

Invariant Physics-Informed Neural Networks for Ordinary Differential Equations

Shivam Arora and **Alex Bihlo**

*Department of Mathematics and Statistics
Memorial University of Newfoundland
St. John's, NL, A1C 5S7, Canada*

SARORA17@MUN.CA, ABIHLO@MUN.CA

Francis Valiquette

*Department of Mathematics
Monmouth University
West Long Branch, NJ, 07764, USA*

FVALIQUE@MONMOUTH.EDU

Editor: Sayan Mukherjee

Keywords: Differential invariants, Lie point symmetries, moving frames, ordinary differential equations, physics-informed neural networks.

Abstract

Physics-informed neural networks have emerged as a prominent new method for solving differential equations. While conceptually straightforward, they often suffer training difficulties that lead to relatively large discretization errors or the failure to obtain correct solutions. In this paper we introduce *invariant physics-informed neural networks* for ordinary differential equations that admit a finite-dimensional group of Lie point symmetries. Using the method of equivariant moving frames, a differential equation is invariantized to obtain a, generally, simpler equation in the space of differential invariants. A solution to the invariantized equation is then mapped back to a solution of the original differential equation by solving the reconstruction equations for the left moving frame. The invariantized differential equation together with the reconstruction equations are solved using a physics-informed neural network, and form what we call an invariant physics-informed neural network. We illustrate the method with several examples, all of which considerably outperform standard non-invariant physics-informed neural networks.

1. Introduction

Physics-informed neural networks (PINNs) are an emerging method for solving differential equations using deep learning, cf. Lagaris et al. (1998); Raissi et al. (2019). The main idea behind this method is to train a neural network as an approximate solution interpolant for the differential equation. This is done by minimizing a loss function that incorporates both the differential equation and its initial and/or boundary conditions. The method has a particular elegance as the derivatives in the differential equation can be computed using automatic differentiation rather than numerical discretization, which greatly simplifies the solution procedure, especially when solving differential equations on arbitrary surfaces, see, e.g. Tang et al. (2022).

The ease of the discretization procedure in physics-informed neural networks, however, comes at the price of numerous training difficulties, and numerical solutions that are either not particularly accurate, or fail to converge at all to the true solutions of the differential equations. Since training a physics-informed neural network constitutes a non-convex optimization problem, an analysis of failure modes when physics-informed neural networks fail to train accurately is a non-trivial endeavour. This is why several modified training methodologies have been proposed, which include domain decomposition strategies, Jagtap et al. (2020), modified loss functions, Wang et al. (2024), and custom optimization, Bihlo (2024). While all of these strategies, sometimes substantially, improve upon vanilla physics-informed neural networks, none of these modified approaches completely overcome all the inherent training difficulties.

Here we propose a new approach for training physics-informed neural networks, which relies on using Lie point symmetries of differential equations and the method of equivariant moving frames to simplify the form of the differential equations that have to be solved. This is accomplished by first projecting the differential equation onto the space of differential invariants to produce an *invariantized differential equation*. The solution to the invariantized equation is then mapped back to the solution of the original equation by solving a system of first order differential equations for the left moving frame, called *reconstruction equations*. The invariant physics-informed neural network architecture proposed in this paper consists of simultaneously solving the invariantized differential equation and the reconstruction equations using a physics-informed neural network. The method proposed is entirely algorithmic, and can be implemented for any system of differential equations that is strongly invariant under the action of a group of Lie point symmetries. Since many equations of physical relevance admit a non-trivial group of Lie point symmetries, the proposed method is potentially a viable path for improving physics-informed neural networks for many real-world applications. The idea of projecting a differential equation into the space of invariants and then reconstructing its solution is reminiscent of the recent work by Vaquero et al. (2023), where the authors consider Hamiltonian systems with symmetries, although the tools used in our paper and in Vaquero et al. (2023) to achieve the desired goals are very different. Moreover, in our approach we do not assume that our equations have an underlying symplectic structure.

To simplify the theoretical exposition, we focus on the case of ordinary differential equations in this paper. We show using several examples that the proposed approach substantially improves upon the numerical results achievable with vanilla physics-informed neural networks. Applications to partial differential equations will be considered elsewhere.

The paper is organized as follows. We first review relevant work on physics-informed neural networks and symmetry-preserving numerical methods in Section 2. In Section 3 we introduce the method of equivariant moving frames and review how it can be used to solve ordinary differential equations that admit a group of Lie point symmetries. Building on Section 3, we introduce a version of invariant physics-informed neural network in Section 4. We illustrate our method with several examples in Section 5. The examples show that our proposed invariant physics-informed neural network formulation can yield better numerical results than its non-invariant version. A short summary and discussion about potential future research avenues concludes the paper in Section 6.

2. Previous work

Physics-informed neural networks were first proposed in Lagaris et al. (1998), and later popularized through the work of Raissi et al. (2019). The main idea behind physics-informed neural networks is to train a deep neural network to directly approximate the solution to a system of differential equations. This is done by defining a loss function that incorporates the given system of differential equations, along with any relevant initial and/or boundary conditions. Crucially, this turns training physics-informed neural networks into a multi-task, non-convex optimization problem that can be challenging to minimize, cf. Krishnapriyan et al. (2021). There have been several solutions proposed to overcome the training difficulties and improve the generalization capabilities of physics-informed neural networks. These include modified loss functions, McClenny and Braga-Neto (2023); Wang et al. (2024), meta-learned optimization, Bihlo (2024), domain decomposition methods, Bihlo and Popovych (2022); Jagtap et al. (2020), and the use of operator-based methods, Brecht et al. (2023); Lu et al. (2021).

The concepts of symmetries and transformation groups have also received considerable attention in the machine learning community. Notably, the equivariance of convolutional operations with respect to spatial translations has been identified as a crucial ingredient for the success of convolutional neural networks, Cohen and Welling (2016). The generalization of this observation for other types of neural network layers and other transformation groups has become a prolific subfield of deep learning since. For example, see Finzi et al. (2021) for some recent results.

Here we do not consider the problem of endowing a neural network with equivariance properties but rather investigate the question of whether a better formulation of a given differential equation can help physics-informed neural networks better learn a solution. As we will be using symmetries of differential equations for this re-formulation, our approach falls within the framework of geometric numerical integration, see e.g. Blanes and Casas (2016). The problem of symmetry-preserving numerical schemes, in other words the problem of designing discretization methods for differential equations that preserve the symmetries of differential equations, has been studied extensively over the past several decades, see Dorodnitsyn (1991); Shokin (1983); Yanenko and Shokin (1973), for some early work on the topic. Invariant discretization schemes have since been proposed for finite difference, finite volume, finite element and meshless methods, see, for example, Bihlo (2013); Bihlo et al. (2022); Bihlo and Nave (2014); Bihlo and Valiquette (2017, 2019); Budd and Dorodnitsyn (2001); Kim (2007); Olver (2001); Rebelo and Valiquette (2013, 2015).

3. Method

In this section we introduce the theoretical foundations on which the invariant physics-informed neural network framework is based. In order to fix some notation, we begin by recalling certain well-known results pertaining to symmetries of differential equations, and refer the reader to Bluman et al. (2010); Bluman and Kumei (1989); Hydon (2000); Olver (1993) for a more thorough exposition. Within the field of group analysis, the use of moving frames to solve differential equations is not well-known. Therefore, the main purpose of this section is to introduce this solution procedure. We note that, in contrast to the

approach proposed in Mansfield (2010, Chapter 6), we avoid the introduction of computational variables. Instead, all computations are based on the differential invariants of the prolonged group action. Our approach is a simplified version of the algorithm presented in Thompson and Valiquette (2015), which deals with partial differential equations admitting infinite-dimensional symmetry Lie pseudo-groups.

3.1 Invariant differential equations

As mentioned in the introduction, in this paper we limit our attention to the case of ordinary differential equations. Thus, given a $(q + 1)$ -dimensional manifold M , with $q \geq 1$, let $J^{(n)} = J^{(n)}(M, 1)$ denote the n^{th} order (extended) jet bundle consisting of equivalence classes of 1-dimensional curves $C \subset M$ under the equivalence relation of n^{th} order contact. For a detailed exposition of jet spaces, we refer the reader to Olver (1995, Chapter 4). Introducing the local coordinates $z = (t, u) = (t, u^1, \dots, u^q)$ on M , we consider t to be the independent variable and $u = (u^1, \dots, u^q)$ to be the dependent variables. Accordingly, the n^{th} order jet space $J^{(n)}$ is parametrized by $z^{(n)} = (t, u^{(n)})$, where $u^{(n)}$ denotes all the derivatives $u_j^\alpha = u_{t^j}^\alpha$ of order $0 \leq j \leq n$, with $\alpha = 1, \dots, q$.

Now, let G be an r -dimensional Lie group (locally) acting on M :

$$(T, U) = Z = g \cdot z = g \cdot (t, u), \quad \text{where} \quad g \in G. \quad (1)$$

Since group transformations preserve contact, see Olver (1995, Chapter 4), the group action (1) induces a prolonged action

$$Z^{(n)} = g \cdot z^{(n)} \quad (2)$$

on the n^{th} order jet space $J^{(n)}$. Coordinate expressions for the prolonged action (2) are obtained by applying the implicit total derivative operator

$$D_T = \frac{1}{D_t(T)} D_t, \quad \text{where} \quad D_t = \frac{\partial}{\partial t} + \sum_{j=0}^{\infty} \sum_{\alpha=1}^q u_{j+1}^\alpha \frac{\partial}{\partial u_j^\alpha}$$

denotes the standard total derivative operator, to the transformed dependent variables U^α :

$$U_j^\alpha = U_{T^j}^\alpha = D_T^j(U^\alpha), \quad \alpha = 1, \dots, q, \quad j \geq 0. \quad (3)$$

We are primarily interested in the action of a Lie group on ordinary differential equations. In the following we use the notation $\Delta(z^{(n)}) = \Delta(t, u^{(n)}) = 0$ to denote a system of differential equations, and use the index notation $\Delta_i(z^{(n)}) = 0$, $i = 1, \dots, l$, to label each equation in the system of equations $\Delta(z^{(n)}) = 0$. If $\Delta(z^{(n)}) = 0$ is a single equation, then we omit the indexing notation.

Definition 1 A nondegenerate¹ ordinary differential equation $\Delta(z^{(n)}) = 0$ is said to be strongly invariant under the prolonged action (2) of a connected local Lie group of transformations G if and only if

$$\Delta(g \cdot z^{(n)}) = 0 \quad \text{for all} \quad g \in G$$

near the identity element.

1. A differential equation is nondegenerate if at every point in its solution space it is both locally solvable and of maximal rank, (Olver, 1993, Definition 2.70).

Remark 2 *Strong invariance is more restrictive than the usual notion of symmetry, where invariance is only required to hold on the solution space. In the following, we require strong invariance to guarantee that our differential equation is an invariant function.*

Invariance is usually stated in terms of the infinitesimal generators of the group action. To this end, let

$$\mathbf{v}_\kappa = \xi_\kappa(t, u) \frac{\partial}{\partial t} + \sum_{\alpha=1}^q \phi_\kappa^\alpha(t, u) \frac{\partial}{\partial u^\alpha}, \quad \kappa = 1, \dots, r, \quad (4)$$

be a basis for the Lie algebra \mathfrak{g} of infinitesimal generators of the group action G . The prolongation of the vector fields (4), induced from the prolonged action (3), is given by

$$\mathbf{v}_\kappa^{(n)} = \xi_\kappa(t, u) \frac{\partial}{\partial t} + \sum_{j=0}^n \sum_{\alpha=1}^q \phi_\kappa^{\alpha,j}(t, u^{(j)}) \frac{\partial}{\partial u_j^\alpha}, \quad \kappa = 1, \dots, r, \quad (5)$$

where the coefficients of the prolonged vector fields are computed using the prolongation formula

$$\phi_\kappa^{\alpha,j} = D_t^j(\phi_\kappa^\alpha - \xi_\kappa u_1^\alpha) + \xi_\kappa u_{j+1}^\alpha, \quad \kappa = 1, \dots, r, \quad \alpha = 1, \dots, q, \quad 0 \leq j \leq n,$$

which can be found in Olver (1993, Eq. 2.50). The vector fields (5) provide a basis for the Lie algebra of prolonged infinitesimal generators $\mathfrak{g}^{(n)}$.

Let $F: J^{(n)} \rightarrow \mathbb{R}$ be a differential function. As explained in Olver (1993, Section 1.3), the infinitesimal change of F under the flows generated by the prolonged vector fields (5) is given by

$$\mathbf{v}_\kappa^{(n)}[F(z^{(n)})] = \xi_\kappa \frac{\partial F}{\partial t} + \sum_{j=0}^n \sum_{\alpha=1}^q \phi_\kappa^{\alpha,j} \frac{\partial F}{\partial u_j^\alpha}, \quad \kappa = 1, \dots, r.$$

Then, at the infinitesimal level, the strong invariance notion introduced in Definition 1 is equivalent to the following proposition.

Proposition 3 *A nondegenerate ordinary differential equation $\Delta(z^{(n)}) = 0$ is strongly invariant under the prolonged action of a connected local Lie group of transformations G if and only if*

$$\mathbf{v}_\kappa^{(n)}[\Delta_i(z^{(n)})] = 0, \quad \kappa = 1, \dots, r, \quad i = 1, \dots, l,$$

where $\mathbf{v}_1, \dots, \mathbf{v}_r$ is a basis of infinitesimal generators for the group of transformations G .

Proposition 3 follows from the fact that for a strongly invariant system of differential equations $\Delta(z^{(n)}) = 0$, each function $\Delta_i(z^{(n)})$, $i = 1, \dots, l$, is a differential invariant function (Olver, 1993, Thm 2.8).

Remark 4 *As one may observe, we do not include the initial conditions*

$$u^{(n-1)}(t_0) = u_0^{(n-1)} \quad (6)$$

when discussing the symmetry of the differential equation $\Delta(z^{(n)}) = \Delta(t, u^{(n)}) = 0$. This is customary when studying symmetries of differential equations. Of course, the initial conditions are necessary to select a particular solution and when implementing numerical simulations.

We finish the section with two definitions introducing regularity assumptions on the prolonged group action that will guarantee that the solution procedure discussed in the next sections is valid.

Definition 5 *The (prolonged) group action of G on $J^{(n)}$ is said to be semi-regular if all the orbits have the same dimension. A semi-regular group action is regular if, in addition, each point $z^{(n)} \in J^{(n)}$ has arbitrarily small neighborhoods whose intersection with each orbit is a connected subset thereof.*

Definition 6 *The prolonged action of G is said to be transversed to the solution space*

$$S^{(n)} = \{z^{(n)} \in J^{(n)} \mid \Delta(z^{(n)}) = 0\}$$

of the differential equation $\Delta(z^{(n)}) = 0$ if at every point $z^{(n)} \in S^{(n)}$, the intersection of the Lie algebra of prolonged infinitesimal generators $\mathfrak{g}^{(n)}|_{z^{(n)}}$ at $z^{(n)}$ intersects the tangent space $T_{z^{(n)}}S^{(n)}$ trivially so that $\mathfrak{g}^{(n)} \cap TS^{(n)} = \{0\}$.

3.2 Invariantization

In this section we assume that $\Delta(z^{(n)}) = 0$ is a nondegenerate differential equation, which is strongly invariant under the prolonged action of an r -dimensional Lie group G acting regularly on $J^{(n)}$. We now explain how to use the method of equivariant moving frames to “project” the differential equation onto the space of differential invariants. For the theoretical foundations of the method of equivariant moving frames, we refer the reader to the foundational papers by Fels and Olver (1999); Kogan and Olver (2003), and the textbook by Mansfield (2010).

Definition 7 *A right moving frame is a map $\rho: J^{(n)} \rightarrow G$ that satisfies the G -equivariance condition*

$$\rho(g \cdot z^{(n)}) = \rho(z^{(n)}) \cdot g^{-1} \tag{7}$$

for all $g \in G$ where the prolonged action is defined. Taking the group inverse of a right moving frame yields the left moving frame

$$\bar{\rho}(z^{(n)}) = \rho(z^{(n)})^{-1}$$

satisfying the equivariance condition

$$\bar{\rho}(g \cdot z^{(n)}) = g \cdot \bar{\rho}(z^{(n)}).$$

To guarantee the existence of a moving frame, we need to introduce the notion of freeness of the (prolonged) group action.

Definition 8 *A Lie group G acting on $J^{(n)}$ is said to act freely if for all $z^{(n)} \in J^{(n)}$ the isotropy group at $z^{(n)}$ given by*

$$G_{z^{(n)}} = \{g \in G \mid g \cdot z^{(n)} = z^{(n)}\}$$

is trivial, that is $G_{z^{(n)}} = \{e\}$. The Lie group G is said to act locally freely if $G_{z^{(n)}}$ is a discrete subgroup of G for all $z^{(n)} \in J^{(n)}$.

Theorem 9 *A moving frame exists in some neighborhood of a point $z^{(n)} \in \mathbf{J}^{(n)}$ if and only if the prolonged action of G is (locally) free and regular near $z^{(n)}$.*

The proof of Theorem 9 can be found in Fels and Olver (1999, Thm 4.4).

Remark 10 *In general, Theorem 9 might only hold on a G -invariant submanifold $\mathcal{V}^{(n)} \subset \mathbf{J}^{(n)}$. In this case we would restrict Definitions 7 and 8 to $\mathcal{V}^{(n)}$. To simplify the discussion, we assume that $\mathcal{V}^{(n)} = \mathbf{J}^{(n)}$ in the subsequent considerations.*

A moving frame is obtained by selecting a (regular) cross-section $\mathcal{K} \subset \mathbf{J}^{(n)}$ to the orbits of the prolonged action.

Definition 11 *A (local) cross-section is a submanifold $\mathcal{K} \subset \mathbf{J}^{(n)}$ of codimension $r = \dim G$ such that \mathcal{K} intersects each orbit transversally. The cross-section is regular if \mathcal{K} intersects each orbit at most once.*

Keeping with most applications, and to simplify the exposition, we assume \mathcal{K} is a regular coordinate cross-section obtained by setting r coordinates of the jet $z^{(n)}$ to constant values:

$$z^{a_\kappa} = c^\kappa, \quad \kappa = 1, \dots, r. \quad (8)$$

Under the assumptions of Theorem 9, the right moving frame at $z^{(n)}$ is the unique group element mapping $z^{(n)}$ onto the cross-section \mathcal{K} specified by (8). This transformation is obtained by solving the *normalization equations*

$$Z^{a_\kappa} = g \cdot z^{a_\kappa} = c^\kappa, \quad \kappa = 1, \dots, r,$$

for the group parameters $g = (g^1, \dots, g^r)$, yielding the right moving frame ρ . Given a right moving frame, there is a systemic procedure for constructing differential invariant functions.

Definition 12 *Let $\rho: \mathbf{J}^{(n)} \rightarrow G$ be a right moving frame. The invariantization of the differential function $F: \mathbf{J}^{(n)} \rightarrow \mathbb{R}$ is the differential invariant function*

$$\iota(F)(z^{(n)}) = F(\rho(z^{(n)}) \cdot z^{(n)}). \quad (9)$$

Remark 13 *The fact that (9) is a differential invariant function follows from the G -equivariant property (7) for the right moving frame.*

Applying the invariantization map ι introduced in Definition 12 to $z^{(n)}$ componentwise yields the differential invariants

$$\iota(z^{(n)}) = \rho(z^{(n)}) \cdot z^{(n)},$$

which can be used as coordinates for the cross-section \mathcal{K} . In particular, the invariantization of the coordinates used to define the cross-section in (8) are constant

$$\iota(z^{a_\kappa}) = c^\kappa, \quad \kappa = 1, \dots, r,$$

and are called *phantom invariants*. The remaining invariantized coordinates are called *normalized invariants*.

In light of Theorem 5.32 in Olver (2009), assume there are $q + 1$ independent normalized invariants

$$H, I^1, \dots, I^q, \quad (10)$$

such that locally

$$I^\alpha = I^\alpha(H), \quad \alpha = 1, \dots, q,$$

are functions of the invariant H , and generate the algebra of differential invariants. This means that any differential invariant function can be expressed in terms of (10) and their invariant derivatives with respect to D_H . In the following we let $I^{(n)}$ denote the derivatives of $I = (I^1, \dots, I^q)$ with respect to H , up to order n .

Assuming the differential equation $\Delta(z^{(n)}) = 0$ is strongly invariant and its solutions are transverse to the prolonged action, this equation, once invariantized, yields a differential equation in the space of differential invariants

$$\iota(\Delta)(t, u^{(n)}) = \Delta_{\text{Inv}}(H, I^{(k)}) = 0, \quad \text{where } k \leq n. \quad (11a)$$

Initial conditions for (11a) are obtained by invariantizing (6) to obtain

$$I^{(k-1)}(H_0) = I_0^{(k-1)}. \quad (11b)$$

Example 1 *To illustrate the concepts introduced thus far, we use Schwarz' equation*

$$\frac{u_{ttt}}{u_t} - \frac{3}{2} \left(\frac{u_{tt}}{u_t} \right)^2 = F(t), \quad (12)$$

where, for simplicity, we assume that $F(t)$ is a continuous function. The general solution to (12) can be found in Hille (1976, Section 10). According to Ovsienko and Tabachnikov (2009), the Schwarz derivative $\{u, t\} = u_{ttt}/u_t - (3/2)(u_{tt}/u_t)^2$ first appeared in the treatise by Lagrange (1779). Over time the Schwarz derivative has found applications in complex analysis, one-dimensional dynamics, Teichmüller theory, integrable systems, and conformal field theory. Equation (12) admits a three-dimensional Lie group of point transformations given by

$$T = t, \quad U = \frac{\alpha u + \beta}{\gamma u + \delta}, \quad \text{where } g = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \in SL(2, \mathbb{R}), \quad (13)$$

so that $\alpha\delta - \beta\gamma = 1$. A cross-section to the prolonged group action

$$\begin{aligned} U_T &= D_t(U) = \frac{u_t}{(\gamma u + \delta)^2}, \\ U_{TT} &= D_t(U_T) = \frac{u_{tt}}{(\gamma u + \delta)^2} - \frac{2\gamma u_t^2}{(\gamma u + \delta)^3}, \\ U_{TTT} &= D_t(U_{TT}) = \frac{u_{ttt}}{(\gamma u + \delta)^2} - \frac{6\gamma u_t u_{tt}}{(\gamma u + \delta)^3} + \frac{6\gamma^2 u_t^3}{(\gamma u + \delta)^4}, \end{aligned}$$

is given by

$$\mathcal{K} = \{u = 0, u_t = \sigma, u_{tt} = 0\} \subset \mathcal{V}^{(n)} \subset \mathcal{J}^{(n)}, \quad (14)$$

where $\sigma = \text{sign}(u_t)$ and $\mathcal{V}^{(n)} = \{z^{(n)} \in \mathbf{J}^{(n)} \mid u_t \neq 0\}$ with $n \geq 2$. Solving the normalization equations

$$U = 0, \quad U_T = \sigma, \quad U_{TT} = 0,$$

for the group parameters, taking into account the unitary constraint $\alpha\delta - \beta\gamma = 1$, yields the right moving frame

$$\alpha = \pm \frac{1}{\sqrt{|u_t|}}, \quad \beta = \mp \frac{u}{\sqrt{|u_t|}}, \quad \gamma = \pm \frac{u_{tt}}{2|u_t|^{3/2}}, \quad \delta = \pm \frac{2u_t^2 - uu_{tt}}{2|u_t|^{3/2}}, \quad (15)$$

where the sign ambiguity comes from solving the normalization $U_T = \sigma$, which involves the quadratic term $(\gamma u + \delta)^2$. Invariantizing the third order derivative u_{ttt} produces the normalized differential invariant

$$\iota(u_{ttt}) = \frac{u_{ttt}}{(\gamma u + \delta)^2} - \frac{6\gamma u_t u_{tt}}{(\gamma u + \delta)^3} + \frac{6\gamma^2 u_t^3}{(\gamma u + \delta)^4} \Big|_{(15)} = \sigma \left(\frac{u_{ttt}}{u_t} - \frac{3}{2} \left(\frac{u_{tt}}{u_t} \right)^2 \right). \quad (16)$$

In terms of the general theory previously introduced, we have the invariants

$$H = t, \quad I = \frac{u_{ttt}}{u_t} - \frac{3}{2} \left(\frac{u_{tt}}{u_t} \right)^2. \quad (17)$$

Since the independent variable t is an invariant, instead of using H , we use t in the following computations. The invariantization of Schwarz' equation (12) yields the algebraic equation

$$I = F(t). \quad (18)$$

Since the prolonged action is transitive on the fibers of each component $\{(t, u, u_t, u_{tt}) \mid u_t > 0\} \cup \{(t, u, u_t, u_{tt}) \mid u_t < 0\} = \mathcal{V}^{(2)}$, any initial conditions

$$u(t_0) = u_0, \quad u_t(t_0) = u_t^0, \quad u_{tt}(t_0) = u_{tt}^0, \quad (19)$$

is mapped, via the invariantization map ι , to the identities

$$0 = 0, \quad \sigma = \sigma, \quad 0 = 0.$$

3.3 Recurrence relations

In this section we introduce the recurrence relations for the normalized differential invariants, and explain how the invariantized equation (11a) can be derived symbolically, without requiring the coordinate expressions for the right moving frame ρ or the invariants $(H, I^{(n)})$.

A key observation is that the invariantization map ι and the exterior differential, in general, do not commute

$$\iota \circ d \neq d \circ \iota.$$

The extent to which these two operations fail to commute is encapsulated in the recurrence relations. To state these equations we need to introduce the (contact) invariant one-form

$$\varpi = \iota(dt) = \rho^*(D_t(T)) dt,$$

which comes from invariantizing the horizontal one-form dt , and where ρ^* denotes the right moving frame pull-back. We refer the reader to Kogan and Olver (2003) for more details.

Given a Lie group G , assume its elements $g \in G$ are given by a faithful representation. Then the *right Maurer–Cartan form* is given by

$$\mu = dg \cdot g^{-1}. \quad (20)$$

The pull-back of the Maurer–Cartan form (20) by a right moving frame ρ yields the invariant matrix

$$\nu = d\rho \cdot \rho^{-1} = [I_{ij}] \varpi, \quad (21)$$

where the invariants I_{ij} are called Maurer–Cartan invariants.

Proposition 14 *Let $F: J^{(n)} \rightarrow \mathbb{R}$ be a differential function and $\rho: J^{(n)} \rightarrow G$ a right moving frame. We then have the recurrence relation*

$$d[\iota(F)] = \iota[dF] + \sum_{\kappa=1}^r \iota[\mathbf{v}_{\kappa}^{(n)}(F)] \nu^{\kappa}, \quad (22)$$

where ν^1, \dots, ν^r is a basis of normalized Maurer–Cartan forms extracted from (21).

In particular, substituting for F in (22) the jet coordinates $z^{(n)} = (t, u^{(n)})$ yields the recurrence relations

$$\begin{aligned} d[\iota(t)] &= \varpi + \sum_{\kappa=1}^r \iota(\xi_{\kappa}) \nu^{\kappa}, \\ d[\iota(u_j^{\alpha})] &= \iota(u_{j+1}^{\alpha}) \varpi + \sum_{\kappa=1}^r \iota(\phi_{\kappa}^{\alpha,j}) \nu^{\kappa}, \end{aligned} \quad (23)$$

where ξ_{κ} , $\phi_{\kappa}^{\alpha,j}$ are the coefficients of the prolonged vector fields in (5). The recurrence relations for the jet coordinates (8) specifying the coordinate cross-section \mathcal{K} lead to r linear equations for the normalized Maurer–Cartan forms ν^1, \dots, ν^r . Solving those equations and substituting the result back in (23) yields symbolic expressions for the differential of the normalized invariants without requiring the coordinate expressions for the moving frame ρ or $\iota(t)$, $\iota(u^{(n)})$. More generally, substituting the expressions for the normalized Maurer–Cartan forms in (22) gives the symbolic expression for the differential of any invariantized differential function F .

Example 2 *Continuing Example 1, we differentiate the group action (13) with respect to the group parameters α , β , γ , respectively, recalling the unitary constraint $\alpha\delta - \beta\gamma = 1$, and evaluate the result at the identity transformation $\alpha = \delta = 1$, $\gamma = \beta = 0$ to obtain the basis of infinitesimal generators*

$$\mathbf{v}_1 = \frac{\partial}{\partial u}, \quad \mathbf{v}_2 = u \frac{\partial}{\partial u}, \quad \mathbf{v}_3 = u^2 \frac{\partial}{\partial u}. \quad (24)$$

The prolongation of those vector fields, up to order 2, is given by

$$\begin{aligned}
 \mathbf{v}_1^{(3)} &= \frac{\partial}{\partial u}, \\
 \mathbf{v}_2^{(3)} &= u \frac{\partial}{\partial u} + u_t \frac{\partial}{\partial u_t} + u_{tt} \frac{\partial}{\partial u_{tt}}, \\
 \mathbf{v}_3^{(3)} &= u^2 \frac{\partial}{\partial u} + 2uu_t \frac{\partial}{\partial u_t} + 2(u_t^2 + uu_{tt}) \frac{\partial}{\partial u_{tt}}.
 \end{aligned} \tag{25}$$

Computing the recurrence relations (23) for t , u , u_t , and u_{tt} yields

$$\begin{aligned}
 d[\iota(t)] &= \varpi, \\
 d[\iota(u)] &= \iota(u_t)\varpi + \nu^1 + \iota(u)\nu^2 + \iota(u)^2\nu^3, \\
 d[\iota(u_t)] &= \iota(u_{tt})\varpi + \iota(u_t)\nu^2 + 2\iota(u)\iota(u_t)\nu^3, \\
 d[\iota(u_{tt})] &= \iota(u_{ttt})\varpi + \iota(u_{tt})\nu^2 + 2[\iota(u_t)^2 + \iota(u)\iota(u_{tt})]\nu^3.
 \end{aligned} \tag{26}$$

We note that the coefficients of the correction terms involving the normalized Maurer–Cartan forms ν^1, ν^2, ν^3 in the recurrence relations (26) are obtained by invariantizing the coefficients of the prolonged vector fields (25).

Recalling the cross-section (14) and the invariants (16), (17), we make the substitutions $\iota(t) = t$, $\iota(u) = 0$, $\iota(u_t) = \sigma$, $\iota(u_{tt}) = 0$, $\iota(u_{ttt}) = \sigma I$ into (26) and obtain

$$dt = \varpi, \quad 0 = \varpi + \nu^1, \quad 0 = \nu^2, \quad 0 = I\sigma\varpi + 2\nu^3.$$

Solving the last three equations for the normalized Maurer–Cartan forms yields

$$\nu^1 = -\sigma\varpi, \quad \nu^2 = 0, \quad \nu^3 = -\frac{I\sigma}{2}\varpi.$$

In matrix form we have

$$\nu = \begin{bmatrix} 2\nu^2 & \nu^1 \\ -\nu^3 & -2\nu^2 \end{bmatrix} = \begin{bmatrix} 0 & -\sigma \\ \frac{1}{2}\sigma I & 0 \end{bmatrix} \varpi = \begin{bmatrix} 0 & -\sigma \\ \frac{1}{2}\sigma F(t) & 0 \end{bmatrix} \varpi, \tag{27}$$

where we used the algebraic relationship (18) originating from the invariantization of Schwarz' equation (12). The, perhaps, unexpected coefficients in the Maurer–Cartan matrix (27), namely $2\nu^2$ and $-\nu^3$, originate from the fact that when introducing the basis of infinitesimal generators (24) we scaled \mathbf{v}_2 by $1/2$ and \mathbf{v}_3 by -1 .

3.4 Reconstruction

Let $I(H)$ be a solution to the invariantized differential equation (11a) with initial conditions (11b). In this section we explain how to reconstruct the solution to the original equation $\Delta(x, u^{(n)}) = 0$ with initial conditions (6). To do so, we introduce the reconstruction equations for the left moving frame $\bar{\rho} = \rho^{-1}$ given by

$$d\bar{\rho} = -\bar{\rho} \cdot d\rho \cdot \bar{\rho} = -\bar{\rho}\nu, \tag{28}$$

where ν is the normalized Maurer–Cartan form introduced in (21). As we have seen in Section 3.3, the invariantized Maurer–Cartan matrix ν can be obtained symbolically using

the recurrence relations for the phantom invariants. On the other hand, the Maurer–Cartan invariants I_{ij} in (21) can be expressed in terms of H , the solution $I(H)$ to the invariantized initial value problem (11), and its derivatives. Thus, equation (28) yields a first order system of differential equations for the group parameters depending on the independent variable H . Integrating (28), we obtain the left moving frame that sends the invariant curve $(H, I(H))$ to the original solution

$$(t(H), u(H)) = \bar{\rho}(H) \cdot \iota(t, u)(H). \quad (29)$$

The transversality of the prolonged group action implies that the derivative $t_H \neq 0$. Assuming $t_H > 0$, the initial conditions to the reconstruction equations (28) are given by

$$\bar{\rho}(H_0) = \bar{\rho}_0 \quad \text{such that} \quad \bar{\rho}_0 \cdot \iota(t_0, u_0^{(n-1)}) = (t_0, u_0^{(n-1)}). \quad (30)$$

If $t_H < 0$, one can always reparametrize the solution so that the derivative becomes positive.

The solution (29) is a parametric curve with the invariant H serving as parameter. From a numerical perspective, this is sufficient to graph the solution. Though, we note that since $t_H \neq 0$, it is possible, by the implicit function theorem, to invert $t = t(H)$ to express the invariant $H = H(t)$ in terms of t and recover the solution $u = u(H(t))$ as a function of t .

Example 3 *The left moving frame*

$$\bar{\rho} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \in SL(2, \mathbb{R})$$

that will send the invariant solution (18) to the original solution $u(t)$ of Schwarz' equation (12) with initial conditions (19) must satisfy the reconstruction equations

$$\begin{bmatrix} \alpha_t & \beta_t \\ \gamma_t & \delta_t \end{bmatrix} = - \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} 0 & -\sigma \\ \frac{1}{2}\sigma F(t) & 0 \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} 0 & \sigma \\ -\frac{1}{2}\sigma F(t) & 0 \end{bmatrix}, \quad (31a)$$

with the initial conditions

$$\delta_0 = \pm \frac{1}{\sqrt{|u_t^0|}}, \quad \beta_0 = \pm \frac{u_0}{\sqrt{|u_t^0|}}, \quad \gamma_0 = \mp \frac{u_{tt}^0}{2(u_t^0)^{3/2}}, \quad \alpha_0 = \pm \sqrt{|u_t^0|} \mp \frac{u_0 u_{tt}^0}{2(|u_t^0|)^{3/2}}. \quad (31b)$$

Once the reconstruction equations (31) are solved, the solution to Schwarz' equation (12) is

$$u(t) = \bar{\rho} \cdot 0 = \frac{\beta}{\delta}. \quad (32)$$

3.5 Summary

Let us summarize the algorithm for solving an ordinary differential equation $\Delta(t, u^{(n)}) = 0$ admitting a group of Lie point symmetries G using the method of moving frames.

1. Select a cross-section \mathcal{K} to the prolonged action.
2. Choose $q + 1$ independent invariants H, I^1, \dots, I^q from $\iota(t, u^{(n)})$ that generate the algebra of differential invariants, and assume $I^1(H), \dots, I^q(H)$ are functions of H .

3. Invariantize the differential equation $\Delta(t, u^{(n)}) = 0$ and use the recurrence relations (23) to write the result in terms of H and $I^{(k)}$ to obtain the differential equation $\Delta_{\text{Inv}}(H, I^{(k)}) = 0$.
4. Solve the equation $\Delta_{\text{Inv}}(H, I^{(k)}) = 0$ subject to the initial conditions (11b).
5. A parametric solution to the original equation $\Delta(t, u^{(n)}) = 0$ is given by $\bar{\rho}(H) \cdot \iota(t, u)(H)$, where the left moving frame $\bar{\rho}(H)$ is a solution of the reconstruction equation (28) subject to the initial conditions (30).

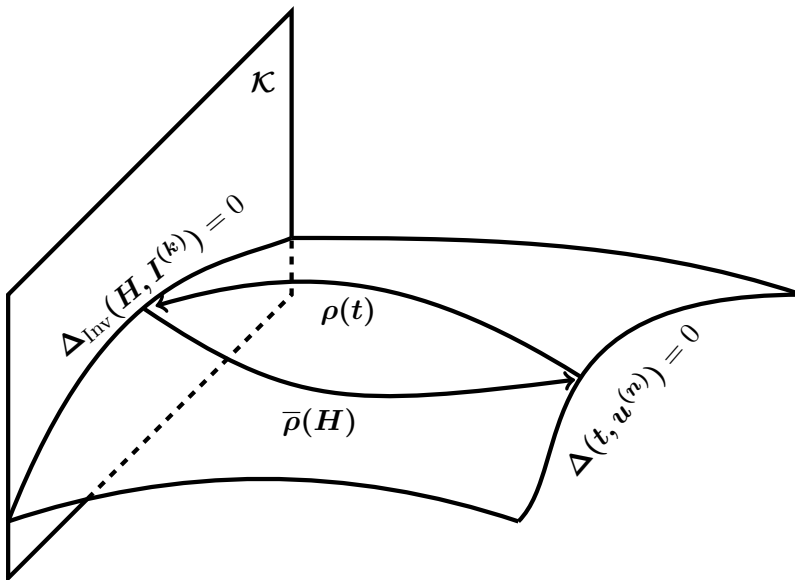


Figure 1: Solving a differential equation using moving frames.

4. Invariant physics-informed neural networks

Before introducing our invariant physics-informed neural network, we recall the definition of the standard physics-informed loss function that needs to be minimized when solving ordinary differential equations. To this end, assume we want to solve the ordinary differential equation $\Delta(t, u^{(n)}) = 0$ subject to the initial conditions $u^{(n-1)}(t_0) = u_0^{(n-1)}$ on the interval $[t_0, t_f]$. We recall that the ordinary differential equation $\Delta(t, u^{(n)}) = 0$ can also be a system of differential equations.

First we introduce the collocation points $\{t_i\}_{i=0}^N$ sampled randomly over the interval $[t_0, t_f]$ with $t_0 = t_0 < t_1 < \dots < t_N = t_f$. Then, a neural network of the form $u_{\theta}(t) = \mathcal{N}_{\theta}(t)$, parameterized by the parameter vector θ , is trained to approximate the solution of the differential equation, i.e. $u_{\theta}(t) \approx u(t)$, by minimizing the physics-informed loss function

$$\mathcal{L}(\theta) = \mathcal{L}_{\Delta}(\theta) + \alpha \mathcal{L}_{\text{I.C.}}(\theta) \quad (33a)$$

with respect to θ , where

$$\mathcal{L}_\Delta(\theta) = \sum_{i=0}^N \|\Delta(t_i, u_\theta^{(n)}(t_i))\|_{\ell^2}^2 \quad (33b)$$

is the *differential equation loss* and $\|\cdot\|_{\ell^2}$ is the ℓ^2 -norm,

$$\mathcal{L}_{\text{I.C.}}(\theta) = \|u_\theta^{(n-1)}(t_0) - u_0^{(n-1)}\|_{\ell^2}^2 \quad (33c)$$

is the *initial condition loss*, and $\alpha \in \mathbb{R}^+$ is a hyper-parameter to re-scale the importance of both loss functions. We note that the differential equation loss is the mean squared error of the differential equation evaluated at the collocation points $\{t_i\}_{i=0}^N \subset [t_0, t_f]$ over which the numerical solution is sought. The initial condition loss is likewise the mean squared error between the true initial conditions and the initial conditions approximated by the neural network. We note in passage that the initial conditions could alternatively be enforced as a hard constraint in the neural network, see, e.g. Brecht et al. (2023); Lagaris et al. (1998), in which case the physics-informed loss function would reduce to $\mathcal{L}_\Delta(\theta)$ only.

The physics-informed loss function (33) is minimized using gradient descent, usually using the Adam optimizer, Kingma and Ba (2014), but also more elaborate optimizers can be employed, Bihlo (2024). The particular elegance of the method of physics-informed neural networks lies in the fact that the derivatives $u_\theta^{(n)}$ of the neural network solution approximation are computed using *automatic differentiation*, Baydin et al. (2018), which is built into all modern deep learning frameworks such as JAX, PyTorch, or TensorFlow.

Similar to the above standard physics-informed neural network, an invariant physics-informed neural network is a feed-forward neural network approximating the solution of the invariantized differential equation and the reconstruction equations for the left moving frame.

In light of the five step process given in Section 3.5, assume the invariantized equation $\Delta_{\text{Inv}}(H, I^{(k)}) = 0$ and the reconstruction equation $d\bar{\rho} = -\bar{\rho}\nu$ have been derived. Introduce an interval of integration $[H_0, H_f]$ over which the numerical solution is sought, and consider the collocation points $\{H_i\}_{i=0}^N \subset [H_0, H_f]$, such that $H_N = H_f$. The neural network has to learn a mapping between H and the functions

$$I_\theta(H) \quad \text{and} \quad \bar{\rho}_\theta(H),$$

where $I_\theta(H)$ denotes the neural network approximation of the differential invariants $I(H)$ solving (11a), and $\bar{\rho}_\theta(H)$ is the approximation of the left moving frame $\bar{\rho}(H)$ solving the reconstruction equations (28). We note that the output size of the network depends on the numbers of invariants $I(H)$ and the size of the symmetry group via $\bar{\rho}(H)$.

The network is trained by minimizing the invariant physics-informed loss function consisting of the invariantized differential equation loss and the reconstruction equations loss defined as the sum of mean squared errors

$$\mathcal{L}_{\Delta_{\text{Inv}}, \bar{\rho}}(\theta) = \sum_{i=0}^N (\|\Delta_{\text{Inv}}(H_i, I_\theta^{(k)}(H_i))\|_{\ell^2}^2 + \|d\bar{\rho}_\theta(H_i) + \bar{\rho}_\theta(H_i)\nu(H_i, I_\theta^{(\kappa)}(H_i))\|_{\ell^2}^2). \quad (34)$$

We note that there is an abuse of notation in the mean squared error of the reconstruction equations. Technically, $d\bar{\rho} + \bar{\rho}\nu$ is a differential one-form in dH . Thus, when computing the ℓ^2 -norm, we implicitly only consider the components of that one-form.

We supplement the loss function (34) with the initial conditions (11b) and (30) by considering the invariant initial conditions loss function

$$\mathcal{L}_{\text{I.C.}}(\boldsymbol{\theta}) = \|I_{\boldsymbol{\theta}}^{(k-1)}(H_0) - I_0^{(k-1)}\|_{\ell^2}^2 + \|\bar{\rho}_{\boldsymbol{\theta}}(H_0) - \bar{\rho}_0\|_{\ell^2}^2.$$

The final invariant physics-informed loss function is thus given by

$$\mathcal{L}_{\text{Inv}}(\boldsymbol{\theta}) = \mathcal{L}_{\Delta_{\text{Inv}, \bar{\rho}}}(\boldsymbol{\theta}) + \alpha \mathcal{L}_{\text{I.C.}}(\boldsymbol{\theta}),$$

where $\alpha \in \mathbb{R}^+$ is again a hyper-parameter rescaling the importance of the equation and initial condition losses.

5. Examples

In this section we implement the invariant physics-informed neural network procedure introduced in Section 4 for several examples. We also train a standard physics-informed neural network to compare the solutions obtained. For both models we use feed-forward neural networks minimizing the invariant loss function and standard PINN loss function, respectively. For the sake of consistency, all networks used throughout this section have 5 layers, with 40 nodes per layer, and use the hyperbolic tangent as activation functions. For most examples, the loss stabilizes at fewer than 3,000 epochs, but for uniformity we trained all models for 3,000 epochs. All examples use 200 collocation points. The numerical errors of the two neural network solutions are obtained by comparing the numerical solutions to the exact solution, if available, or to the numerical solution obtained using `odeint` in `scipy.integrate`. We also compute the mean square error over the entire interval of integration for all examples together with the standard deviation averaged over 5 runs. These results are summarized in Table 1. Finally, the point-wise square error plots for each example are provided to show the error varying over the interval of integration.

Example 4 *As our first example, we consider the Schwarz equation (12), with $F(t) = 2$. For the numerical simulations, we used the initial conditions*

$$u_0 = u_{tt}^0 = 0, \quad u_t^0 = 1, \tag{35}$$

in (19) with $t_0 = 0$. According to (18) the invariantization of Schwarz' equation yields the algebraic constraint $I = 2$. Thus, the loss function (34) will only contain the reconstruction equations (31a). Namely,

$$\alpha_t + \beta = \beta_t - \alpha = \gamma_t + \delta = \delta_t - \gamma = 0, \tag{36a}$$

where we used the fact that $\sigma = 1$. Substituting (35) into (31b), yields the initial conditions

$$\delta_0 = \alpha_0 = \pm 1, \quad \beta_0 = \gamma_0 = 0 \tag{36b}$$

for the reconstruction equations. In our numerical simulations we worked with the positive sign. Once the reconstruction equations have been solved, the solution to the Schwarz equation is given by the ratio (32). The solution is integrated on the interval $t \in [0, \pi]$. Error plots for the solutions obtained via the invariant PINN and the standard PINN implementations are given in Figure 2. These errors are obtained by comparing the numerical solutions to the exact solution $u(t) = \tan(t)$. Clearly, the invariant implementation is substantially more precise near the vertical asymptote at $t = \pi/2$. The reason for this improvement originates from the fact that the invariant PINN implementation seeks to solve the system of linear ODEs (36) to approximate the exact solutions $\beta(t) = \sin(t)$ and $\delta(t) = \cos(t)$, which are bounded functions, while the standard PINN implementation attempts to approximate the unbounded function $u(t) = \tan(t) = \beta(t)/\delta(t)$ using the Schwarz equation (12).

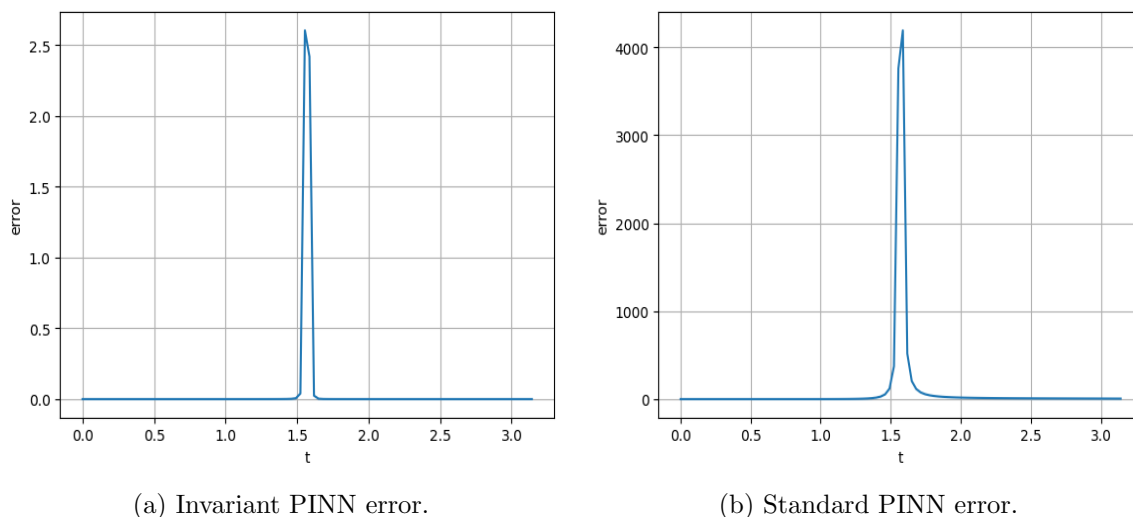


Figure 2: Time series of the squared error for the Schwarz equation (12).

Example 5 As our second example, we consider the logistic equation

$$u_t = u(1 - u) \tag{37}$$

occurring in population growth modeling. Equation (37) admits the one-parameter symmetry group

$$T = t, \quad U = \frac{u}{1 + \epsilon u e^{-t}}, \quad \text{where } \epsilon \in \mathbb{R}.$$

Implementing the algorithm outlined in Section 3.5, we choose the cross-section $\mathcal{K} = \{u = 1\}$. This yields the invariantized equation

$$I = \iota(u_t) = 0.$$

The reconstruction equation is

$$\epsilon_t = I = 0, \tag{38}$$

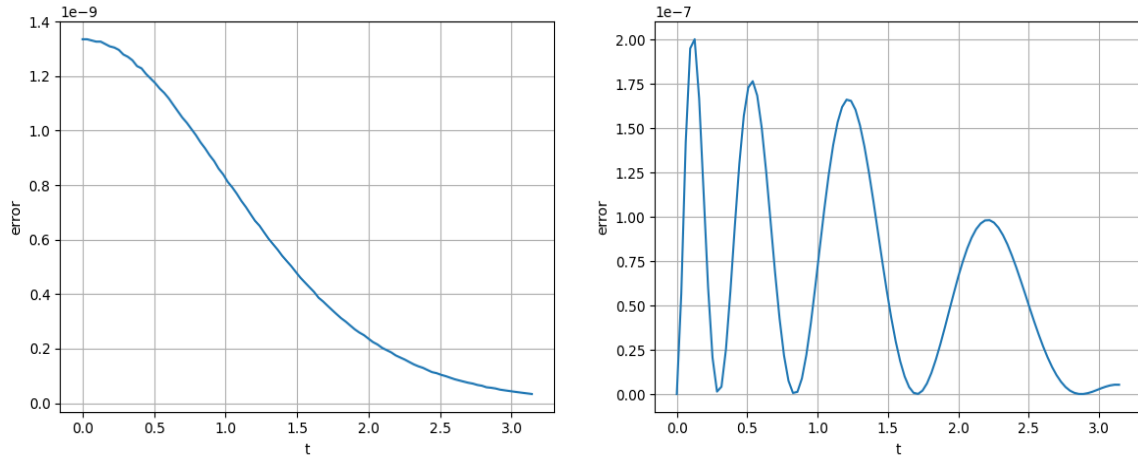
subject to the initial condition

$$\epsilon(t_0) = \left(\frac{1 - u_0}{u_0} \right) e^{t_0},$$

where $u_0 = 0.5$ and our interval of integration is $[0, \pi]$. The solution to the logistic equation is then given by

$$u(t) = \frac{1}{1 + \epsilon e^{-t}}. \quad (39)$$

As Figure 3 illustrates, the error incurred by the invariant PINN model is significantly smaller than the standard PINN implementation, by about a factor of more than 100, when compared to the exact solution (39).



(a) Invariant PINN error.

(b) Standard PINN error.

Figure 3: Time series of the squared error for the logistic equation (37).

Example 6 We now consider the driven harmonic oscillator

$$u_{tt} + u = \sin(t^a), \quad (40)$$

which appears in inductor-capacitor circuits, Serway and Jewett (2003). In the following we set $a = 0.99$, which yields bounded solutions close to resonance occurring when $a = 1$. The differential equation (40) admits the two-dimensional symmetry group of transformations

$$T = t, \quad U = u + \alpha \sin(t) + \beta \cos(t), \quad \text{where } \alpha, \beta \in \mathbb{R}.$$

A cross-section to the prolonged action is given by $\mathcal{K} = \{u = u_t = 0\}$. The invariantization of (40) yields

$$I = \iota(u_{tt}) = \sin(t^a).$$

The reconstruction equations are

$$\alpha_t = \sin(t^a) \cos(t), \quad \beta_t = -\sin(t^a) \sin(t), \quad (41)$$

with initial conditions

$$\alpha(t_0) = u_0 \sin(t_0) + u_t^0 \cos(t_0), \quad \beta(t_0) = u_0 \cos(t_0) - u_t^0 \sin(t_0),$$

where, in our numerical simulations, we set $u_0 = u_t^0 = 1$ and integrate over the interval $[0, 10]$. Given a solution to the reconstruction equations (41), the solution to the driven harmonic oscillator (40) is

$$u(t) = \alpha(t) \sin(t) + \beta(t) \cos(t).$$

Figure 4 shows the error for the invariant PINN implementation and the standard PINN approach compared to the solution obtained using the Runge–Kutta method `odeint` as provided in `scipy.integrate`. As in the previous two examples, the invariant version yields substantially better numerical results than the standard PINN method.

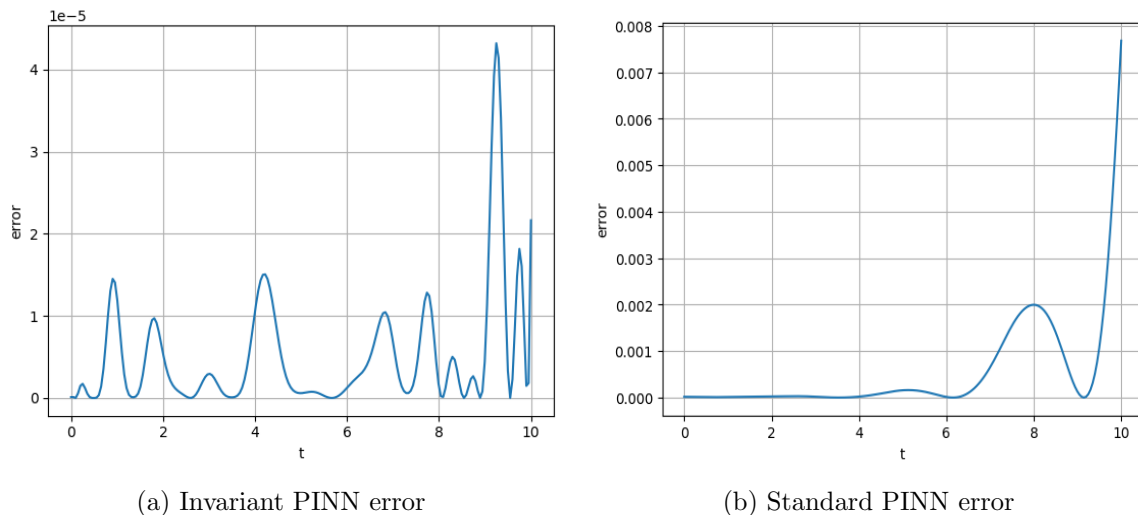


Figure 4: Time series of the squared error for the driven harmonic oscillator (40).

Example 7 We now consider the second order ordinary differential equation

$$u_{tt} = \exp[-u_t] \tag{42}$$

with an exponential term. Equation (42) admits a three-dimensional symmetry group action given by

$$T = e^\epsilon t + a, \quad U = e^\epsilon u + \epsilon e^\epsilon x + b,$$

where $a, b, \epsilon \in \mathbb{R}$. In the following, we only consider the one-dimensional group

$$T = e^\epsilon t, \quad U = e^\epsilon u + \epsilon e^\epsilon x.$$

We note that in this example the independent variable is not invariant as in the previous examples. A cross-section to the prolonged action is given by $\mathcal{K} = \{u_t = 0\}$. Introducing the invariants

$$H = \ln \left[\frac{1}{1 - \iota(t)} \right] = \ln \left[\frac{1}{1 - tu_t} \right], \quad I = \iota(u) = \exp[-u_t](u - tu_t),$$

the invariantization of the differential equation (42) reduces to the first order linear equation

$$I_H + I = e^{-H} - 1. \quad (43a)$$

The reconstruction equation for the left moving frame is simply

$$\epsilon_H = 1. \quad (43b)$$

In terms of ϵ and I , the parametric solution to the original differential equation (42) is

$$t = e^\epsilon(1 - e^{-H}), \quad u = e^\epsilon(I + \epsilon(1 - e^{-H})). \quad (44)$$

The solution to (42) is known and is given by

$$u(t) = (t + c_1) \ln(t + c_1) - t + c_2, \quad (45)$$

where c_1, c_2 are two integration constants. For the numerical simulations, we use the initial conditions

$$I_0 = \exp[-u_t^0](u_0 - t_0 u_t^0), \quad \epsilon_0 = u_t^0,$$

where $u_0 = u(t_0)$, $u_t^0 = u_t(t_0)$ with $t_0 = 0$, and $c_1 = \exp(-5)$, $c_2 = 0$ in (45). The interval of integration $[H_0, H_f]$ is given by

$$H_0 = \ln \left[\frac{1}{1 - t_0 u_t^0} \right], \quad H_f = \ln \left[\frac{1}{1 - t_f u_t^f} \right], \quad (46)$$

where $u_t^f = u_t(t_f)$ and $t_f = 2$. We choose the interval of integration given by (46), so that when t is given by (44) it lies in the interval $[0, 2]$.

Figure 5 shows the error obtained for the invariant PINN model when compared to the exact solution (42), and similarly for the non-invariant PINN model. As in all previous examples, the invariant version drastically outperforms the standard PINN approach.

Example 8 As our final example, we consider a system of first order ODEs

$$u_t = -u + (t + 1)v, \quad v_t = u - tv. \quad (47)$$

This system admits a two-dimensional symmetry group of transformations given by

$$T = t, \quad U = \alpha u + \beta t, \quad V = \alpha v + \beta,$$

where $\alpha > 0$ and $\beta \in \mathbb{R}$. Working with the cross-section $\mathcal{K} = \{u = 1, v = 0\}$, the invariantization of (47) yields

$$I = \iota(u_x) = -1, \quad J = \iota(v_x) = 1.$$

The reconstruction equations are

$$\alpha_t = \alpha(1 + t), \quad \beta_t = \alpha \quad (48)$$

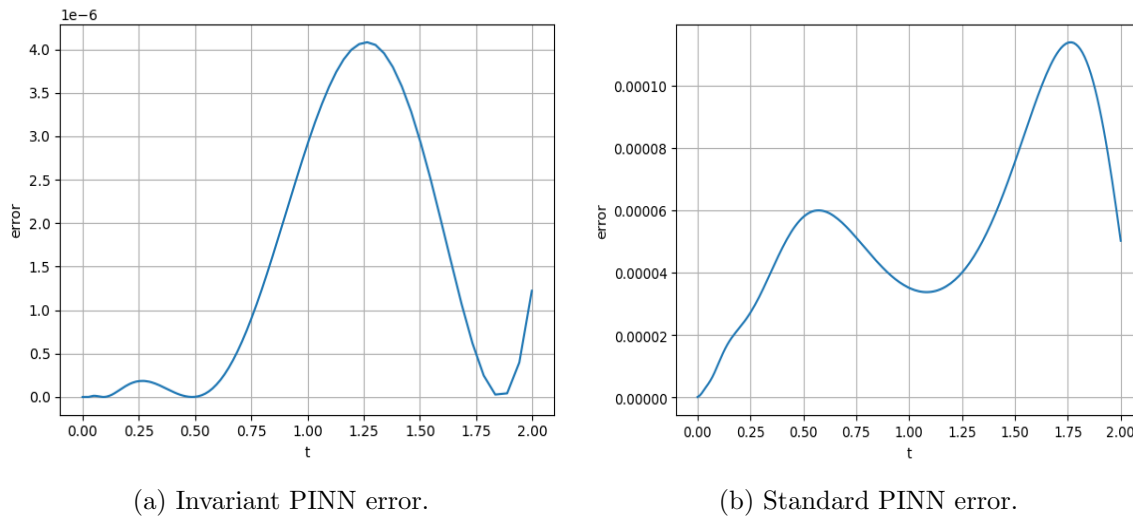


Figure 5: Time series of the squared error for the exponential equation (42).

subject to the initial conditions $\alpha_0 = 1$, $\beta_0 = 1$, corresponding to the initial conditions $u_0 = v_0 = 1$, when $t_0 = 0$. In our numerical simulations we integrated over the interval $[0, 2]$. The solution to (47) is then given by

$$u(t) = \alpha(t) + t\beta(t), \quad v(t) = \beta(t).$$

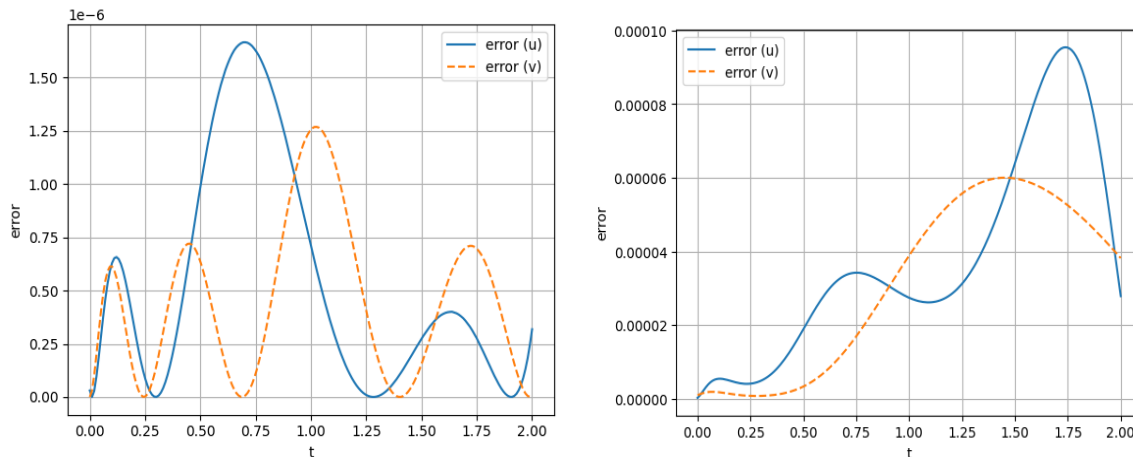
As in all previous examples, comparing the numerical solutions to the exact solution

$$u(t) = \sqrt{\frac{2}{\pi}} c e^{-(t+1)^2/2} + ct \operatorname{erf}\left(\frac{t+1}{\sqrt{2}}\right) + kt, \quad v(t) = c \operatorname{erf}\left(\frac{t+1}{\sqrt{2}}\right) + k,$$

with $c = (\sqrt{2/\pi} \exp(-1/2))^{-1}$ and $k = 1 - c \operatorname{erf}(1/\sqrt{2})$, where $\operatorname{erf}(t) = 2/\sqrt{\pi} \int_0^t e^{-x^2} dx$ is the standard error function, we observe in Figure 6 that the invariant version of the PINN model considerably outperforms its non-invariant counterpart.

6. Summary and conclusions

In this paper we have introduced the notion of invariant physics-informed neural networks. These combine physics-informed neural networks with symmetry methods for differential equations to simplify the form of the differential equations that have to be solved. In turn, this simplifies the loss function that has to be minimized. For example, in the case of the Schwarz equation considered in Examples 4, the third order nonlinear differential equation (12) with $F(t) = 2$, is replaced by the system of first order linear reconstruction equations (36a). Similarly, the nonlinear logistic equation (37) in Example 5 is replaced by the linear reconstruction equation (38). The same phenomenon occurs in Example 7, where the nonlinear equation (42) is substituted by the system of linear equations (43). Finally, in Example 8 the coupled system of linear differential equations (47) is replaced by the triangular system of differential equations (48). Our numerical tests show that the solutions obtained with the (simplified) invariant models outperformed their non-invariant



(a) Invariant PINN error.

(b) Standard PINN error.

Figure 6: Time series of the squared error for the system of equations (47).

counterparts. Table 1 summarizes the examples considered in the paper and shows that the invariant PINNs drastically improve over vanilla PINNs for all examples considered. Quoting Perlis (1982) “Symmetry is a complexity-reducing concept . . .” which in the context of the current paper is, we conjecture, at the source of the numerical improvements.

Example	Invariant PINN	Vanilla PINN
Schwarz (12)	$5.1 \cdot 10^{-2} \pm 4.1 \cdot 10^{-2}$	96.89 ± 0.81
Logistic (37)	$2.6 \cdot 10^{-10} \pm 1.8 \cdot 10^{-10}$	$3 \cdot 10^{-2} \pm 4.6 \cdot 10^{-5}$
Harmonic (40)	$5.8 \cdot 10^{-6} \pm 4.8 \cdot 10^{-6}$	$1.4 \cdot 10^{-5} \pm 2.3 \cdot 10^{-5}$
Exponential (42)	$3.2 \cdot 10^{-7} \pm 2.2 \cdot 10^{-7}$	$2.6 \cdot 10^{-5} \pm 3.6 \cdot 10^{-5}$
System (47)	$9.4 \cdot 10^{-7} \pm 2.0 \cdot 10^{-7}$	$6.6 \cdot 10^{-6} \pm 7.9 \cdot 10^{-6}$

Table 1: Mean square error with standard deviation averaged over five runs for all examples considered in Section 5.

The proposed method is fully algorithmic and as such can be applied to any system of differential equations that is strongly invariant under the prolonged action of a group of Lie point symmetries. It is worth noting that the work proposed here parallels some of the work on invariant discretization schemes which, for ordinary differential equations, also routinely outperform their non-invariant counterparts. We have observed this to also be the case for physics-informed neural networks.

Lastly, while we have restricted ourselves to the case of ordinary differential equations, our method extends to partial differential equations as well. Though, when considering partial differential equations, it is not sufficient to project the differential equation onto the space of differential invariants as done in this paper. As explained in Thompson and Valiquette (2015), integrability conditions among the differential invariants must also be

added to the invariantized differential equation. In the multivariate case, the reconstruction equations (28) will then form a system of first order partial derivatives for the left moving frame. Apart from these modifications, invariant physics-informed neural networks can also be constructed for partial differential equations, which will be investigated elsewhere.

Acknowledgement

This research was undertaken, in part, thanks to funding from the Canada Research Chairs program and the NSERC Discovery Grant program. The authors also acknowledge support from the Atlantic Association for Research in the Mathematical Sciences (AARMS) Collaborative Research Group on *Mathematical Foundations for Scientific Machine Learning*.

References

- A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind. Automatic differentiation in machine learning: a survey. *J. Mach. Learn. Res.*, 18:Paper No. 153, 2018.
- A. Bihlo. Invariant meshless discretization schemes. *J. Phys. A: Math. Theor.*, 46(6):062001 (12 pp), 2013.
- A. Bihlo. Improving physics-informed neural networks with meta-learned optimization. *Journal of Machine Learning Research*, 25(14):1–26, 2024.
- A. Bihlo and J.-C. Nave. Convecting reference frames and invariant numerical models. *J. Comput. Phys.*, 272:656–663, 2014.
- A. Bihlo and R. O. Popovych. Physics-informed neural networks for the shallow-water equations on the sphere. *J. of Comput. Phys.*, 456:111024, 2022.
- A. Bihlo and F. Valiquette. Symmetry-preserving numerical schemes. In *Symmetries and Integrability of Difference Equations*, pages 261–324. Springer, 2017.
- A. Bihlo and F. Valiquette. Symmetry-preserving finite element schemes: An introductory investigation. *SIAM J. Sci. Comput.*, 41(5):A3300–A3325, 2019.
- A. Bihlo, J. Jackaman, and F. Valiquette. Invariant variational schemes for ordinary differential equations. *Stud. Appl. Math.*, 148(3):991–1020, 2022.
- S. Blanes and F. Casas. *A Concise Introduction to Geometric Numerical Integration*, volume 23 of *Mono-graphs and Research Notes in Mathematics*. CRC Press, Boca Raton, 2016.
- G. W. Bluman, A. F. Cheviakov, and S. C. Anco. *Application of symmetry methods to partial differential equations*, volume 168 of *Applied Mathematical Sciences*. Springer, New York, 2010.
- G.W. Bluman and S. Kumei. *Symmetries and differential equations*. Springer, New York, 1989.
- R. Brecht, D. R. Popovych, A. Bihlo, and R. O. Popovych. Improving physics-informed deeponets with hard constraints. *arXiv preprint arXiv:2309.07899*, 2023.
- C. Budd and V. A. Dorodnitsyn. Symmetry-adapted moving mesh schemes for the nonlinear Schrödinger equation. *J. Phys. A*, 34(48):10387–10400, 2001.
- T. Cohen and M. Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.
- V. A. Dorodnitsyn. Transformation groups in mesh spaces. *J. Sov. Math.*, 55(1):1490–1517, 1991.

- M. Fels and P. J. Olver. Moving coframes. II. Regularization and theoretical foundations. *Acta Appl. Math.*, 55:127–208, 1999.
- M. Finzi, M. Welling, and A. G. Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In *ICML*, pages 3318–3328. PMLR, 2021.
- E. Hille. *Ordinary Differential Equations in the Complex Domain*. John Wiley & Sons, New York, 1976.
- P. E. Hydon. *Symmetry methods for differential equations*. Cambridge University Press, Cambridge, 2000.
- A. D. Jagtap, E. Kharazmi, and G. E. Karniadakis. Conservative physics-informed neural networks on discrete domains for conservation laws: applications to forward and inverse problems. *Comput. Methods Appl. Mech. Eng.*, 365:113028, 2020.
- P. Kim. Invariantization of numerical schemes using moving frames. *BIT Numerical Mathematics*, 47(3): 525–546, 2007.
- D.P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. URL <https://api.semanticscholar.org/CorpusID:6628106>.
- I. A. Kogan and P. J. Olver. Invariant Euler–Lagrange equations and the invariant variational bicomplex. *Acta Appl. Math.*, 76(2):137–193, 2003.
- A. Krishnapriyan, A. Gholami, S. Zhe, R. Kirby, and M. W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. *Advances in Neural Information Processing Systems*, 34: 26548–26560, 2021.
- I. E. Lagaris, A. Likas, and D. I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE Trans. Neural Netw.*, 9(5):987–1000, 1998.
- J.L. Lagrange. Sur la construction des cartes géographiques. *Oeuvres complètes, tome 4, Nouveaux mémoires de l'académie royale des sciences et belles-lettres de Berlin*, pages 637–692, 1779.
- L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.*, 3(3):218–229, 2021.
- E. L. Mansfield. *A practical guide to the invariant calculus*, volume 26. Cambridge University Press, 2010.
- L.D. McClenny and U.M. Braga-Neto. Self-adaptive physics-informed neural networks using a soft attention mechanism. *J. Comp. Phys.*, 474:111722, 2023.
- P. J. Olver. *Application of Lie groups to differential equations*, volume 107 of *Graduate Texts in Mathematics*. Springer–Verlag, New York, 1993.
- P. J. Olver. *Equivalence, invariants, and symmetry*. Cambridge University Press, Cambridge, 1995. ISBN 0521478111.
- P. J. Olver. Geometric foundations of numerical algorithms and symmetry. *Appl. Algebra Engrg. Comm. Comput.*, 11(5):417–436, 2001.
- P. J. Olver. *Equivalence, invariants and symmetry*. Cambridge University Press, Cambridge, 2009.
- V. Ovsienko and S. Tabachnikov. “What is ... the Schwarzian derivative?”. *AMS Notices*, 56:34–36, 2009.
- Alan J. Perlis. Special feature: Epigrams on programming. *SIGPLAN*, 17:7–13, 1982.
- M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378:686–707, 2019.

- R. Rebelo and F. Valiquette. Symmetry preserving numerical schemes for partial differential equations and their numerical tests. *J. of Diff. Eq. and Appl.*, 19(5):738–757, 2013.
- R. Rebelo and F. Valiquette. Invariant discretization of partial differential equations admitting infinite-dimensional symmetry groups. *J. Difference Equ. Appl.*, 21(4):285–318, 2015.
- R.A. Serway and J.W. Jewett. *Physics for Scientists and Engineers*. Brooks/Cole, 2003.
- Y.I. Shokin. *The method of differential approximation*. Springer, 1983.
- Z. Tang, Z. Fu, and S. Reutskiy. An extrinsic approach based on physics-informed neural networks for pdes on surfaces. *Mathematics*, 10(16):2861, 2022.
- R. Thompson and F. Valiquette. Group foliation of differential equations using moving frames. *Forum of Math.: Sigma*, 3:e22, 2015.
- M. Vaquero, J. Cortés, and D.M. de Diego. Symmetry preservation in hamiltonian systems: Simulation and learning, 2023.
- S. Wang, S. Sankaran, and P. Perdikaris. Respecting causality for training physics-informed neural networks. *Comp. Methods Appl. Mech. Engrg.*, 421:116813, 2024.
- N. N. Yanenko and Y. I. Shokin. The group classification of difference schemes for the system of equations of gas dynamics. *Tr. Mat. Inst. Akad. Nauk SSSR*, 122:85–97, 1973.