant NODEs in terms of the differential invariants of the action of G on M, based on Lie theory for symmetries of differential equations, which provides an efficient parameterisation of the space of equivariant vector fields in a way that is agnostic to both the manifold M and the symmetry group G. Second, we construct augmented manifold NODEs through embeddings into equivariant flows, and show that they are universal approximators of equivariant diffeomorphisms on any connected M. Furthermore, we show that the augmented NODEs can be incorporated in the geometric framework and parametrised using higher order differential invariants. Finally, we consider the induced action of G on different fields on M and show how it generalises previous work, e.g., continuous normalizing flows, to equivariant models in any geometry.

Keywords: neural ODEs, manifolds, augmentation, equivariance, differential geometry, differential invariants, symmetries of differential equations, geometric deep learning

1. Introduction

Recent years have seen a 'geometrisation' of neural networks, driven by the need to deal with data defined on non-Euclidean domains, for example, graphs or manifolds. The field of geometric deep learning (Bronstein et al., 2017; Bronstein et al., 2021; Gerken et al., 2023) aims to incorporate the geometry of data in a foundational mathematical description of neural networks. This endeavour has been extremely successful in the context of group equivariant convolutional neural networks (CNNs), in which the symmetries of the underlying data are built into the neural network using techniques from group theory and representation theory.

Another development in the same spirit explores the connection to differential equations and dynamical systems in the limit of infinitely deep networks. When formulated in terms of the dynamics propagating information through the network, the learning problem becomes amenable to powerful numerical techniques for differential equations and a rich theory of dynamical systems. In particular, neural ordinary differential equations (NODEs) have received considerable attention since they were proposed in their current form by Chen et al. (2018).

NODE models were originally conceived by considering the continuous dynamical systems obtained in the limit of infinitely deep residual neural networks (He et al., 2016), building on previous similar constructions considered by E (2017); Haber and Ruthotto (2018); Ruthotto and Haber (2020). The dynamical system describes the evolution of a state $u(t) \in \mathbb{R}^n$ according to a governing ordinary differential equation (ODE)

$$\dot{u}(t) = \phi(u, t)\,, \tag{1}$$

where $\phi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is a vector valued function parameterising the dynamics. The NODE model is then given by the map $h : \mathbb{R}^n \to \mathbb{R}^n$ defined by evolving the input $x = u(0)$ to the output $h(x) = u(1)$. In $\mathbb{R}^n$, this can be obtained by direct integration of the vector field ϕ as

$$h(x) = x + \int_0^1 \phi(u, t)\,\mathrm{d}t\,.$$

Chen et al. (2018) parametrised ϕ using a fully connected feed-forward neural network, and the backpropagation of gradients by solving an adjoint ODE related to (1) was used to enable learning of ϕ, extending previous results by Stapor et al. (2018). We emphasise that the vector field ϕ can be parametrised by any machine learning model and is not restricted to be a neural network. However, 'neural ODE' is an established term in the literature—in particular in the context of geometrical constructions such as manifolds and equivariance relevant for our work. Therefore, we opt to use it here, interpreting 'neural' more generally as 'learnable' or 'parametrised'.

NODE models possess several attractive theoretical properties. Of particular conceptual interest is the bijectivity of h, bestowed by the uniqueness of integral curves guaranteed by the Picard-Lindelöf theorem, which allows the application to generative models of probability densities (Rezende and Mohamed, 2015; Chen et al., 2018).

A natural generalisation of neural differential equations in geometric deep learning is to consider ODEs on a smooth manifold M, rather than $\mathbb{R}^n$, to accommodate non-Euclidean data. Such manifold neural ODEs were introduced in co0,ncurrent works by Falorsi and Forré

(2020); Lou et al. (2020); Mathieu and Nickel (2020). The situation where the data exhibits symmetries under some group G of transformations naturally entails the construction of equivariant NODEs which were obtained by Köhler et al. (2020) for the Euclidean setting and more recently by Katsman et al. (2021) for Riemannian manifolds M.

In this paper, we develop the mathematical foundations of equivariant manifold NODEs using their description in terms of differential geometry. In the past few years, the field of neural ODEs has seen several interesting developments in different directions such as novel model constructions, efficient implementations and competitive performance on applied problems. In §1.2, we provide an overview of the most relevant previous works and how they relate to our results. Our aim in the present paper is to complement these recent developments by providing a geometric framework for manifold neural ODEs, incorporating Lie theory of symmetries of differential equations. We then use the framework to analyse the theoretical modelling and approximation capabilities of manifold NODEs and to provide geometrical insight into their properties.

1.1 Summary of results

We first provide some mathematical background on differential geometry and Lie theory in §2. In §3, we then construct a geometric framework for NODEs on arbitrary smooth manifolds M equivariant under the action of a connected Lie group G. In particular, we generalise previous constructions by removing assumptions on a metric structure on M and establishing a stronger version of the fundamental relationship between the equivariance of the NODE model $h : M \to M$ and the corresponding ODE on M (Theorem 10). In addition, we provide a novel way of describing the NODE model space, based on the classical Lie theory of symmetries of differential equations, by parameterising equivariant vector fields ϕ on M using differential invariants of the action of G (Theorem 11).

Using the geometric framework, we then investigate the approximation capabilities of manifold NODEs in §4. In particular, we show how NODEs can be augmented in a way that respects the manifold structure of M by embedding the diffeomorphism $h : M \to M$ in a flow on the tangent bundle TM (Theorem 15). We prove that the augmented NODE models are universal approximators when M is connected (Corollary 16). We proceed to show that the embedding into TM is compatible with the induced action of G on TM (Theorem 19), which proves that equivariant NODEs are universal approximators of equivariant diffeomorphisms (Corollary 20). Furthermore, we show how the space of augmented equivariant NODEs can be parametrised in terms of higher-order differential invariants of G (Theorem 21).

Finally, in §5 we use the induced action of the diffeomorphism h to construct NODE models for different types of fields on M. In particular, we consider explicitly equivariant scalars, densities (Theorem 24) and vector fields on M (Theorem 25).

The main contributions of the present paper are:

- We develop a geometric framework for NODEs on an arbitrary smooth manifold M, equivariant with respect to the action of a connected Lie group G. In particular, we use the differential invariants of G to parametrise the space of equivariant manifold NODEs. Our general method is outlined in two algorithms presented in §3 and §4.

- We show how NODEs can be augmented within our geometric framework using higher-order differential invariants in a way that respects both the manifold structure of M and the equivariance under G. We prove that the resulting (equivariant) models are universal approximators for (equivariant) diffeomorphisms $h : M \to M$ of connected manifolds.
- We show how our framework can be used to model different kinds of equivariant densities and fields on M, using the induced action of the diffeomorphism h.

1.2 Related work

In this section, we provide an overview of important previous results in the literature related to equivariant manifold NODEs in general and our constructions and results in particular. In line with the focus of the present paper, we emphasise theoretical developments and contextualise our results where appropriate.

Equivariant continuous normalizing flows Normalizing flows are generative models that construct complicated probability distributions by transforming a simple probability distribution (e.g., the normal distribution) using a sequence of invertible maps (Rezende and Mohamed, 2015; Chen et al., 2019). Similar to residual networks, the sequence can be viewed as an Euler discretisation. This connection was used by Chen et al. (2018) to show that NODEs can model infinite sequences of infinitesimal normalizing flows, resulting in continuous normalizing flows (CNFs). This construction has been further developed in several directions, e.g., by Grathwohl et al. (2019); Onken et al. (2021).

Equivariant continuous normalizing flows on Euclidean space $\mathbb{R}^n$ were introduced by Köhler et al. (2020), extending previous constructions by Satorras et al. (2021); Köhler et al. (2019); Rezende et al. (2019). The prior and target densities in these normalizing flows share the same symmetries, provided that the maps are equivariant, which is equivalent to equivariance of the vector field generating the flow.

Manifold neural ODEs The mathematical framework required for extending NODEs to manifolds was developed by Falorsi and Forré (2020) which also provides a rigorous description of the adjoint method for backpropagation in terms of the canonical symplectic structure on the cotangent bundle. Explicit implementations of both forward and backward integration on Riemannian manifolds using dynamical trivialisations were developed by Lou et al. (2020) based on the exponential map. The application considered for the manifold NODEs in these works is mainly restricted to continuous normalizing flows (Mathieu and Nickel, 2020). In this context, Rezende et al. (2020), who develop finite normalizing flows on tori and spheres, and Katsman et al. (2023), who consider Riemannian residual networks, also deserve to be mentioned.

Closest to the results presented in this paper is the work of Katsman et al. (2021), which considers flows on Riemannian manifolds equivariant under subgroups of isometries. In contrast, our work generalises this construction to a framework that accommodates arbitrary symmetry groups and manifolds, and extends it from densities to more general fields on M without requiring the existence of a metric.

Universality of neural ODEs Since NODEs model invertible transformations as flows on M, which are restricted in that flow lines cannot intersect on M, not all diffeomorphisms

$h : M \to M$ can be approximated using NODEs realised on the manifold M itself. However, any diffeomorphism $h : M \to M$ can be expressed as a flow on some ambient space where M is embedded (Utz, 1981). This idea was first implemented by Dupont et al. (2019) to define augmented NODEs for $M = \mathbb{R}^n$, and further developed by Zhang et al. (2020) to prove that augmented NODEs on $\mathbb{R}^{2n}$ are universal approximators of diffeomorphisms $h : \mathbb{R}^n \to \mathbb{R}^n$. Using augmentation, Bose et al. (2021) extended the universality to equivariant finite normalizing flows on $\mathbb{R}^n$. However, to the best of our knowledge, augmentation of manifold NODEs has not been previously considered in the literature. In this work, we use our geometric framework to construct augmented manifold NODEs and establish their universality for connected M. Furthermore, we generalise the construction to the equivariant case and show that universality persists in the presence of a non-trivial symmetry group G.

Differential invariants in neural networks An unresolved problem in the theory of manifold neural ODEs is to parametrise the space of vector fields, or at least a sufficiently large subset of them, to obtain expressive models (Falorsi and Forré, 2020). In the equivariant setting, this has been accomplished by considering gradient flows of invariant potential functions on Euclidean spaces (Köhler et al., 2020) and Riemannian manifolds (Katsman et al., 2021), respectively. This strategy, while practically appealing, restricts models to conservative vector fields.

Our approach to parameterising the neural ODE models is based on the theory of symmetries of differential equations (see, e.g., the seminal text by Olver (1993)) and uses differential invariants of the action of the symmetry group G on the manifold M. Recently, a similar approach was proposed by Knibbeler (2024) to address the parameterisation of invariant sections of homogeneous vector bundles related to equivariant convolutional neural networks (Kondor and Trivedi, 2018; Cohen et al., 2019; Aronsson, 2022).

The theory of symmetries of differential equations was also used by Akhound-Sadegh et al. (2023) to create physics-informed networks (PINNs), where a loss term incorporating the infinitesimal generators of symmetry transformations is used to enforce approximate invariance of the equations under the action of the symmetry group. Even closer to our approach of using differential invariants is the work of Arora et al. (2024), which constructs solutions to an ODE by first mapping to the space spanned by the differential invariants, learning a solution to the invariantised ODE using a PINN, and reconstructing the corresponding solution to the original equation. In both cases, PINNs are used to learn solutions to specific differential equations. In contrast, our approach uses differential invariants to parametrise the space of equivariant differential equations, and we attempt to learn the one which best approximates the diffeomorphism $h : M \to M$ mapping input data to output data. Somewhat conversely, equivariant CNNs were used by Lagrave and Tron (2022) to learn differential invariants of partial differential equations (PDEs) and use them to derive symmetry-preserving finite difference solution algorithms.

Flow matching An important recent contribution to the field of NODEs in the form of CNFs is the development of the flow matching (FM) framework, which eliminates the need to perform the computationally expensive ODE simulations and gradient evaluations required in the maximum likelihood training approach originally proposed by Chen et al. (2018). The FM paradigm amounts to directly regressing the vector field generating the normalizing flow to the tangent vector of some desired path in the space of probability distributions. This approach was first introduced for the Euclidean case by Lipman et al. (2023), who defined the target probability path and corresponding vector fields by marginalising over probability paths conditioned on the data samples, leading to conditional flow matching (CFM). The method was then further developed by Tong et al. (2024), who considered mixtures of probability paths and vector fields conditioned on general latent variables. In particular, optimal transport CFM (OT-CFM) was obtained by using the Wasserstein optimal transport map as the latent distribution.

The flow matching framework was extended to Riemannian manifolds by Chen and Lipman (2024) at the expense of still requiring ODE simulations for manifolds without closed-form expressions for the geodesics. This result was then applied by Yim et al. (2023); Bose et al. (2024) to construct equivariant models for protein structure generation by considering equivariant CNFs on the manifold SE(3). A general equivariant framework for flow matching in the Euclidean case was obtained by Klein et al. (2023) by minimising over the orbit of the symmetry group in the distance function used in OT-CFM, which amounts to aligning each sample pair along their respective orbits.

In the context of the work we present in this paper, it is important to emphasise that flow matching is a framework for training CNFs rather than a class of NODE models. The FM approach can be directly applied to CNF models—including augmented models—in our framework to provide scalable and efficient training based on the parameterisation of NODE model space in terms of differential invariants.

2. Mathematical preliminaries

In this section, we review some mathematical concepts that are relevant to the remainder of the article. The reader is assumed to have a working knowledge of differential geometry and group theory. For readers unfamiliar with these topics, Nakahara (2003) covers most of the background knowledge needed for this section and provides a physics-based perspective on differential geometry and group theory. More mathematically rigorous texts include the works of Lee (2012) and Helgason (2001).

2.1 Group actions and flows on manifolds

Let G be a finite-dimensional Lie group with Lie algebra $\mathfrak{g}$[1]. The left action $\alpha : G \times M \to M$ of a Lie group G on a finite-dimensional manifold M is an operator endowed with the identity and associativity property of the group structure of G. We introduce the operator $L_g : M \to M$ satisfying $L_g(p) = \alpha(g, p)$ for each $g \in G$ and $p \in M$. The *orbit* of a point $p \in M$ is given by

$$\mathcal{O}_p = \{L_g(p) \,:\, g \in G\}.$$

The action of G is called *semi-regular* if all orbits have the same dimension.

Example 1 *Let $M = \mathbb{R}^2 \setminus \{0\}$ and $G = \mathrm{SO}(2)$. A semi-regular action of G on M is given by rotations in the usual way, i.e.,*

$$L_\epsilon(x, y) = (x\,\cos\epsilon - y\,\sin\epsilon, x\,\sin\epsilon + y\,\cos\epsilon),$$

1. Here $\mathfrak{g}$ denotes the Lie algebra associated with G, i.e., the vector space isomorphic to T_eG.

where $u(t) = (x(t), y(t))$ *and* ϵ *is the arbitrary angle of rotation parametrising* $\mathrm{SO}(2)$. *The fix-point at the origin in* $\mathbb{R}^2$ *is excluded to obtain a semi-regular action of* $\mathrm{SO}(2)$.

A *vector field* ϕ is a smooth assignment of a tangent vector to each $p \in M$ and a *flow* on M is a map $\Phi : \mathbb{R} \times M \to M$ which is an action of the additive group $\mathbb{R}$, i.e., $\Phi(0, p) = p$ and $\Phi(s, \Phi(t, p)) = \Phi(s + t, p)$ for all $p \in M$ and $s, t \in \mathbb{R}$. The flow Φ is generated by the vector field ϕ if for every point $p \in M$ we have

$$\frac{d}{dt}\Phi(t,p)\bigg|_{t=0} = \phi_p \,. \tag{2}$$

Such a flow can be viewed as a solution of an ODE. Let $u : \mathbb{R} \to M$ be the integral curve solving the initial value problem

$$\frac{du}{dt} = \phi_{u(t)} \,, \quad u(0) = p \,.$$

It follows from $\Phi(0, p) = p$, equation (2) and the uniqueness of u, guaranteed by the Picard-Lindelöf theorem, that $u(t) = \Phi(t, p)$. Thus, the flow Φ generated by the vector field ϕ is the collection of all integral curves of ϕ.

Given a vector field ϕ, its corresponding flow is referred to as the *exponentiation* of ϕ:

$$\Phi(t, p) = \exp{(t\phi)}(p).$$

It follows from the properties of a flow that exp satisfies $\exp{(0\phi)}p = p$, $\frac{d}{dt}\exp{(t\phi)}p = \phi_{\exp(t\phi)p}$ and $\exp{(t\phi)}\exp{(s\phi)} = \exp{((s + t)\phi)}$ for all $s, t \in \mathbb{R}$.

Given two flows σ and ψ generated by the vector fields X and Y, respectively, the *Lie derivative* $\mathcal{L}_X Y$ is defined as the rate of change of Y along σ. One can show that the Lie derivative of a vector field is given by the Lie bracket, $\mathcal{L}_X Y = [X, Y]$.

The *push-forward* (or *differential*) of a smooth map $f : M \to N$ between smooth manifolds at a point $p \in M$ is a map $df_p : T_pM \to T_{f(p)}N$ such that, for all differentiable functions h defined on a neighbourhood of $f(p)$,

$$(df_p(X_p))(h) = X_p(h \circ f),$$

where $X_p \in T_pM$. If γ is a curve on M such that $\gamma(0) = p$ and $\dot{\gamma}(0) = X_p$, this can also be written as

$$(df_p(X_p))(h) = \frac{d}{dt}(h \circ f \circ \gamma(t))\bigg|_{t=0} .$$

Throughout this paper, we will primarily consider the push-forward of the left action $L_g : M \to M$ and denote this by $(L_g)_* : T_pM \to T_{L_g p}M$. To make the expression less cumbersome, we suppress the reference to p in the notation $(L_g)_*$.

Let $f : M \to N$ be a map between two manifolds M and N, each equipped with a left group action of G. Then f is called *equivariant* if $f \circ L_g = L_g \circ f$ holds for all $g \in G$. A special case of equivariance is *invariance*, i.e., $f \circ L_g = f$ for all $g \in G$. When determining the equivariance properties of manifold neural ODEs, we will primarily consider equivariance of vector fields, flows and diffeomorphisms. Vector fields and flows were defined above. A diffeomorphism is a smooth bijection $h : M \to M$ with a smooth inverse. Below, we define what equivariance means for each respective object.

Definition 1 *A flow $\Phi : \mathbb{R} \times M \to M$ is equivariant if $L_g\Phi(t,p) = \Phi(t, L_g p)$ for all $t \in \mathbb{R}$, $p \in M$ and $g \in G$.*

Definition 2 *A vector field $\phi : M \to TM$ is equivariant if $\phi_{L_g p} = (L_g)_* \phi_p$ for all $p \in M$ and $g \in G$.*

Definition 3 *A diffeomorphism $h : M \to M$ is equivariant if $L_g \circ h = g \circ L_g$ for all $g \in G$.*

2.2 Symmetries of differential equations

We will now briefly review the general theory of Lie point symmetries of differential equations as originally envisioned by Lie. We follow the geometric description of these symmetries by, e.g., Olver (1993, 1995) and consider a general ODE of the form

$$u^{(k)} = \phi\left(t, u, u^{(1)}, \dots, u^{(k-1)}\right)\,, \quad u^{(k)}(t) = \frac{d^k}{dt^k}u(t)\,,$$

or equivalently $\Delta\left(t, u, u^{(1)}, \dots, u^{(k)}\right) := u^{(k)} - \phi\left(t, u, u^{(1)}, \dots, u^{(k-1)}\right) = 0$.

The solution $u(t)$ of an ODE is a curve on a manifold M depending on a variable t. Let $T \cong \mathbb{R}$ be the space parametrised by the independent variable and M the space of dependent variables. We define the *total space* as the space $E = T \times M$. A smooth curve $u : \mathbb{R} \to M$ has the *graph* $\gamma_u = \{(t, u(t))\} \subset E$. The general theory allows for actions of G on both the independent variables T and the dependent variables M. In the application to NODEs, symmetries are considered only in terms of transformations of the inputs and outputs, and not as acting on the time t parameterising the internal dynamics of the model. Thus, we restrict the left action of G to only the dependent variables M, i.e., the left action on a graph $\{(t, u(t)\} \subset E$ is given by $\{(t, L_g u(t)\}$ for each $g \in G$.

The fundamental objects describing a Lie group G of point transformations acting on the total space E are the induced vector fields X_i, $i = 1, \dots, d = \dim G$, on E defining a basis for the Lie algebra $\mathfrak{g}$. The induced vector fields generate the actions of G via exponentiation, i.e., for each $g \in G$ there is a corresponding vector field X_g such that $L_g p = (\exp X_g)p$ for each $p \in M$. The group G is a symmetry group of Δ if it preserves the solution space of Δ, i.e., for every solution $u(t)$ and group element $g \in G$, the transformed function $L_g(u(t))$ is also a solution to $\Delta = 0$.

The geometric description uses the concept of jet bundles $J^{(k)}E$ which extends the space $E = T \times M$ to include the derivatives $u^{(1)}, \dots, u^{(k)}$. It is customary to write $J^{(k)}E = T \times M^{(k)}$, where $M^{(k)}$ is a manifold of dimension $k \dim M$ containing all the points in M as well as their derivatives up to order k. The graph γ_u of a smooth curve $u : \mathbb{R} \to M$ can be *prolonged* to $J^{(k)}E$ as $\gamma_u^{(k)} = \{(t, u(t), \partial_t u(t), \dots, \partial_t^k u(t))\} \subset J^{(k)}E$. The *prolonged group action* $L_g^{(k)}$ on the graph $\gamma_f^{(k)}$ is then defined as $L_g^{(k)}\gamma_f^{(k)} = \gamma_{L_g f}^{(k)}$. The induced action of X on the derivatives is generated by the *prolonged vector field* $X^{(k)}$, which can be used to express the symmetry condition succinctly in its infinitesimal form

$$X^{(k)}\left(u^{(k)} - \phi\left(t, u, u^{(1)}, \dots, u^{(k-1)}\right)\right)\Big|_{\Delta=0} = 0\,.$$

Note that the restriction of L_g to M implies that $L_g^{(k)}$ and the corresponding generating vector fields are restricted to $M^{(k)}$.

Elementary results from the theory of algebraic equations on manifolds, applied to the manifold $J^{(k)}E$, can then be used to construct the most general ODE which is equivariant under the symmetry group G. The construction is based on *differential invariants*, i.e., functions $I : J^{(k)}E \to \mathbb{R}$ which are invariant under the action of G, which infinitesimally means $X_i^{(k)}(I) = 0$ for $i = 1, \dots, d$. Since a function of a set of invariants is also an invariant, we only need to consider the complete set of functionally independent invariants. The number of such invariants is given by the following theorem.

Theorem 4 (Theorem 2.34, Olver (1995)) *If the prolonged group $G^{(k)}$ of a group G acts semi-regularly on $J^{(k)}E$ with orbit dimension s_k, there are $\dim J^{(k)}E - s_k$ functionally independent local differential invariants of order k.*

Having a complete set of functionally independent differential invariants makes it possible to express the most general G-equivariant ODE.

Theorem 5 (Proposition 2.56, Olver (1993)) *Let G be a Lie group acting semi-regularly on $E = T \times M$ and let $I_1, \dots, I_{\mu_k}$ be a complete set of functionally independent differential invariants of order k. Then the most general ODE of order k for which G is a symmetry group is locally on the form*

$$F(I_1, \dots, I_{\mu_k}) = 0$$

for an arbitrary function $F : \mathbb{R}^{\mu_k} \to \mathbb{R}^n$, where n is the dimension of M.

Example 2 *Let $G = \mathrm{SO}(2)$ act semi-regularly on $M = \mathbb{R} \setminus \{0\}$ as described in Example 1. The induced vector field X generating G is then obtained by differentiating the group action in Example 1 as $X = -y\partial_x + x\partial_y$. The first-order prolongation $X^{(1)}$ is obtained by direct computation (see, e.g., Olver (1993, §2.3)), resulting in*

$$X^{(1)} = -y\partial_x + x\partial_y - \dot{y}\partial_{\dot{x}} + \dot{x}\partial_{\dot{y}}.$$

By Theorem 4, there are four functionally independent first-order differential invariants, $I : J^{(1)}E \to \mathbb{R}$, satisfying $X^{(1)}(I) = 0$. Solving the corresponding characteristic system, they are found to be

$$I_1 = t\,, \quad I_2 = r = \sqrt{x^2 + y^2}\,,$$
$$I_3 = r\dot{r} = x\dot{x} + y\dot{y}\,, \quad I_4 = r^2\dot{\theta} = x\dot{y} - y\dot{x},$$

where (r, θ) are polar coordinates on M. By Theorem 5, the most general first-order ODE for which G is a symmetry group is of the form $H(t, r, r\dot{r}, r^2\dot{\theta}) = 0 \in \mathbb{R}^2$.

Remark 6 *In the above construction, we focus exclusively on the case where the space of independent variables, T, is isomorphic to $\mathbb{R}$. The general setting—which includes the theory for partial differential equations (PDEs)—allows T to be of arbitrary dimension. However, since our framework is concerned with ODEs, including such a general setting would be redundant and require cumbersome notation. See, e.g., Olver (1993, 1995) for a more comprehensive overview.*

3. Geometric framework

In this section, our aim is to establish the appropriate geometric framework for describing manifold NODEs on a smooth manifold M of dimension $\dim M = n$, and in particular their equivariance under a group G of symmetry transformations acting on M. Throughout, we will assume that M is connected and that G is a connected Lie group whose action on M is semi-regular.

Comparing to the general theory described in the previous section, this amounts to restricting the action of G to be trivial on T. This restriction is motivated by the role of $t \in T$ as parameterising the flow of information through the model; we are interested in transformations acting on the state $u(t)$, not transformations acting on the parameter t.

In §3.1, we give a precise definition of NODEs and their equivariance. In §3.2, we generalise previous results on the equivalence between a NODE, its generating vector field and the corresponding flow on M. Finally, §3.3 introduces the notion of differential invariants in the context of NODEs and how they can be used to parametrise a NODE based on Theorem 5.

3.1 Manifold neural ODEs

Manifold neural ODEs are machine learning models defined by a vector field describing how the data evolves continuously over time governed by an ordinary differential equation. The word *manifold* emphasises that the data can be defined on non-Euclidean manifolds. We define manifold neural ODEs in the following way:

Definition 7 (Manifold neural ODE) *Given a point $p \in M$ and a learnable vector field $\phi : M \to TM$, let $u : \mathbb{R} \to M$ be the unique curve solving the Cauchy problem*

$$\dot{u}(t) = \phi_{u(t)}\,, \quad u(0) = p\,.$$

A manifold neural ODE on M is the diffeomorphism $h : M \to M$ defined by $h : u(0) \mapsto u(1)$. The point $p \in M$ is referred to as the input and $h(p) \in M$ as the output.

Remark 8 *The map $h : M \to M$ can, equivalently, be defined by $h : \Phi(0, p) \mapsto \Phi(1, p)$, where Φ is the flow on M generated by a learnable vector field $\phi : M \to TM$.*

A neural network is said to be equivariant with respect to a group G if applying the group action to the input before applying the network gives the same result as applying the group action to the output after applying the network. Definition 7 allows us to define equivariance of a manifold neural ODE in a concise way:

Definition 9 (Equivariant manifold neural ODE) *A manifold neural ODE $h : M \to M$ is said to be G-equivariant if $h \circ L_g = L_g \circ h$ holds for all $g \in G$.*

3.2 Equivariance of neural ODEs

In this section, we introduce Theorem 10 which demonstrates the equivalence between equivariance of a vector field ϕ, its generated flow Φ and the corresponding diffeomorphism h mapping the input to the output. The implication (i) to (iii) in Theorem 10 below was proven for $M = \mathbb{R}^n$ by Köhler et al. (2020) and the equivalence between (i) and (ii) was established for Riemannian manifolds by Katsman et al. (2021). Our result further extends these connections by establishing the equivalence between equivariance of the vector field defining the Cauchy problem, the flow that it generates, and the diffeomorphism $h : M \to M$ defining the corresponding NODE. Furthermore, our construction works for any connected manifold without the requirement of a metric structure on M. The proof by Katsman et al. (2021), in fact, makes no explicit use of the Riemannian metric and establishes the equivalence of (i) and (ii) also in our more general setting. We include this as part of the proof for completeness.

Theorem 10 *Let ϕ be a smooth vector field on the manifold M, $\Phi : \mathbb{R} \times M \to M$ be the flow generated by ϕ and $h : M \to M$ be the diffeomorphism defined by $p = u(0)$, $h(p) = u(1)$, where $u(t)$, $t \in [0, 1]$, is the unique solution to the Cauchy problem*

$$\dot{u}(t) = \phi_{u(t)}\,, \quad u(0) = p\,. \tag{3}$$

Then the following are equivalent:

(i) ϕ is a G-equivariant vector field,

(ii) Φ is a G-equivariant flow,

(iii) h is a G-equivariant diffeomorphism.

Proof We begin by establishing the equivalence between (i) and (ii). Assume that ϕ is G-equivariant. We want to show that $\Phi(t, L_g p) = L_g\Phi(t, p)$ for all $g \in G$, $p \in M$ and for all t. By the properties of flows, we have $\Phi(0, L_g p) = L_g p = L_g\Phi(0, p)$. The equivariance of ϕ (Definition 2) implies

$$\frac{d}{dt} L_g\Phi(t, p) = (L_g)_*\phi_{\Phi(t,p)} = \phi_{L_g\Phi(t,p)},$$

i.e., $L_g\Phi(t, p)$ and $\Phi(t, L_g p)$ solve the same ODE and have the same initial condition. G-equivariance of Φ follows from the uniqueness of the integral curve. Conversely, if Φ is G-equivariant, then

$$\phi_{L_g p} = \frac{d}{dt}\Phi(t, L_g p)\bigg|_{t=0} = \frac{d}{dt} L_g\Phi(t, p)\bigg|_{t=0} = (L_g)_*\phi_p$$

for any point $p \in M$. Thus, (i) and (ii) are equivalent.

We complete the proof by showing that (ii) and (iii) are equivalent. To show that (ii) implies (iii) is trivial, since it follows directly from the fact that $h(p) = \Phi(1, p)$. To show the converse, however, requires some more work.

Suppose that the diffeomorphism h is G-equivariant. We know that the flow generated by ϕ is given by the exponentiation of ϕ, i.e., $\Phi(t, p) = \exp(t\phi)p$. It follows that h can be expressed in terms of the vector field ϕ as

$$h(p) = \exp(\phi)p.$$

Since the group G is a Lie group, there is a vector field X_g on M generating the action of the group element $g \in G$ as $L_g p = \exp(X_g)p$. The map $\Psi : \mathbb{R} \times M \to M$ given by

$$\Psi(s, p) = \exp(sX_g)p$$

defines a flow on M. Equivariance of h, which can be expressed as $h \circ L_g = L_g \circ h$, implies that

$$\exp(\phi)\exp(X_g) = \exp(X_g)\exp(\phi).$$

This is true if and only if the Lie bracket $[\phi, X_g]$ vanishes, or equivalently, the Lie derivative of ϕ along the flow generated by X_g does. A well-known result (see, e.g., Nakahara (2003)) is that this holds if and only if $\Psi(s, \Phi(t, p)) = \Phi(t, \Psi(s, p))$, from which it follows that $L_g\Phi(t, p) = \Phi(t, L_g p)$. This completes the proof. ■

Theorem 10 implies that the neural ODE defined by a G-equivariant vector field has a G-equivariant solution. An alternative formulation of this result, obtained directly from the connection between the flow Φ and the solution $u(t)$ to (3), is that $L_g u(t)$ is also a solution to (3) for every $g \in G$. This is the definition of equivariance occurring in the classical theory of (continuous) symmetries of differential equations.

3.3 Differential invariants and equivariant NODEs

In §2.2, we introduced the notion of differential invariants and saw how they can be useful when determining the most general ODE which is equivariant with respect to some group G. We will now show how this theory can be used to provide the most general parametrisation of a manifold NODE with symmetry group G. Combining Theorem 10 and Theorem 5 entails the following theorem.

Theorem 11 *Let G be a Lie group acting semi-regularly on $E = T \times M$ and let $I_1, \ldots, I_{\mu_k}$ be a complete set of functionally independent first-order differential invariants of order k. Then the most general G-equivariant NODE is (locally) on the form $H(I_1, \ldots, I_{\mu_1}) = 0$ where $H : \mathbb{R}^{\mu_1} \to \mathbb{R}^n$ is an arbitrary function with $n = \dim M$.*

In practice, it is often convenient to use the Implicit Function Theorem to rewrite $H(I_1, \ldots, I_{\mu_1}) = 0$ on the equivalent form (3) to obtain the most general form of the equivariant vector field. In this way the differential invariants of G can be used to parametrise the equivariant vector fields on M, generalising the construction by Katsman et al. (2021) and incorporating it into the geometric framework. We emphasise that there is no additional structural conditions on the function; equivariance is obtained through the use of the differential invariants.

Remark 12 *Below, we will also consider higher order differential invariants in the context of augmented equivariant NODEs. The computations of these invariants by solving $X^{(k)}(I) = 0$ become more extensive in the sense that higher order derivatives $u^{(k)}$ of the state enter, but remain tractable in the sense that $X^{(k)}(I)$ is always linear in the derivatives of the invariants.*

3.4 Application and examples of equivariant NODEs

To demonstrate the framework based on differential invariants we outline a general method for supervised learning in Algorithm 1 and consider its application to two examples where M is, respectively, Euclidean and non-Euclidean.

Algorithm 1: Supervised learning using equivariant manifold NODEs in the differential invariant framework

---

Input: A manifold M, a Lie group G with a semi-regular left action $G \times M \to M$ on M, a dataset $\{p_k, q_k\}_{k=1}^N$, $(p_k, q_k) \in M \times M$, and a loss function $l : M \times M \to \mathbb{R}$.

Goal: Learn a model $h : M \to M$ minimising $\sum_k l(h(p_k), q_k)$.

Part I: Geometric setting

1. Compute the induced vector fields X_i, $i = 1, \dots, d$.
2. Compute the first prolongations $X_i^{(1)}$, $i = 1, \dots, d$.
3. Compute the functionally independent first order differential invariants $I_1, \dots, I_{\mu_1}$ by solving $X_i^{(1)}(I) = 0$, $i = 1 \dots, d$.
4. Construct the most general NODE using Theorem 11 and infer the form of the corresponding vector field ϕ.

Part II: Supervised learning

5. Define NODE model $h : M \to M$ by $h(p) = u(1)$, $\dot{u}(t) = \phi_{u(t)}$, $u(0) = p$.
6. Select a machine learning model and use it to parametrise the unknown functions of the differential invariants $I_1, \dots, I_{\mu_1}$ appearing in ϕ.
7. Select a NODE training algorithm and train to minimise $\sum_k l(h(p_k), q_k)$.

Output: A trained model $h : M \to M$ defined by the learned ϕ.

We emphasise again that steps 1-4 only involve the manifold M and the symmetry group G, but not the dataset. Once the invariants are computed, they can be used to learn models h approximating the diffeomorphism corresponding to every possible dataset. Furthermore, once the most general equivariant NODE on M has been constructed, the training step can be performed using any available methodology (e.g., adjoint sensitivity methods or flow matching on manifolds).

Example 3 *We apply Algorithm 1 for the case $M = \mathbb{R}^2 \setminus \{0\}$ and $G = \mathrm{SO}(2)$, where a semi-regular left action is defined in Example* 1*. Steps 1-3 are performed in Example* 2*. In Step 4, we use that, by Theorem* 5*, the most general first-order ODE for which G is a symmetry group is of the form $H(t, r, r\dot{r}, r^2\dot{\theta}) = 0 \in \mathbb{R}^2$, where (r, θ) denote the polar coordinates in M. Consequently, $\dot{r}$ and $r^2\dot{\theta}$ can be treated as arbitrary functions of r and t; denote these functions by $\dot{r} = U(t, r)$ and $r^2\dot{\theta} = V(t, r)$. Switching to Cartesian coordinates (x, y), we can write the expressions for $\dot{x}$ and $\dot{y}$ explicitly. Defining $A(t, r) = U(t, r)/r$ and*

$B(t,r) = V(t,r)/r^2$ yields

$$\begin{cases} \dot{x} = A(t,r)x - B(t,r)y \\ \dot{y} = B(t,r)x + A(t,r)y \end{cases} . \tag{4}$$

More generally, this result is a consequence of the Implicit Function Theorem. The formulation (4) corresponds to Step 5 in Algorithm 1.

Thus, rotationally equivariant diffeomorphisms of M are obtained from the NODE model by parameterising the space of equivariant vector fields in terms of the functions A and B in (4) using, e.g., neural networks (Step 6 in Algorithm 1), and training amounts to learning approximations of A and B (Step 7 in Algorithm 1). This generalises the constructions appearing in the works of Chen et al. (2018); Lou et al. (2020); Katsman et al. (2021). Again, we emphasise that the neural networks are not required to be equivariant with respect to the symmetry group G; equivariance of the NODE model is obtained through the use of differential invariants to derive the system of ODEs in (4).

To illustrate how the above construction can be used in practice, we consider a rotationally equivariant diffeomorphism on the plane $\mathbb{R}^2$ that can be learned using a NODE of the form (4). Let $h : \mathbb{R}^2 \to \mathbb{R}^2$ be a unit translation in the radial direction, i.e.,

$$h(x,y) = \left(x + \frac{x}{r}, y + \frac{y}{r}\right) ,$$

for $x, y \in \mathbb{R}$ and $r = \sqrt{x^2+y^2}$. We let the data set consist of 64 equally spaced points on $[-2,2] \times [-2,2]$. The functions A and B are modelled using feed-forward neural networks consisting of four linear layers. The neural architecture is given by

```
L(1 → 16) + ReLU → L(16 → 32) + ReLU → L(32 → 16) + ReLU → L(16 → 1),
```

where `L` *denotes a linear (fully connected) layer. No activation is added after the final layer.*

The NODE (4) is defined and trained using the `torchdiffeq` *package (Chen et al., 2018), using the Adam optimiser with mean squared error (MSE) as the loss criterion. All hyperparameters are set to their default values.*

The functions A and B in (4) are plotted as functions of r in Fig. 1 after 300 epochs of training. A flow corresponding to the target diffeomorphism h can be analytically obtained as $(x(t), y(t)) = (x(0) + t\,x(0)/r, y(0) + t\,y(0)/r)$ with tangent vector $(\dot{x}(t), \dot{y}(t)) = (x(0)/r, y(0)/r)$. We observe in Fig. 1 that $A(r)$ approaches the function $a(r) = 1/r$ and $B(r)$ approaches the function $b(r) = 0$ during training, which corresponds to the tangent vector $(x(0)/r, y(0)/r)$.

Remark 13 *In our framework, $I = t$ is always an invariant since we consider the case where G acts trivially on T. In our examples, we consider autonomous NODE systems, in which case the functions A and B are independent of t.*

Remark 14 *In the example above, rotational equivariance is conveniently expressed in terms of the polar coordinates (r, θ). For any one-dimensional symmetry group G we can similarly find canonical coordinates on M defined by G acting by translation in one coordinate and trivially in all others. For higher-dimensional symmetry groups, however, this is not generally possible.*

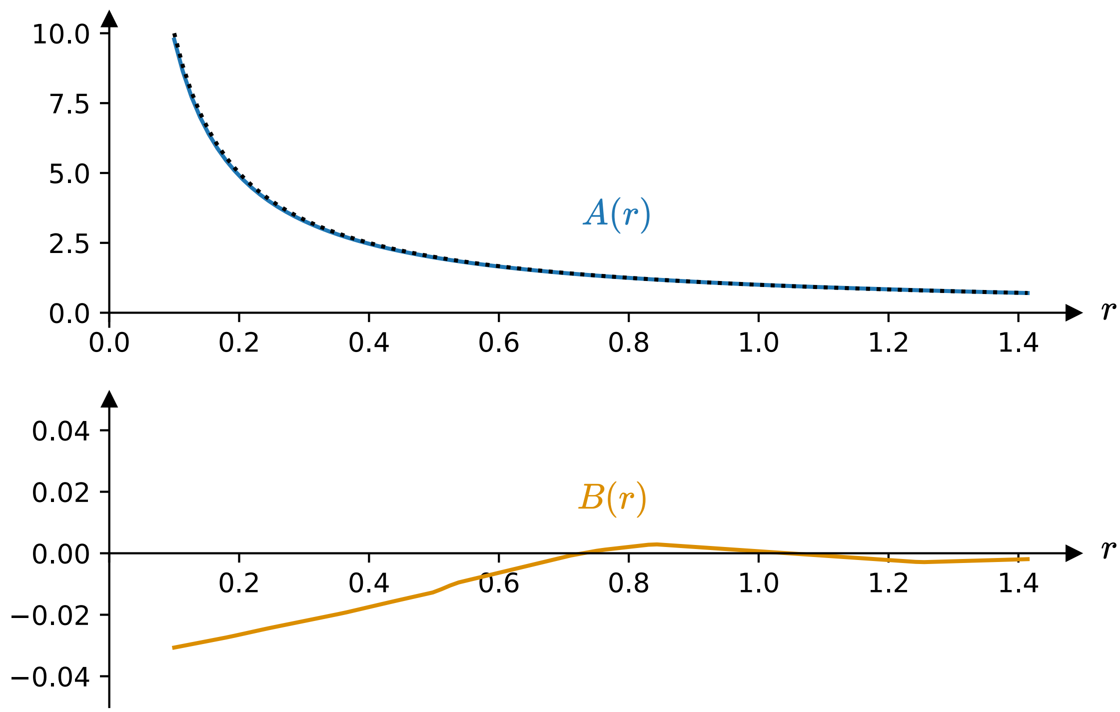


Figure 1: The figure shows $A(r)$ (blue line) and $B(r)$ (orange line) in Example 3 as functions of r after 300 epochs. The dashed line in the top figure illustrates the function $a(r) = 1/r$.

Example 4 *We now consider a non-Euclidean example where $G = \mathrm{SO}(2)$ acts on the sphere $M = S^2$ through rotation around the z-axis. If θ is the polar angle and φ is the azimuthal angle, the action of $\mathrm{SO}(2)$ can, in spherical coordinates, be given by*

$$L_\epsilon(\theta, \varphi) = (\theta, \varphi + \epsilon),$$

where ϵ is the rotation angle. The action is generated by the vector field X, and its prolongation $X^{(1)}$ is given by

$$X^{(1)} = X = \partial_\varphi.$$

Since $J^{(1)}S^2$ is five-dimensional and $\mathrm{SO}(2)$ has dimension one, there are four first-order functionally independent differential invariants, according to Theorem 4, which are readily computed as

$$I_1 = t, \quad I_2 = \theta, \quad I_3 = \dot{\theta}, \quad \text{and} \quad I_4 = \dot{\varphi}.$$

Theorem 11 implies that the most general ODE is given by

$$\begin{cases} \dot{\theta} &= A(t, \theta) \\ \dot{\varphi} &= B(t, \theta) \end{cases}, \tag{5}$$

where A and B are arbitrary functions.

A rotationally equivariant map on S^2 *can be obtained through scaling of the polar angle and shifting of the azimuthal angle, i.e.,*

$$h : (\theta, \varphi) \mapsto (\theta e^{\epsilon}, \varphi + \nu)$$

for some real numbers ϵ *and* ν*, which we take to be* $\epsilon = 1$ *and* $\nu = 0.05$*. We define a NODE of the form* (5) *using neural networks and train it for a set of 400 randomly distributed points on the sphere and their images under the map* h*. The neural architecture, optimiser, loss criterion, and hyperparameters are the same as in Example* 3*.*

In Fig. 2*, we have plotted the resulting* A *and* B *as functions of* θ *after 400 epochs of training. We see that* $A(\theta)$ *is approximately equal to a function* $a(\theta) = \theta$ *and* $B(\theta)$ *is approximately equal to a function* $b(\theta) = 0.05$*. This corresponds to the flow* $(\theta(t), \varphi(t)) = (\theta(0)\, e^t, \varphi(0) + 0.05\, t)$*, with tangent vector* $(\dot{\theta}(t), \dot{\varphi}(t)) = (\theta(t), 0.05)$*, into which the diffeomorphism* h *can be embedded.*

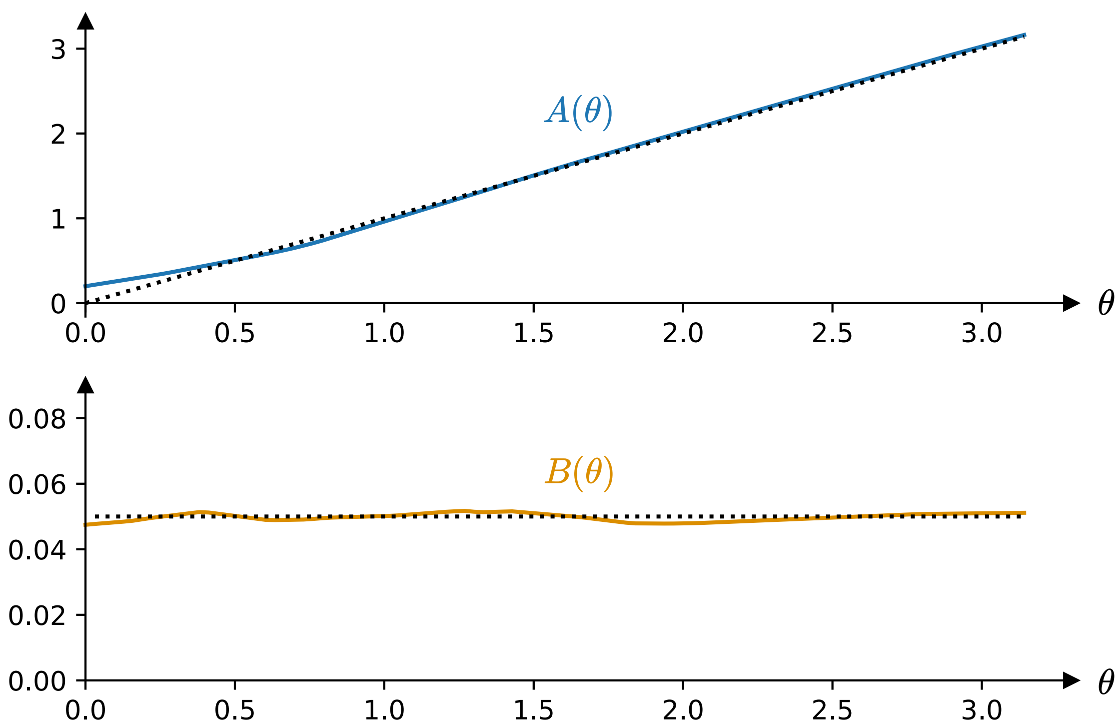


Figure 2: A and B in Example 4 as functions of θ after 400 epochs of training. The dashed lines represent the functions $a(\theta) = \theta$ and $b(\theta) = 0.05$, respectively.

We have now seen two examples where equivariant manifold NODEs can be trained to approximate equivariant diffeomorphisms $h : M \to M$. Next, we turn to the question of whether this is possible for any diffeomorphism on any manifold.

4. Augmentation and universality

A well-known issue with neural ODEs is their inability to learn certain classes of diffeomorphisms. Because the trajectories of the ODE solutions can not intersect, not all diffeomorphisms $h : M \to M$ can be obtained from integral curves on M. The canonical example in the Euclidean setting $M = \mathbb{R}^n$ is the diffeomorphism $h(x) = -x$, which cannot be represented by a NODE. The resolution of this issue for the case $M = \mathbb{R}^n$ was discussed by Dupont et al. (2019), who introduced augmented neural ODEs, and by Zhang et al. (2020), who proved that NODEs on the augmented space $\mathbb{R}^{2n}$ are universal approximators of diffeomorphisms $h : \mathbb{R}^n \to \mathbb{R}^n$. Heuristically, augmenting the state space of the ODE amounts to introducing enough extra dimensions to resolve intersections of the integral curves.

Augmentation of NODEs is equivalent to the problem of embedding the diffeomorphisms $h : M \to M$ in a flow, i.e., the integral curve of an ODE on some ambient space. Utz (1981) showed that any h can be embedded in a flow on an ambient twisted cylinder of dimension $n+1$ obtained as a fibration over M. However, the properties of the ambient space, e.g., the topological class of the fibration, depend on the diffeomorphism. As a consequence, in the setting where the objective is learning h, we must instead consider an embedding into an ambient space that is common for all h.

We first consider augmentation of manifold NODEs for the non-equivariant case, corresponding to a trivial symmetry group G, to establish universality of manifold NODEs on connected M. We then show that the construction is equivariant with respect to the action of G which implies that equivariant manifold NODEs are universal approximators of equivariant diffeomorphisms $h : M \to M$. The geometric perspective offered by our framework is essential to the construction and its subsequent generalisation to include non-trivial symmetry groups G.

4.1 Augmented manifold NODEs

A natural candidate for an ambient space that allows us to resolve intersections of integral curves in the case of manifold NODEs is the tangent bundle TM of the manifold M. The following theorem shows that it is possible to embed any diffeomorphism on M into a flow on the tangent bundle if M is connected, and provides the theoretical basis for augmentation of manifold NODEs.

Theorem 15 *Let M be a connected, smooth manifold and $h : M \to M$ be a diffeomorphism. Then h can be embedded in a flow on TM.*

Proof Because M is connected it is also path-connected, so for every $p \in M$ there exists a path $\gamma_p : [0,1] \to M$ such that $\gamma_p(0) = p$, $\gamma_p(1) = h(p)$. Let Γ_p be the lift of γ_p to TM, defined by

$$\Gamma_p(t) = \left(\gamma_p(t), \frac{d}{dt}\gamma_p(t) \right) . \tag{6}$$

The obstruction to embedding γ_p into a flow on M is the presence of intersections between the paths γ_p, i.e., points where $\gamma_p(\tau) = \gamma_{p'}(\tau')$ for some $\tau, \tau' \in [0,1]$. The paths γ_p and $\gamma_{p'}$ can always be smoothly deformed (by the General Transversality Theorem) to intersect only in isolated points $\tau, \tau' \in [0,1]$ so that

$$\frac{d}{dt}\,\gamma_p(t)\big|_{t=\tau} \neq \frac{d}{dt}\,\gamma_{p'}(t)\big|_{t=\tau'}\,.$$

The corresponding lifts Γ_p and $\Gamma_{p'}$, by construction, never intersect in TM even where γ_p and $\gamma_{p'}$ intersect in M. Consequently, the gradients of Γ_p define a vector field on TM and a corresponding flow $\Phi : \mathbb{R} \times TM \to TM$ into which h is embedded as

$$h(p) = \pi\Phi\left(1, p, \frac{d}{dt}\gamma_p(t)\bigg|_{t=0}\right)\,, \tag{7}$$

where $\pi : TM \to M$ is the tangent bundle projection. ■

Theorem 15 shows that it is possible to augment manifold NODEs by embedding the diffeomorphism h into a flow on TM. The augmented NODE model corresponding to this embedding is the Cauchy problem describing the evolution of a state $U : \mathbb{R} \to TM$, $U(t) = (u(t), \dot{u}(t))$, generated by a vector field $\phi = (\chi, \psi)$ on TM according to

$$\dot{U}(t) = \phi_{U(t)}\,, \quad U(0) = (u(0), \dot{u}(0))\,, \tag{8}$$

where we require that $\chi(u, \dot{u}) = \dot{u}$, so that $\phi_{(u,\dot{u})} = (\dot{u}, \psi(u, \dot{u}))$, in order for ϕ to be consistent with tangent vectors to a lift $\dot{U}(t) = (\dot{u}(t), \ddot{u}(t))$. The diffeomorphism $h : M \to M$ described by the augmented NODE model is obtained by $p = u(0)$, $h(p) = u(1) = \pi U(1)$, and we may take $\dot{u}(0) = 0$ as a convenient initial condition for $\dot{u}$. By Theorem 15, the resulting augmented manifold NODE can approximate any diffeomorphism on M.

Corollary 16 *Augmented manifold NODEs are universal approximators of diffeomorphisms* $h : M \to M$.

This result generalises the construction by Zhang et al. (2020) to arbitrary manifolds M and provides an explanation of the geometric origin of the augmentation. The idea behind the augmented construction is illustrated in Fig. 3.

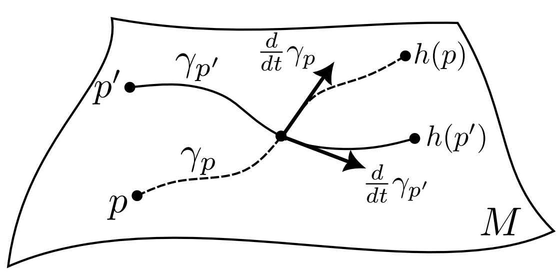


Figure 3: Intersection of paths γ_p and $\gamma_{p'}$.

We note that the augmented manifold NODE is equivalent to the second-order ODE $\ddot{u}(t) = \psi(u, \dot{u})$ on M, a fact that will prove useful when we now proceed to consider augmenting equivariant manifold NODEs.

Remark 17 *The lift Γ_p can alternatively be described in terms of a section of the first jet bundle $J^{(1)}$. Specifically, the condition $\chi(u, \dot{u}) = \dot{u}$ corresponds to restricting the vector field ϕ to $J^{(1)}$. However, for the purpose of describing flows on M, and the subsequent construction of augmented NODEs, the formulation in terms of the tangent bundle TM and the lift (6) is more convenient.*

Remark 18 *Connectedness of M is a strict limitation in any flow-based model, since integral curves can never span disconnected components. An equivalent statement to the one made above is that manifold NODEs are universal approximators for diffeomorphisms restricted to the individual connected components of M.*

4.2 Augmented equivariant manifold NODEs

In order to extend augmentation to equivariant manifold NODEs we first show that the construction in Theorem 15 is equivariant with respect to the action of G on TM induced by the action on M.

Theorem 19 *Let G be a Lie group acting on a connected, smooth manifold M and let $h : M \to M$ be an equivariant diffeomorphism satisfying $h(L_g p) = L_g h(p)$ for every $p \in M$ and $g \in G$. Then h can be embedded in a G-equivariant flow on TM.*

Proof Let $p \in M$, γ_p be a path from p to $h(p)$ and Γ_p be the lift (6). For every $L_g p$ in the orbit of p we construct the corresponding path as $\gamma_{L_g p} = L_g \gamma_p$, which satisfies $\gamma_{L_g p}(0) = L_g p$ and $\gamma_{L_g p}(1) = L_g h(p) = h(L_g p)$, as required by the equivariance of h. The action of G on M induces an action on the lift in TM as

$$L_g \Gamma_p(t) = \left(L_g \gamma_p(t), (L_g)_* \frac{d}{dt} \gamma_p(t) \right) . \tag{9}$$

From the definition of $\gamma_{L_g p}$, we have

$$(L_g)_* \frac{d}{dt} \gamma_p(t) = \frac{d}{dt} L_g \gamma_p(t) = \frac{d}{dt} \gamma_{L_g p}(t) \, ,$$

which implies that $L_g \Gamma_p = \Gamma_{L_g p}$.

The vector tangent to the lift Γ_p in TM is

$$\frac{d}{dt} \Gamma_p(t) = \left(\frac{d}{dt} \gamma_p, \frac{d^2}{dt^2} \gamma_p \right) .$$

Considering the induced action of G in (9) and its push-forward, we have

$$(L_g)_* \frac{d}{dt} \Gamma_p = \left(\frac{d}{dt} \gamma_{L_g p}, \frac{d^2}{dt^2} \gamma_{L_g p} \right) = \frac{d}{dt} \Gamma_{L_g p} \, . \tag{10}$$

Consequently, by Theorem 10, the flow $\Phi : \mathbb{R} \times TM \to TM$ generated by the gradients of Γ_p is equivariant, or equivalently, compatible with the induced action of G on the lift Γ_p.

Finally, since $\pi \circ L_g = L_g \circ \pi$ by construction, the embedding of h in (7) preserves equivariance:

$$\begin{aligned} h(L_g p) &= \pi\Phi\left(1, L_g p, \frac{d}{dt}\,\gamma_{L_g p}(t)\big|_{t=0}\right) \\ &= \pi L_g \Phi\left(1, p, \frac{d}{dt}\,\gamma_p(t)\big|_{t=0}\right) = L_g h(p)\,. \end{aligned}$$

■

The lift Γ_p and its tangent vector are illustrated in Fig. 4.

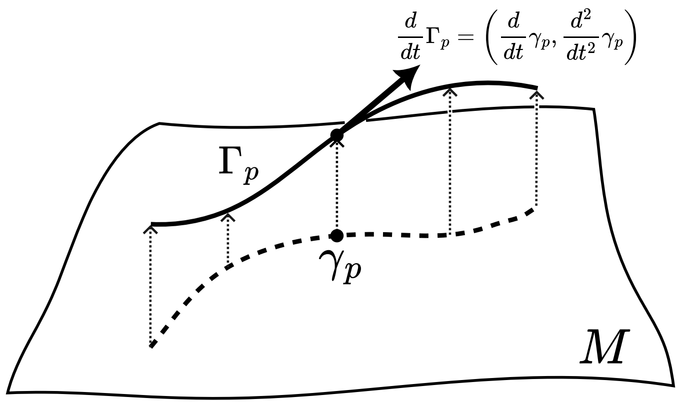


Figure 4: The lift Γ_p of γ_p and the vector $\frac{d}{dt}\Gamma_p$ tangent to TM.

Theorem 19 guarantees that any equivariant diffeomorphism $h : M \to M$ can be embedded in an equivariant flow on TM, and that augmentation can be extended to equivariant manifold NODEs. To incorporate equivariance under the induced action of G on TM into the augmented manifold NODE in (8), we require that the vector field ϕ is equivariant. Theorem 10 then ensures that the solution to (8) is an equivariant flow $\Phi : \mathbb{R} \times TM \to TM$, which can approximate any equivariant diffeomorphism $h : M \to M$ by the embedding in Theorem 19.

Corollary 20 *Augmented G-equivariant manifold NODEs are universal approximators of G-equivariant diffeomorphisms $h : M \to M$.*

From the observation that the augmented manifold NODE is equivalent to the second-order ODE $\ddot{u}(t) = \psi(u, \dot{u})$ on M, it follows that differential invariants of G can be used to parametrise the equivariant vector fields $\phi = (\dot{u}, \psi(u, \dot{u}))$ on TM tangent to lifts $U(t)$, analogously to Theorem 11. From our construction, and Theorem 10, it is clear that the solution $u(t)$ is G-equivariant, in the sense that $L_g u(t)$ is also a solution to $\ddot{u} = \psi(u, \dot{u})$, if and only if the flow Φ is G-equivariant. Consequently, the most general form of an augmented equivariant NODE can be deduced from Theorem 5.

Theorem 21 *Let G be a Lie group acting semi-regularly on $E = T \times M$ and let $I_1, \ldots, I_{\mu_2}$ be a complete set of functionally independent second-order differential invariants. Then, the most* *general augmented G-equivariant NODE on M is (locally) of the form $H(I_1, \ldots, I_{\mu_2}) = 0$, where $H : \mathbb{R}^{\mu_2} \to \mathbb{R}^n$ is an arbitrary function with $n = \dim M$.*

In this way, the second-order differential invariants parametrise equivariant vector fields restricted to lifts, and equivalently the space of augmented equivariant manifold NODEs.

Remark 22 *The geometric objects appearing in the equivariant augmentation can, as before, be understood in terms of jet bundles. Enforcing equivariance of ϕ corresponds to prolonging the action of G to sections of the second jet bundle $J^{(2)}$, which are equivalent to lifts of Γ_p to the tangent bundle of TM.*

4.3 Application and examples of augmented equivariant NODEs

The general method outlined in Algorithm 1 can be modified to incorporate augmentation, resulting in Algorithm 2. To illustrate the above construction of equivariant augmentation and its universality properties, we return to the setting of Example 3 and consider the problem of learning a topologically non-trivial equivariant diffeomorphism.

Example 5 *Let M and G be as in Example 3 and consider the augmented NODE $\dot{U}(t) = \phi_{U(t)}$, where $U(t) = (x(t), y(t), \dot{x}(t), \dot{y}(t))$. To parametrise the space of such models, equivariant under G, we consider second-order differential invariants of G. These are obtained from the second prolongation of the vector field X, which is given by*

$$X^{(2)} = -y\partial_x + x\partial_y - \dot{y}\partial_{\dot{x}} + \dot{x}\partial_{\dot{y}} - \ddot{y}\partial_{\ddot{x}} + \ddot{x}\partial_{\ddot{y}}.$$

By solving the infinitesimal condition $X^{(2)}(I) = 0$, we obtain the first-order invariants $I_1, \ldots, I_4$ in Example 3 and the additional second-order invariants[2]

$$I_5 = r^3\ddot{r}\,, \quad I_6 = r^4\ddot{\theta}.$$

By Theorem 5, any second-order G-equivariant ODE on M can be written as

$$H(t, r, r\dot{r}, r^2\dot{\theta}, \ddot{r}, \ddot{\theta}) = 0 \in \mathbb{R}^2$$

for some arbitrary function H. The equivalent expression in terms of $\ddot{u} = \psi(u, \dot{u})$ is given by

$$\begin{cases} \ddot{x} = A(t, r, r\dot{r}, r^2\dot{\theta})\dot{x} - B(t, r, r\dot{r}, r^2\dot{\theta})\dot{y} \\ \ddot{y} = B(t, r, r\dot{r}, r^2\dot{\theta})\dot{x} + A(t, r, r\dot{r}, r^2\dot{\theta})\dot{y} \end{cases},$$

where A and B are two arbitrary functions. Note that $\ddot{u}(t) = (\ddot{x}(t), \ddot{y}(t))$ depends non-trivially on both u and $\dot{u}$ through A and B. In terms of (8), this implies that the most general vector field ϕ in (8) restricted to lifts in TM is

$$\begin{aligned} \phi_{U(t)} = \Big(&\dot{x}(t), \dot{y}(t), A(t, r, r\dot{r}, r^2\dot{\theta})\dot{x}(t) - B(t, r, r\dot{r}, r^2\dot{\theta})\dot{y}(t), \\ &B(t, r, r\dot{r}, r^2\dot{\theta})\dot{x}(t) + A(t, r, r\dot{r}, r^2\dot{\theta})\dot{y}(t)\Big). \end{aligned} \tag{11}$$

2. Explicit expressions for I_5 and I_6 in terms of the Cartesian coordinates (x, y) are suppressed for the sake of brevity.

Algorithm 2: Supervised learning using augmented equivariant manifold NODEs in the differential invariant framework

Input: A manifold M, a Lie group G with a left action $G \times M \to M$ on M, a dataset $\{p_k, q_k\}_{k=1}^N$, $(p_k, q_k) \in M \times M$, and a loss function $l : M \times M \to \mathbb{R}$.

Goal: Learn a model $h : M \to M$ minimising $\sum_k l(h(p_k), q_k)$.

Part I: Geometric setting

1. Compute the induced vector fields X_i, $i = 1, \ldots, d$.
2. Compute the second prolongations $X_i^{(2)}$, $i = 1, \ldots, d$.
3. Compute the functionally independent second order differential invariants $I_1, \ldots, I_{\mu_2}$ by solving $X_i^{(2)}(I) = 0$, $i = 1 \ldots, d$.
4. Construct the most general augmented NODE using Theorem 21 and infer the form of the corresponding augmented vector field $\phi = (\dot{u}, \psi(u, \dot{u}))$ on TM in (8).

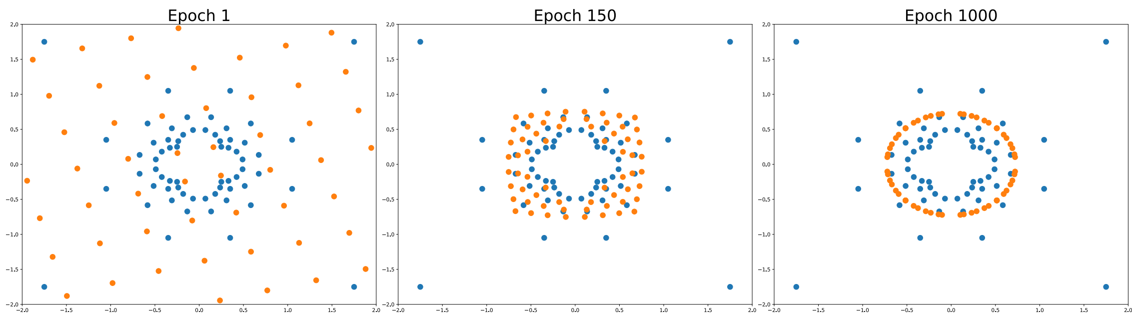
Part II: Supervised learning

5. Augment the dataset to $((p_k, 0), (q_k, 0)) \in TM \times TM$, $k = 1, \ldots, N$.
6. Define NODE model $h : M \to M$ by $h(p) = \pi U(1)$, $\dot{U}(t) = \phi_{U(t)}$, $U(0) = (p, 0)$.
7. Select a machine learning model and use it to parametrise the unknown functions of the differential invariants $I_1, \ldots, I_{\mu_2}$ appearing in ϕ.
8. Select a NODE training algorithm and train to minimise $\sum_k l(h(p_k), q_k)$.

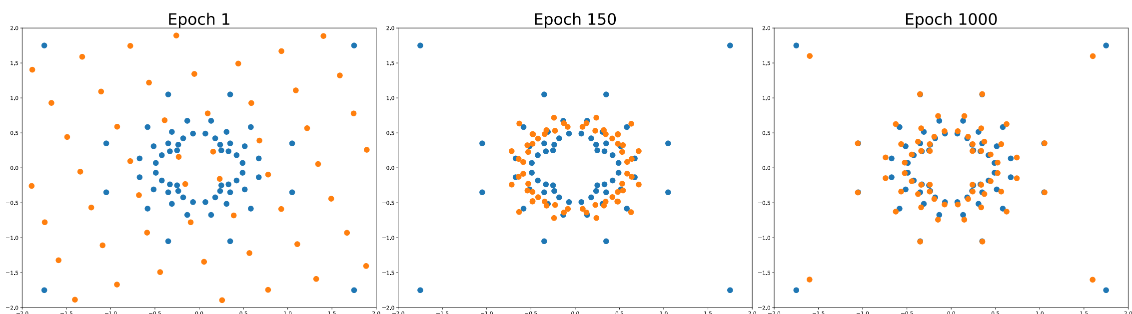
Output: A trained model $h : M \to M$ defined by the learned ϕ.

This is our desired parameterisation of equivariant vector fields in the augmented NODE, where A and B can again be approximated using feed-forward neural networks and the output is obtained by the projection π onto the first two components of $U(t)$. The neural architecture given for A and B in Example 3 *needs to be altered to have input and output dimension 2. We use the same optimiser, loss criterion and hyperparameters as in Examples 3 and 4.*

An example of rotationally equivariant diffeomorphism that cannot be approximated by a non-augmented NODE is the mapping $h : (r, \theta) \mapsto (1/r, \theta)$. We let the data set consist of 100 equally spaced points on $[-2, 2] \times [-2, 2]$ and use NODE models of the forms (4) *and* (11) *and train them to map the input points to their respective images under h. The performance of the non-augmented NODE and augmented NODE are compared in Fig.* 5*. In the non-augmented case, the best performance is obtained when the points are transformed to a circle minimising the total distance to the target points. Crossing this circle is not possible due to the limitation of flow lines not being able to intersect. The augmented case, however, performs much better, with the output points almost overlapping the target points after* 1000 *epochs of training.*



(a) Non-augmented case.



(b) Augmented case.

Figure 5: In Example 5, we train a non-augmented NODE model and an augmented NODE model to approximate the equivariant diffeomorphism $h(r, \theta) = (1/r, \theta)$. The orange dots represent the resulting outputs by the models, while the blue dots represent the target values. In (a), the neural ODE has not been augmented. We see that, as a result of the solution curves not being able to cross, the outputs freeze on a circle minimising the distance to the targets after around 1000 epochs. In (b), the NODE model has been augmented to the tangent bundle. As a result, the outputs are able to better approximate the targets.

To further illustrate the differences between the non-augmented and augmented model, we plot the flow lines from $t = 0$ *to* $t = 1$ *for sample points with the same angular coordinate* θ*. Fig. 6 shows the flow lines* $r(t)$ *from* $t = 0$ *to* $t = 1$ *in the non-augmented case after 1000 epochs of training. We see that the curves appear to converge to a specific value of* r *(around* $r = 0.8$*), not intersecting. In Fig. 7, we show the flow lines* $r(t)$ *and* $\dot{r}(t)$ *in the augmented case after 1000 epochs of training. We see that an intersection occurs in the left plot in Fig. 7, while there are no intersections in the plot on the right-hand side.*

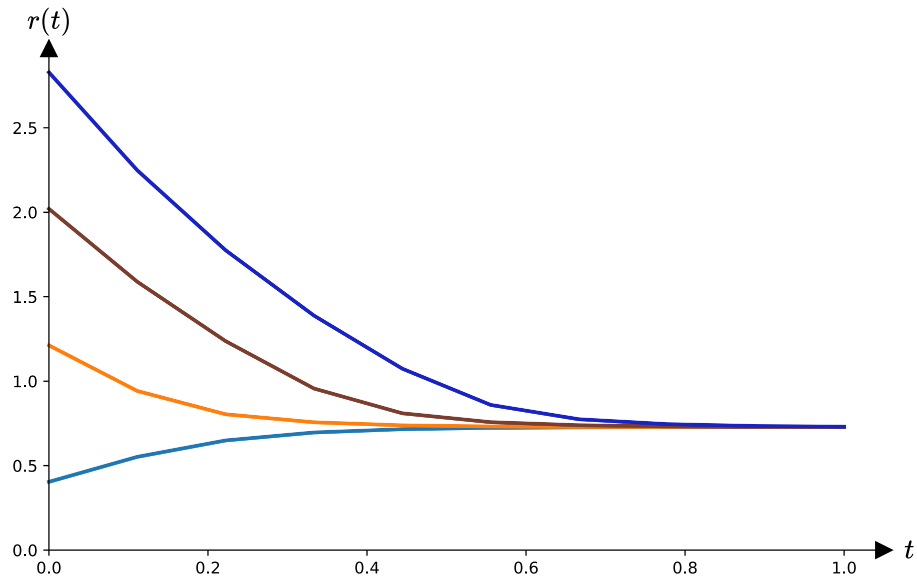


Figure 6: Flow lines of the non-augmented NODE model in Example 5 after 1000 epochs at $\theta \approx 0.25\pi$. The curves converge to a specific value of r, but do not cross.

This example illustrates the increased performance obtained using an augmented neural ODE to approximate topologically non-trivial diffeomorphisms. Fig. 7 shows that lifting the solution curves to the tangent bundle TM *allows for intersections on the base space* M*, providing a more accurate solution.*

5. Equivariant fields and densities

Having constructed the geometric framework for equivariant manifold NODEs and their augmentations, we now discuss how they are used to construct models acting on different objects on M. More specifically, this amounts to describing the induced action of the diffeomorphism $h : M \to M$ on such objects, and their transformation properties under the action of G. We consider functions, densities and vector fields, but the principle can be applied to any kind of geometrical object. The resulting models allow learning, e.g., vector valued quantities on M, in analogy with steerable group equivariant CNNs (see Gerken et al. (2023)).

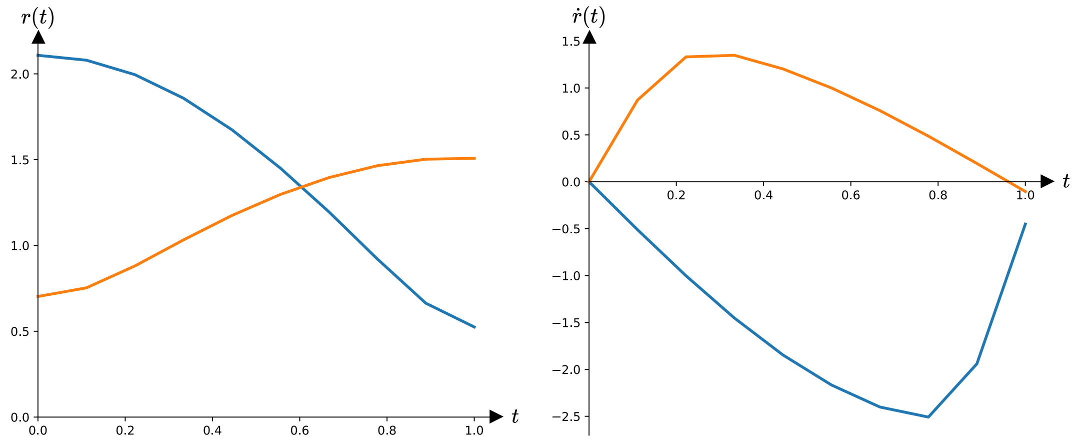


Figure 7: Flow lines r and $\dot{r}$ of the augmented NODE model in Example 5 at $\theta \approx 0.6\pi$. The model has been trained for 1000 epochs. Intersection in the left figure is allowed since the paths do not simultaneously cross in the figure to the right, which means that no intersections occur in the tangent bundle TM.

5.1 Equivariant functions and densities

For a function $f : M \to \mathbb{R}$, or scalar field, the map induced by the diffeomorphism $h : M \to M$ is simply $f_h = f \circ h^{-1}$. Symmetry of a scalar function f_h amounts to invariance, meaning $f(L_g p) = f(p)$. That equivariance of h and f implies equivariance of f_h can be shown in one line as

$$f_h(L_g p) = f(h^{-1}(L_g p)) = f(L_g h^{-1}(p)) = f_h(p).$$

Next, we consider a (probability) density ρ on M, in which case the induced action of h is obtained as

$$\rho_h(p) = \rho\left(h^{-1}(p)\right) |\det J_{h^{-1}}(p)| \,, \tag{12}$$

where the Jacobian determinant accounts for the induced action on the volume element on M.

It has been shown by Köhler et al. (2020); Katsman et al. (2021) that, given a density ρ, invariant under the isometry group G, and a diffeomorphism $h : M \to M$ equivariant under G (or some subgroup $H < G$), the transformed density ρ_h is also G-invariant (or H-invariant). This makes it possible to construct continuous normalizing flows which preserve the symmetries of the densities.

The following theorem generalises the results on invariant densities by Köhler et al. (2020); Katsman et al. (2021) to the case of arbitrary groups acting on M, which means that the Jacobian determinant corresponding to the change in the volume element is not necessarily unity. Equivariance under G in the context of densities amounts to preservation of the probability associated with the infinitesimal volume element. Consequently, equivariance of ρ is given by the following definition.

Definition 23 *The density ρ on M is equivariant if*

$$\rho(p) = \rho(L_g^{-1} p) \, |\det J_{L_g^{-1}}(p)|$$

for every $g \in G$ and $p \in M$.

Theorem 24 *Let ρ be a G-equivariant density on M. If $h : M \to M$ is a G-equivariant diffeomorphism, then the induced density ρ_h is G-equivariant.*

Proof Equivariance of h means $L_g \circ h = h \circ L_g$ and it is straightforward to show that h^{-1} is also G-equivariant. From this, and the equivariance of ρ, it follows that

$$\rho(h^{-1}(p)) = \rho(h^{-1}L_g^{-1}(p))\, |\det J_{L_g^{-1}}(h^{-1}(p))|. \tag{13}$$

By a rearrangement of (12),

$$\rho(h^{-1}L_g^{-1}(p)) = \rho_h(L_g^{-1}p)\frac{1}{|\det J_{h^{-1}}(L_g^{-1}p)|}. \tag{14}$$

Finally, we use (12), combine (13) and (14), and use the chain rule to obtain

$$\rho_h(p) = \rho_h(L_g^{-1}p)\frac{|\det J_{L_g^{-1}\circ h^{-1}}(p)|\,|\det J_{L_g^{-1}}(p)|}{|\det J_{h^{-1}\circ L_g^{-1}}(p)|} = \rho_h(L_g^{-1}p)\,|\det J_{L_g^{-1}}(p)|.$$

Thus, $\rho_h(p) = \rho_h(L_g^{-1}p)\,|\det J_{L_g^{-1}}(p)|$ for every $g \in G$ and $p \in M$, and consequently ρ_h is equivariant. ■

5.2 Equivariant vector fields

Next, we consider a vector field $V : M \to TM$ and the induced action of the diffeomorphism $h : M \to M$ obtained as the push-forward of V along h. At the point $p \in M$, the transformed vector field V_h is given by

$$V_h(p)[f] = V(h^{-1}(p))[f \circ h],$$

where f is an arbitrary differentiable function defined on a neighbourhood of p. Note that, for clarity, we use $V(p)$ to denote the evaluation of V at p in this section. Equivariance of a vector field V amounts to left invariance under G in the usual sense, i.e., $V(L_g p) = (L_g)_* V(p)$ for every $g \in G$ and $p \in M$, see Definition 2. Similarly to the objects previously considered, the vector field V_h inherits the equivariance of V and h.

Theorem 25 *Let V be a G-equivariant vector field on M, and $h : M \to M$ be G-equivariant diffeomorphism. Then V_h is G-equivariant.*

Proof We want to show that $V_h(L_g p) = (L_g)_*(V_h(p))$ for all $g \in G$ and $p \in M$. Let f be an arbitrary real function defined on a neighbourhood of $L_g p$. It follows by the definition of the push-forward and the equivariance of h^{-1} that

$$V_h(L_g p)[f] = V(L_g(h^{-1}(p)))[f \circ h].$$

By the equivariance of V and the definition of the push-forward, this in turn equals $(L_g)_* V_h(p)[f]$. Since f is arbitrary, $V_h(L_g p) = (L_g)_* V_h(p)$ for all $p \in M$ and $g \in G$, which completes the proof. ■

To illustrate the principle of the induced action and the way equivariance of $h : M \to M$ preserves the symmetry of the object being transformed, we visualise the different quantities appearing in Theorem 25 in Fig. 8.

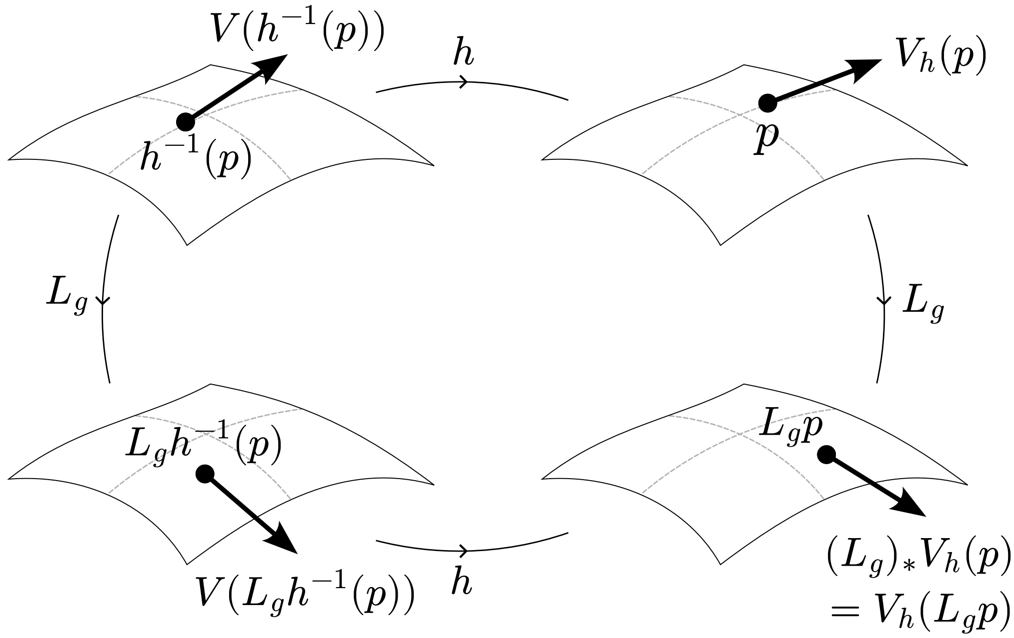


Figure 8: Illustration of the construction in Theorem 25. The vector field $V : M \to TM$ and the diffeomorphism $h : M \to M$ are assumed to be G-equivariant.

The results in this section allow for the construction of G-equivariant models based on NODEs for the different kinds of geometric objects on M considered. The approach uses the diffeomorphism $h : M \to M$ to induce a transformation, which defines the action of the model and is possible to generalise to other types of differential geometric objects. In particular, the induced action on vectors generalises to tensor fields on M.

6. Conclusion

In this paper, we develop a novel geometric framework for equivariant manifold neural ODEs based on the classical theory of symmetries of differential equations. In particular, for any smooth, connected manifold M and connected Lie symmetry group G acting semi-regularly on M, we establish the equivalence of the different notions of equivariance related to the Cauchy problem and show how the space of equivariant NODEs can be efficiently parametrised using the differential invariants of G. This generalises the idea of Katsman et al. (2021) to use gradients of invariant functions by exploiting the fundamental role of differential invariants in the theory of symmetries of differential equations.

Subsequently, we show how the manifold M, and the neural ODEs, can be augmented to obtain models which are universal approximators of diffeomorphisms $h : M \to M$.

Our construction extends previous work in the Euclidean case (Zhang et al., 2020) to the manifold setting, and provides the appropriate geometric framework to establish universality of augmented manifold NODE models, meaning that they are capable of learning any diffeomorphism $h : M \to M$. Furthermore, we show that the augmentation is equivariant with respect to a non-trivial symmetry group G acting on the manifold, and that universality persists, meaning that augmented equivariant manifold NODEs are universal approximators of equivariant diffeomorphisms $h : M \to M$. In the equivariant case, we also show how the augmentation fits in our geometric framework by parameterising the space of augmented equivariant manifold NODEs using second-order differential invariants.

Finally, following previous works on manifold NODEs, we discuss how different kinds of objects can be modelled using the induced action of h on different types of fields on M. As a prominent example of learning fields beyond densities, we consider vector fields, which are found in many physical applications. Our framework could, for example, be used for learning vector fields on the sphere, which has applications, e.g., in omnidirectional vision (Coors et al., 2018), weather forecasting (Linander et al., 2025) and cosmological applications (Defferrard et al., 2020).

In our formulation, all calculations involving the group G are concerned with determining the differential invariants $I_1, \ldots, I_\mu$, which is analytically tractable due to the fact that $X^{(k)}$ is a linear operator. There are several computer algebra software packages dedicated to symmetries of differential equations, which can be used to perform these calculations (see Hereman (1997) for a review of packages implemented in various common commercial software programs). The invariants are computed prior to implementation and training and require no discretisation of the group G. Compared to group equivariant CNNs, integration over G or G/H in the convolutional layers is replaced by computing derivatives on $J^{(1)}$ in our framework (see Olver (1993, Prop. 2.53)).

It is interesting to note that the notion of symmetry transformations of differential equations typically only requires a local action of G, as opposed to the global action on homogeneous spaces $M = G/H$ emphasised by Kondor and Trivedi (2018). In particular, in contrast to the group equivariant networks, there is no integration over the group G. In fact, the only computations using the group G required during training and inference are straightforward function evaluations of the invariants. Since the invariants only depend on the manifold and symmetry group, they only need to be computed once for each combination of M and G and can then be applied in NODE models of any dataset with the corresponding differentiable structure and symmetry group.

The structure of the invariants $I_1, \ldots, I_\mu$ can be used to impose further constraints on the model by excluding a subset of the I_k in the construction of the ODE $\dot{u}(t) = \phi(I_1, \ldots, I_\mu)$. This corresponds to enlarging the symmetry group, and the constraints are 'physically' meaningful since they are invariant under the action of G. A very interesting direction for future work would be to investigate this approach systematically and, more generally, to explore the connections between invariants in our geometric framework and conservation laws in the physical sciences (Greydanus et al., 2019; Lutter et al., 2019; Toth et al., 2020; Cranmer et al., 2020).

An interesting aspect, which we have not explored, is related to the approximation properties of the neural networks (or more generally machine learning models) that are obtained by discretising NODEs. Universality of NODEs in the sense of Corollaries 16 and 20 guarantees the existence of a flow into which an arbitrary diffeomorphism can be embedded, and Theorems 11 and 21 show that this flow can be described using the differential invariants. However, the implications for the approximation capabilities of, e.g., the residual networks obtained (He et al., 2016; Chen et al., 2018) by a finite discretisation of the NODE using a finite width neural network to parametrise ϕ are not obvious. Investigating such properties for equivariant models on manifolds would be an interesting application of the geometric framework presented here.

To understand the merits and limitations compared to other network architectures, the performance that can be obtained by equivariant manifold NODE models—both in our framework and those of others—in realistic applications requires significant further investigation. In this work, we focus on the mathematical foundations of equivariant manifold NODEs and their augmentation, but there is a substantial body of experimental results in the literature (see §1.2) showing that NODE models are practically feasible and viable as approximators. In future work, we plan to continue the exploration of the modelling capabilities of NODEs for different types of applications and fields on topologically non-trivial manifolds M with symmetries.

Acknowledgements

We thank Jan Gerken for helpful discussions. The work of E.A., D.P., and F.O. was supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. We thank the JMLR referees for insightful comments and suggestions which have improved the final version of the manuscript.

References

Tara Akhound-Sadegh, Laurence Perreault-Levasseur, Johannes Brandstetter, Max Welling, and Siamak Ravanbakhsh. Lie point symmetries and physics informed networks. *Advances in Neural Information Processing Systems*, 36:42468–42481, 2023.

Jimmy Aronsson. Homogeneous vector bundles and G-equivariant convolutional neural networks. *Sampling Theory, Signal Processing, and Data Analysis*, 20, 2022.

Shivam Arora, Alex Bihlo, and Francis Valiquette. Invariant physics-informed neural networks for ordinary differential equations. *Journal of Machine Learning Research*, 25:1–24, 2024.

Avishek Joey Bose, Mila Marcus Brubaker, and Ivan Kobyzev. Equivariant finite normalizing flows. *Arxiv e-prints arXiv:2110.08649*, 2021.

Avishek (Joey) Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian Fatras, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael Bronstein, and Alexander Tong. SE(3) stochastic flow matching for protein backbone generation. In *Proceedings of the 12th International Conference on Learning Representations (ICLR)*, 2024.

Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Velickovic. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *Arxiv e-prints arXiv:2104.13478.*, 2021.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *Proceedings of the 12th International Conference on Learning Representations (ICLR)*, 2024.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 31, 2018.

Ricky T. Q. Chen, Jens Behrmann, David Duvenaud, and Jörn-Henrik Jacobsen. Residual flows for invertible generative modeling. *Advances in Neural Information Processing Systems*, 32, 2019.

Taco S. Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral CNN. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 1321–1330. PMLR, 2019.

Benjamin Coors, Alexandru Paul Condurache, and Andreas Geiger. SphereNet: Learning spherical representations for detection and classification in omnidirectional images. In *Proceedings of the 15th European Conference on Computer Vision (ECCV)*, 2018.

Miles Cranmer, Sam Greydanus, Stephan Hoyer, Peter Battaglia, David Spergel, and Shirley Ho. Lagrangian neural networks. In *Proceedings of the Deep Differential Equations Workshop, ICLR 2020*, 2020.

Michaël Defferrard, Martino Milani, Frédérick Gusset, and Nathanaël Perraudin. DeepShere: A graph-based spherical CNN. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, 2020.

Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural ODEs. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Weinan E. A proposal on machine learning via dynamical systems. *Commun. Math. Stat.*, 5: 1–11, 2017.

Luca Falorsi and Patrick Forré. Neural ordinary differential equations on manifolds. In *Proceedings of the INNF+ Workshop of the International Conference on Machine Learning (ICML)*, 2020.

Jan E. Gerken, Jimmy Aronsson, Oscar Carlsson, Hampus Linander, Fredrik Ohlsson, Christoffer Petersson, and Daniel Persson. Geometric deep learning and equivariant neural networks. *Artificial Intelligence Review*, 56:14605–14662, 2023.

Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. FFJORD: Free-form continuous dynamics for scalable reversible generative models. In *Proceedings of the 7th International Conference on Learning Representations (ICLR 2019)*, 2019.

Sam Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Eldad Haber and Lars Ruthotto. Stable architectures for deep neural networks. *Inverse Problems*, 34:014004, 2018.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

Sigurdur Helgason. *Differential Geometry, Lie Groups and Symmetric Spaces*. American Mathematical Society, 2001.

Willy Hereman. Review of symbolic software for Lie symmetry analysis. *Mathematical and Computer Modelling*, 25:115–132, 1997.

Isay Katsman, Aaron Lou, Derek Lim, Qingxuan Jiang, Ser-Nam Lim, and Christopher De Sa. Equivariant manifold flows. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Isay Katsman, Eric Chen, Sidhanth Holalkere, Anna Asch, Aaron Lou, Ser-Nam Lim, and Christopher De Sa. Riemannian residual neural networks. In *Advances in Neural Information Processing Systems*, 2023.

Leon Klein, Andreas Krämer, and Frank Noé. Equivariant flow matching. In *Advances in Neural Information Processing Systems*, 2023.

Vincent Knibbeler. Computing equivariant matrices on homogeneous spaces for geometric deep learning and automorphic Lie algebras. *Advances in Computaitonal Mathematics*, 50, 2024.

Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *Proceedings of the 35th International Conference on Machine Learning*, pages 2747–2755. PMLR, 2018.

Jonas Köhler, Leon Klein, and Frank Noé. Equivariant flows: sampling configurations for multi-body systems with symmetric energies. In *Proceedings of the Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019)*, 2019.

Jonas Köhler, Leon Klein, and Frank Noé. Equivariant flows: Exact likelihood generative learning for symmetric densities. In *Proceedings of the 37th International Conference on Machine Learning*, pages 5361–5370. PMLR, 2020.

Pierre-Yves Lagrave and Eliot Tron. Equivariant neural networks and differential invariants theory for solving partial differential equations. *Physical Sciences Forum*, 5, 2022.

John M. Lee. *Introduction to Smooth Manifolds.* Springer New York, NY, 2012.

Hampus Linander, Christoffer Petersson, Daniel Persson, and Jan E. Gerken. PEAR: Equal area weather forecasting on the sphere. *Arxiv e-prints arXiv:2505.17720*, 2025.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. In *Proceedings of the 11th International Conference on Learning Representations (ICLR)*, 2023.

Aaron Lou, Derek Lim, Isay Katsman, Leo Huang, Qingxuan Jiang, Ser Nam Lim, and Christopher M. De Sa. Neural manifold ordinary differential equations. *Advances in Neural Information Processing Systems*, 33, 2020.

Michael Lutter, Christian Ritter, and Jan Peters. Deep lagrangian networks: Using physics as model prior for deep learning. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, 2019.

Emile Mathieu and Maximilian Nickel. Riemannian continuous normalizing flows. *Advances in Neural Information Processing Systems*, 33, 2020.

Mikio Nakahara. *Geometry, Topology and Physics.* CRC Press, 2003.

Peter J. Olver. *Applications of Lie Groups to Differential Equations.* Springer-Verlag, 1993.

Peter J. Olver. *Equivalence, Invariants and Symmetry.* Cambridge University Press, 1995.

Derek Onken, Samy Wu Fung, Xingjian Li, and Lars Ruthotto. OT-Flow: Fast and accurate continuous normalizing flows via optimal transport. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI-21)*, 2021.

Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 1530–1538. PMLR, 2015.

Danilo Jimenez Rezende, Sébastien Racanière, Irina Higgins, and Peter Toth. Equivariant hamiltonian flows. In *Proceedings of the Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019)*, 2019.

Danilo Jimenez Rezende, George Papamakarios, Sébastien Racanière, Michael S. Albergo, Gurtej Kanwar, Phiala E. Shanahan, and Kyle Cranmer. Normalizing flows on tori and spheres. *Arxiv e-print arXiv:2002.02428*, 2020.

Lars Ruthotto and Eldad Haber. Deep neural networks motivated by partial differential equations. *J. Math. Imaging Vis*, 62:352–364, 2020.

Victor Garcia Satorras, Emiel Hoogeboom, Fabian B. Fuchs, Ingmar Posner, and Max Welling. E(n) equivariant normalizing flows. *Advances in Neural Information Processing Systems*, 34, 2021.

Paul Stapor, Fabian Fröhlich, and Jan Hasenauer. Optimization and profile calculation of ODE models using second order adjoint sensitivity analysis. *Bioinformatics*, 34:151, 2018.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions of Machine Learning Research*, 03, 2024.

Peter Toth, Danilo Jimenez Rezende, Andrew Jaegle, Sébastien Racanière, Aleksandar Botev, and Irina Higgins. Hamiltonian generative networks. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, 2020.

W. Roy Utz. The embedding of homeomorphisms in continuous flows. *Topology Proceedings*, 6:159–177, 1981.

Jason Yim, Andrew Campbell, Andrew Y. K. Foong, Michael Gastegger, José Jiménez-Luna, Sarah Lewis, Victor Garcia Satorras, Bastiaan S. Veeling, Regina Barzilay, Tommi Jaakkola, and Frank Noé. Fast protein backbone generation with SE(3) flow matching. In *Proceedings of the Machine Learning in Structural Biology Workshop (NeurIPS 2023)*, 2023.

Han Zhang, Xi Gao, Jacob Unterman, and Tom Arodz. Approximation capabilities of neural ODEs and invertible residual networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 11086–11095. PMLR, 2020.