LEARNING CONTINUOUS GROUP SYMMETRIES FOR SIMPLIFYING AND SOLVING DIFFERENTIAL EQUATIONS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

İLKER GÜRCAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

MARCH 2025

Approval of the thesis:

LEARNING CONTINUOUS GROUP SYMMETRIES FOR SIMPLIFYING AND SOLVING DIFFERENTIAL EQUATIONS

submitted by **İLKER GÜRCAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Naci Emre Altun
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Halit Oğuztüzün
Head of Department, **Computer Engineering** _____

Assoc. Prof. Dr. Ramazan Gökberk Cinbiş
Supervisor, **Computer Engineering, METU** _____

Examining Committee Members:

Prof. Dr. Pınar Karagöz
Computer Engineering, METU _____

Assoc. Prof. Dr. Ramazan Gökberk Cinbiş
Computer Engineering, METU _____

Assoc. Prof. Dr. Elif Vural
Electrical-Electronics Engineering, METU _____

Assoc. Prof. Dr. Lale Özkahya
Computer Engineering, Hacettepe University _____

Assist. Prof. Dr. Gökhan Koray Gültekin
Electrical-Electronics Engineering, AYBU _____

Date:06.03.2025

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: İlker Gürcan

Signature :

ABSTRACT

LEARNING CONTINUOUS GROUP SYMMETRIES FOR SIMPLIFYING AND SOLVING DIFFERENTIAL EQUATIONS

Gürcan, İlker
Ph.D., Department of Computer Engineering
Supervisor: Assoc. Prof. Dr. Ramazan Gökberk Cinbiş

March 2025, 144 pages

Symmetry groups play a fundamental role in understanding the behavior of mathematical and physical systems governed by differential equations. This work presents a novel framework for learning continuous, **non-trivial** group symmetries in the context of Lie point transformations. The core idea of the proposed approach is to construct the symmetry group G by learning its infinitesimal action on a relevant jet space, given by the symmetry generators spanning its Lie algebra. Another critical component of this approach is the construction of the exponential map, $\exp$, which allows for the generation of any group element $g \in G$ from the learned $\mathfrak{g}$ and is implemented using invertible residual networks. In contrast to previous algebraic or analytical methods typically implemented in symbolic libraries, our iterative approach discovers the representation of infinitesimal symmetries on a given jet space through learning directly from input data of a given dynamic system. Moreover, our involutivity conditions are **decoupled** from the involutivity of the contact system describing the dynamic system. In our experiments, we validate the learned vector fields forming $\mathfrak{g}$ by demonstrating their ability to generalize to multiple solution domains beyond the training domain. Notably, we show that a canonical vector field for

the wave equation lies within the span of the learned $\mathfrak{g}$. Our framework has potential applications in discovering G-invariant coframes to simplify differential equation systems that are solvable by neural architectures. Additionally, this study aims to introduce the concept of continuous symmetry for analytic mathematical objects to the machine learning community.

Keywords: Group symmetries, Cartan equivalence, Moving (co)frame, Differential equations, Machine learning

ÖZ

SÜREKLİ SİMETRİ GRUPLARININ DİFERANSİYEL DENKLEMLERİN ÇÖZÜMÜ VE SADELEŞTİRİLMESİ AMACIYLA ÖĞRENİLMESİ

Gürcan, İlker
Doktora, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Ramazan Gökberk Cinbiş

Mart 2025 , 144 sayfa

Simetri grupları, diferansiyel denklemlerle tanımlanan matematiksel ve fiziksel sistemlerin davranışını anlamada temel bir rol oynar. Bu çalışma, Lie nokta dönüşümleri bağlamında **önem arz eden** sürekli grup simetrilerini öğrenmek için yeni bir çerçeve sunmaktadır. Önerilen yaklaşımın temel fikri, simetri grubu G'yi, Lie cebirini oluşturan simetri üreteçlerinin ilgili jet uzayı üzerindeki sonsuz küçük etkisini öğrenerek inşa etmektir. Bu yaklaşımın bir diğer kritik bileşeni, öğrenilen $\mathfrak{g}$'den herhangi bir $g \in G$ grup elemanını üretmeyi sağlayan üstel haritanın ($\exp$) oluşturulmasıdır; bu harita, terslenebilir artık ağlar kullanılarak gerçeklenmektedir. Geleneksel olarak sembolik kütüphanelerde uygulanan cebirsel veya analitik yöntemlerin aksine, önerilen yinelemeli yöntem, verilen bir dinamik sistemin giriş verilerinden doğrudan öğrenme yoluyla, ilgili jet uzayı üzerindeki sonsuz küçük simetri üreteçlerinin bir temsilini bulmaktadır. Ayrıca, öngörülen involütivite koşullarımız, dinamik sistemi tanımlayan kontak sisteminin involütivitesinden **ayrışmaktadır**. Deneylerimizde, öğrenilen vektör alanlarının $\mathfrak{g}$'yi oluşturduğunu doğruluyor ve eğitim alanının ötesindeki çözüm alanlarına genelleştirilebildiklerini gösteriyoruz. Özellikle, dalga denklemi için

bir kanonik vektör alanının öğrenilen $\mathfrak{g}$'nin doğrusal uzayında bulunduğunu ortaya koyuyoruz. Önerilen yöntem, diferansiyel denklem sistemlerinin çözülme sürecini sadeleştirmek için ilgili G-değişmez eş-çerçevesini keşfetmeye olanak tanıyabilir ve sinir ağlarıyla çözülebilir sistemlerin analizine katkı sağlayabilir. Ayrıca, bu çalışma, sürekli simetri kavramını analitik matematiksel nesneler bağlamında makine öğrenmesi topluluğuna tanıtmayı amaçlamaktadır.

Anahtar Kelimeler: Grup simetrileri, Cartan benzerliği, Hareketli (eş)çerçeveler, Diferansiyel denklemler, Makina öğrenme

To my wife, my family, my advisor, and Prof. Peter J. Olver for their unwavering support and guidance.

ACKNOWLEDGMENTS

İlker Gürcan was financially supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) and the Council of Higher Education (YÖK) in Türkiye by means of granting 2211-C and 100/2000 scholarships, respectively. The authors also extend their gratitude to TÜBİTAK for granting access to the National Grid System of Türkiye (TRUBA), which was instrumental in meeting the computational requirements during the development and testing of the associated software.

İlker Gürcan gratefully acknowledges the continuous support of Prof. Peter J. Olver over the past three years. His contributions included providing crucial theoretical insights, resolving numerous challenges related to the advanced mathematical background required for this research, and offering guidance on the state-of-the-art studies in continuous group symmetries of PDEs and G-coframe construction.

During the preparation of this work İlker Gürcan used ChatGPT in order to refine the grammar, syntax, and formal tone of the manuscript. After using this tool/service, İlker Gürcan reviewed and edited the content as needed and takes full responsibility for the content of the publication.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ALGORITHMS

ALGORITHMS

LIST OF ABBREVIATIONS

DE	Differential equations
ODE	Ordinary differential equation
PDE	Partial differential equation
FEM	Finite element method
FDM	Finite difference method
iResNET	Invertable residual network
EDS	Exterior differential system
VFS	Vector field system
M-C	Maurer-Cartan
PINN	Physics informed neural-network
IC	Initial condition
BC	Boundary condition
IBC	Inital & boundary conditions
KKT	Karush-Kuhn-Tucker

CHAPTER 1

INTRODUCTION

With advancements in computational power and the extensive data collected from dynamical systems, various computational methods have been developed to solve differential equations (DEs), including finite element methods (FEMs) [2], finite difference methods (FDMs) [3], single-step methods, and multi-step methods. However, these methods can quickly become computationally prohibitive, depending on factors such as the linearity of the DEs, the dimensionality of the domain, or the smoothness of the domain (see Section 11.6 in [4]). An important aspect of dynamical systems is learning the intrinsic, continuous symmetries that govern them. Identifying these group symmetries can lead to lower-dimensional problems [5], facilitate the translation of solutions between different basins within a given phase space [6], or enable the approximation of unknown densities of specified datasets that are equivariant under local group symmetries [7] and controlled by predefined dynamics. Moreover, the intersection of numerical solutions of DEs and the discovery of continuous symmetries for dynamical systems is the application of group symmetries to DEs in both theoretical and practical contexts [8].

We introduce the targeted problem in the following four subsections. First, we provide a concise definition and an example to illustrate how group symmetries can be employed to aid in solving DEs. Second, we will distinguish between two types of group symmetries: *trivial* and *non-trivial* (see Chapter 3.2.3 in [9]). Third, we shortly describe the significance of effectiveness and transitivity of the group action (see Definitions 12 and 13 resp.) and of the involutivity of a given differential system (see Section 3.5), both of which play an important role in the design of our framework. Additionally, another component in defining our problem is the *class of diffeomor-*

phisms that represent the *action* of the symmetry group G. This research involves learning about *Lie point* transformations, which is a subset of *contact* transformations (see Definition 22 and Chapter 4 in [10]), and will be deferred to Section 3. Fourth, we add a short subsection to clarify issues regarding accuracy, scalability, and applicability of our framework, in comparison to existing computational algebraic methods. Finally, we will outline the structure of this paper before concluding the section.

1.1 An Introduction to the Equivalence Problem

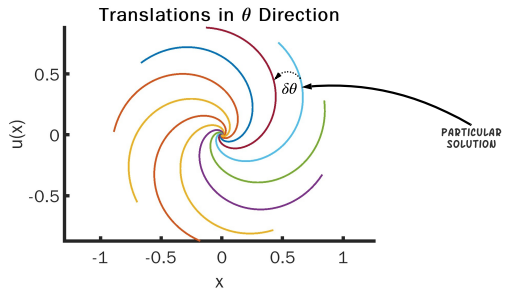

Figure 1.1: Various solutions of $\Delta^{(1)}$ in (1.1) under rotational symmetry. The graph of other solutions are derived from the particular solution colored in light blue.

Most scientific phenomena can be modeled by DEs; thus, characterizing the solutions of these DEs is crucial in science, engineering, and economics. In this context, "characterizing solutions" is defined as follows:

Definition 1. *Let $u = f(x)$ be a solution for a given set of DEs, $\Delta^{(s)}\left(x, u^{(s)}\right) = 0$, where $\mathbb{N} \ni s$ is the order of $\Delta^{(s)}$ and $u^{(s)}$ represents all partial derivatives of u w.r.t. x up to s. Suppose that $\Gamma_f = \{x, f(x)\}$ be the graph of f. If there exists a diffeomorphism $\phi : (x, u) \to (\hat{x}, \hat{u})$; s.t. $\Gamma_{\hat{f}} = \left\{\hat{x}, \hat{f}(\hat{x})\right\}$ is also the graph of another solution for $\Delta^{(s)}$, then $f(x)$ is equivalent to $\hat{f}(\hat{x})$ w.r.t. ϕ and is denoted by $\Gamma_f \cong_\phi \Gamma_{\hat{f}}$.*

In other words, we seek an "appropriate" collection of transformations that map one solution of $\Delta^{(s)}$ to another within their domain of definition. This collection is referred to as a Lie group (see Definition 3). Although these definitions may seem ab-

stract, their deeper meaning will become clear, once all required mathematical tools are explained coherently in Section 3. To illustrate this concept, consider the following simple, first-order ordinary differential equation (ODE) [11]. Let $p = 1$ and $q = 1$ represent the number of independent and dependent variables, respectively, for the jet space $J^{(1)}(M) = \mathcal{X} \times \mathcal{U}^{(1)}$ (see Definition 14), with coordinates given by (x, u, u_x), where u_x is partial derivative of u w.r.t. x and $M = \mathcal{X} \times \mathcal{U}$. Then:

$$\Delta^{(1)}(x, u, u_x) = u_x - \frac{u^3 + x^2 u - u - x}{x^3 + u^2 x + u - x} = 0 \tag{1.1}$$

defines a manifold $\Sigma^{(1)} \subset J^{(1)}(M)$. Although it appears challenging to separate and solve this ODE, recognizing that $SO(2)$, rotation in $x - u$ plane, is a *symmetry* for $\Delta^{(1)}$ (see Figure 1.1), it simplifies the problem significantly. The previously difficult equation now becomes:

$$\frac{d\theta}{dr} = \frac{1}{r(1 - r^2)}, \tag{1.2}$$

where (θ, r) is polar coordinates on M. There are two important takeaways from (1.2):

(i) The number of variables is reduced by one: $(x, u, u_x) \to (\theta_r, r)$, where $\theta_r = {}^{d\theta}/_{dr}$.

(ii) The dependent variable θ_r is separated from the independent variable r.

These two concepts form the foundation for studying continuous group symmetries to simplify and solve DEs. A parallel can be drawn between kernel methods in statistical analysis and the process of learning group symmetries for DE systems, as both approaches aim to represent the problem in a simpler space to mitigate the effects of the *curse of dimensionality* [12]. We refer reader to Chapter 4, 6, and 7 in [8] for further application of such group symmetries to variational problems, Hamiltonion Systems, and evolution equations respectively.

1.2 Type of the Symmetry

As discussed at the beginning of this section, defining the type of symmetry we are targeting is crucial, as it outlines the scope and purpose of our problem. First, we

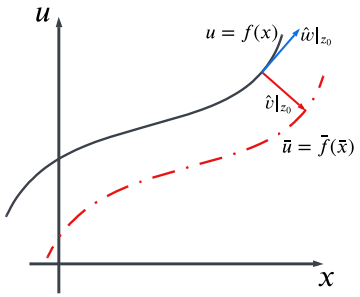

Figure 1.2: Non-trivial and trivial symmetry types generated by vector fields $\hat{v}|_x$ and $\hat{w}|_x$, respectively.

explain what is meant by the "type" of a symmetry group G. However, a precise mathematical definition will not be provided until Section 3.4, as it necessitates several other mathematical tools as prerequisites. Fortunately, this intuitive explanation should clarify the core concept for a wide range of audiences.

Definition 2. *Let $\hat{v}|_x \in \mathfrak{X}(M)$ be a vector field on the manifold M. If the graph of any solution $u = f(x)$ is translated to the graph of a new solution $\hat{u} = \hat{f}(\hat{x})$ under the flow generated by $\hat{v}|_x$, then the one-parameter symmetry group associated with $\hat{v}|_x$ (see Definition 10) is of non-trivial type. Conversely, if this flow leaves Γ_f invariant, then the symmetry is of trivial type. See Figure 1.2 for an illustration.*

Non-trivial symmetries (see Definition 26) are associated with canonical coordinates in which a given DE system becomes separable, as demonstrated in the example in Section 1.1. Therefore, we can conclude that (θ, r, θ_r) represents the canonical coordinate system for $\Delta^{(1)}$ as presented in (1.1). Trivial symmetries are ineffective in simplifying DE systems in regard to the two remarks listed at the end of Section 1.1. On the other hand, non-trivial symmetries provide these canonical coordinates (see Chapter 2 in [11] and "Equivalence and Trivial Symmetries" section in Chapter 5 in [8] for more information). Thus, the objective of this study is to identify such non-trivial symmetries that also form a (local) Lie group.

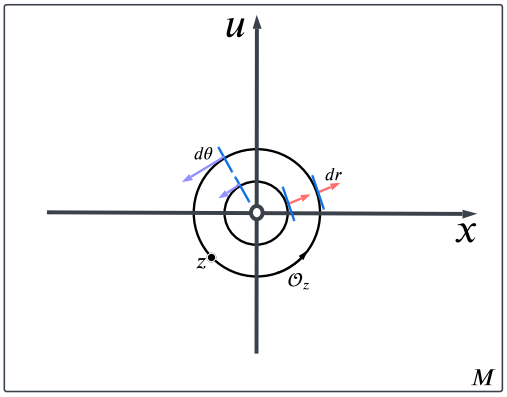

Figure 1.3: The orbits (circles centered at the origin) of $G = SO(2)$ are isomorphic to each other in $M = \mathbb{R}^2 \setminus \{\mathbf{0}\}$ due to transitive group action when the action is restricted to a particular orbit $\mathcal{O}_{z=(x,u)}$. $= \Omega \{d\theta, dr\}$ is a coframe (see Remark 6) on M

1.3 Transitive and Effective Actions & Involutivity of Differential Systems

The concepts of *effective* and *transitive* group actions (see Definition 12 and Definition 13 resp.), along with the *involutivity* of exterior differential systems (see Section 3.5), are pivotal in both determining a complete invariant coframe (see Section 5 in [13] and Chapter 8 in [10]) Ω on $J^{(s)}(M)$ w.r.t. a given equivalence ϕ and the integrability of DE systems within their defined spaces (see Chapters 14 and 15 in [10], as well as Chapter 4 in [14]). Although we do not introduce the notion of *Cartan equivalence*, under a specified ϕ, we strongly encourage reader to go through the first four Chapters and Example 4 in Chapter 5 in [15] as Cartan's equivalence problem is the root of any concept discussed in this study. Transitive and locally effective group actions enable the uniform treatment of equivalent coframes, when they are restricted to different integral submanifolds (see Definition 37). Specifically, each orbit of G, induced by the equivalence ϕ, possesses the same maximal dimension, and these orbits are isomorphic to one another as manifolds: $\mathcal{O}z_1 \cong \mathcal{O}z_2$, where $z_1 \neq z_2 \in M$. An illustration of a transitive group action is provided in Figure 1.3. Regarding the involutivity of a given exterior differential ideal $\mathcal{I}$ (see Section 3.5), which ensures the consistency of the vector space tangent to an integral submanifold N, it is instrumental in establishing the existence/unique of N. However, the existence/uniqueness of N for non-simply-generated exterior differential systems (see Definition 35) neces-

sitates a more comprehensive analysis of such differential systems (refer to Cartan's involutivity test in Chapters 11 and 15 of [10]). Readers, familiar with the existence and uniqueness theorems for ODEs (see Theorem 2.8.1 in [16]), may draw parallels with the Frobenius' or Cartan-Kähler theorem (see Theorem 14.1 & 15.7 in [10] resp.). The transitivity of a group action also significantly impacts the involutivity of a given coframe. When restricted to an orbit where G acts transitively, the structure equations (see Chapter 3 in [15]) associated with a coframe Ω on M exhibit constant *essential torsion* coefficients. These coefficients correspond to the terms of the form $\theta^j \wedge \theta^k$ in (15.26) of [10]. This transitivity simplifies the process of establishing the involutivity of Ω, reducing it to verifying the involutivity of a rank-0 coframe (see Theorem 14.26 in [10]), provided the integrability conditions defined by the Cartan-Kähler theorem are satisfied.

In summary, this research assumes transitive, once restricted to a certain orbit of G, and locally effective group actions with every orbit being locally *transverse*. If these conditions are not met, the framework may fail to function properly, as elaborated in Section 4. The transitivity condition is also crucial when selecting data points for sample datasets (see Section 5 for further details).

1.4 Accuracy, Scalability, and Applicability

Algebraic methods, relying on symbolic manipulations, focus on obtaining precise, pointwise solutions for symmetry generators by solving the ***determining equations*** exactly (see the discussion in Section 2 for further details). In contrast, our data-oriented framework employs neural networks to learn approximate solutions across the entire domain, offering a global perspective. While algebraic methods provide exact solutions or an analytic approximation in the vicinity of specific points[17], they require re-computation of initial conditions and Taylor approximation steps for each new point, potentially leading to inefficiencies in scenarios necessitating solutions over extensive domains. Since their solutions are exact, closed-form mathematical expressions, the accuracy comparison with the approach, we propose, is not reasonable.

Algebraic methods also face significant scalability challenges as the number of determining equations can explode with the number of independent/dependent variables and the nonlinearity of the system. For example, even for relatively simple systems like the heat equation in one dimension, there are already ten linear PDEs to solve (see Example 2.41 in [8]). For the sake of illustration, if one more spatial dimension was added to the linear heat equation, then it would have become:

$$u_t = u_{xx} + u_{yy}. \tag{1.3}$$

Assume that any symmetry generator for (1.3) is in the following form (see (3.18)):

$$\hat{v} = \hat{\xi}(x, y, t, u)\frac{\partial}{\partial x} + \hat{\eta}(x, y, t, u)\frac{\partial}{\partial y} + \hat{\tau}(x, y, t, u)\frac{\partial}{\partial t} + \hat{\phi}(x, y, t, u)\frac{\partial}{\partial u}. \tag{1.4}$$

Then, using the Matlab's Symlink library[18], the following determining equations are obtained:

$$\begin{aligned}
\hat{\tau}_u = \hat{\xi}_u = \hat{\eta}_u &= 0,\\
\hat{\phi}_{uu} = \hat{\tau}_{xu} = \hat{\tau}_{yu} &= 0,\\
\hat{\xi}_{xu} + \hat{\eta}_{yu} &= 0,\\
\phi_{xu} - \hat{\xi}_{xx} - \hat{\xi}_{yy} + \hat{\tau}_t - 2\hat{\xi}_x &= 0,\\
\hat{\phi}_{yu} - \hat{\eta}_{xx} - \hat{\eta}_{yy} + \hat{\tau}_t - 2\hat{\eta}_y &= 0,\\
\hat{\phi}_{tu} - \hat{\xi}_{tx} - \hat{\eta}_{ty} + \hat{\tau}_t - \hat{\xi}_x - \hat{\eta}_y &= 0,\\
\hat{\phi}_t - \hat{\phi}_{xx} - \phi_{yy} &= 0,\\
\hat{\tau}_x = \hat{\tau}_y &= 0,\\
\hat{\tau}_t - 2\hat{\xi}_x &= 0,\\
\hat{\tau}_t - 2\hat{\eta}_y &= 0.
\end{aligned} \tag{1.5}$$

These fifteen determining equations must be solved simultaneously to obtain the coefficient functions of the various symmetry generators $\hat{v}$ in (1.4). As evident from the comparison with the one-dimensional case, which requires solving only ten determining equations, the complexity of the system grows significantly with an increasing number of independent variables. This highlights the rapid escalation in computational difficulty as the dimensionality of the base manifold increases. For more complex systems, the number of determining equations can reach the order of hundreds or even thousands, rendering algebraic approaches computationally intractable.

In contrast, our framework exhibits an approximately **linear** growth in the number of learnable parameters with respect to the number of independent and dependent variables, i.e., the dimension of the base manifold M. Furthermore, the presence of non-linearities in the DE system does not alter the fundamental structure of our method; at most, it may lead to a slight increase in optimization time.

Regarding the applicability of algebraic approaches, these methods rely on the availability of a closed-form system of determining equations expressed in terms of differential polynomials of the coefficient functions. In the absence of such a system, algebraic methods fail entirely. By contrast, our data-driven approach does not require explicit knowledge of the determining equations and remains applicable even when the underlying system is non-polynomial, partially unknown, or highly complex. As discussed in Sections 2 and 6, an example of this applicability can be seen in Neural-ODE-based architectures [19], which can learn the dynamics of a system in its phase space from observed data. Since such learned dynamics are first-order ODEs expressed in terms of phase-space variables and the time variable t, the output of a Neural-ODE model can serve as an input to our framework for symmetry discovery. This, in turn, enables the construction of an **involutive** G-coframe (see Example 4.12 in [13] for constructing a G-coframe, provided that the symmetry group G is known), ultimately facilitating the simplification of the ODE system before attempting to solve it. Reformulating the system in this manner not only enhances the accuracy of the solution but also reduces the computational complexity associated with its numerical integration.

1.5 Structure of the Paper

The remainder of this paper is organized as follows: Section 2 presents a comprehensive literature review; Section 3 discusses the mathematical foundations; Section 4 outlines the methodology we developed; Section 5 details the experimental results; and Section 6 offers concluding remarks.

In Section 2, we begin by examining methods that adopt an algebraic approach, noting that our framework in Section 4 is analytical due to the intrinsic nature of neural

networks. The section then covers significant advancements in machine learning related to modeling and solving differential equations (DEs) and concludes with recent studies leveraging group symmetries to improve the quality of solutions for DE systems.

Section 3 provides the essential mathematical components required to formulate our problem and establish the necessary notation. Readers familiar with applying group symmetries to DE systems may skim this section, focusing on the **notation** and **assumptions** (typically located toward the end of each subsection) used throughout this work. The final two sections of this chapter are dedicated to the prospective continuation of the research presented in this thesis and are essential for ensuring the completeness of this study. The notation introduced here will also be used in **future relevant studies**. In summary, this chapter serves as a concise reference for fundamental concepts, which the reader may revisit as needed throughout the paper. Additionally, it includes several remarks to clarify potential ambiguities arising from varied terminologies across sources, a challenge the authors have already encountered.

Section 4 describes our contributions to continuous symmetry learning. It begins by modeling the Lie algebra (see Definition 7). The second subsection discusses the implementation of involutivity for the vector field system (see Definition 39) $\mathcal{V}$, that is generated by the action of non-trivial symmetry generators (see Definition 26) for the DE system of interest. In the final subsection, we address how to construct any group action from the learned infinitesimal group action that is associated with $\mathfrak{g}$.

Section 5 presents the experimental results of the developed framework. It begins by detailing the input data, then demonstrates how the generated dataset validates the learning process against a DE system with known non-trivial group symmetries, as well as assessing the generalization of learned symmetries across distinct domains.

In Section 6, we summarize our contributions to symmetry learning using machine learning tools and outline future directions. The scope of the current study is restricted to geometric subgroups of symmetries for DE systems under fundamental involutivity conditions. We provide a sequential roadmap for future research based on these findings. Moreover, we include various advantages of the proposed framework over

existing computational algebraic and analytical approaches.

CHAPTER 2

LITERATURE REVIEW

The development of symbolic computation libraries such as MAPLE, REDUCE, and Mathematica in the 1980s led to the creation of various iterative, algebraic algorithms. These algorithms derive the determining equations by applying the infinitesimal symmetry condition specified in (3.35). The resulting system is an overdetermined, linear, and homogeneous system formulated within the R-module of differential operators, $D_{\mathbb{K}}(R)[\boldsymbol{\xi}, \boldsymbol{\varphi}]$ (see [20]). Here, R represents a differential ring of polynomials in jet space coordinates, defined over the field $\mathbb{K}$ and closed under partial differentiation with respect to these coordinates (see Section 3.2). The variables $(\boldsymbol{\xi}, \boldsymbol{\varphi})$ correspond to the component functions of the symmetry generators (see Section 3.1), which act as indeterminates in the module $D_{\mathbb{K}}(R)[\boldsymbol{\xi}, \boldsymbol{\varphi}]$. The system is then transformed into a canonical (standard) form, which is the involutive completion of a reduced form. The algorithm for achieving this standard form comprises two main steps: reducing the system to a simpler form and completing the reduced form by adding integrability conditions to achieve involutivity (as implemented in the **AutoReduce** and **CompleteSystem** algorithms, respectively, in [21]). Reid et al. further enhanced these methods in [17] by developing techniques that automatically determine the initial conditions needed to obtain a Taylor approximation of the solution to the determining equations, in addition to implementing the *standard form algorithm*. These algorithms for obtaining a closed-form or Taylor approximation of $(\boldsymbol{\xi}, \boldsymbol{\varphi})$, the component functions of symmetry generators, trace their roots to the pioneering work of Maurice Janet in the early 20$^{\text{th}}$ century, as summarized in [22]. Janet's approach establishes a profound connection between monomials in the polynomial ring $\mathbb{K}[x^1, \dots, x^n, \dots, u^{\alpha}_{x^{I_k}}, \dots]$ and elements of the ring of polynomial differential operators $\mathcal{D}(\partial_{x^1}, \dots, \partial_{x^n}, \dots, \partial_{u^{\alpha}_{x^{I_k}}}, \dots)$, which act on the functions de-

fined in a given jet space, where I_k is a multi-index into independent variables of order k (see Section 4 in [22]). This parallelism forms the basis for treating determining equations algebraically, leveraging techniques like reduction to a standard form and systematic elimination to solve the overdetermined linear system effectively. Unlike the approach in this study, which learns the non-trivial symmetry generators, that are intrinsic to a given input dataset governed by a DE system; those algebraic methods do not differentiate between trivial and non-trivial symmetries (see Definition 26). For a comprehensive theoretical and computational background on the *modern algebraic approach* to solving determining equations, we refer the reader to the textbook "Involution: The Formal Theory of Differential Equations and Its Applications in Computer Algebra" by Werner M. Seiler [23]. In addition to its relevance for determining equations of DE systems, this book serves as an excellent resource for learning computational algebra methods applicable to solving various types of DE systems.

Another collection of methods, which bears a resemblance to our framework (though still an algebraic approach), is based on contact systems (see Definition 38) and the Cartan-Kähler theorem (see Theorem 15.7 in [10]). Hartley et al. developed such an algorithm in [24] by converting a given system of DEs into a contact system and constructing a chain of regular integral submanifolds (see Definition 37) by ensuring the involutivity condition as dictated by Frobenius' theorem (see Theorem 5), at each step of extending the chain. However, this approach may fail in certain pathological cases where the DE system is not involutive (see the discussion towards the end of Chapter 2 in [24]). To address such cases, they improved their method for non-involutive systems in [25]. It is important to note that the concept of "involutivity" in this type of algorithm is distinct from the involutivity introduced by Janet et al. [22] for the previously discussed algebraic algorithms. In essence, involutivity in the contact system is a geometric property ensuring integrability of the differential structure, while involutivity in Janet's standard form ensures that the determining equations are algebraically consistent and computationally complete. The studies, in this category, convert a given DE system to an exterior differential system (see Definition 36) and construct a chain of integral elements, that are tangent to integral submanifolds. However, our research focuses on the **complementary** integral submanifold, which is the foliation of the flow generated by non-trivial, contact preserving (see Definitions 26

and 22) symmetry generators for the DE system of interest. For a comprehensive survey of algorithms developed in symbolic algebra systems, the reader is referred to Chapter 13 in [26].

Emerging methods for solving DEs have become increasingly data-driven. Of these, three approaches have garnered significant attention: *(i)* Physics-Informed Neural Networks (PINNs) [27], *(ii)* Koopman Operator Theory-based models [28], & *(iii)* Neural ODE-like frameworks [19]. PINNs seek to approximate a solution $\tilde{u}_\theta$ to the equation $\Delta^{(s)} = 0$, where θ represents a set of parameters learned to approximate $\tilde{u}_\theta$. PINNs exploit the *universal approximation* property of neural networks (NNs) [29] and incorporate initial/boundary conditions as constraints. In contrast, Koopman Operator-based methods primarily address nonlinear dynamics within a given phase space [30]. Originally introduced by Bernard Koopman in 1931 [28], this method experienced a resurgence following an in-depth analysis of its spectral properties [31]. Since then, it has drawn considerable interest from the machine learning community. Although a nonlinear system can typically only be linearly approximated in a local neighborhood, Koopman theory posits that a transformed phase space can be identified in which the dynamics of interest become linear. Finally, Neural ODE-like methods, such as the one described in [19], aim to learn continuous dynamics, in contrast to residual networks, which are designed to learn discrete-time dynamics. A recent study further refines this approach by learning the manifolds on which the dynamics occur through constructing the underlying atlas [32]. This approach is based on the premise that a nonlinear manifold lacks a *global parameterization* and therefore requires more than one coordinate chart. Consequently, nonlinear dynamics, which cannot be accurately captured by linear manifold learning methods, are reconstructed from data with high precision.

The study of continuous groups and their numerical applications has traditionally been confined to the realms of mathematics and the physical sciences. Historically, the engineering community has exhibited limited interest in continuous symmetries. However, there has been a recent shift toward recognizing their relevance in addressing engineering problems. This interest, in particular, was sparked by Mallat's work in [5], which introduced a theoretical framework explaining how convolutional neural networks (CNNs) learn group symmetries through their spectral properties. As for the

realm of DE systems, most existing studies focus on leveraging known group symmetries of a given DE system to improve problem-solving frameworks. For instance, Mialon et al. [33] demonstrated that enforcing symmetry constraints enhances the quality of representations for objects in computer vision tasks. Similarly, Akhoundi et al. [34] showed that predefined group symmetries improve the accuracy of solutions computed via PINNs. In another notable contribution, Finzi et al. [35] developed neural network layers that are equivariant under predefined general linear groups, such as $SO(2)$ and $SO(3)$ [1], and applied this framework to dynamical systems. Of particular relevance to our work is the research by Otto et al. [36], which extends beyond enforcing known symmetries to also extracting novel symmetries inherent in a given dynamical system dataset. However, in contrast to the present study, their architecture is restricted to general linear groups acting on vector bundles. Despite this limitation, their framework is a valuable tool for learning equivariant Koopman operators acting on *linearized* subspaces.

CHAPTER 3

MATHEMATICAL FOUNDATIONS

In this chapter, we present the essential mathematical background required to formulate our target problem in Sections 4 and 5. While this chapter will consolidate relevant information, it will not merely reiterate what is already available in the literature. Consequently, we will frequently direct the reader to pertinent resources for more comprehensive details. Additionally, any propositions, theorems, or lemmas requiring proof will have their proofs provided in the Appendix A, ensuring that the main discussion remains uncluttered.

We begin by introducing the necessary notations, definitions, and mathematical tools from the fields of differential geometry, group theory, modern algebra, calculus, and functional analysis. Next, we provide a precise definition of Cartan's equivalence problem, as our framework ultimately aims to address a subproblem within this context. Finally, we discuss the moving coframes method developed by Olver et al. through a series of articles. Unlike Cartan's framework, which is an algebraic framework, moving coframes is an analytical tool.

Remark 1 (Notational Convention)**.** *In this chapter, we establish the notational conventions that will be consistently applied throughout this paper, including in the preceding two chapters. Any terminology or notation introduced earlier in the text may not be redefined later, and all subsequent references will assume the initial definition.*

3.1 Lie Groups

In this subsection, we present the continuous group terminology that will be utilized throughout this study. The notation, definitions, and related concepts closely follow

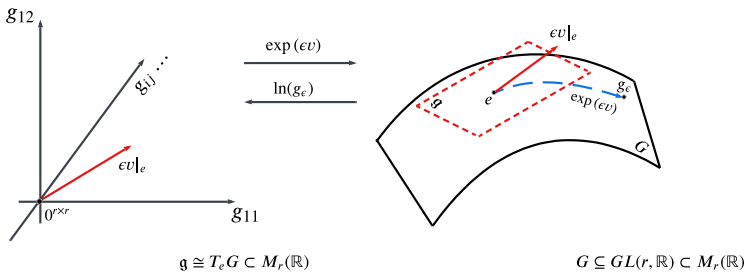

Figure 3.1: An illustration of various concepts discussed in Section 3.1. A sketch of the general linear group $GL(r,\mathbb{R})$. The coordinates g_{ij} belong to the space of $r \times r$ matrices denoted by $M_r(\mathbb{R})$. $GL(r,\mathbb{R})$ represents the continuous set of invertible matrices, and hence $GL(r,\mathbb{R}) \subset M_r(\mathbb{R})$. The exponential map $\exp$ in (3.10) and the logarithm map ln facilitate the transition between the Lie algebra $\mathfrak{g}$ (see Definition 7) and $GL(r,\mathbb{R})$ (refer to Chapter 2.3 in [1] for more details on the ln map).

those provided in [10], [1], and [37]. Broadly defined, a *Lie group* is a group that has continuous group operations. Formally:

Definition 3. *A Lie group G is a smooth manifold equipped with continuous group multiplication, $(g,h) \to g \cdot h$, and inversion, $g \to g^{-1}$, operations that satisfy the group axioms, specifically:*

(i) the existence of a unique identity element $e \in G$,

(ii) the associativity of group multiplication, $(g \cdot h) \cdot k = g \cdot (h \cdot k)$, and

(iii) the existence of an inverse element $g^{-1} \in G$ for every $g \in G$.

For readers unfamiliar with continuous group theory, a Lie group can be thought of as a topological set of invertible, continuous transformations, closed under the composition of those transformations. Additionally, G can act on objects other than itself via its representation on the space where those objects are defined.

Definition 4. *The action of a Lie group on a smooth manifold M is defined by a diffeomorphism $\Psi : G \times M \to M$ and satisfies the following:*

$$\begin{aligned} \Psi(e,x) &= x, \text{ and} \\ \Psi(g \cdot h, x) &= \Psi(g, \Psi(h,x)), \end{aligned} \tag{3.1}$$

where $x \in M$ and $g, h \in G$.

For brevity, we will also denote the group action by $\Psi(g, x) \equiv g \cdot x$. It is common for Lie group multiplication and/or its action on M to be undefined globally (see Example 1.27/(c) in [8]). Therefore, this research will focus on *local* group actions given by $g \in G$, such that g is sufficiently close to e (in a topological sense). We define the local Lie group action in the vicinity of $(e, x) \in G \times M$ as $G_x \coloneqq \{g \,|\, (g, x) \in \mathscr{U}\}$, where $\{e\} \times M \subset \mathscr{U} \subset G \times M$ is an open set. This locality definition induces an open submanifold $M_g \coloneqq \{x \,|\, (g, x) \in \mathscr{U}\}$.

Every Lie group is associated with a Lie algebra, denoted by $\mathfrak{g}$, which is isomorphic to the tangent space at the identity element, specifically T_eG. Before providing a rigorous definition, it is necessary to define two types of group multiplication and the concept of a vector field on any tangent bundle, such as TG.

Definition 5. *Left multiplication by $g \in G$, denoted by $L_g : G \to G$, is a diffeomorphism on G and is given by:*

$$L_g(h) \coloneqq g \cdot h, , \forall, h \in G. \tag{3.2}$$

Similarly, right multiplication, denoted by $R_g(h)$, is defined as:

$$R_g(h) \coloneqq h \cdot g^{-1}. \tag{3.3}$$

The inverse is required to satisfy the group action composition law in (3.1) *(i.e., the second equation).*

We denote a vector field on any manifold M by $v \in \mathfrak{X}(M)$, where $\mathfrak{X}(M) \subset TM$ represents the set of all vector fields on M. To specify the value of v at a point $x \in M$, we use the notation $v|_x$. Thus, we treat any vector field as a *section* on the tangent bundle TM (see Chapter 7 in [38]). The final component is the *right invariance* of a given vector field v on G:

Definition 6. *Let $R_{g*} : T_hG \to T_{h \cdot g}G$ denote the pushforward for R_g (see Vector Fields and Smooth Maps in Chapter 8 in [39] for a detailed description of the pushforward map). Then, a right invariant vector field $v|_h$ on G is one that remains invariant under right translations:*

$$R_{g^*}(v|_h) = v|_{h \cdot g}. \tag{3.4}$$

In other words, the right action of G on v does not translate v to a new vector field on G; but only changes its point of evaluation.

In light of these definitions:

Definition 7. *A Lie algebra $\mathfrak{g}$ is a vector space spanned by right-invariant vector fields $\{v_1, \ldots, v_r\}$ on the Lie group G. It is equipped with an anti-symmetric, bilinear multiplication operation called the Lie bracket, denoted as follows:*

$$[v_i, , v_j] = -[v_j, , v_i] = \sum_k c_{ij}^k v_k \in \mathfrak{g}, \tag{3.5}$$

where r is the dimension of G, denoted by $\dim(G) = r$ and c_{ij}^k are constant, called structure constants for G. The operation $[\,,\,]$ does not satisfy an associativity law; however, it adheres to the Jacobi identity:

$$[u, [v, w]] + [w, [u, v]] + [v, [w, u]] = 0. \tag{3.6}$$

Moreover, $\mathfrak{g}$ is isomorphic to T_eG (see Proposition 29.1.11 in [37]).

The right invariance of $\mathfrak{g}$ and its isomorphism to T_eG will be pivotal in Section 4, as learning $\mathfrak{g}$ is effectively equivalent to learning G. This study will focus on the former.

Remark 2 (The Notation for Vector Fields)**.** *From this point onward, we distinguish between vector fields on G and those on the manifold M on which G acts. Thus, we denote vector fields in $\mathfrak{X}(G) \subset TG$ by v and those in $\mathfrak{X}(M) \subset TM$ by $\hat{v}$.*

Any $v \in \mathfrak{X}(G)$ has a corresponding representation $\hat{v}$ on M. To describe the connection between them as well as the one between $\mathfrak{g}$ and G, we introduce the *exponential map*, a crucial construct in Lie group theory. First, we define how v acts on any other vector field $w \in T_eG$:

Definition 8. *Let the component functions of $v|_e \in T_eG$ be given by:*

$$v|_e = \sum_{j=1}^{r} \zeta^j(e) \frac{\partial}{\partial g^j}, \tag{3.7}$$

where $\{\partial/\partial g^j\}_j$ denotes the basis tangent vectors on TG, which coincide with coordinate flows on G (see Sections 1.3 and 1.4 in [8]). Similarly, let $w = \sum_{j=1}^{r} \xi^j(e) \partial/\partial g^j \in_e$

G. Then, the action of v on w at e is as follows:

$$v(w)|_e = \sum_{k=1}^{r} \left\{ \sum_{j=1}^{r} \zeta^j \frac{\partial \xi^k}{\partial g^j} \right\}_{\Big|_e} \frac{\partial}{\partial g^k}. \tag{3.8}$$

This may be considered as the directional derivative of w in the direction of v at the point $e \in G$. Furthermore, the Lie bracket defined in Definition 7 becomes:

$$[v_i, v_j]\,|_e = v_i(v_j)|_e - v_j(v_i)|_e, \tag{3.9}$$

which measures the commutativity of two vector fields.

It is worth noting that, in some references, a vector field might be regarded as a "derivation" (cf. Definition 12.45 in [40] for more details). On the other hand, the Lie bracket in (3.9) is essential for defining various tools required in this research. However, it also carries geometric significance, and we encourage readers to consult Figure 5 and Theorem 1.33 in [8]. We now introduce the first important application of a vector field:

Definition 9. *The exponential map $\exp : T_eG \to G$ is given by:*

$$\exp\left(\epsilon v|_e\right) \coloneqq e + \epsilon v|_e + \frac{\epsilon^2}{2!} v\left(v\right)|_e + \ldots = \sum_{i=0}^{\infty} \frac{\epsilon^i}{i!} v^i|_e, \tag{3.10}$$

where $\epsilon \in \mathbb{R}$, and $v^i|_e$ denotes the action of v on itself i times.

One may interpret $\exp$ as constructing a group element in the vicinity of e. This tool will be implemented by invertible residual network (iResNET) [41] to generate local group transformations from any $\boldsymbol{v} \in \mathfrak{g}$ in Section 4.4. What makes $\exp$ exceptionally significant is its ability to construct any $g \in G$, provided that e and g are *path-connected* (see Section 1.3.2 in [1]), by utilizing a diffeomorphism known as the *normal coordinate map* (see (1.40) in [8]):

Theorem 1. *Let $\mathfrak{g} = \{v_1, \ldots, v_r\}$ for G with $\dim(G) = r$. Assuming G is connected, any $g_{\boldsymbol{\epsilon}} \in G$ sufficiently close to $e \in G$ can be written as:*

$$g_{\boldsymbol{\epsilon}} = \boldsymbol{\Phi}(\boldsymbol{v}) = \exp\left(\epsilon^1 v_1\right) \cdot \ldots \cdot \exp\left(\epsilon^r v_r\right), \tag{3.11}$$

where $\boldsymbol{v} = \sum_j \epsilon^j v_j$, $\boldsymbol{\epsilon} = [\epsilon^1, \ldots, \epsilon^r]^t \in \mathbb{R}$, and the map $\boldsymbol{\Phi} : \mathfrak{g} \to G$ is called the normal coordinate map. Furthermore, any $g \in G$ can be expressed as:

$$g = \exp\left(\epsilon^{i_1} v_{i_1}\right) \cdot \ldots \cdot \exp\left(\epsilon^{i_k} v_{i_k}\right), \tag{3.12}$$

where $k \in \mathbb{N}$ is arbitrarily large and $1 \leq i_j \leq r$.

We provide a proof of Theorem 1 in Appendix A.1. For an illustration of the terminology discussed thus far, see Figure 3.1. The *logarithm map*, $\ln$, for $GL(r, \mathbb{R})$, is not easily defined for a Lie group other than $GL(r, \mathbb{R})$. However, the reader should keep in mind *Ado's Theorem* (see Theorem in 3.17.7 in [42]), particularly when considering any finite dimensional G locally and and any of its element, in the vicinity of e, given by $\mathbf{\Phi}$ map. Fortunately, $\ln$ is not required for implementation in this study.

Before introducing Definition 8, our aim in this extensive introduction was to establish a connection between the vector field v on G and its representation $\hat{v}$ on M. With the necessary groundwork laid, we can now state this relationship explicitly:

Definition 10. *Let $v \in \mathfrak{g}$ and $f : M \to \mathbb{R}$ be a smooth, scalar-valued function. The action of v on f is given by:*

$$\hat{v}(f) = \frac{d}{d\epsilon} f\left(\exp\left(\epsilon v\right) \cdot x\right)\Big|_{\epsilon=0}, \tag{3.13}$$

where $x \in M$, $\epsilon \in \mathbb{R}$, and the group action on x is defined by (3.1).

Equation (3.13) describes the action of any vector field on a scalar function on M. This concept can be extended by replacing $f \in C^{\infty}(M)$ with any other analytical object on M, thereby defining the action of $\hat{v} \in TM$ on that object. Using (3.13), the actions on M induced by mathematical constructs in (3.8), (3.9), and (3.10) could be found out.

Remark 3 (Representation Theory is Concerned with Linear Actions of G)**.** *Linear actions of a Lie group G on a vector space V play a fundamental role in various applications within physics and engineering. As a result, mathematicians typically use the term "representation" to refer to the action of G on V by means of a set of invertible matrices. This perspective constitutes a significant branch of group theory, commonly known as representation theory [1]. In contrast, the term "representation" in this thesis is employed in a broader sense: it encompasses both linear and nonlinear actions of G on a manifold M, where M is not necessarily a vector space. As such, our usage extends beyond the classical, linear context of representation theory.*

Before concluding this subsection, we need to define various types of group actions, as these will be referenced in the subsequent sections. For a concise explanation, see Chapter 2 in [13]. Here, we present only the definitions essential for this research.

Definition 11. *A group action $\Psi : G \times M \to M$ is called free if $\Psi(g, x) = x$ holds only for $g = e \in G$ and $\forall\, x \in M$. In other words, all stabilizers (isotropy groups) are trivial: $S_x = \{e\}$, $\forall x \in M$.*

However, a milder condition can be imposed on G:

Definition 12. *Let $S_x = \{g \in G \,|\, g \cdot x = x\}$ be the stabilizer at $x \in M$. The group G is said to act effectively if:*

$$S \coloneqq \bigcap_{x \in M} S_x = \{e\}\,. \tag{3.14}$$

While a free action is inherently effective, the converse is not necessarily true. For example, consider $G \subset SO(3)$, representing rotations about the z-axis and let $M = \mathbb{S}^2$ (the two-dimensional sphere embedded in $\mathbb{R}^3$). Now, if $\{e\} \neq S_M \subset G$ is the global isotropy group, then G can be replaced with G/S_M to obtain an effective group action on M without losing any pertinent information. Note that this quotient group is well-defined because S_M is a normal subgroup of G (see Theorem 2.12 in [10]). Since our focus is on locally defined groups and actions, we are primarily interested in groups with locally effective action. In the context of this study, we replace global G and S_M with their local counterparts. The significance of local effectiveness will become clear in Section 4, but our primary interest in effective group action lies in the following theorem (cf. Theorem 2.62 in [10]):

Theorem 2. *Let $\{v_1, \ldots, v_r\} = \mathfrak{g}$ for a group G acting locally effectively on M. Then, $\mathfrak{g} \cong \mathcal{V}$, where $\mathcal{V}$ is an involutive vector field system (see Definition 39) on M. Hence, the set $\{\hat{v}_1, \ldots, \hat{v}_r\}$, defined by* (3.13)*, forms a basis in TM.*

Therefore, any G, that acts locally effectively, avoids the complications arising from linear dependencies within the set $\{\hat{v}_1, \ldots, \hat{v}_r\}$ (see Definition 27). The final class of group actions that we will consider is known as "transitive":

Definition 13. *A group action $\Psi : G \times M \to M$ is said to be transitive if it possesses a single orbit; that is, for any $x, y \in M$, $\exists\, g \in G$ such that $g \cdot x = y$.*

Transitivity is a crucial property for defining homogeneous spaces, which are constructed using the orbits of an isotropy subgroup $H \leqslant G$ on M, with $M \cong G/H$ by selecting a base point $x_0 \in M$ (see Section 4 in Chapter II of [43]). In addition to that, transitivity will be instrumental in Section 4.2, where it will aid in simplifying the involutivity conditions. It is important to note that G/H is not necessarily a quotient group, as H is not required to be a normal subgroup. Additionally, readers should note that a free action does not necessarily imply transitivity. For instance, consider the action of the group $\mathbb{S}^1$ on $\mathbb{S}^3 \subset \mathbb{C}^2$, defined by $e^{i\theta} \cdot (z_1, z_2) = (e^{i\theta} z_1, e^{i\theta} z_2)$. As discussed in [44], this action is free but not transitive, as the group $\mathbb{S}^1$ preserves the fibration structure of $\mathbb{S}^3$ (Hopf fibration) and does not connect every pair of points on $\mathbb{S}^3$. Therefore, both **locally effective** and **transitive** (once G's action is restricted to a particular orbit $\mathcal{O}_x$) properties must be satisfied in the context of this research.

3.2 Jet Space and the Prolongation

The purpose of this subsection is to familiarize the reader with the subspace $\Sigma^{(s)}$, on which the symmetry group G acts and within which the DE system of interest is defined. We begin by defining the jet space $J^{(s)}(M)$ over M (see Chapter 2 in [8] for a brief introduction and Sections 2.1 and 2.2 in [23] for a thorough treatment), as $\Sigma^{(s)}$ is a subspace of $J^{(s)}(M)$. Following this, we discuss the method of *prolongation* for multi-valued functions and vector fields on M. This process extends objects represented in M to objects in $J^{(s)}(M)$; while preserving the *contact* condition (see Proposition 1).

In simple terms, the jet space $J^{(s)}(M)$ is constructed by augmenting the graph of a multi-valued function with its partial derivatives, up to a certain order $s \in \mathbb{N}$, as coordinates.

Definition 14. *Let $\boldsymbol{u} = (u^1, \ldots, u^\alpha, \ldots, u^q) \in \mathcal{U}$ and $\boldsymbol{x} = (x^1, \ldots, x^i, \ldots, x^p) \in \mathcal{X}$ represent the sets of dependent and independent coordinates, respectively, such that $M = \mathcal{X} \times \mathcal{U}$, where $p, q \in \mathbb{N}$. Let $u^\alpha_{x^{i_1} \ldots x^{i_k}}$ denote the partial derivatives of u^α with respect to $x^{i_1}, \ldots, x^{i_k}$, where $k = 1, \ldots, s$ represents the order of differentiation. Then, $J^{(s)}(M) = \mathcal{X} \times \mathcal{U}^{(s)}$ is called the jet space of order s defined over M, where*

$\mathcal{U}^{(s)} = \mathcal{U} \times \mathcal{U}^1 \times \ldots \times \mathcal{U}^s$, $\mathcal{U}^k = \left\{u^\alpha_{x^{I_k}}\right\}^\alpha_{I_k}$ *is the subspace spanned by the* k^{th} *order partial derivatives of* $\boldsymbol{u}$*, and* $I_k = (i_1 \ldots i_k)$ *is the multi-index of order* k *into* $\boldsymbol{x}$*, s.t.* $i_1 \leq \ldots \leq i_k$.

Remark 4 (Geometric vs Coordinate View of Jet Space)**.** *The jet space definition in Definition 14 is the coordinate-based framework. However, in the geometric framework of exterior differential systems (see Section 3.5), the partial derivatives,* $u_{x^{I_k}}$*, are replaced by independent jet variables (e.g.,* p, q, r*) to treat the jet space as a fiber bundle over the base manifold* $M = \mathcal{X} \times \mathcal{U}$*. The two frameworks are equivalent on integral submanifolds of the jet space (see Definition 37), as the geometric perspective enforces the derivative relationships only when restricted to these submanifolds. Specifically, the pullback of the contact forms (see Definition 20) to an integral submanifold ensures that the geometric constraints align with the partial derivative relations, e.g.* $p^\alpha_i = {}^{\partial u^\alpha}/_{\partial x^i}$*. This equivalence underscores the consistency of the two definitions while allowing the geometric framework to encode additional structural properties of the system.*

We will always assume that all partial derivatives exist (see Theorem 5.13 in [40]) wherever they are defined in M and commute. Thus:

$$\frac{\partial}{\partial x^i}\frac{\partial}{\partial x^j} = \frac{\partial}{\partial x^j}\frac{\partial}{\partial x^i}, \tag{3.15}$$

where $i \neq j$ and $i, j = 1, \ldots, p$. Consequently, we remove all redundant dimensions from $J^{(s)}(M)$ (refer to Section 2.3 in [8] for a discussion on its dimension). With the definition of $J^{(s)}(M)$ established, we can introduce the manifold induced by a given DE system:

Definition 15. *Let* $\Delta^{(s)} : J^{(s)}(M) \to \mathbb{R}^n$ *be a map such that* $\Delta^{(s)}(\boldsymbol{x}, \boldsymbol{u}^{(s)}) = 0$ *represents a DE system of order* $s \in \mathbb{N}$ *defined in* $J^{(s)}(M)$ *and is of maximal rank (see Theorem 1.13 [8]). Then,* $\Delta^{(s)} = 0$ *defines a regular submanifold* $\Sigma^{(s)} \subset J^{(s)}(M)$ *of dimension* $m - n$*, where* $m = \dim\left(J^{(s)}(M)\right)$ *and* $n \leq m$.

In the context of Definition 15, the geometric view of jet space is utilized and hence, all partial derivatives of order $0 \leq k \leq s$ are considered to be independent variables on $J^{(s)}(M)$. For the example we presented in Section 1.1, $J^{(1)}(M) = \mathcal{X} \times \mathcal{U}^{(1)}$ with

$\mathcal{X} = \{x\}, \mathcal{U} = \{u\}$, and $m = 3$. Since $n = 1$, $\dim(\Sigma^{(1)}) = 2$ under the assumption of regularity.

Suppose that $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$ is a solution to a given $\Delta^{(s)} = 0$, has no vertical tangent, and therefore, by the Implicit Function Theorem (see Theorem 10.26 in [40]), the graph $\Gamma_{\boldsymbol{f}} = \{\boldsymbol{x}, \boldsymbol{f}(\boldsymbol{x})\}$ defines a submanifold $S_{\boldsymbol{f}} \subset M$. We must then establish a method to

(i) prolong $\Gamma_{\boldsymbol{f}}$ to $\Gamma_{\boldsymbol{f}}^{(s)}$ such that $S_{\boldsymbol{f}}^{(s)} \subset \Sigma^{(s)}$, and

(ii) define $x \to \Gamma_{\boldsymbol{f}}^{(s)}$ as a section in the jet bundle $J^{(s)}(M) \to \mathcal{X}$.

Remark 5 ($J^{(s)}(M) \to \mathcal{X}$ as a Special Jet Bundle)**.** *Note the second condition above. The bundle is not standard jet bundle $J^{(s)}(M) \to M$, as pr$^{(s)}(f)$ is a section over $J^{(s)}(M)$ provided that f only depends on independent variables $\boldsymbol{x}$.*

Before addressing these two objectives, we should define the *prolongation* operator pr$^{(s)}$ and its behavior on a smooth function:

Definition 16. *Let $f \in C^\infty(\mathcal{X})$ be scalar-valued. Then, its s^{th} prolongation is given by:*

$$pr^{(s)}(f)\,[\boldsymbol{x}] = \left(f(\boldsymbol{x}), \frac{\partial f}{\partial x^{i_1}}(\boldsymbol{x}), \frac{\partial^2 f}{\partial x^{i_1}\partial x^{i_2}}(\boldsymbol{x}), \ldots, \frac{\partial^s f}{\partial x^{I_s}}(\boldsymbol{x})\right), \tag{3.16}$$

where I_s is the multi-index of order s (see Definition 14), $i_j \in (1, \ldots, p)$, and $j = 1, \ldots, s$. If $f \in C^\infty(\mathcal{X})$ (i.e., a multi-valued function from $\mathcal{X}$ to $\mathcal{U}$ with $\dim(\mathcal{U}) = q$), then its prolongation is simply applying (3.16) *to each $f^\alpha(\boldsymbol{x})$ for $\alpha = 1, \ldots, q$.*

Now, to satisfy the second condition, (ii), simply let $\Gamma_{\boldsymbol{f}}^{(s)} = \left\{\boldsymbol{x}, \text{pr}^{(s)}(\boldsymbol{f})\,[\boldsymbol{x}]\right\}$. As for the first condition, we will address it after stating the symmetry condition for DE systems in Section 3.4.

As a section of the tangent bundle $TM \to M$, the vector field $\hat{v}$ defined in Definition 10 can be naturally prolonged to the jet tangent bundle $TJ^{(s)}(M) \to J^{(s)}(M) \to M$. To compactly define the prolongation of $\hat{v}$, we first introduce the total derivative operator D_i on M:

$$D_i = \frac{\partial}{\partial x^i} + \sum_\alpha \sum_I u^\alpha_{x^I x^i} \frac{\partial}{\partial u^\alpha_I}, \tag{3.17}$$

where the order of the multi-index I, denoted by $|I|$, is arbitrarily large. However, we can safely assume that $|I| \leq s$, corresponding to the maximum order s of the system under consideration. The repeated application of the total derivative operator D_i is denoted by $D_{I_k} = Di_k \dots D_{i_1}$, where I_k is a multi-index of order k. Before presenting the formula for the prolongation of a vector field, we introduce a key quantity that encodes significant information about the vector field. This quantity will be utilized in various contexts throughout this research.

Definition 17. *Let the component functions of $\hat{v} \in TM$ for $M = \mathcal{X} \times \mathcal{U}$, which is induced by* (3.13)*, be defined as:*

$$\hat{v}|_z = \sum_{i=1}^{p} \hat{\xi}^i(z) \frac{\partial}{\partial x^i} + \sum_{\alpha=1}^{q} \hat{\varphi}^\alpha(z) \frac{\partial}{\partial u^\alpha}, \tag{3.18}$$

where $z = (\boldsymbol{x}, \boldsymbol{u}) \in M$. Then, the characteristic of $\hat{v}$ (see (2.48) in [8]) is given by

$$Q^\alpha(z^{(1)}) = \hat{\varphi}^\alpha(z) - \sum_{i=1}^{p} u^\alpha_{x^i}(\boldsymbol{x}) \hat{\xi}^i(z), \tag{3.19}$$

where $z^{(1)} = (\boldsymbol{x}, \boldsymbol{u}^{(1)})$ and $\alpha = 1, \dots, q$.

Theorem 3. *Let the component functions and the characteristic of $\hat{v}$ be as in* (3.7) *and* (3.19) *respectively. Then, the s^{th} prolongation of $\hat{v}$ is given by:*

$$pr^{(s)}(\hat{v})|_{z^{(s)}} = \hat{v}^{(s)}|_{z^{(s)}} = \sum_{i=1}^{p} \hat{\xi}^i(z) \frac{\partial}{\partial x^i} + \sum_{\alpha=1}^{q} \sum_{I} \hat{\varphi}^\alpha_I(z^{(|I|)}) \frac{\partial}{\partial u^\alpha_I}, \tag{3.20}$$

where $0 \leq |I| \leq s$ is the order of the multi-index into $\boldsymbol{x}$ (see Definition 14), and

$$\hat{\varphi}^\alpha_I(z^{(|I|)}) = D_I Q^\alpha(z^{(1)}) + \sum_{i=1}^{p} u^\alpha_{x^I x^i}(\boldsymbol{x}) \hat{\xi}^i(z), \tag{3.21}$$

with $D_I = D_{i_k} \dots D_{i_1}$, $k = 1, \dots, s$, and $i_j \in (1, \dots, p)$.

The proof of Theorem 3 can be found in Theorem 2.36 of [8]. Note that $\hat{v}^{(s)}|_{z^{(s)}} \in TJ^{(s)}(M)$, and (3.17) suggests that D_i is also a vector field. Consequently, the interpretation of $\hat{v}$ in Definition 8 as a directional derivative is consistent with the interpretation of D_i is a vector field. However, observe that D_i in (3.17) is not defined in $TJ^{(s)}(M)$ as it has coefficients in $J^{(s+1)}(M)$. It is given in a special tangent space $\mathfrak{X}_{\text{loc}}\left((\pi^{s+1}_s)^*\left(TJ^{(s)}(M)\right)\right)$, where $(\pi^{s+1}_s)^*$ is the pullback of the canonical projection in (3.22) and $\mathfrak{X}_{\text{loc}}(\cdot)$ denotes the local sections (i.e. vector fields) of the specified tangent bundle (see Remark 2.1.8 in [23]).

We highlight the following canonical projection for a given jet space, which establishes a connection between analytical objects, such as the prolonged vector field in Theorem 3 and the prolonged scalar function in Definition 16, of different jet orders. Specifically, this projection maps jet spaces of higher orders to those of lower orders, as described in (2.20) of [8]:

$$\pi_s^t(z^{(t)}) = z^{(s)}, \tag{3.22}$$

where $z^{(t)} \in J^{(t)}(M)$ and $z^{(s)} \in J^{(s)}(M)$ for $t > s \geq 0$. This projection can also be applied when $z^{(t)}$ represents either a symmetry generator, $\mathrm{pr}^{(t)}(\hat{v}) \in TJ^{(t)}(M)$, or a scalar function, $\mathrm{pr}^{(t)}(\boldsymbol{f}) \in C^\infty(J^{(t)}(M))$. In computational differential algebra, a similar projection is also employed to project the prolonged DE system $\Delta^{(s+1)} = 0$, represented by $\Sigma^{(s+1)}$. The resulting space $\Sigma_{(s)}^{(s+1)} \coloneqq \pi_s^{s+1}\left(\Sigma^{(s+1)}\right)$ incorporates any newly generated integrability conditions of order s. For further details, we refer the reader to [23], particularly Example 2.3.9.

3.3 Contact Forms and Transformations

Thus far, we have discussed vector fields on various tangent spaces. In this subsection, we introduce one of the core objects relevant to the equivalence of solutions of a given DE system $\Delta^{(s)} = 0$, which are dual to $\boldsymbol{p}$-vector fields in the *alternating graded-algebra* on TM. These objects are referred to, in the differential geometry literature, as (differential) $\mathbf{p}$-forms, defined in the alternating graded-algebra on the cotangent bundle T^*M. They are particularly essential for describing Cartan's equivalence [15] and the method of moving coframes [45], especially when $p = 1, 2$. The study of $\mathbf{p}$-forms is a rich area of mathematics, and we direct the reader to Sections 26.2 and 26.3 for their algebraic properties, and to Section 28.5 for their definition and application on smooth manifolds, as detailed in [37]. In the final part of this subsection, our primary focus will be on *contact* forms and their prolongation to higher-order jet cotangent bundles. These forms are crucial for constructing the invariant coframe in case of DE system and for determining whether a given section, $\mathcal{X} \to J^{(s)}(M)$, results from the prolongation of a function $\boldsymbol{f}(\boldsymbol{x})$, a topic that we shall explore shortly.

Remark 6 (Frames and Coframes)**.** *We do not define the "frame" and "coframe" notions from differential geometry; but one may safely assume that they consist of*

(smooth) ordered basis and dual basis for respective vector bundles (see Chapter 8 and 10 in [39]).

We begin by introducing a set of notations. We denote the cotangent bundle over M by T^*M and the jet cotangent bundle by $T^*J^{(s)}(M)$. Furthermore, $\Omega^{(s),l}(M) := \Lambda^l(T^*J^{(s)}(M))$ represents the space of l-forms on $T^*J^{(s)}(M)$. Then, we can infer that $\Omega^{(s),0}(M)$ is simply composed of smooth scalar functions in $C^\infty(J^{(s)}(M))$. Specifically, the definition of one-forms on $T^*J^{(s)}(M)$ is as follows:

Definition 18. *Let $z^{(s)} = (\boldsymbol{x}, \boldsymbol{u}^{(s)}) \in J^{(s)}(M)$, and let $\{dx^1, \dots, dx^p, \dots, du_{x^{I_k}}, \dots\} = \left\{dz^{(s),i}\right\}^i = \Omega^{(s),1}(M)$, such that $dz^{(s),i}\left(\partial/\partial z^{(s),j}\right) = \delta^i_j$, where δ^i_j is the Kronecker delta function, $\left\{\partial/\partial z^{(s),j}\right\}_j$ forms a basis for $TJ^{(s)}(M)$, I_k is a multi-index of order k into $\boldsymbol{x}$, $k = 1, \dots, s$, and $i = 1, \dots, m = \dim(J^{(s)}(M))$. Then, $\Omega^{(s),1}(M)$ is called the space of one forms on M. In particular, any other one-form $\omega|_{z^{(s)}} \in \Omega^{(s),1}(M)$ can be expressed as a linear combination of these basis one-forms:*

$$\omega|_{z^{(s)}} = \sum_{i=1}^{m} P_i(z^{(s)})\, dz^{(s),i}, \tag{3.23}$$

where $P_i : J^{(s)}(M) \to \mathbb{R}$ are smooth functions.

To define higher-order forms from lower-order ones, we require the *exterior derivative* operator $d : \Omega^{(s),l}(M) \to \Omega^{(s),l+1}(M)$, which maps an l-form to an $(l+1)$-form:

Definition 19. *Let $\omega|_{z^{(s)}}$ be an l-form. Then, the operator d acts on $\omega|_{z^{(s)}}$ as:*

$$d(\omega)|_{z^{(s)}} := \sum_{i=1}^{m} dP_i(z^{(s)}) \wedge dz^{(s),i} \in \Omega^{(s),l+1}(M), \tag{3.24}$$

where $\wedge$ denotes the wedge product (refer to the discussion on exterior algebra in Section 26.3 of [37]). Furthermore, it holds that $d^2 = 0$; i.e., applying d two or more times to the same form yields 0.

We can now define contact forms:

Definition 20. *A contact form θ is a specialized type of one-form that annihilates prolonged functions (see Definition 16). Thus, its defining property is:*

$$\theta\Big|_{\Gamma_{\boldsymbol{f}}^{(s)}} = 0, \tag{3.25}$$

where $\Gamma_{\boldsymbol{f}}^{(s)}$ denotes the prolonged graph of a multi-valued function $\boldsymbol{f}$ on $\mathcal{X}$ (see Definition 16) and $s \geq 1$.

An equivalent definition of a contact form can be given using the *pullback* of θ by $pr^{(s)}(\boldsymbol{f})$. To establish this, we first define the pullback of a given l-form:

Definition 21. *Let $F : M \to N$ be a smooth map between any two manifolds, and let $\bar{\omega}|_{\bar{z}} = \sum_{J_l} \bar{P}_{J_l}(\bar{z}) d\bar{z}^{J_l} \in \Omega^l(N)$, where $d\bar{z}^{J_l} = d\bar{z}^{j_1} \wedge \ldots \wedge d\bar{z}^{j_l}$ forms a basis for l-forms, with $1 \leq j_1 < \ldots < j_l \leq \dim(N)$ for the multi-index $J_l = (j_1 \ldots j_l)$, and $\bar{z} = F(z)$. The pullback of $\bar{\omega}|\bar{z}$ by F is then given by:*

$$\begin{aligned} F^*(\bar{\omega}|_{\bar{z}}) &= \sum_{J_l} \bar{P}_{J_l}(F(z)) \frac{\partial F(z)^{j_1}}{\partial z^{i_1}} dz^{i_1} \wedge \ldots \wedge \frac{\partial F(z)^{j_l}}{\partial z^{i_l}} dz^{i_l} \\ &= \underbrace{\sum_{J_l} \bar{P}_{J_l}(F(z)) \frac{\partial F(z)^{j_1}}{\partial z^{i_1}} \cdots \frac{\partial F(z)^{j_l}}{\partial z^{i_l}}}_{=P_{I_l}(z)} \underbrace{dz^{i_1} \wedge \ldots \wedge dz^{i_l}}_{=dz^{I_l}} \\ &= \sum_{I_l} P_{I_l}(z) dz^{I_l} = \omega|_z \in \Omega^l(M), \end{aligned} \tag{3.26}$$

where sums over corresponding i_k indices are implied, with the $\sum_{i_k=1}^{\dim(M)}$ operators omitted in the first and second equalities for all $k = 1, \ldots, l$, and in the final equality, $1 \leq i_1 < \ldots < i_l \leq \dim(M)$ for the multi-index $I_l = (i_1 \ldots i_l)$.

Remark 7 (Einstein Notation)**.** *Henceforth, if any summation operator is omitted for a particular index, it is understood that corresponding upper and lower indices are to be summed over, e.g.:*

$$\sum_i P_i dz^i = P_i dz^i. \tag{3.27}$$

Note that M and N in Definition 21 can be replaced by jet spaces without altering the final result. Consequently, the equivalent defining property for θ in (3.25) is:

$$\left[pr^{(s)}(\boldsymbol{f})\right]^* (\theta) = 0, \tag{3.28}$$

where $pr^{(s)}$ is the prolongation operator as defined in Definition 16.

The following contact forms are known as *basic* contact forms:

$$\theta_I^\alpha = du_I^\alpha - \sum_{i=1}^{p} u_{x^I x^i}^\alpha(\boldsymbol{x}) dx^i \in \Omega^{(s+1),1}(M), \tag{3.29}$$

where $\alpha = 1, \ldots, q = \dim(\mathcal{U})$, $p = \dim(\mathcal{X})$, $I = (i_1 \ldots i_s)$ is a multi-index of order s into $\boldsymbol{x}$, $s \geq 1$, and $M = \mathcal{X} \times \mathcal{U}$.

Remark 8 (Order of a Contact Form)**.** *Note that, s^{th} order contact form θ_I^α in* (3.29) *is defined on $J^{(s+1)}(M)$, one order higher than of the contact form itself. For instance, the 0^{th} order contact form $\theta^\alpha = du^\alpha - \sum_i u^\alpha_{x^i} dx^i$ is defined on $J^{(1)}(M)$, because u_{x^i} coordinates are involved in its definition.*

Using basic contact forms, any other contact form θ can be expressed (see Theorem 4.23 in [10]):

$$\theta = \sum_I \sum_\alpha P_I^\alpha(z^{(s)}) \theta_I^\alpha, \tag{3.30}$$

where $z^{(s)} \in J^{(s)}(M)$. Thus, the set $\{\theta_I^\alpha\}_I^\alpha$ forms a subspace $\mathcal{C}^{(s)}(M) \subset T^*J^{(s)}(M)$. This subspace $\mathcal{C}^{(s)}(M)$ allows us to decompose $T^*J^{(s)}(M)$ into *horizontal* and *vertical* cotangent subbundles. Let $\mathcal{C}^{(s)}(M)$ denote the vertical cotangent subbundle, then its complement, the horizontal subbundle $\mathcal{H}(M)$, is given by $\{dx^1, \ldots, dx^p\}$ (see (8.2) in [13]). Hence, we obtain the decomposition:

$$T^*J^{(s)}(M) = \mathcal{H}(M) \oplus \mathcal{C}^{(s)}(M). \tag{3.31}$$

The final element necessary for our discussion is the interaction between contact forms and Lie group transformations, beginning with *contact invariant* transformations:

Definition 22. *Let G be a local Lie group acting on M, with its prolonged group of transformations on $J^{(s)}(M)$ denoted by $G^{(s)}$ and induced by* (3.20) *and* (3.12)*. For any $\theta \in \mathcal{C}^{(s)}$, we define G as a group of contact-preserving transformations if $g^{(s)*}(\theta) \in \mathcal{C}^{(s)}$ for all $g^{(s)} \in G^{(s)}$, where $g^{(s)*}$ is pullback by the prolonged group action (see Definition 21).*

In other words, if a contact form remains invariant under the translation of a group, such group actions are termed *contact transformations*. These transformation groups are of particular interest in this study, as the lift of a solution $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$ (via prolongation in Definition 16) to $J^{(s)}(M)$ is determined by a *contact system*, which is a specific type of exterior differential system (see Definition 36) with the independence

condition $dx^1 \wedge \ldots \wedge dx^p \neq 0$, owing to the contact form property in (3.25). The independence condition implies that any submanifold $S_{\boldsymbol{f}}^{(s)} \subset J^{(s)}(M)$ parameterized by $x^1, \ldots, x^p$ and associated with $\Gamma_{\boldsymbol{f}}^{(s)}$ is fully transverse. The contact system's definition is provided in Definition38.

To ensure that G is a group of contact transformations, an infinitesimal *contact condition* must be satisfied (see 4.55 and 4.56 in [10] for proof):

Proposition 1. *Let $\hat{v}^{(1)} := pr^{(1)}(\hat{v})$ (see Theorem 3). Suppose that $\theta \in \mathcal{C}^{(s)}$ is an arbitrary contact form. Then, the condition in Definition 22 is satisfied iff:*

$$\hat{v}^{(1)}(\theta) = \nu \in \mathcal{C}^{(s)}. \tag{3.32}$$

The condition in (3.32) *also translates into:*

$$\frac{\partial \hat{\varphi}^{\alpha}}{\partial u_j^{\beta}} - \sum_i u_i^{\alpha} \frac{\partial \hat{\xi}^i}{\partial u_j^{\beta}} = 0, \ \forall\, \alpha, \beta, \text{ and } j, \tag{3.33}$$

where $i, j = 1, \ldots, p$, and $\alpha, \beta = 1, \ldots, q$. Alternatively, in terms of the characteristics $\boldsymbol{Q} = (Q^1, \ldots, Q^{\alpha}, \ldots, Q^q)$ (see Definition 17):

$$\frac{\partial Q^{\alpha}}{\partial u_i^{\beta}} + \hat{\xi}^i \delta_{\beta}^{\alpha} = 0, \ \forall\, \alpha, \beta, \text{ and } i, \tag{3.34}$$

where δ_{β}^{α} is the delta function.

The condition (3.33) (or equivalently (3.34)) will be one of several conditions that must be satisfied in the construction of a future model covering Lie contact transformations. However, in this particular study, grounded in a data-oriented framework, will concentrate on learning a particular subset of transformations with notable **geometric significance** in physics and engineering: *Lie point* transformations and any symmetry generator $\hat{v}$ of a given set of Lie point transformations implicitly satisfies (3.34), thanks to any of its Q^{α}'s **linear dependence** in u_i^{α}.

Lie point transformations constitute a significant portion of contact transformations. According to Theorem 4.32 in [10], every contact transformation for $q > 1$ arises from the prolongation (see Theorem 3) of a Lie point transformation. However, when $q = 1$, there may exist first-order contact transformations that are not Lie point transformations. The distinction between these transformations is summarized by the following additional condition:

Definition 23. *Let $\hat{v}$ be given by* (3.18)*. If ξ^i and φ^α do not depend on $u^\alpha_{x^{I_k}}$ (where k^{th} order partial derivatives of u^α are considered) for all $i = 1, \ldots, p$, $\alpha = 1, \ldots, q$, and $k \geq 1$, then the transformation induced by $\hat{v}$ is called a Lie point transformation. Conversely, if these component functions of $\hat{v}$ depend only on $u^\alpha_{x^i}$ in addition to $(\boldsymbol{x}, \boldsymbol{u})$ $\forall\, i = 1, \ldots, p$, then the transformation is termed a Lie contact transformation. Therefore, Lie point symmetries are a subset of Lie contact transformations.*

Transformations that involve dependencies of order higher than or equal to 2 are referred to as *generalized* transformations. However, they do not constitute symmetries of geometric PDEs. Furthermore, symmetries for DE systems, that are covered by Lie point and contact transformations, can even be involved in analyzing solitons, such as the Korteweg-de Vries (KdV) equations (see Example 2.44 in [8]), that are not only governed by Lie groups; but also by *non-local* and *discrete* symmetries, which can either be defined by generalized symmetries or be approximated by Lie pseudo-groups. Therefore, Lie point and contact symmetries are powerful tools across wide variety of DE systems.

3.4 Symmetry and Invariance for Differential Equations

This subsection marks the point where we formally introduce the concept of a symmetry group for a given differential equation (DE) system $\Delta^{(s)} = 0$. We proceed to define invariant objects on $J^{(s)}(M)$, distinguishing between an invariant function, a differential invariant, and a lifted differential invariant. A clear understanding of these three notions of invariance is essential for grasping the subsequent discussions.

The symmetry condition for a given subvariety $\Delta^{(s)} = 0$ (a DE system within the scope of our research) can be formulated in terms of either the ordinary group action (see Definition 4) or the infinitesimal action (see Definition 8) of a group G on the subvariety.

Definition 24. *Suppose that $\Sigma^{(s)} = \{z^{(s)} \in J^{(s)}(M) \mid \Delta^{(s)}(z^{(s)}) = 0\}$ is a submanifold of maximal rank (see Theorem 1.13 in [8]). The group G is a symmetry group for $\Delta^{(s)} = 0$ if $g^{(s)} \cdot \Sigma^{(s)} \subseteq \Sigma^{(s)}$, $\forall\, g^{(s)} \in G^{(s)}$, where $g^{(s)}$ denotes the prolonged group transformation.*

Alternatively, the infinitesimal symmetry condition is provided by Theorem 2.31 in [8]:

Definition 25. *Let $\Delta^{(s)} = 0$ be the subvariety of interest. Let $\{v_1, \ldots, v_r\} = \mathfrak{g}$ (see Definition 7) represent the Lie algebra of G. Then, G is a symmetry group for the subvariety if:*

$$pr^{(s)}(\hat{v}_i)\left[\Delta^{(s)}\right]_{|_{\Delta^{(s)}=0}} = 0, \ \forall\, i = 1, \ldots, r, \tag{3.35}$$

where $\hat{v}_i$ denotes the representation of v_i on M (see Definition 8), and $pr^{(s)}$ is the prolongation operator (see Theorem 3).

Remark 9. *It is important to note that we do not assert that the group G in either Definition 24 or Definition 25 is necessarily a point or contact transformation group. These conditions simply imply that any transformation given by $g^{(s)} \in G^{(s)}$ leaves $\Sigma^{(s)}$ invariant.*

Now, consider that $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$ is a solution to a given DE system $\Delta^{(s)} = 0$. In this context, it is important to distinguish between two different types of symmetries. The first type of symmetry is one that transforms $pr^{(s)}\left[\boldsymbol{f}\right](\boldsymbol{x})$ into a new solution $pr^{(s)}\left[\bar{\boldsymbol{f}}\right](\bar{\boldsymbol{x}})$. The second type is a symmetry that leaves $\Gamma^{(s)}_{\boldsymbol{f}}$ (see Definition 16) invariant. As can be readily observed, the definition provided in Definition 24 (or equivalently, the one in Definition 25) encompasses both types of symmetry. To differentiate between those types, we introduce the following infinitesimal condition (see the section on Equivalence and Trivial Symmetries in Chapter 5 of [10] and Section 3.2.3 in [9]):

Definition 26. *Let $pr^{(s)}(\hat{v})|_{z^{(s)}}$ be the prolonged representation of a vector field $v \in \mathfrak{g}$, as induced by* (3.13)*, on the jet space $J^{(s)}(M)$. We refer to $\hat{v}$ as the generator of a trivial one-parameter symmetry group if its characteristic vector $\boldsymbol{Q}$, defined in* (3.19)*, vanishes on $\Gamma^{(1)}_{f}$, that is:*

$$Q^{\alpha}(z^{(1)})|_{\Gamma^{(1)}_{\boldsymbol{f}}} = 0, \ \forall, \alpha = 1, \ldots, q. \tag{3.36}$$

Otherwise, $\hat{v}$ is said to be the generator of a non-trivial one-parameter symmetry group.

To facilitate the discussion of various types of invariant objects, we first introduce the concept of stable orbit dimension for G, which provides insights into the number of

differential invariants that may be obtained. Besides, it will aid in formulating the stopping criteria for the algorithm discussed in Section 4.

Definition 27. *Let* $\dim(G) = r$. *Denote by* $o_t \leq r$ *the maximal orbit dimension of* $G^{(t)}$ *on* $J^{(t)}(M)$. *The maximal orbit dimension* o_t *eventually stabilizes at a particular order* s_G, *such that* $o_t = o_{s_G} = r \, \forall t \geq s_G$ *(see (5.5) in [10]). This dimension* o_{s_G} *is termed the stable orbit dimension, and* s_G *is referred to as the order of stabilization.*

In this context, the property of locally effective group action (see Definition 12) is also crucial. If $G^{(s_G)}$ acts locally effectively on $J^{(s_G)}(M)$, then $o_{s_G} = r = \dim(G)$ (see Theorem 5.11 in [10]). Since the group is assumed to act locally effectively, the number of differential invariants is given by

$$i_{s_G} = \dim(J^{(s_G)}(M)) - \dim(G) = p + q\binom{p + s_G}{s_G} - r, \tag{3.37}$$

where p and q denote the number of independent and dependendent variables, respectively. For further details, see (9.6) in [13]. If $G^{(s)}$ has not yet reached its stable orbit dimension, then r can be replaced with the maximal orbit dimension o_s over $J^{(s)}(M)$.

While this work focuses exclusively on learning non-trivial Lie point transformations that satisfy particular involutivity conditions, defining the concept of an invariant is crucial for understanding the role of *invariant coordinates* in **reformulating** and **simplifying** DE systems. In the context of algebraic equations, an invariant is a function $I : M \to \mathbb{R}$ that remains unchanged under the action of a symmetry group G. For DEs, such functions are referred to as *differential invariants*.

Definition 28. *Let* $G^{(s)}$ *be a symmetry group for* $\Delta^{(s)} = 0$ *acting on* $J^{(s)}(M)$. *A smooth function* $I^{(s)} : J^{(s)}(M) \to \mathbb{R}$ *is called a differential invariant if*

$$I^{(s)}(g^{(s)} \cdot z^{(s)}) = I^{(s)}(z^{(s)}) \tag{3.38}$$

where $z^{(s)} = (\boldsymbol{x}, \boldsymbol{u}^{(s)})$. *Alternatively, this condition can be expressed in infinitesimal terms as:*

$$\hat{v}_i^{(s)}(I^{(s)}) = 0, \ \forall \, i = 1, \ldots, r, \tag{3.39}$$

where $\hat{v}_i^{(s)}$ *represents* $v_i \in \mathfrak{g}$ *in* $J^{(s)}(M)$, *and* $\dim(G) = r$.

Another type of invariance involves a scenario where the algebraic form of the invariant function remains unchanged, although its output value may vary. This is known as lifted invariance (see Definition 3.3 in [13]).

Definition 29. *Let G be a symmetry group for $\Delta^{(s)} = 0$ acting on $J^{(s)}(M)$, and $\mathscr{L}^{(s)} : J^{(s)}(M) \times G \to J^{(s)}(M)$ be a differential function, given by:*

$$\mathscr{L}^{(s)} = \Psi(z^{(s)}, g) = g^{(s)} \cdot z^{(s)}, \tag{3.40}$$

where Ψ is the group action as defined in Definition 4. Then, $\mathscr{L}^{(s)}$ is called the lifted invariant.

The final type of invariance relevant to the future research is the *invariant function*:

Definition 30. *Let $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x}) \in C^{\infty}(M)$ with $\Gamma_{\boldsymbol{f}}^{(s)}$ being its prolonged graph. If $g^{(s)} \cdot \Gamma_{\boldsymbol{f}}^{(s)} \subseteq \Gamma_{\boldsymbol{f}}^{(s)}$, then $\boldsymbol{f}(\boldsymbol{x})$ is called an invariant function.*

3.5 Exterior Differential Systems and Involutivity

The content of this section is essential for understanding the concept of "involutivity" as discussed in Sections 4.2. It provides a summary of one of two main forms of involutivity, namely the Frobenius theorem (the other form will briefly be discussed in the end of this subsection), and includes an example of an exterior differential system known as a *Pfaffian* system. We assume the geometric view of jet space (see Remark 4) in this subsection, unless any specified $\boldsymbol{p}$-form is pulled-back to an integral submanifold.

Remark 10 (Omitting Jet Order s)**.** *For the sake of generalization, we omit the jet space order s from the notation for the space of $\mathbf{p}$-forms, denoted $\Omega^{(s),l}(M)$ (see the discussion preceding Definition 18). This omission is justified, as the following definitions are not limited to systems of DEs. However, we will reintroduce the order s into the notation when the context specifically returns to DEs.*

Broadly speaking, *involutivity* refers to a set of conditions that ensure the *integrability* of a system. Involutivity is closely related to ideals generated by $\mathbf{p}$-forms or, equivalently, by vector field systems (see Chapter 14, Section 1 in [10]). To formalize the discussion, a series of mathematical objects should be introduced:

Definition 31. *Let $\Omega^* = \bigoplus_{l=0}^{\infty} \Omega^l$ be a graded algebra of* $\mathbf{p}$*-forms closed under the wedge product $\wedge$ (see Theorem 26.3.6 in [37]). Any subspace $\mathcal{I}^l \subset \Omega^l$, consisting solely of forms of degree l, is called a homogeneous subspace.*

Definition 32. *An algebraic ideal $\mathcal{I}_{alg}$ is defined as a direct sum of homogeneous subspaces $\mathcal{I}^l$ (i.e., $\mathcal{I}_{alg} = \bigoplus_{l=0}^{\infty} \mathcal{I}^l$), such that:*

(i) $\omega^1 + \omega^2 \in \mathcal{I}_{alg}$*, if* $\omega^1, \omega^2 \in \mathcal{I}^l \subset \mathcal{I}_{alg}$ *for any* $l \geq 0$*,*

(ii) $f\,\omega \in \mathcal{I}_{alg}$ *if* $\omega \in \mathcal{I}_{alg}$ *and* $f \in \Omega^0$ *(i.e., multiplication by a smooth scalar function), and*

(iii) $\omega \wedge \alpha \in \mathcal{I}_{alg}$ *if* $\omega \in \mathcal{I}_{alg}$ *for any arbitrary* $\mathbf{p}$*-form* α *(not necessarily a homogeneous form, i.e.,* $\alpha \in \Omega^l$ *for some* $l \in \mathbb{N}$*).*

Definition 33. *An algebraic ideal $\mathcal{I}_{alg}$ is termed a differential ideal if it is closed under the exterior derivative operator d, as defined in Definition 19. In other words, if $\omega \in \mathcal{I}_{diff}$, then $d\omega \in \mathcal{I}_{diff}$.*

Differential ideals are also referred to as "closed" ideals, and this closedness property will be integral to describing the involutivity of the Pfaffian system associated with a given DE system. Furthermore, both types of ideals can be succinctly defined if they have a finite number of generators:

Definition 34. *Let $\mathcal{G} = \{\omega^1, \ldots, \omega^L\} \subset \mathcal{I}$ be a subset of* $\mathbf{p}$*-forms. If any $\beta \in \mathcal{I}$ can be expressed as:*

$$\beta = \sum_{l=1}^{L} \omega^l \wedge \alpha^l, \tag{3.41}$$

then $\mathcal{I}$ is said to admit the generator $\mathcal{G}$, where α^l is a $\mathbf{p}$*-form such that deg(ω^l) + deg(α^l) = deg(β). For brevity, we denote this by $\mathcal{I} \coloneqq \{\omega^1, \ldots, \omega^L\}$ and will subsequently omit reference to $\mathcal{G}$.*

Remark 11 (Generators of $\mathcal{I}_{\text{diff}}$)**.** *Since $d\omega \in \mathcal{I}_{diff}$ for any $\omega \in \mathcal{I}_{diff}$ by definition, we will not include the forms $d\omega$ explicitly in the generator of any $\mathcal{I}_{diff}$.*

This concise definition will be particularly valuable when defining the concept of an *integral submanifold* (see Definition 37). The last ingredient necessary to describe higher-level notions is so-called *simply-generated exterior ideal*:

Definition 35. *An ideal $\mathcal{I} = \{\omega^1, \ldots, \omega^{m-p}\}$ is called simply-generated, if each of its generators $\omega^i \in \Omega^{(s),1}$ is a one-form.*

This is crucial because, in the systems of interest, the exterior differential system associated with a vector field system of constant rank can be defined by such a simply-generated ideal. For additional insights into the mathematical tools introduced, thus far, in this subsection, readers are directed to Chapter 13 of [10] and Appendix B.4 of [14].

The following two definitions provide a foundation for understanding the involutivity of any equivalence problem, extending beyond differential equation (DE) systems.

Definition 36. *Let M be a manifold of dimension m. A differential system on M with an independence condition $0 \neq \alpha \in \Omega^p(M)$, associated with a differential ideal $\mathcal{I}_{diff} \subset \Omega^*$, is called an exterior differential system (EDS), where $p \leq m$.*

Definition 37. *An integral submanifold $N \subset M$ of an EDS is an immersion (i.e., $\iota : N \to M$ where ι_* maps TN injectively, though $\iota(N)$ itself may not be injective) such that the differential ideal $\mathcal{I}_{diff}$ of the EDS vanishes on N; that is, $\omega|_N := i^*(\omega) = 0$, for all $\omega \in \mathcal{I}_{diff}$, where $\dim(N) = p$.*

With the concept of an integral submanifold now introduced, the reader may appreciate the significance of ideal generators in Definition 34. If all generators of $\mathcal{I}$ vanish on N, then $\mathcal{I}$ also vanishes on N (see Lemma 13.4 in [10]).

Before illustrating these concepts with an example from a DE system, we note an aspect of the homogeneous space $\mathcal{I}^0$ (see discussion related to (13.1) in [10]):

Remark 12 (Excluding $\mathcal{I}^0$)**.** *$\mathcal{I}^0$ is a subspace of smooth scalar functions on M (i.e., $\in C^\infty(M)$). Whenever we define an EDS, it is implicitly restricted to a submanifold specified by the vanishing set of such functions. For example, if a DE system is expressed as $\Delta^{(s),i} = 0$, where $i = 1, \ldots, n$, then $\mathcal{I}_{diff} := \mathcal{I}_{diff}|_{\Sigma^{(s)}}$, with $\Sigma^{(s)}$ as defined in Definition 24.*

We revisit the example introduced in Section 1.1. As noted in Remark 12, since $\Delta^{(1)}(x, u, u_x) = 0$ in (1.1), we have $\left\{\Delta^{(1)}\right\} = \mathcal{I}^0$. For $\mathcal{I}_{\text{diff}}$, we first define a specific

EDS in which the prolongation of any solution graph $\Gamma_{\boldsymbol{f}}$, using the operator $pr^{(s)}$ as specified in Definition 16, is the integral submanifold.

Definition 38. *Let the jet space of order s, its coordinates, and the base manifold M be as defined in Definition 14. Suppose $\Omega^{(0),p} \ni \alpha \coloneqq dx^1 \wedge \ldots \wedge dx^p \neq 0$ represents the independence condition in Definition 36, and $\mathcal{I}^0 = \left\{\Delta^{(s),1}, \ldots, \Delta^{(s),n}\right\}$ defines a system of DEs. Then, $\mathcal{I}_{diff}$, given by all basic contact forms in* (3.29) *and restricted to $\Sigma^{(s)}$, specifies an EDS called a contact system (see (1.29) in [14]). Contact systems, without any restriction to a particular submanifold, are also Pfaffian systems (see Definition 5.1.1 in [14]), as they consist solely of one-forms with $\alpha \neq 0$.*

Remark 13 (Pfaffian System Clarification)**.** *A DE system of Pfaffian type should not be conflated with the fact that any contact system is Pfaffian. Suppose $\mathcal{I}_{diff}$, together with an independence condition $\Omega^p(M) \ni \alpha \neq 0$, forms a contact system. Since any $d\theta \in \mathcal{I}_{diff}$ for any contact form $\theta \in \mathcal{I}_{diff}$, it is called a Pfaffian system. However, once this contact system is restricted (or pulled back) to a DE system, $\mathcal{I}_{diff}$ may no longer be closed, and thus, the resulting system may cease to be Pfaffian. To determine if a DE system (restated in the form of a contact system) is Pfaffian, it must satisfy the following condition for any $\theta^i \in \mathcal{I}_{diff}$ when restricted to $\Sigma^{(s)}$:*

$$\theta^1 \wedge \ldots \wedge \theta^L \wedge d\theta^i = 0. \tag{3.42}$$

Now, let $A(x,u) = \frac{u^3+x^2u-u-x}{x^3+u^2x+u-x}$. Returning to the example in the introduction, according to Definition 38, we have $\mathcal{I}_{\text{diff}} = \{\theta_1 \coloneqq du - u_x dx, \theta_2 \coloneqq du_x - dA\}$, where the second one-form is imposed by (1.1). Furthermore, from Definition 33, $d\theta_1 \in \Omega^{(1),2}$ should an element of $\mathcal{I}_{\text{diff}}$, while $d\theta_2 = 0$ due to $d^2 = 0$ as specified in Definition 19. However, as indicated in Remark 11, we do not include any $d\beta$ for $\beta \in \mathcal{I}_{\text{diff}}$ in the generator subset $\mathscr{G}$ of $\mathcal{I}_{\text{diff}}$. Thus, θ_1 and θ_2 are the only generators and the independence condition is $dx \neq 0$. Let $\iota : \Sigma^{(1)} \to J^{(1)}(M)$ be the immersion, where $\Sigma^{(1)}$ is the regular submanifold describing $\Delta^{(1)} = 0$ in (1.1). Then:

$$\begin{aligned} \iota^*(\theta_1) &= du - A(x,u)\, dx \quad \text{and} \\ \iota^*(d\theta_1) &= d\,\iota^*(\theta_1) \\ &= -d\,A(x,u) \wedge dx \\ &= \left(\frac{\partial A}{\partial x}(x,u)\, dx + \frac{\partial A}{\partial u}(x,u)\, du\right) \wedge dx, \end{aligned} \tag{3.43}$$

where d is the exterior derivative operator defined in Definition 19, ι^* is the pull-back by the immersion ι (see Definition 21), and the second equality follows from (8.8) in [10]. Thus, $\iota^*(\theta_1) \wedge \theta_2 \wedge \iota^*(d\theta_1) = 0$, which implies that the DE system in (1.1) is Pfaffian by (3.42). Specifically, $\iota^*(\theta_1) \wedge \iota^*(d\theta_1) = 0$. Finally, the EDS for our simple DE system is represented as follows:

$$\begin{aligned} \Delta^{(1)}(x, u, u_x) &= 0, \\ \mathcal{I}_{\text{diff}} &= \{\theta_1, \theta_2\}, \\ \alpha &= dx \neq 0. \end{aligned} \tag{3.44}$$

The rest of this subsection expands on the concept of involutivity in the context of vector field systems (VFS) and EDS, including Frobenius' theorem, which guarantees the existence of integral submanifolds when the system is Pfaffian.

Remark 14 (Applicability of Frobenius' Theorem to $J^{(s)}(M)$)**.** *In the remainder of this subsection, we intentionally refrain from explicitly specifying the jet space $J^{(s)}(M)$, which itself constitutes a manifold, as the discussion applies to any manifold M. Additionally, we highlight the following commutation relation, which is essential for understanding how involutivity calculation on M extends to $J^{(s)}(M)$ (see Theorem 2.39 in [8]):*

$$pr^{(s)}([\hat{v}_j, \hat{v}_k]) = \left[pr^{(s)}(\hat{v}_j), pr^{(s)}(\hat{v}_k)\right]. \tag{3.45}$$

Although the equality in (3.45) *holds universally, it is important to note that $[\hat{v}_i, \hat{v}_j] \in \mathcal{V}$ does not necessarily imply $pr^{(s)}([\hat{v}_i, \hat{v}_j]) \in pr^{(s)}(\mathcal{V})$. This is because the prolongation of $\Delta^{(s_0)}$ (see Section 4.3) to $\Delta^{(s)}$, for $s > s_0$, may introduce additional constraints that must be satisfied.*

We begin with the following definition:

Definition 39. *Let $\mathcal{V} = \{\hat{v}_1, \ldots, \hat{v}_p\} \subset TM$ be a vector space closed under multiplication by any $f \in C^\infty(M)$. Then, $\mathcal{V}$ is called a vector field system (VFS). Involutivity of a VFS, as specified here, requires that the Lie bracket between any pair of vector fields in the system remains within the VFS, ensuring closure (see Definition 13.20 in [10]):*

$$[\hat{v}_j, \hat{v}_k]\,|_z = \sum_i a^i_{jk}(z)\hat{v}_i|_z \in \mathcal{V}, \tag{3.46}$$

where $a^i_{jk}(z)$ are smooth scalar functions on M and $i, j, k = 1, \ldots, p$.

Remark 15 ($\mathcal{V}$ is not a $\mathfrak{g}$)**.** *$\mathcal{V}$ in Definition39 is not necessarily a Lie algebra $\mathfrak{g}$, as defined in Definition 7. This distinction means that the structure constants $a^i_{jk}(z)$ are not required to be constants, which would otherwise be necessary in a Lie algebra.*

The next definition formalizes the dual relationship between an EDS and a VFS:

Definition 40. *Let $\mathcal{V}$, as in Definition 39, be an involutive system. Then, the simply-generated EDS of constant rank $m - p$ (see Remark 16) determined by $\mathcal{I}_{diff}$ is its corresponding dual, if:*

$$\omega|_z(\hat{v}_j|_z) = 0, \ \forall\, \omega \in \mathcal{I}_{diff} \text{ and } j = 1, \ldots, p, \tag{3.47}$$

where $z \in M$.

Analogous to the existence and uniqueness theorem for the first-order ODEs (see Theorem 2.8.1 in [16]), the involutivity of a VFS or the closure of its simply-generated dual EDS implies the existence of an integral submanifold. This is formalized as follows:

Theorem 4 (Frobenius' Theorem for a VFS)**.** *Let $\mathcal{V}$ be a VFS of constant rank p in an open neighborhood $U \subset M$ containing $z \in U$. Then, there exists an integral submanifold $N_U \subset U$ of dimension p passing through z, and $\mathcal{V}$ is called integrable at z if and only if $\mathcal{V}$ is involutive (see Definition 39).*

Equivalently, this condition can be expressed in terms of EDS:

Theorem 5 (Frobenius' Theorem for an EDS)**.** *Let $\mathcal{I}$ be a simply-generated ideal (see Definition 35) of constant rank $m - p$. Then, $\mathcal{I}$ is called p-integrable if and only if $\mathcal{I}$ is closed (i.e., forms a differential ideal).*

We refer the reader to Theorem 14.1 and Proposition 13.22 in [10] for the proofs of these theorems.

Remark 16 (Constant Rank and Independence Condition)**.** *The reader may interpret the "constant rank" condition in Theorems 4 and 5 as the independence condition $\alpha \neq 0$ on a given integral submanifold described in Definition 36.*

This subsection focuses primarily on contact systems as defined in Definition 38, which meet the involutivity condition stated in Frobenius' theorem (Theorem 4, equivalently Theorem 5) for the sake of an introduction to "how a DE system is represented in terms of an EDS". It is important to note again, however, that not all differential equations can be expressed as a simply-generated EDS (see Remark 13). Addressing more complex systems often requires the Cartan-Kähler theorem (see Chapter 15 in [10]). For readers interested in an example of a non-Pfaffian DE system, the wave equation in three spatial dimensions, with no additional constraint imposed (e.g. radially symmetric dependent variable), is a good candidate and its EDS can be constructed using REDUCE's *pde2eds* procedure in **EDS** package[46]. It will introduce higher order compatibility conditions, which in turn, indicate a not-simply-generated EDS. This particular study employs the wave equation in two spatial dimensions as the sample DE system in Section 5.1 and in Appendix B, it is also shown to be of Pfaffian type

Fortunately, the framework proposed in this research, as detailed in Section 4, is unaffected by whether the DE system under consideration is Pfaffian. This is because the framework specifically targets non-trivial, contact-preserving symmetries that map the graph of one solution to that of an entirely new solution. By assumption, the action of such non-trivial symmetries is locally effective (see Definition 12). Consequently, the complementary submanifold $\mathcal{N}^{(s)^\perp} \subset \Sigma^{(s)}$ generated by the flow of these symmetry generators, which are in involution, is transverse to the integral submanifolds defined by the contact system. The associated VFS corresponding to $\mathfrak{g}$ must satisfy involutivity in the sense of Frobenius and preserve the contact structure. If these conditions are not met, the action of $G^{(s)}$ would disrupt the contact forms characterizing the DE system, and its orbits would fail to form a foliation aligned with the flow of these non-trivial symmetry generators.

3.6 Cartan Equivalence

This and the following sections present the mathematical tools essential for constructing the G-coframe [45, 13]. In subsequent studies, the framework established in Chapter 4 and the results obtained in Chapter 5 will be utilized for this purpose. Ac-

cordingly, these sections serve as a reference guide for the forthcoming developments.

Cartan's equivalence is a powerful and versatile tool that defines continuous equivalence between various mathematically significant objects, such as metrics, sets of DEs, or collections of differential geometric entities. We strongly recommend that readers, even those familiar with the subject, refer to Chapter 1 of [15] for a concise overview of various equivalence problems. Owing to the foundational work of Sophus Lie, Jean Gaston Darboux, Élie Cartan, and many of their successors, the state-of-the-art *equivalence framework* is well-established. However, its scope is too vast to cover comprehensively, even when confined to the domain of DEs.

The primary aim of this section is to introduce its application to DEs with a particular focus on the study in the follow-up work of this thesis. We will refrain from detailing the three main steps of Cartan's algorithm, namely absorption, normalization, and prolongation, as these are seamlessly integrated into the algorithm we are planning to employ-the method of moving coframes (see Section 3.7). However, it is important to note that the method of moving coframes does indeed incorporate several steps from Cartan's algorithm. Therefore, we strongly encourage the reader to consult Chapter 4 of [15] for a thorough understanding of absorption (particularly the section concerning Lie algebra-compatible absorption) and normalization (specifically the part related to the reduction of structure group theorem). As for a detailed discussion on prolongation, refer to Chapter 12 of [10].

Remark 17 (Cartan's Prolongation Procedure)**.** *It is crucial to clarify that the term "prolongation," as used in Cartan's algorithm, is unrelated to the prolongation procedure discussed in Section 3.2.*

3.6.1 The Equivalence and Its Lifting Procedure

We previously introduced a glimpse of the equivalence problem in Definition 1 concerning a set of DEs. However, Cartan's equivalence problem is defined using one-forms (see Section 3.3).

Definition 41 (Cartan's Equivalence)**.** *Let U and V be two subsets of M (where M is any smooth manifold, although it will be $J^{(s)}(M)$ in our targeted problem) with*

$U \cap V \neq \varnothing$. Let $\boldsymbol{\omega}_U = \{\omega_U^1, \ldots, \omega_U^m\}$ and $\boldsymbol{\Omega}_V = \{\omega_V^1, \ldots, \omega_V^m\}$ be two coframes (see Remark 6) on U and V, respectively, where $m = \dim(M)$. If there exists a diffeomorphism $\phi : U \to V$, such that:

$$\phi^*(\boldsymbol{\Omega}_V) = \boldsymbol{\omega}_U, \tag{3.48}$$

where ϕ^* denotes the pullback (see Definition 21) of the coframe $\boldsymbol{\Omega}_V$ on V to U via ϕ, then $\boldsymbol{\Omega}_V$ is said to be equivalent to $\boldsymbol{\omega}_U$ under ϕ, denoted by $\boldsymbol{\omega}_U \cong_\phi \boldsymbol{\Omega}_V$.

Remark 18 (Non-empty Intersection)**.** *The condition $U \cap V \neq \varnothing$ is not essential for Definition 41; however, this fact will be employed when defining the bundle transition functions on a principal bundle and the connection one-form in Section 3.6.2.*

Cartan recognized that to "construct" an equivalence problem, as outlined in Definition 41, it is necessary to extend the equivalence space M to a (trivial) principal bundle $M \times G$, where G is referred to as the *structure group* (see Definition 42), which incorporates bundle transition functions (see Definition 45). Providing a sample case will aid in understanding the concept of "extension".

Let $z^{(1)} = (x, u, u_x) \in U$ and $Z^{(1)} = (X, U, U_X) \in V$ be two coordinate systems for subsets $U, V \subset J^{(1)}(M)$, where $M = \mathcal{X} \times \mathcal{U}$, $\mathcal{X} = \{x\}$, and $\mathcal{U} = \{u\}$. Suppose that two (ODEs) are given in these respective coordinates:

$$\begin{aligned} \Delta^{(1)} &= u_x - f(x, u) = 0, \\ \bar{\Delta}^{(1)} &= U_X - F(X, U) = 0, \end{aligned} \tag{3.49}$$

where f and F are smooth scalar functions on M (noting that they do not depend on u_x and U_X respectively, see the *Cauchy-Kovalevskaya* Theorem in Chapter 2 [8]). First, we require that the integral curves defined by the contact forms $\omega_U^3 \coloneqq du - u_x dx$ and $\Omega_V^3 \coloneqq dU - U_X dX$ (see Definition 20 and the following discussion) are mapped to each other under a diffeomorphism ϕ:

$$\phi^*(\Omega_V^3) = w\, \omega_U^3, \tag{3.50}$$

where w is a scalar function on $J^{(1)}(M)$ (see (4.53) in [10] for an explanation of why the transformation of contact forms takes this specific form). Notice that ϕ may be interpreted as a transformation of coordinate systems. Consequently, its Jacobian

matrix induces a transformation between the one-form sets $(\omega_U^1, \omega_U^2) \coloneqq (fdx, du)$ and $(\Omega_V^1, \Omega_V^2) \coloneqq (FdX, dU)$:

$$\phi^* \begin{pmatrix} \Omega_V^1 \\ \Omega_V^2 \end{pmatrix} = \begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix} \begin{pmatrix} \omega_U^1 \\ \omega_U^2 \end{pmatrix}, \tag{3.51}$$

where u and v are scalar functions on $J^{(1)}(M)$. The off-diagonal elements are equal to zero, as we assume that the transformation is *fiber-preserving*, a common assumption for most physical systems (see Example 9.3 in [10] for fiber-preserving transformations in the context of second-order differential equations). Combining (3.50) and (3.51) leads to the following equivalence problem:

$$\phi^* \begin{pmatrix} \Omega_V^1 \\ \Omega_V^2 \\ \Omega_V^3 \end{pmatrix} = \underbrace{\begin{pmatrix} u & 0 & 0 \\ 0 & v & 0 \\ 0 & 0 & w \end{pmatrix}}_{=(\gamma_{VU})^{-1} \in G} \begin{pmatrix} \omega_U^1 \\ \omega_U^2 \\ \omega_U^3 \end{pmatrix}. \tag{3.52}$$

The equivalence in (3.52) can be further simplified due to the system in (3.52) being overdetermined; but it is not a priority for the scope of this research. We refer the interested reader to Example 3 in Chapter 1 of [15] for a complete analysis of this example. Lastly, we suggest the reader should observe that the sample ODE in Section 1.1 is also subject to a similar equivalence problem.

We can observe that the matrix $(\gamma_{VU})^{-1}$ in (3.52) governs the transition between the specified coframes in the example given in (3.49). Strictly speaking, such transformations form a special linear Lie group G:

Definition 42. *Let $G \subset GL(m, \mathbb{R})$, where $GL(m, \mathbb{R})$ denotes the general linear group of dimension m over the real field (see Chapters 1-5 in [1] for a comprehensive study of general linear groups), and let the equivalence be as defined in Definition 41. The general form of the equivalence problem is then given by:*

$$\phi^*(\boldsymbol{\Omega}_V) = \left(\gamma_{VU}(z^{(s)})\right)^{-1} \boldsymbol{\omega}_U, \tag{3.53}$$

where $\gamma_{VU} : J^{(s)}(M) \to G$ is a G-valued function, and G is called the structure group of the equivalence with respect to ϕ. When $G = \{e\}$, the equivalence problem reduces to (3.48)*.*

Remark 19 (G is not the symmetry group of (3.53))**.** *The symmetry group associated with a given system of DEs, as defined in Definition 24, is distinct from the structure group G (see Definition 8.21 in [10] for symmetries of a coframe). While the symmetry group is induced by the diffeomorphism ϕ, the structure group primarily serves to model additional constraints imposed on the coframe elements by the nature of the equivalence problem (see Example 9.3 in [10] for an illustration of these constraints in practice).*

Remark 20 (Why is $G \subset GL(m, \mathbb{R})$?)**.** *Note that ϕ is a diffeomorphism, and its Jacobian matrix is involved in its pullback operator for the one-forms that describe the equivalence problem in Definition 41. Since the the composition of two Jacobian matrices is just another Jacobian matrix for ϕ, G is a collection of invertible matrices that preserves the equivalence.*

Remark 21 (Why the Inverse of γ_{VU}?)**.** *Our action on the principal bundle $\mathscr{P}$ in Definition 43 is defined as a right action. However, we often reference [15], where the author assumes a left action on the fibers of $\mathscr{P}$. To maintain consistency with our framework, we convert Gardner's left action into a right action by taking the inverse of γ_{VU} (see Definition 5 for more information).*

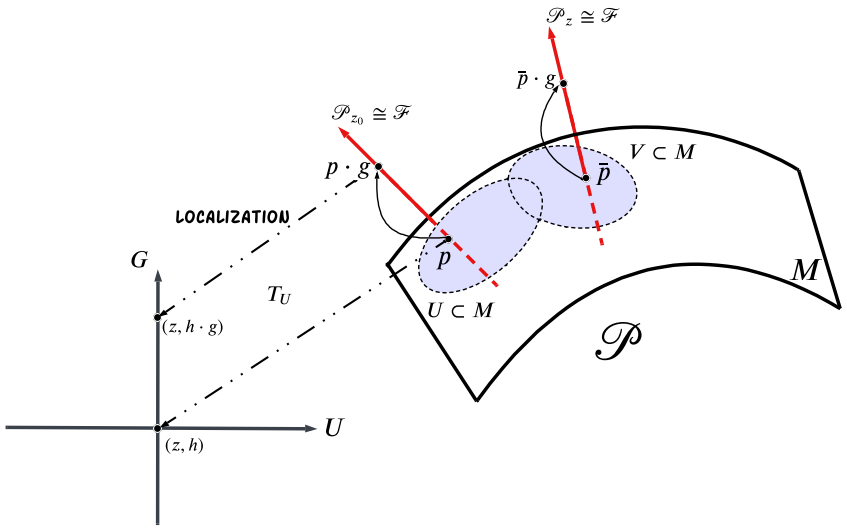

Figure 3.2: A depiction of a principal bundle and of its localization (see Definitions 43 and 44, respectively). Each fiber $\mathscr{P}_z$ is isomorphic to a given subspace $\mathscr{F}$, $\forall\, z \in M$.

A straightforward comparison of (3.48) and (3.53) reveals that the presence of a non-

trivial G (i.e., $G \neq \{e\}$) is the ultimate difference. This leads us to the question: "*How can* (3.53) *be transformed into the form stated in* (3.48)*?*" Only after this transformation will we be able to apply all the theorems, propositions, and lemmas developed for the equivalence problem in Definition 41 (see Chapter 8 in [10] for the theorems, propositions, and lemmas related to the equivalence in (3.48)).

To address the question posed in the previous paragraph, we first require additional definitions and notations.

Definition 43. *A principal G-bundle $\mathscr{P}$ is a fiber bundle $\pi : \mathscr{P} \to M$ (see Part X in [37] for an extensive examination of principal fiber bundles, as many relevant notions are drawn from it) with a continuous right action $\Psi : \mathscr{P} \times G \to \mathscr{P}$ that preserves its fibers, i.e.:*

$$\mathscr{P}_z \cdot g \in \mathscr{P}_z, \ \forall\, g \in G, \tag{3.54}$$

where M is the base space, $\mathscr{P}_z$ denotes the fiber over $z \in M$, and G is a topological group. Furthermore, G is assumed to act freely and transitively (see Definitions 11 and 13, respectively). Therefore, $\mathscr{P}_z$ is isomorphic to G (as a manifold), $\forall\, z \in M$ (see Figure 3.2).

Remark 22 (Notational Convention for $\mathscr{P}$)**.** *We will denote any principal bundle $\mathscr{P}$ as $\mathscr{P}(M, G)$ to explicitly specify its base space and fiber-preserving topological group in order; however, we may omit (M, G) if it is clear from the context.*

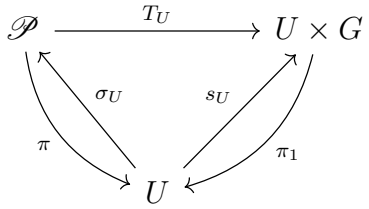

Figure 3.3: The commuting diagram that shows the localization of a principle bundle $\mathscr{P}$ (see Definition 44 for details). σ_U is a **local** section on $\mathscr{P}$; while s_U is a section on its trivialized bundle $U \times G$.

However, this global definition is insufficient for the specific problem under consideration, necessitating the localization (trivialization) of $\mathscr{P}$, similar to the localization of G and its action in Section 3.1 (see the discussion following Definition 4).

Definition 44. *A principal bundle $\mathscr{P}$ is called locally trivial if there exists an open subset $U \subset M$ for every $z \in M$, and a diffeomorphism $T_U : \mathscr{P} \to U \times G$, such that $T_U(p) = (\pi(p), \tau_U(p))$, where $p \in \mathscr{P}$, $\pi : \mathscr{P} \to U$ is the projection map satisfying $\pi(p \cdot g) = \pi(p) \in U$, and $\tau_U(p) \in G$. Since $\mathscr{P}_z \cong G$ for any $z \in U$, it follows that $\tau_U(p \cdot g) = \tau_U(p) \cdot g$. The trivialized bundle $U \times G$ is equipped with the following continuous right action:*

$$(z, h) \cdot g = (z, h \cdot g), \tag{3.55}$$

where $z \in U$, $h, g \in G$. Moreover, the diagram in Figure 3.3 commutes, ensuring that $\pi_1 \circ T_U = \pi$.

$\mathscr{P}$ could also be considered as a smooth manifold and hence, for the reader unfamiliar with *local coordinate chart* notion on a manifold, *localization* could be thought of as expressing $\mathscr{P}$ in flat coordinates (i.e. in a Euclidean space) to do calculus. The local sections $\sigma_U : U \to \mathscr{P}$ and $\sigma_V : V \to \mathscr{P}$ (see the diagram in Figure 3.3 and Definition 34.1.8 in [37]) will be essential for the development of our framework, as the moving frame (see Definition 55) is also a local section on $\mathscr{P}$. Since $\sigma_U(z) \in \mathscr{P}_z$:

$$\sigma_U(z) \cdot g \in \mathscr{P}_z, \ \forall\, g \in G. \tag{3.56}$$

Now, consider another section σ_V defined on $V \subset M$, such that $U \cap V \neq \varnothing$ as in Definition 41. Let $z \in U \cap V$. There must exist a $\gamma_{VU} \in G$ such that:

$$\sigma_U(z) \cdot \gamma_{VU} = \sigma_V(z), \tag{3.57}$$

since both $\sigma_U(z)$ and $\sigma_V(z)$ belong to $\mathscr{P}_z$ (note the property of any section σ: $\pi \circ \sigma = \mathbb{1}_M$). Using the localization as specified in Definition 44:

Definition 45. *Let T_U and T_V be the localizations of $\mathscr{P}(M, G)$ over subsets U and V, respectively. Assume that $z \in U \cap V$. Without loss of generality, let $\sigma_U(z) = p = T_U^{-1}(z, e) \in \mathscr{P}_z$ and $\sigma_V(z) = \bar{p} = T_V^{-1}(z, e) \in \mathscr{P}_z$. Therefore, $\tau_U(p) = \tau_V(\bar{p}) = e = \tau_V(p) \cdot g$, for a $g \in G$, implying $\tau_V(p) = g^{-1}$, where the right equivariance of τ_V defined in Definition 44 is employed. By utilizing the fact $\bar{p} = p \cdot g$:*

$$\begin{aligned} p = \sigma_U(z) \cdot \tau_U(p) = \sigma_V(z) \cdot g^{-1} = \sigma_V(z) \cdot \tau_V(p) \\ \implies \sigma_U(z) \cdot \tau_U(p) \cdot (\tau_V(p))^{-1} = \sigma_V(z). \end{aligned} \tag{3.58}$$

Let $\gamma_{VU}(p) \coloneqq \tau_U(p) \cdot (\tau_V(p))^{-1}$. Then, γ_{VU} is called the bundle transition function.

Remark 23 (γ_{VU} only Depends on $z \in M$)**.** *Although γ_{VU} appears to be a function of $p \in \mathscr{P}$, it is, in fact, only a function of $z \in U$ (see Proposition 34.1.5 in [37]). This is consistent with* (3.53) *in Definition 42.*

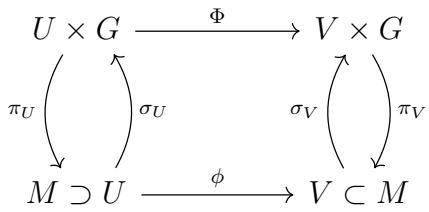

Figure 3.4: Lifting the equivalence problem in (3.53). The equivalence for σ_U and σ_V sections is provided by Φ (see (3.63)).

Now, we are prepared to articulate the "lifted" equivalence necessary to achieve the desired transformation of our previously proposed problem. Assume that $\mathscr{P}$ is a principal bundle with fibers isomorphic to the structure group G, where the initial form of G is provided by the equivalence problem itself (e.g., the γ_{VU} matrix in (3.52). We refer the reader to check out for the first example in Chapter 5 in [15]). Let $\boldsymbol{\omega}_U$ and $\boldsymbol{\Omega}_V$ be two coframes on U and V as defined in Definition 41. Suppose that we have already obtained the trivial principal bundles $U \times G$ and $V \times G$ through the localizations T_U and T_V, respectively, as given in Definition 44. We begin by lifting these coframes to their corresponding principal bundles:

Definition 46. *Suppose that:*

$$\begin{aligned} \omega^i|_{(z,\tau_U)} &= \left(\tau_U(z)^{-1}\right)^i_j \, \omega^j_U|_z, \\ \Omega^i|_{(\bar{z},\tau_V)} &= \left(\tau_V(\bar{z})^{-1}\right)^i_j \, \Omega^j_V|_{\bar{z}}, \end{aligned} \tag{3.59}$$

where $\tau_U, \tau_V \in G$, $z \in U$, and $\bar{z} \in V$ (see Remark 7 for Einstein notation and Remark 21 for inverting τ_U and τ_V). Then, both $\boldsymbol{\omega}$ and $\boldsymbol{\Omega}$ are called lifted coframes on the corresponding trivial principal bundles (see Chapter 2 in [15] for intricate details of lifted coframes).

It is now evident that $\boldsymbol{\omega}$ and $\boldsymbol{\Omega}$ are defined over $U \times G$ and $V \times G$ respectively, with $U \cap V \neq \varnothing$. Our goal is to establish an equivalence between these two collection of forms, such that:

$$\Phi^*(\boldsymbol{\Omega}) = \boldsymbol{\omega}, \tag{3.60}$$

where $\Phi : U \times G \to V \times G$ is a diffeomorphism between the two bundles (see Definition 21 for pulling back any $\mathbf{p}$-form). This partially answers the question we posed earlier, as we sought a way to transform (3.53) into a form analogous to (3.48). However, we still need to determine the form and properties that Φ must satisfy. We require Φ to fulfill the following criteria:

(i) The lifted equivalence in (3.60) must hold for the specified lifted coframes in Definition 46.

(ii) Φ should encompass the equivalence ϕ. Specifically, satisfying the Φ-equivalence must imply the ϕ-equivalence, thereby ensuring the commutativity of the diagram in Figure 3.4 (i.e., $\pi_V \circ \Phi \circ \sigma_U = \phi$).

The following definition for Φ ensures that these properties are satisfied (cf. Proposition on page 11 in [15]):

$$\Phi(p) = (\phi(z), \tau_U(z)^{-1} \cdot \gamma_{VU}(z)) = (\bar{z}, \tau_V(\bar{z})), \tag{3.61}$$

where $\gamma_{VU}(z) \in G$ is the bundle transition function defined in (3.58). Observe that the second function is equal to $(\tau_V(z))^{-1}$ (cf. Definition 45). Indeed:

$$\begin{aligned} \Phi^*(\boldsymbol{\Omega}) &= \Phi^*(\tau_V(\bar{z}) \cdot \boldsymbol{\Omega}_V) \\ &= \tau_U(z)^{-1} \cdot \gamma_{VU}(z) \cdot \phi^*(\boldsymbol{\Omega}_V) \\ &= \tau_U(z)^{-1} \cdot \gamma_{VU}(z) \cdot (\gamma_{VU}(z))^{-1} \cdot \boldsymbol{\omega}_U \\ &= \tau_U(z)^{-1} \cdot \boldsymbol{\omega}_U \\ &= \boldsymbol{\omega}, \end{aligned} \tag{3.62}$$

where (3.53) is used to derive the third equality. Note that $\sigma_U(z) = (z, \tau_U(z))$ and $\sigma_V(\bar{z}) = (\bar{z}, \tau_V(\bar{z}))$ are two sections over $U \cap V$ (see the diagram in Figure 3.4 and do not confuse these sections with the section in Figure 3.3, as these are relabeling s_U and s_V in Figure 3.3 with σ_U and σ_V respectively). Thus:

$$\Phi^*(\sigma_V) = \sigma_U. \tag{3.63}$$

We have succinctly presented the core aspects of the Cartan Equivalence problem. However, for a more comprehensive understanding, we strongly recommend that the reader consults the references cited throughout this section. Nevertheless, the introduction provided here should be sufficient for the purposes of this study.

3.6.2 Structure Equations, Connection and Torsion

We have reached a significant milestone in Chapter 3, where we introduce the fundamental elements of Cartan's algorithm (see Chapters 4, 5, and 7 in [14]). The mathematical objects discussed in this section are universally applicable, regardless of the specific algorithm used to implement a given equivalence problem, whether in practical or theoretical contexts. Consequently, these concepts are also integral to the moving coframes approach discussed in Section 3.7.

Remark 24 ($\mathscr{P}$ is Trivialized)**.** *In this subsection, the assumption is that $\mathscr{P}$ is already trivialized through the process explained in Definition 44, and hence $\mathscr{P}$ denotes $M \times G$ and is equipped with a right multiplication specified in* (3.55)*.*

We begin by introducing a crucial set of one-forms defined on T^*G, where G is the structure group defined in Definition 42:

Definition 47. *Let $\{\mu^1, \dots, \mu^r\}$ be a basis for T^*G, such that $\mu^i(v_j) = \delta^i_j$, where $\mathfrak{g} = \{v_1, \dots, v_r\}$ (see Definition 7) and $r = \dim(G)$. The defining property of these forms is:*

$$\mu^i|_g(w|_g) \coloneqq w^i|_e, \ \forall\, i = 1, \dots, r, \tag{3.64}$$

where $w|_g \in T_gG$ and $g \in G$. These forms are known as Maurer-Cartan (M-C) forms. Utilizing (3.3)*, we have:*

$$\mu^i|_g(w|_g) \coloneqq R_{g^{-1}*}\,[w|_g] = w^i|_e, \ \forall\, i = 1, \dots, r, \tag{3.65}$$

where $R_{g^{-1}*}$ denotes the pushforward associated with right multiplication by g^{-1} , as defined in Definition 6.

In the case where $G \subset GL(m, \mathbb{R})$, the M-C forms take the form:

$$\mu^i_j \coloneqq dg^k_j \cdot (g^{-1})^i_k, \tag{3.66}$$

resulting in an $r \times r$ matrix, where dg^k_j are the basis elements for T^*G (see Remark 7).

Another essential set of one-forms, known as *connection one-forms*, is defined on principal fiber bundles (see Definition 43). These forms play a pivotal role in the analysis of structure equations and are central to numerous physical applications (see

Chapters 35, 36, and 37 in [37]). Connection one-forms are also crucial in the context of this research, as the moving frame (see Definition 55) is a *vertical lift* (see Lemma 1, which directly implies the vertical lift). To proceed, we first introduce the following definition (see (34.3) in [37]):

Definition 48. *Let $\hat{v} \in \mathfrak{X}(\mathscr{P})$ (see Remark 2). Suppose that $\tau \circ c_{\hat{v}} : \mathbb{R} \to G$ defines a flow in G, such that $\frac{dc_{\hat{v}}}{d\epsilon}(\epsilon) = \hat{v}|_{c_{\hat{v}}(\epsilon)}$. If $\tau \circ c_{\hat{v}}(0) = e$, then $\hat{v}$ is called the fundamental vector field (see the equation before (34.5) in [37] for a discussion on how this condition may fail to hold). This is the converse of the statement in Definition 8, and since $\mathfrak{g} \cong T_eG$, $\exists!\ v \in \mathfrak{g}$ such that:*

$$[\tau_*]\,|_{c_{\hat{v}}(0)}(\hat{v}) = v|_e, \tag{3.67}$$

where τ_* denotes the pushforward of the map τ (cf. (3.4)*).*

Next, we describe a new action on $\hat{v}$ in Definition 48:

Definition 49. *Let $g \in G$ and $v \in \mathfrak{g}$ be related to $\hat{v} \in \mathfrak{X}(\mathscr{P})$ as defined in Definition 48. Suppose $\mathscr{P}$ is equipped with a left action $\Psi : G \times \mathscr{P} \to \mathscr{P}$ (though in our case, it is a right action). The adjoint action $Ad : G \times \mathfrak{g} \to \mathfrak{g}$ is then defined as follows:*

$$Ad_g(v) \coloneqq [R_{g^{-1}*} \circ L_{g*}]\,(v), \tag{3.68}$$

where Ad_g denotes the adjoint action of G on $v \in \mathfrak{g}$, and L_{g*} and $R_{g^{-1}*}$ are the pushforwards for the left and right multiplications by g and g^{-1}, respectively, as given in (3.2) *and* (3.3)*. The notation Ad_g represents the pushforward of the following inner automorphism of G (i.e., $Ad_g \equiv I_{g*}$):*

$$I_g(\exp(\epsilon\, v)) = g \cdot \exp(\epsilon\, v) \cdot g^{-1}, \tag{3.69}$$

where exp *is the exponential map defined in* (3.10)*. Since Ad_g is a Lie algebra isomorphism (see Definition 29.1.25 in [37]), the relationship between* (3.68) *and* (3.69) *is given by Corollary 29.1.27 in [37] as:* $\exp(\epsilon\, Ad_g(v)) = I_g(\exp(\epsilon\, v))$.

The induced action of Ad on $\hat{v}$ via (3.8) *is then defined as:*

$$\begin{aligned}
\widehat{Ad}_g(\hat{v})|_p &\coloneqq \frac{d}{d\epsilon}\Psi(\exp(\epsilon\, Ad_g(v))\, p)\Big|_{\epsilon=0} \\
&= [\Psi_{g*}]\,|_{g^{-1}\cdot p}\,(\hat{v}|_{g^{-1}\cdot p}) \\
&= [\Psi_{g*}(\hat{v})]\,|_p,
\end{aligned} \tag{3.70}$$

where $p \in \mathscr{P}$ and Ψ_{g*} is the pushforward of the left action given by g on $\mathscr{P}$. The final equality follows from the fact that $\Psi_{g*}|_{g^{-1}\cdot p} : T_{g^{-1}\cdot p}\mathscr{P} \to T_p\mathscr{P}$ (see Proposition 29.1.34 in [37] for the complete proof). Since in our case, $\mathscr{P}$ is equipped with a right action as in Definition 43, substituting g^{-1} for g (see Definition 4 for the relationship between left and right multiplications) gives:

$$\widehat{Ad}_g(\hat{v}|_p) = [\Psi_{g^{-1}*}(\hat{v})]\,|_p, \tag{3.71}$$

or equivalently:

$$\widehat{Ad}_{g^{-1}}(\hat{v}|_p) = [\Psi_{g*}(\hat{v})]\,|_p. \tag{3.72}$$

Remark 25 (Adjoint Action for $GL(m,\mathbb{R})$)**.** *When $G \subset GL(m,\mathbb{R})$, the expressions $Ad_g(v)$ and $g \cdot v \cdot g^{-1}$ can be used interchangeably, as v, g, and g^{-1} are all matrices. Differentiating* (3.69) *with respect to ϵ and evaluating the result at $\epsilon = 0$ yields a matrix multiplication by g on the left and by g^{-1} on the right. This observation justifies the notation $g \cdot v \cdot g^{-1}$.*

Observe that $\hat{v}|_p$ in Definition 48 is tangent to $\mathscr{P}_z$ (see Definition 43). Consequently, we define the following one-form:

Definition 50. *Let $\boldsymbol{\eta}|_p : T_p\mathscr{P} \to T_eG$ be a G-valued one-form on $\mathscr{P}$. If $\boldsymbol{\eta}$ satisfies:*

(i) $\boldsymbol{\eta}|_p(\hat{v}|_p) = v|_e$*, where v and $\hat{v}$ are as defined in Definition 48, and*

(ii) $\Psi_g^*(\boldsymbol{\eta}|_p)\,[\hat{v}|_p] = Ad_{g^{-1}} \circ \boldsymbol{\eta}|_p(\hat{v}|_p) = Ad_{g^{-1}}(v|_e)$*, where Ψ_g^* is the pullback of the right action given by $g \in G$ (see Definition 21) and $Ad_{g^{-1}}$ is the adjoint action as defined in* (3.68)*,*

then $\boldsymbol{\eta}$ is called the connection one-form.

Remark 26 (M-C forms are connection one-forms)**.** *Observe that any M-C form $\mu \in T^*G$ in Definition 47 satisfies two conditions specified in Definition 50 and hence, is a connection one-form.*

Remark 27 ($\boldsymbol{\eta}$ as a Transformation)**.** *It is important to note that, unlike the one-forms $\omega \in \Omega^{(s),1}$ discussed thus far, $\boldsymbol{\eta}$ does not yield a scalar. Instead, it maps $\hat{v}|_p \in T_p\mathscr{P}$ to $v|_e \in T_eG$. Therefore, it should be regarded as a linear transformation rather than a linear map.*

The property *(ii)* in Definition 50 states that the right action of G on $\mathscr{P}$, as given in (3.72), is preserved for right-invariant vector fields in $\mathfrak{g}$. Moreover, the form $\boldsymbol{\eta}$ induces a decomposition of the tangent bundle $T\mathscr{P}$ into horizontal $\mathbf{H}(\mathscr{P})$ and vertical $\mathbf{V}(\mathscr{P})$ subspaces. Notice that for any $\hat{w}|_p \in T_p\mathscr{P}$, property *(i)* in Definition 50 implies:

$$\hat{w}|_p = \hat{v}|_p + \hat{w}_H|_p, \tag{3.73}$$

where $\boldsymbol{\eta}|_p(\hat{w}|_p) = \boldsymbol{\eta}|_p(\hat{v}|_p) = v|_e$ and $\boldsymbol{\eta}|_p(\hat{w}_H|_p) = 0$, with v and $\hat{v}$ as defined in Definition 48. From this, we deduce the following:

Definition 51. *Let $\boldsymbol{\eta}$ be the connection one-form defined on $\mathscr{P}$. Then:*

$$\mathbf{H}(\mathscr{P}) \coloneqq \{\hat{w} \in T\mathscr{P} \mid \boldsymbol{\eta}|_p(\hat{w}|_p) = 0, \forall\, p \in \mathscr{P}\} \tag{3.74}$$

is called the horizontal subspace of $T\mathscr{P}$. The complement of $\mathbf{H}(\mathscr{P})$ is known as the vertical subspace of $T\mathscr{P}$ and is denoted by $\mathbf{V}(\mathscr{P}) \coloneqq \overline{\mathbf{H}(\mathscr{P})}$, or equivalently:

$$\mathbf{V}(\mathscr{P}) \coloneqq \{\hat{w} \in T\mathscr{P} \mid \pi_*(\hat{w}|_p) = 0, \forall\, p \in \mathscr{P}\}, \tag{3.75}$$

where $\pi_* : T\mathscr{P} \to TM$ is the pushforward for the projection map as defined in Definition 43. Therefore, $T\mathscr{P} = \mathbf{H}(\mathscr{P}) \oplus \mathbf{V}(\mathscr{P})$.

Since $\mathfrak{X}(\mathscr{P})$ consists of sections in $T\mathscr{P}$, the subspaces defined in Definition 51 are applicable to $\mathfrak{X}(\mathscr{P})$ when those subspaces are restricted to $\mathfrak{X}(\mathscr{P})$.

Remark 28 (Splitting $T^*J^{(s)}(M)$ vs. $T\mathscr{P}$)**.** *It is important not to confuse the decomposition $T^*J^{(s)}(M) = \mathcal{H}(M) \oplus \mathcal{C}^{(s)}(M)$ (as discussed prior to* (3.31)*) with the decomposition $T\mathscr{P} = \mathbf{H}(\mathscr{P}) \oplus \mathbf{V}(\mathscr{P})$ as given in Definition 51. The connection one-form $\boldsymbol{\eta}$ is also employed to split $T^*\mathscr{P}$, as demonstrated in Proposition 2.*

The primary objective of the preceding definitions is to establish a method for "connecting" two nearby fibers since any action by $g \in G$ moves a point $p \in \mathscr{P}$ only along the fiber $\mathscr{P}_z$ where p resides, with $z \in M$. Hence, a mechanism is required to traverse **across** fibers.

Let $\mathscr{P}$ be localized as in Definition 44 and depicted in Figure 3.3. Suppose that σ_U and σ_V are two local sections on $U \times G$ and $V \times G$, respectively, as defined in (3.57). Additionally, let $\hat{w} \in \mathfrak{X}(M)$ and $z \in U \cap V$. The connection one-form $\boldsymbol{\eta}$

can then be expressed in the following *localized* form using sections σ_U and σ_V (for a complete proof, see Section 34.2.1 in [37], noting that the inverse of the bundle transition function $g_{uv}(x)$ should be applied):

$$\boldsymbol{\eta}_V|_z(\hat{w}|_z) = L_{(\gamma_{VU}(z))^{-1}*} \circ d\gamma_{VU}|_z(\hat{w}|_z) + Ad_{(\gamma_{VU}(z))^{-1}} \circ \boldsymbol{\eta}_U|_z(\hat{w}|_z) \tag{3.76}$$

where $L_{(\gamma_{VU}(z))^{-1}*} \circ d\gamma_{VU}|_z(\hat{w}|_z)$ represents the pullback of the left invariant M-C form by γ_{VU} as per Definition 45. Here, $\boldsymbol{\eta}_V = \sigma_V^*(\boldsymbol{\eta})$, and $\boldsymbol{\eta}_U = \sigma_U^*(\boldsymbol{\eta})$, with $Ad_{(\gamma_{VU}(z))^{-1}}$ denoting the adjoint action as defined in (3.68) by γ_{VU}. Consequently, $\boldsymbol{\eta}_U|_z : T_zU \to T_eG$ and $\boldsymbol{\eta}_V|_z : T_zV \to T_eG$ serve as localized connection one-forms on U and V, respectively. The sections σ_U and σ_V are often referred to as "lifts", as they elevate any problem defined on M to one on $\mathscr{P}$ (see Definition 1.5.3 and the exercises in Chapters 1–2 of [14]). A specific lift that connects fibers horizontally with respect to a given connection one-form $\boldsymbol{\eta}$ plays a fundamental role in this analysis. As derived from (3.76), such a lift is characterized by the condition that the localized connection one-form vanishes on the generator $\hat{w}_H$ of a horizontally lifted flow F_H, i.e., $\boldsymbol{\eta}_V(\hat{w}_H) = 0$. This property leads to the following result:

Lemma 1. *Let $F_{\hat{w}} : [0,1] \to U$ be the integral curve (or flow) generated by a vector field $\hat{w} \in \mathfrak{X}(U)$. Suppose that for a local section σ_U of $U \times G$, the curve*

$$F_L \coloneqq \sigma_U \circ F_{\hat{w}} \tag{3.77}$$

defines a lifted trajectory. Then, there exists a unique curve $F_H : [0,1] \to \mathscr{P}$ such that

1. *$\pi(F_H(0)) = F_{\hat{w}}(0)$,*
2. *$F_H = F_L \cdot g$ for some function $g : [0,1] \to G$,*
3. *$\pi_*(\hat{w}_H) = \hat{w}$,*

where $\hat{w}_H \in \mathbf{H}(\mathscr{P})$ is the generator of the flow F_H, and the function g satisfies the DE:

$$\begin{aligned} \boldsymbol{\eta}_U|_{F_{\hat{w}}}(\hat{w}|_{F_{\hat{w}}}) &= -R_{g^{-1}*} \circ dg|_{F_{\hat{w}}}(\hat{w}|_{F_{\hat{w}}}) \\ &= -\hat{\boldsymbol{\mu}}|_{F_{\hat{w}}}(\hat{w}|_{F_{\hat{w}}}), \end{aligned} \tag{3.78}$$

where $\boldsymbol{\eta}_U|F_{\hat{w}}$ is the localized connection one-form and $\hat{\boldsymbol{\mu}}|_{F_{\hat{w}}}$ is the right-invariant M-C form as given in (3.65)*, both pulled back along the flow $F_{\hat{w}}$. The curve F_H is referred to as a horizontally lifted curve.*

This result will be used to establish that the moving frame section $\sigma^{(s)}$, introduced in Definition 56, constitutes a vertical lift. A proof of Lemma 1 is provided in Proposition 3.1 of [47].

By applying the adjoint action $Ad_{g^{-1}}$ to both sides of (3.78), rearranging terms, and invoking (3.76), we obtain

$$Ad_{g^{-1}} \circ \boldsymbol{\eta}_U|_{F_{\hat{w}}}(\hat{w}|_{F_{\hat{w}}}) + L_{g^{-1}*} \circ dg|_{F_{\hat{w}}}(\hat{w}|_{F_{\hat{w}}}) = \boldsymbol{\eta}_V|_{F_{\hat{w}}}(\hat{w}|_{F_{\hat{w}}}) = 0, \tag{3.79}$$

Here, similar to the transition function in (3.57), $\boldsymbol{\eta}_V$ represents the localized connection one-form associated with the flow F_H. Thus, equation (3.79) encapsulates the defining property of a horizontal lift, ensuring that it establishes a structured connection between nearby fibers in $\mathscr{P}$.

Another critical application of the connection one-forms $\boldsymbol{\eta}_U$ (or $\boldsymbol{\eta}_V$) from (3.76) is their role in spanning the cotangent bundle $T^*(U \times G)$ (or $T^*(V \times G)$), in conjunction with the lifted one-forms $\boldsymbol{\omega}$ (or $\boldsymbol{\Omega}$), which are *semi-basic* (see Appendix B.3 in [14]) as described in (3.59):

Proposition 2. *Let $\boldsymbol{\eta}_U \in T^*(U \times G)$ and $\boldsymbol{\omega} \in T^*(U \times G)$ be the local connection one-forms from* (3.76) *and the lifted one-forms from* (3.59)*, respectively. Then, $T^*(U \times G)$ is spanned by the components of $\boldsymbol{\eta}_U$ and $\boldsymbol{\omega}$.*

The proof of Proposition 2 is presented in Appendix A.2 (see also Exercise 8.3.2/1 in [14]). This proposition will be instrumental in deriving the structure equations discussed in (3.80).

With the necessary background now established, we can proceed to discuss the structure equations. These equations will provide all the essential information needed to address the equivalence problem introduced in (3.60) in Section 3.6. Let $\boldsymbol{\omega}$ denote the lifted one-forms generating a coframe on U related to the equivalence problem in (3.59). Additionally, let $\boldsymbol{\pi} = \{\pi^1, \ldots, \pi^r\}$ form a basis for T^*G. By applying

the exterior derivative operator d as defined in Definition 19 and omitting evaluation points, we obtain:

$$\begin{aligned} d\,\omega^i &= d\left[\left(\tau_U^{-1}\right)^i_j\,\omega^j_U\right] \\ &= d\left(\tau_U(z)^{-1}\right)^i_j \wedge \omega^j_U + \left(\tau_U(z)^{-1}\right)^i_j\, d\,\omega^j_U \\ &= d\left(\tau_U(z)^{-1}\right)^i_j\,\tau_U(z)^j_k \wedge \left(\tau_U(z)^{-1}\right)^k_l\,\omega^l_U + \left(\tau_U(z)^{-1}\right)^i_j\, d\,\omega^j_U \\ &= \mu^i_k|_{\tau_U^{-1}} \wedge \omega^k + (T_U)^i_{jk}(z,g)\,\omega^j \wedge \omega^k \\ &= A^i_{k\kappa}\pi^\kappa \wedge \omega^k + (T_U)^i_{jk}(z,g)\,\omega^j \wedge \omega^k, \end{aligned} \tag{3.80}$$

where $\mathbb{1} = \tau_U(z)^j_k\,(\tau_U(z)^{-1})^k_l$ is inserted to obtain the fourth equality, $\mu^i_k|_{\tau_U^{-1}}$ is the M-C form in (3.65), and $A^i_{k\kappa}$ is an $m \times m$ constant matrix for $\kappa = 1, \ldots, r$ (see Exercise on page 19 in [15] for why it is a constant matrix). As for the second term in the second equality, $d\,\omega^j_U = K^j_{kl}(z)\,\omega^k_U \wedge \omega^l_U$, where K^j_{kl} is a set of smooth functions, since $\boldsymbol{\omega}_U$ forms a coframe on U. When multiplied by $(\tau_U(z)^{-1})^i_j$, they become semi-basic (see Appendix B.3 in [14]) two-forms and can thus be represented by the lifted coframe $\boldsymbol{\omega}$ on $U \times G$, which, themselves, are semi-basic too. For further details regarding (3.80), refer to (3) in Chapter 3 of [15].

We now turn our attention to analyzing the structure equations in (3.80), continuing through the end of this section. We begin by examining the set of functions $(T_U)^i_{jk}(z,g)$, known as torsion coefficients (or, as we shall see shortly, the torsion of the connection). These coefficients are of both physical and theoretical significance. In Euclidean space, as described in (3.15), partial derivative operators acting on a smooth function commute, except in pathological cases. However, this commutativity generally does not hold when differentiating a section on a principal bundle. Torsion measures the *failure* of this commutation relation of differential operators for a given frame on a fiber bundle (see Section 8.2 in [38] for an introduction to the torsion tensor and its relation with connection one-forms in Definition 50). Let $\boldsymbol{\mathcal{D}} = \{\mathcal{D}_1, \ldots, \mathcal{D}_m\}$ be the frame dual to the coframe $\boldsymbol{\omega}$ in (3.80). Since the lifted forms $\boldsymbol{\omega}$ are sections on $\mathscr{P}$, the coefficients $(T_U)^i_{jk}$ measure the failure of the commutation of $\mathcal{D}_j$ and $\mathcal{D}_k$ along the direction of $\mathcal{D}_i$. Therefore, torsion represents an obstacle that needs to be addressed.

By Proposition 2, there should be an equivalent representation of the structure equa-

tions in (3.80) using the components of $\boldsymbol{\eta}_U$ and $\boldsymbol{\omega}$. Indeed, we have:

$$d\,\omega^i = {\eta_U}^i_k \wedge \omega^k + \left(T_{\boldsymbol{\eta}_U}\right)^i_{jk}(z,g)\,\omega^j \wedge \omega^k, \tag{3.81}$$

where $\boldsymbol{\eta}_U$ represents the local connection one-forms in (3.76), and $\left(T_{\boldsymbol{\eta}_U}\right)^i_{jk}$ is known as the torsion of the connection (see Proposition 8.3.3 in [14] for the proof). A simple comparison of (3.80) and (3.81) reveals:

$$\begin{aligned} \left({\eta_U}^i_k - A^i_{k\kappa}\pi^\kappa\right) \wedge \omega^k &\equiv 0 \bmod \left\{\omega^1, \ldots, \omega^m\right\} \\ \implies {\eta_U}^i_k - A^i_{k\kappa}\pi^\kappa &\equiv 0 \bmod \{\boldsymbol{\omega}\}, \end{aligned} \tag{3.82}$$

where $\bmod\{\boldsymbol{\omega}\} \coloneqq \bmod\{\omega^1, \ldots, \omega^m\}$ indicates that we are considering the statement modulo the algebraic ideal $\mathcal{I}_{\text{alg}}$ generated by the coframe $\boldsymbol{\omega}$ (see Definition 32), and Cartan's Lemma (see Lemma A.1.9 in [14]) is applied to obtain the second equivalence statement. Equation (3.82) is remarkable in that it asserts that any connection one-forms $\boldsymbol{\eta}_U$ differ from a set of M-C forms on G by a combination of one-forms in $\boldsymbol{\omega}$:

$$\begin{aligned} {\eta_U}^i_k &= A^i_{k\kappa}\pi^\kappa + \tilde{\nu}^i_{kl}\omega^l \text{ or} \\ {\eta_U}^i_k &= A^i_{k\kappa}(\pi^\kappa + \nu^\kappa_l\omega^l), \end{aligned} \tag{3.83}$$

where $\tilde{\nu}^i_{kl} = A^i_{k\kappa}\nu^\kappa_l$ is a set of smooth functions on $\mathscr{P}$. Since, by Remark 26, any M-C form $\mu^i_k = A^i_{k\kappa}\pi^\kappa$ is a connection one-form and hence, any two connection one-forms differ by a linear combination of elements of the coframe $\boldsymbol{\omega}$ (at a given $p \in \mathscr{P}$). Although we have utilized the coframe on $U \subset M$, the same reasoning applies to $\boldsymbol{\Omega}$ on $V \subset M$ as well.

To define a **unique** connection one-form $\boldsymbol{\eta}_U$, the ambiguity, caused by $\bar{\nu}^i_{kl}\omega^l$ in (3.83), should be resolved. This operation is called *absorption*. Substituting xx into (3.81) results in:

$$d\,\omega^i = A^i_{k\kappa}\pi^\kappa \wedge \omega^k + \left[\left(T_{\boldsymbol{\eta}_U}\right)^i_{jk}(z,g) + \left(A^i_{k\kappa}\nu^\kappa_j - A^i_{j\kappa}\nu^\kappa_k\right)\right]\,\omega^j \wedge \omega^k \tag{3.84}$$

due to anti-symmetric property of two forms, where $j < k$. To eliminate the inessential torsions, i.e. $\left(T_{\boldsymbol{\eta}_U}\right)^i_{jk}(z,g)$ satisfying

$$\left(T_{\boldsymbol{\eta}_U}\right)^i_{jk}(z,g) = -\left(A^i_{k\kappa}\nu^\kappa_j - A^i_{j\kappa}\nu^\kappa_k\right), \tag{3.85}$$

we solve (3.85) for ν^κ_j. Note that the number of variables in the system given by (3.85) is mr (notice the number of ν^κ_j functions). However, the rank of the system on the right-hand side is constrained by $\dim(G) = r$, meaning only the subset

$\left\{A^{i_1}_{k_1\kappa}\pi^\kappa, \ldots, A^{i_r}_{k_r\kappa}\pi^\kappa\right\}$ is linearly independent (recall that i and k are index into M-C forms in Definition 47). Consequently, $\boldsymbol{A}^{i_k}_{j_k\kappa}$ forms an $r \times r$ matrix of rank r (see the discussion in Chapter 3 of [15]).

Definition 52. *Solving* (3.85) *eliminates the so-called inessential torsion coefficients (refer to the L map in Chapter 4 of [15] and (10.14) in [10] for further details) and uniquely defines the connection one-form:*

$$\eta_U{}^i_k = A^i_{k\kappa}\left(\pi^\kappa + \nu^\kappa_j \omega^j\right) \tag{3.86}$$

on $U \times G$ for remaining $\left(T_{\boldsymbol{\eta}_U}\right)^i_{jk}(z, g)$ called ***essential torsion coefficients****, that do not satisfy* (3.85)*, and is denoted by $(T_E)^i_{jk}(z, g)$. This procedure is known as Cartan's absorption method and constitutes the next step after formulating the equivalence problem in* (3.60)*.*

We revisit the lifted equivalence problem introduced in (3.60) to elucidate the role of the structure equations in addressing it. Cartan made a crucial observation that the pullback of a diffeomorphism commutes with the exterior derivative operator d as defined in Definition 19 (see (8.8) in [10]):

$$d\,\Phi^*(\boldsymbol{\Omega}) = d\,\boldsymbol{\omega} = \Phi^*(d\,\boldsymbol{\Omega}). \tag{3.87}$$

In addition to the structure equations for the coframe on U provided in (3.80), we now present a corresponding set of structure equations for the coframe on $V \times G$:

$$d\,\Omega^i = A^i_{k\kappa}\bar{\pi}^\kappa \wedge \Omega^k + (T_V)^i_{jk}(\bar{z}, \bar{g})\,\Omega^j \wedge \Omega^k \tag{3.88}$$

where $\{\bar{\pi}^1, \ldots, \bar{\pi}^r\}$ forms another basis on G arising from the diffeomorphism Φ. Notably, the same constant matrices $A^i_{k\kappa}$ are used; the rationale for this is that $A^i_{k\kappa}$ is a constant matrix and, under the pullback of Φ, it remains unchanged. Moreover, the essential torsion coefficients are also transferred to $V \times G$ without being affected by Φ^* (see the diagram on page 32 in [15]), i.e.

$$\Phi^*((T_V)^i_{jk}(\bar{z}, \bar{g})) = (T_E)^i_{jk}(z, g). \tag{3.89}$$

Finally, since the absorption process in Definition 52 uniquely defines $\boldsymbol{\eta}_U$, using the observation in (3.87) results in

$$\Phi^*(\bar{\pi}^\kappa) = \pi^\kappa + \nu^\kappa_j \omega^j. \tag{3.90}$$

Recall the discussion in Definition 42. The third major step in Cartan's algorithm is to reduce the structure group G to the trivial group $\{e\}$. It is important to note that some of the essential torsion coefficients $(T_E)^i_{jk}$ explicitly depend on the group parameters g, and their level curves correspond to the flows generated by Φ in (3.60).

Definition 53. *Let $\left\{(T_E)^{i_l}_{j_l k_l}\right\}_l$ be a subset of essential torsion coefficients for the equivalence problem in* (3.60)*, where $(T_E)^{i_l}_{j_l k_l}$ explicitly depends on $g \in G$ for all $l \leq r$. To express a group parameter g as an invariant combination of $(z, \boldsymbol{h})$, we require:*

$$(T_E)^{i_l}_{j_l k_l}(z, g, \boldsymbol{h}) = C_g, \tag{3.91}$$

where $\boldsymbol{h}$ represents group parameters other than g, z is the coordinate vector on M, and C_g is an "appropriate" constant (see Example 9.9 in [10] and Example 4 in Lecture 5 of [15] for a discussion on what constitutes an appropriate constant). We then say that g is "normalized", effectively reducing the equivalence problem in (3.60) *to a simpler one, such that $G_{(1)} \subset G$ with $\dim(G_{(1)}) + 1 = \dim(G)$, where $G_{(1)}$ is the new structure group.*

We proceed by iteratively applying the absorption and normalization procedures, systematically reformulating the structure equations to address the reduced equivalence problem at each step (see Algorithm 1). Upon reaching a stage where no further parameters remain for normalization, three possible scenarios may arise:

(i) The reduced equivalence problem, which is governed by a simply generated EDS (see Definition 35), reduces to the form given in (3.48). In this scenario, the system is involutive, as guaranteed by Frobenius' theorem (see Theorem 5).

(ii) The reduced equivalence problem, which is not governed by a simply generated EDS, reduces to the form given in (3.48). Here, the system is involutive according to the Cartan-Kähler theorem (see Chapter 15 in [10]).

(iii) The coframe $\widetilde{\boldsymbol{\omega}}$ for the reduced equivalence problem still depends on some group parameters, yet no non-constant essential torsion coefficient $(\widetilde{T_E})^{i_l}_{j_l k_l}$ remains to normalize these parameters, leaving the system not involutive.

The determination of the correct outcome among these scenarios is made using a procedure known as *Cartan's involutivity test* (see Chapter 11 and the discussion in

the final section of Chapter 15 in [10]). The first case, when dealing with a system of DEs, refers to the Pfaffian systems (see Section 3.5), as illustrated in (3.44) for (1.1). In contrast, the condition for involutivity in (ii) is more challenging to compute, as it requires Cartan-Kähler Theorem. The final scenario, (iii), is addressed through a method called *prolongation* (see Remark 17). In this context, the principal bundle $\mathscr{P}$ on which the equivalence problem is formulated has a dimension less than that of the structure group G (i.e., $m < r$) and therefore requires extension.

3.7 Moving Coframes

The **moving coframes method** is a powerful **analytical** tool for decomposing the cotangent bundle of a localized jet space bundle $\mathscr{P}^{(s)} = G \times \big(U \subset J^{(s)}(M)\big)$ in a G-invariant manner. In short, it provides a systematic approach for constructing a unique G-coframe on $\mathscr{P}^{(s)}$. Before proceeding to the core discussion, we emphasize an important conceptual distinction, as highlighted in Remark 19.

In Cartan's equivalence method (Section 3.6), the group G is referred to as the structure group (Definition 42), which acts on the vector bundle T^*M, where M is the base manifold on which the original equivalence problem (Definition 41) is formulated. In contrast, the symmetry group H acts on M to preserve the underlying **geometric structure**—for example, the prolonged graph of a particular solution $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$. Consequently, G encodes additional **degrees of freedom** inherent in the formulation of the equivalence problem, as exemplified in (3.52).

Cartan's reduction process (Definition 53) eliminates this extraneous gauge freedom, leaving behind a uniquely determined geometry on M. At this stage, the equivalence problem is expressed entirely in terms of transformations of M under the action of H. However, in the moving coframes method, there is no structure group in the sense described in Section 3.6. Instead, we denote the symmetry group itself as G and define the principal bundle of interest as $\mathscr{P}^{(s)}$ to maintain consistency with the notation used in foundational works such as [45, 13, 48].

As the reader may recall, the cotangent bundle $T^*J^{(s)}(M)$ is initially split into horizontal and vertical subbundles, $\mathcal{H}(M)$ and $\mathcal{C}^{(s)}(M)$, respectively, as given in (3.31).

However, the decomposition $\mathcal{H}(M) \oplus \mathcal{C}^{(s)}(M)$ is not invariant under the action of G. Moreover, as in Definition 51, the cotangent bundle $T^*\mathscr{P}^{(s)}$ must also be decomposed into G-invariant subspaces. We discuss it in Section 3.7.3.

To address this issue, the moving coframes method [45, 13] provides an alternative to Cartan's equivalence framework by constructing a G-coframe on $T^*J^{(s)}(M)$. The key ingredients in this approach are the notions of regular cross-sections and locally free group actions (see Definitions 54 and 11, respectively), which enable the construction of a G-coframe through a simpler normalization procedure than that required in Cartan's method. Furthermore, unlike Cartan's approach, the moving coframes method bypasses the need for the inherently algebraic Cartan test (Chapter 11 in [10]) to verify the involutivity of the constructed coframe.

The remainder of this section introduces the moving coframes method within the context of this thesis. A fundamental distinction from Cartan's method is that moving coframes do not lift the equivalence problem to a vector bundle whose fibers are isomorphic to a structure group. Instead, the method utilizes the symmetry group action itself to lift the equivalence problem from $J^{(s)}(M)$ to the localized principal bundle $\mathscr{P}^{(s)}$. The discussion proceeds as follows:

(i) We first describe this lifting procedure in detail.

(ii) Next, we introduce the key mathematical structures required for the construction of a G-coframe via moving coframes, namely the cross-section, moving frame, and moving frame section.

(iii) We then demonstrate how the cotangent bundle $T^*\mathscr{P}^{(s)}$ can be decomposed into its G-invariant components.

(iv) Finally, we discuss how moving coframes ensure the involutivity of a coframe on $T^*J^{(s)}(M)$.

Remark 29 (The Base Manifold in Section 3.7)**.** *In Section 3.6 and in any referenced article, there is no mention of a particular base manifold and it is generally denoted by M. However, in this section, we specifically state that the base manifold of $\mathscr{P}^{(s)}$ is the jet space bundle $J^{(s)}(M)$ as defined in Section 3.2.*

3.7.1 Right Regularization and Lifting Procedure

Let $\mathscr{P}^{(s)} = G \times J^{(s)}(M)$ be a trivial principal bundle (see Definition 43). The right action on this bundle differs from the right action described in (3.54). The purpose of this modified right action is to lift the action of G to $\mathscr{P}^{(s)}$, thereby eliminating irregularities introduced by the original group action of G on $J^{(s)}(M)$ (see the discussion of such irregularities in Chapter 3 of [13]). The right action $\Psi_R : \mathscr{P}^{(s)} \times G \to \mathscr{P}^{(s)}$ is defined as

$$\Psi_R((z^{(s)}, g), h) \coloneqq (h^{(s)} \cdot z^{(s)}, g \cdot (h^{(s)})^{-1}) \tag{3.92}$$

and is referred to as the right regularization of the action of G.

The concept of a lifted invariant was briefly introduced in Definition 29. We now elaborate on this notion in greater detail. Lifting an analytical object (e.g., a vector field or a scalar function) defined on $J^{(s)}(M)$ to $\mathscr{P}^{(s)}$ is a straightforward process when the fundamental right lifted invariant is given by

$$\mathscr{L}_R^{(s)} = \Psi(g, z^{(s)}) = g^{(s)} \cdot z^{(s)}, \tag{3.93}$$

where $\Psi(g, z^{(s)})$ denotes the group action in (3.1). The quantity $\mathscr{L}_R^{(s)}$ is invariant under Ψ_R as defined in (3.92). The level sets of $\mathscr{L}_R^{(s)}$, given by $\left[\mathscr{L}_R^{(s)}\right]^{-1}\left\{z^{(s)}\right\}$, correspond to the orbits of G in $\mathscr{P}^{(s)}$ passing through $(z^{(s)}, e)$ (see Proposition 3.5 in [45] for further details).

To lift a scalar function $F : J^{(s)}(M) \to \mathbb{R}$, it suffices to consider the composition

$$F \circ \mathscr{L}_R^{(s)} : \mathscr{P}^{(s)} \to \mathbb{R}. \tag{3.94}$$

It is important to distinguish between a differential invariant $I^{(s)} : J^{(s)}(M) \to \mathbb{R}$, as introduced in Definition 28, and the lifted invariant $F \circ \mathscr{L}_R^{(s)}$. After lifting an ordinary differential invariant via (3.94), its group differential satisfies

$$d_G\left[I^{(s)} \circ \mathscr{L}_R^{(s)}\right] = 0, \tag{3.95}$$

whereas the group differential of $F \circ \mathscr{L}_R^{(s)}$, i.e., $d_G\left[F \circ \mathscr{L}_R^{(s)}\right]$, is nontrivial and both this result and d_G are thoroughly discussed in Theorem 3.10 of [13]. The moving coframe approach exploits this distinction to differentiate $I^{(s)}$ on $J^{(s)}(M)$ from an arbitrary lifted invariant $F \circ \mathscr{L}_R^{(s)}$ on $\mathscr{P}^{(s)}$.

Although this thesis does not delve into the details of lifting vector fields on $J^{(s)}(M)$, the principle underlying (3.94), combined with the component functions of $\mathrm{pr}^{(s)}(\hat{v})$ in (3.20), provides insight into how this process is achieved. For a comprehensive discussion on the lifting procedure for vector fields and p-forms, we refer the reader to Chapter 5 in both [49] and [48], respectively.

3.7.2 Cross-Section, Moving Frame, & Moving Frame Section

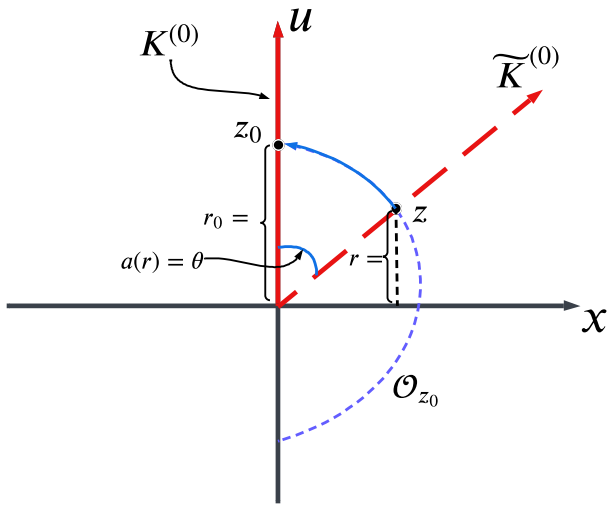

Figure 3.5: Identity cross-section $K^{(0)} = (e, r)$ and moving frame $\rho^{(0)}$, that generates the group action $\rho^{(0)}(z) \in SO(2)$, such that $\rho^{(0)}(z) \cdot (\theta, r) \in K^{(0)}$, where $(\theta, r) \in \widetilde{K}^{(0)}$. z_0 is the base point through which $\mathcal{O}_{z_0}$ passes.

In this subsection, we introduce the fundamental tools that motivated our use of the moving coframe method within our analytical framework. These mathematically elegant structures operate in harmony, enabling us to remain within the realm of **analysis** without resorting to algebraic techniques.

Suppose that G acts regularly (see Definition 1.26 in [8]) and that the dimension of its orbits is equal to r. Then, by Theorem 2.23 in [10], there exists a rectifying coordinate system in a sufficiently small neighborhood $U \subset J^{(s)}(M)$, given by $(t^1, \ldots, t^r, y^1, \ldots, y^{m-r})$, where $\boldsymbol{t} \in G$ corresponds to the coordinates along the group orbits, and $\boldsymbol{y} \in Y$ are invariant coordinates under the action of G (see Example

2.7 in [13]). The coordinates $\boldsymbol{y}$ thus define a special subspace of $J^{(s)}(M)$.

Definition 54. *Let G act semi-regularly on $J^{(s)}(M)$ with $\dim(\mathcal{O}_{z^{(s)}}) = r$, where $\mathcal{O}_{z^{(s)}}$ is the orbit of G with $z^{(s)} \in U \subset J^{(s)}(M)$. The orbits of G, then, intersect an $(m-r)$-dimensional submanifold $K^{(s)} \subset U$ transversally. This submanifold $K^{(s)}$ is called a (local) cross-section (see Figure 3.5). If, in addition, $K^{(s)}$ intersects each orbit $\mathcal{O}_{z^{(s)}}$ at most once, then $K^{(s)}$ is called a regular cross-section. In rectifying coordinates, a cross-section is given by $K^{(s)} = (a(\boldsymbol{y}), \boldsymbol{y})$, where $a : Y \to G$ is a smooth function, i.e., $K^{(s)}$ is the graph of a function in invariant coordinates. If $a(y) = e$, then $K^{(s)}$ is called the identity cross-section.*

The cross-section $K^{(s)}$ plays a crucial role in the *normalization* procedure within the moving coframe method (see Example 4.12 in [13] and Example 10 in [48]).

As demonstrated in Chapter 5, our dataset consists of the prolonged graph of a given solution $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$, denoted by $\Gamma_{\boldsymbol{f}}^{(s)}$ (see Section 3.2 for further details on $\Gamma_{\boldsymbol{f}}^{(s)}$). In the context of our research, we observe that $K^{(s)} \equiv \Gamma_{\boldsymbol{f}}^{(s)}$. This key observation serves as the foundation for constructing the associated G-coframe in future work, leveraging the involutive vector field system $\mathcal{V}^{(s)}$, which represents the Lie algebra $\mathfrak{g}$ within $TJ^{(s)}(M)$ and is obtained through the methodology outlined in Chapter 4.

Remark 30 ($\Gamma_{\boldsymbol{f}}^{(s)}$ and the Identity Cross-Section)**.** *In this work and future studies, the identity cross-section $K^{(s)}$, as defined in Definition 54, will always be taken to correspond to the prolonged solution graph $\Gamma_{\boldsymbol{f}}^{(s)}$.*

In any exposition on Cartan's equivalence method, the lifting procedure plays a central role, whether formulated as in Section 3.6 or as described in 3.7.1. The moving frame section $\sigma^{(s)}$ constitutes a particular type of such a lift, endowed with several beneficial properties that will be discussed in the remainder of Section 3.7. To define $\sigma^{(s)}$, we first introduce a smooth function that gives the moving coframe method its name.

Definition 55. *Let $\rho^{(s)} : J^{(s)}(M) \to G$ be a smooth function satisfying the following two conditions:*

1. *$\rho^{(s)}$ is G-equivariant, meaning that:*

$$\rho^{(s)}(g^{(s)} \cdot z^{(s)}) = \rho^{(s)}(z^{(s)}) \cdot (g^{(s)})^{-1}. \tag{3.96}$$

2. *$\rho^{(s)}$ satisfies the compatible lift condition (see [45]):*

$$\rho^{(s)}(z^{(s)}) \cdot z^{(s)} = z_0^{(s)}, \tag{3.97}$$

*where $z_0^{(s)}$ is a **base point** such that $z_0^{(s)} \in K^{(s)}$, where $K^{(s)}$ is the relevant cross-section as defined in Definition 54.*

A smooth function $\rho^{(s)}$ satisfying these properties is called a (right) moving frame.

Theorem 4.4 in [13] guarantees the existence of a moving frame $\rho^{(s)}$, while Theorem 4.5 in [13] states that $\rho^{(s)}$ provides a complete set of differential invariants on $J^{(s)}(M)$, given by

$$\boldsymbol{I}^{(s)} = \rho^{(s)}(z^{(s)}) \cdot z^{(s)} \in K^{(s)}. \tag{3.98}$$

We illustrate the role of $\rho^{(s)}$ in Figure 3.5, where we take $x = 0$ and $u > 0$ as the identity cross-section, for the case $s = 0$. In this example, (θ, r) (polar coordinates in 2D) serve as the **canonical coordinates** for the action of $G = SO(2)$ on the $x - u$ plane. Locally, the bundle structure is given by $\mathscr{P}^{(s)} = SO(2) \times Y$, where $Y = \{r\}$. The orbits of G are transverse to $K^{(s)}$, and $K^{(s)}$ intersects each orbit exactly once. For any $K^{(0)} = (a(r), r)$, corresponding to a specific base point $z_0^{(0)} \coloneqq z_0 = (0, u_0)$, we have

$$a(r) = \begin{cases} \tan^{-1}\left(\sqrt{\frac{r_0^2}{r^2} - 1}\right), & u > 0 \\ \pi/2, & u = 0 \\ \pi/2 + \tan^{-1}\left(\frac{r}{\sqrt{r_0^2 - r^2}}\right), & u < 0, \end{cases} \tag{3.99}$$

where $r_0 = |u_0|$ and $r = |u|$. While this formulation applies to the half-plane $x \geq 0$, a similar expression holds for $x < 0$. For further details, we refer the reader to Example 4.7 in [13].

Analogous to the lifting procedure in Cartan's equivalence method (see Figure 3.4), the moving frame section is a key component of the moving coframes approach.

Definition 56. *Let $\rho^{(s)}$ be a moving frame as defined in Definition 55. Then, the section $\sigma^{(s)} : J^{(s)}(M) \to \mathscr{P}^{(s)}$, defined by*

$$\sigma^{(s)}(z^{(s)}) = (z^{(s)}, \rho^{(s)}(z^{(s)})), \tag{3.100}$$

is called a moving frame section.

The right-regularized action in (3.92) establishes that the section $\sigma^{(s)}$ is G-equivariant, as follows:

$$\begin{aligned} \sigma^{(s)}(z^{(s)}) \cdot h &= (h^{(s)} \cdot z^{(s)}, \rho^{(s)}(z^{(s)}) \cdot (h^{(s)})^{-1}) \\ &= (h^{(s)} \cdot z^{(s)}, \rho^{(s)}(h^{(s)} \cdot z^{(s)})) \\ &= \sigma^{(s)}(h^{(s)} \cdot z^{(s)}), \end{aligned} \tag{3.101}$$

where the second equality follows from Property 1 in Definition 55. Furthermore, the combination of (3.100) and (3.93) yields the mapping

$$\boldsymbol{I}^{(s)} = \mathscr{L}^{(s)} \circ \sigma^{(s)} : J^{(s)}(M) \to \mathbb{R}^m, \tag{3.102}$$

which is referred to as *invariantization* and $m = \dim(J^{(s)}(M))$. This result reiterates the conclusion obtained in (3.98).

Recalling the discussion of the horizontal lift F_H from Lemma 1, let $\sigma^{(s)}$ correspond to the section σ_U in the lemma. Then, applying (3.79) (omitting evaluation points) yields

$$Ad_{g^{-1}} \circ \boldsymbol{\eta}_U(\hat{w}) - \hat{\boldsymbol{\mu}}(\hat{w}) = 0, \tag{3.103}$$

where $\boldsymbol{\eta}_U$ denotes the localized connection one-form associated with $\sigma^{(s)}$, defined as $\boldsymbol{\eta}_U \coloneqq (\sigma^{(s)})^*(\boldsymbol{\eta}) = \boldsymbol{\eta} \circ \sigma_*^{(s)}$ and $\hat{\boldsymbol{\mu}}$ is right invariant M-C form (cf. Proposition 2.44 in [10] for converting left-invariant Lie algebra to its right-invariant counterpart). Since $\sigma^{(s)}$ is a G-equivariant section and using the property of connection one-forms given in (ii), (3.103) simplifies to

$$\begin{aligned} (Ad_{g^{-1}} \circ \boldsymbol{\eta}) \circ \sigma_*^{(s)}[\hat{w}] - \hat{\boldsymbol{\mu}}(\hat{w}) = 0 \\ \implies \sigma_*^{(s)}[\hat{w}] = \hat{\boldsymbol{\mu}}(\hat{w}). \end{aligned} \tag{3.104}$$

(3.104) establishes that $\sigma^{(s)}$ constitutes a **vertical** lift, meaning that it maps any $\hat{w} \in TJ^{(s)}(M)$ to $w \in TG$. Consequently, $\sigma_*^{(s)}$ defines a map between the Lie algebra $\mathfrak{g}$ and its representation in the subbundle $\mathcal{V}^{(s)} \subset TJ^{(s)}(M)$.

3.7.3 G-Invariant Decomposition of $T^*\mathscr{P}^{(s)}$

This section provides a brief discussion on how moving coframes decompose an invariant coframe $\Upsilon_G^{(s)}$ on $\mathscr{P}^{(s)}$ into its invariant subspaces with respect to the action of G. We begin by defining the relevant coframes while clarifying two key concepts to prevent potential misunderstandings.

The notion of an algebraic ideal $\mathcal{I}_{\text{alg}}$, as introduced in Definition 32, is closely related to that of a coframe.

Definition 57. *Let $\Upsilon^{(s)} = \{\omega^1, \dots, \omega^{m-r}\} \subset \Omega^{(s),1}$, where $m = \dim(J^{(s)}(M))$ and r is the dimension of the symmetry group G that generates self-equivalences of $\Upsilon^{(s)}$ (see Definition 8.21 in [10]). The set $\Upsilon^{(s)}$ is called a coframe if it forms a module over $C^\infty(J^{(s)}(M))$ in the cotangent bundle $T^*J^{(s)}(M)$. The rank of $\Upsilon^{(s)}$ at a point $z^{(s)} \in J^{(s)}(M)$, denoted by $r_\Upsilon(z^{(s)})$, is given by the dimension of the subspace spanned by $\{\omega^1, \dots, \omega^{m-r}\}$ at $z^{(s)}$. Thus, it satisfies the inequality:*

$$r_\Upsilon(z^{(s)}) \leq m - r. \tag{3.105}$$

*Furthermore, if $r_\Upsilon(z^{(s)})$ is constant throughout $T^*J^{(s)}(M)$, then $\Upsilon^{(s)}$ is said to be a **semi-regular** coframe. If, in addition, its classifying submanifold of order s (see Definition 5.3 in [45]), specified by its differential invariants (obtained as per Definition 5.9 in [13]), is an embedded submanifold in the relevant ambient space, then $\Upsilon^{(s)}$ is called a **regular** coframe.*

Remark 31. *(Local Regularity) The regularity condition in Definition 57 is related to the rank of the Jacobian matrix of the differential invariants $\boldsymbol{I}^{(s)}$ associated with $\Upsilon^{(s)}$, which form a functionally independent set. While verifying the constancy of the Jacobian matrix rank across $J^{(s)}(M)$ is sufficient for semi-regularity, establishing regularity additionally requires checking the injectivity and continuity of $\boldsymbol{I}^{(s)}$, as well as confirming that the Jacobian matrix attains maximal rank. Since we primarily consider sufficiently small, open subsets of $T^*J^{(s)}(M)$, the full regularity of $\Upsilon^{(s)}$ is generally unnecessary. Consequently, throughout this thesis and in future studies, we will always assume semi-regularity and that the Jacobian matrix of $\boldsymbol{I}^{(s)}$ attains maximal rank within a sufficiently small open neighborhood $U \subset J^{(s)}(M)$, ensuring local regularity.*

In accordance with Definitions 32 and 57, the algebraic ideal $\mathcal{I}_{\text{alg}}$ generated by $\Upsilon^{(s)}$ is given by $\mathcal{I}_{\text{alg}} = \{\omega^1, \ldots, \omega^{m-r}\}$. This ideal imposes additional structure on $\Upsilon^{(s)}$; for instance, the closeness of $\mathcal{I}_{\text{alg}}$ implies the involutivity of $\Upsilon^{(s)}$ in the sense of the Frobenius or Cartan-Kähler theorems (see Chapters 14 and 15 of [10], respectively).

Another potential source of confusion concerns the role of the symmetry group G of $\Upsilon^{(s)}$. In differential geometry and the study of continuous symmetries, some researchers may omit the M-C forms $\boldsymbol{\mu}$, which span T^*G (see Definition 47), from the definition of $\Upsilon^{(s)}$ when focusing purely on the geometric aspects of the equivalence problem's invariant structures. In contrast, others include the M-C forms $\boldsymbol{\mu}$ to emphasize the self-equivalences of $\Upsilon^{(s)}$. In particular, when employing the moving coframes method, $\boldsymbol{\mu}$ is appended to $\Upsilon^{(s)}$ because the structure equations of $\Upsilon^{(s)}$, and consequently its involutivity, become significant.

Definition 58. *Let $\Upsilon^{(s)}$ be a coframe in $T^*J^{(s)}(M)$ with symmetry group G of dimension r, satisfying:*

$$(g^{(s)})^*(\bar{\omega}^i) = \omega^i, \tag{3.106}$$

for $i = 1, \ldots, m-r$ and $g \in G$. Furthermore, let $\boldsymbol{\mu}$ denote the M-C forms of G, and define the coframe $\tilde{\Upsilon}^{(s)} \coloneqq \{\boldsymbol{\omega}, \boldsymbol{\mu}\}$. If $\tilde{\Upsilon}^{(s)}$ is an involutive, (locally) regular coframe, then it is called a G-coframe and is denoted by $\Upsilon_G^{(s)}$. The rank of $\Upsilon_G^{(s)}$ coincides with that of $\Upsilon^{(s)}$.

This extended discussion is necessary for two reasons. First, textbooks may obscure the distinction between and the similarity of $\Upsilon^{(s)}$ and the algebraic ideal $\mathcal{I}_{\text{alg}}$ it generates. Second, the regularity condition plays a crucial role in ensuring the involutivity of a G-coframe, as discussed in Section 3.7.4.

The goal is to decompose the cotangent bundle $T^*\mathscr{P}^{(s)}$ in an invariant manner using moving coframes. Since the cotangent space of the symmetry group, $T^*G = \{\mu^1, \ldots, \mu^r\}$, is naturally invariant, the key question is whether the remaining part of $T^*\mathscr{P}^{(s)}$ can be spanned by utilizing lifted invariants.

To achieve this, we introduce lifted coordinates for independent (x^i) and dependent

(u^α) variables as:

$$\begin{aligned} X^i &\coloneqq \mathscr{L}_R^{(s),i}, \quad i = 1, \dots, p \\ U_I^\alpha &\coloneqq \mathscr{L}_R^{(s),p+\alpha}, \quad \alpha = 1, \dots, q,\ 0 \le |I| \le s, \end{aligned} \tag{3.107}$$

where $\mathscr{L}_R^{(s)} = \Psi(g, z^{(s)})$ as in (3.93).

Two types of differentials, namely horizontal and vertical differentials, denoted by d_H and d_V, respectively:

$$\begin{aligned} d_H(L) &\coloneqq \sum_{i=1}^{p} D_i(L)\, dx^i \\ d_V(L) &\coloneqq \sum_{I}^{0\le|I|\le s} \sum_{\alpha=1}^{q} \frac{\partial L}{\partial u_I^\alpha}\, \theta_I^\alpha, \end{aligned} \tag{3.108}$$

where $L = F \circ \mathscr{L}_R^{(s)}$ is a lifted scalar function, D_i is total derivative operator in (3.17), and θ_I^α is basic contact form of order $|I|$ in (3.29) (see also Remark 8). These combine into the jet differential:

$$d_{\mathrm{J}} \coloneqq d_H + d_V, \tag{3.109}$$

that acts on coordinates of $J^{(s)}(M)$. This leads to a natural invariant splitting of one-forms on $J^{(s)}(M)$, thanks to Proposition 10.11 in [13]:

$$\begin{aligned} &d_{\mathrm{J}} X^i \quad \text{and} \\ &\Theta_I^\alpha \coloneqq d_{\mathrm{J}} U_I^\alpha - \sum_{i=1}^{p} U_{I,i}^\alpha d_{\mathrm{J}} X^i, \end{aligned} \tag{3.110}$$

where $U_{I,i}^\alpha = \mathscr{E}_i(U_I^\alpha)$, $\mathscr{E}_i \coloneqq {}^{\partial}\!/_{\partial X^i}$ is the *invariant differential operator* in (10.10) in [13] (see also Therorem 10.12 in [13]). From these, we define two invariant subspaces:

$$\begin{aligned} \mathbf{H}(\mathscr{P}^{(s)}) &\coloneqq \left\{ d_{\mathrm{J}} X^i \right\}_{i=1}^{p} \\ \mathbf{V}(\mathscr{P}^{(s)}) &\coloneqq \{\Theta_I^\alpha\}_{\alpha=1,\, I}^{q,0\le|I|\le s}. \end{aligned} \tag{3.111}$$

Thus, the invariant decomposition of $T^*\mathscr{P}^{(s)}$ is given by:

$$T^*\mathscr{P}^{(s)} \equiv \mathbf{H}(\mathscr{P}^{(s)}) \oplus \mathbf{V}(\mathscr{P}^{(s)}) \oplus T^*G. \tag{3.112}$$

This decomposition ensures that the moving coframes method provides a structured way to analyze the geometry of $\mathscr{P}^{(s)}$, while maintaining invariance under G, as

$$\Upsilon_G^{(s)} = \left\{ \mathbf{H}(\mathscr{P}^{(s)}), \mathbf{V}(\mathscr{P}^{(s)}), \boldsymbol{\mu} \right\}. \tag{3.113}$$

3.7.4 Involutivity of a Coframe on $T^*J^{(s)}(M)$

In Section 3.5, we discussed EDS and the conditions under which a given EDS becomes involutive. Our primary focus was on involutivity via Frobenius' Theorem (Theorem 4), with a brief mention of the Cartan–Kähler Theorem, which ensures the involutivity of an EDS composed of p-forms with $p > 1$. In the case of an EDS generated by $\Upsilon^{(s)}$ (constructed via the moving coframes method), ensuring involutivity is more straightforward. In Sections 5 and 6 of [13], the authors provide a detailed analysis of involution for a G-coframe constructed via the moving coframes method. In contrast, the present section provides only the necessary background for the reader to understand how moving coframes facilitate replacing *Cartan's involutivity test* with a more analytically tractable approach, making it more accessible for integration into machine learning frameworks.

In the classical Cartan framework, the standard procedure involves constructing a contact ideal (see Definition 38) for a given DE system without **prior knowledge** of nontrivial, contact-preserving symmetries—i.e., those that preserve the contact ideal while mapping one solution to another. Subsequently, conditions for involutivity are determined using computational algebraic methods [24, 25]. However, the framework we propose inherently provides prior knowledge of *intrinsic* symmetries (see Section 4.1). This enables the learning of an involutive distribution associated with each solution $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$ via Theorem 4, as discussed in Section 4.2. The foliation induced by this distribution defines a submanifold complementary to the integral submanifold (see Definition 37), denoted by $\mathcal{N}^{(s)^\perp}$. This distribution represents nontrivial, contact-preserving symmetry generators within $T^*J^{(s)}(M)$. Nevertheless, as highlighted in Definition 58, any G-coframe $\Upsilon_G^{(s)}$ comprises not only the M-C forms of G but also includes $\boldsymbol{\omega}$, the coframe invariant under the action of G. Consequently, the closure of the algebraic ideal $\mathcal{I}_{\text{alg}}$, induced by $\Upsilon_G^{(s)} = \{\boldsymbol{\omega}, \boldsymbol{\mu}\}$, must be ensured.

By definition, $\Upsilon_G^{(s)}$ must be both regular and involutive. Throughout this chapter, we emphasize that G acts locally effectively on $J^{(s)}(M)$. This ensures that the kernel of the Lie algebra representation $\psi : \mathfrak{g} \to \mathcal{W}_{\boldsymbol{f}}^{(s)}$ on $J^{(s)}(M)$ is trivial, as established by Theorem 2 (see Chapter 4 of [1] for a detailed discussion on representations). Consequently, this guarantees the involutivity of $\mathcal{W}_{\boldsymbol{f}}^{(s)}$. However, we enforce a stricter

constraint in Section 4.1, namely $\mathscr{C}_{\text{ind}}$, suffices to ensure locally free action (see Exercise 2.69 in [10]), which, by default, implies locally effective action. Specifically, the symmetry generators in $\mathcal{V}^{(s)}$ must be pointwise linearly independent in $TJ^{(s)}(M)$.

Remark 32 (Lack of a Complete Symmetry Set and Involutivity)**.** *Suppose that $s < s_G$ (see Definition 27). In this case, while $\mathfrak{g}$ may still form a Lie algebra, it lacks a complete set of symmetry generators. Nevertheless, the isomorphism $\mathfrak{g} \cong_{\psi} \mathcal{W}_{\boldsymbol{f}}^{(s)}$ holds, ensuring that $\left[\mathcal{W}_{\boldsymbol{f}}^{(s)}, \mathcal{W}_{\boldsymbol{f}}^{(s)}\right] \subseteq \mathcal{W}_{\boldsymbol{f}}^{(s)}$.*

Regarding the involutivity of $\Upsilon^{(s)}$, the requirement is more intricate; however, the moving coframes method simplifies it. Lemma 6.4 of [45] provides an analytical guarantee for the closure of $\mathcal{I}_{\text{alg}}$. This result hinges on the **existence** of a moving frame $\rho^{(s)}$, established by Theorem 6.5 of [13]. The requirements for such existence are discussed in Section 3.7.2, where G must act locally *(i)* regularly and *(ii)* freely, due to Theorem 4.5 in [13]. Since, in Section 3.7.2, the cross-section $K^{(s)} \equiv \Gamma_{\boldsymbol{f}}^{(s)}$ associated with a given solution $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$ is assumed to be regular, and regular actions admit regular cross-sections (see Definition 2.6 of [13]), the regularity of $K^{(s)}$ (see Definition 54) implies that G acts regularly. However, we cannot rely solely on the assumption that the given dataset constitutes a regular cross-section, as the proposed framework lacks a mechanism to distinguish between a regular and an irregular cross-section. During the construction of $\mathcal{W}_{\boldsymbol{f}}^{(s)} \subset TJ^{(s)}(M)$, which is dual to $\hat{\boldsymbol{\mu}} \subset T^*J^{(s)}(M)$, in Section 4.2, we impose the condition $\left[\mathcal{W}_{\boldsymbol{f}}^{(s)}, \mathcal{W}_{\boldsymbol{f}}^{(s)}\right] \subseteq \mathcal{W}_{\boldsymbol{f}}^{(s)}$. A natural question arises as to how this condition informs the proposed model of the regularity of G's action. A regular action requires that the orbit dimension remains constant across an open subset $U \subset J^{(s)}(M)$ and that each orbit contains a connected subset at any given point $z^{(s)} \in U$. As previously discussed, the constraint $\mathscr{C}_{\text{ind}}$ in (4.8) ensures the constancy of the orbit dimension. On the other hand, as established by (3.11) in Theorem 1, any local flow $F_{z_0^{(s)}} : \mathbb{R}^r \to J^{(s)}(M)$ passing through $z_0^{(s)}$ at $\boldsymbol{\epsilon} = 0 \in \mathbb{R}^r$ can be expressed as

$$F_{z_0^{(s)}}(\boldsymbol{\epsilon}) = \boldsymbol{\Phi}(\boldsymbol{v}) \cdot z_0^{(s)}. \tag{3.114}$$

By the *Baker-Campbell-Hausdorff* formula (Theorem 5.3 in [1]), the term $\boldsymbol{\Phi}(\boldsymbol{v})$ can be expressed as a series of Lie brackets. Consequently, enforcing the closure of $\mathcal{W}_{\boldsymbol{f}}^{(s)}$ under the Lie bracket operation yields an integrable action on $J^{(s)}(M)$, which, in

turn, guarantees the existence of a connected subset of the orbit for $0 < \|\boldsymbol{\epsilon}\| \leq \delta \in \mathbb{R}$, where $\delta > 0$ is a small constant.

The final question concerning involutivity pertains to the *stabilization order* $s \in \mathbb{N}$ of involutivity for $\Upsilon_G^{(s)}$. The concept of "stabilization order" of G, s_G, as introduced in Definition 27, refers to the order at which the orbits of G on $J^{(s)}(M)$ attain their maximal dimension. One may ask whether this order coincides with the order at which $\Upsilon^{(s)}$ becomes involutive. When s reaches the "order of involutivity", every differential invariant $I^{(s)}$ can be expressed in terms of fundamental invariants $\boldsymbol{I}^{(s)}$, which describe the classifying submanifold of $\Upsilon^{(s)}$ (see Section 5 of [13] for further details). Furthermore, Theorem 9.13 of [13] states that if the orbits of G attain their maximal dimension at s_G, then there exists a *contact-invariant* coframe in $T^*J^{(s)}(M)$, allowing every $I^{(s)}$ to be expressed in terms of $\boldsymbol{I}^{(s)}$. Consequently, both stabilization orders describe the same notion of "stability". In the case of Lie point symmetries, the stabilization order coincides with the order of the original DE system, i.e., $s_G = s_0$.

Remark 33 (Stabilization Order of Lie Point Symmetries)**.** *The fundamental component functions of Lie point symmetries, as defined in* (3.18)*, do not depend on fiber coordinates corresponding to the partial derivatives of u^α. Thus, point symmetries cannot generate involutivity conditions for orders $s > s_0$.*

Remark 34 (Moving Coframes Theory for Lie pseudo-groups)**.** *The fundamental theory of moving coframes for infinite dimensional groups is also established by Olver et. al in [49, 48].*

In conclusion, Section 3.7 provides only an introductory perspective on a sophisticated topic in differential geometry and continuous symmetries.

CHAPTER 4

METHODOLOGY

In this section, we present our contributions to the problem of learning/discovering non-trivial, contact preserving group symmetries for a given system of DEs. As a preliminary step, we propose a sub-algorithm designed to identify non-zero Lie point symmetries (see Remark 36 and Definition 23). Subsequently, we ensure the involutivity and non-triviality (see Definition 26) of the VFS associated with $\mathfrak{g}$ (see Definitions 39 and 7) and a particular solution $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$, as this property is essential for constructing an involutive G-coframe on $J^{(s)}(M)$ [13, 49]. The involutivity requirement arises from the need to pull back the *Maurer-Cartan* forms, which span T^*G, to $T^*J^{(s)}(M)$ via a moving frame section $\sigma^{(s)} : J^{(s)}(M) \to G \times J^{(s)}(M)$ (see Definitions 47 and 56).

As outlined in Section 3.5, the neural architecture developed in this research learns only the involutive VFS corresponding to non-trivial, contact-preserving symmetry generators, whose involutivity is governed by Theorem 4. At first glance, using an existing sample solution dataset as a *cross-section* (see Definition 54) may appear paradoxical, as the primary goal of studying the EDS associated with DE systems is to identify the integral submanifolds representing their solutions. However, in the context of machine learning, it is standard practice to gather a dataset capturing the output of a dynamical system, often expressed as the graph of a solution $\boldsymbol{f}$, $\Gamma_{\boldsymbol{f}}$. Accordingly, leveraging this sample dataset in conjunction with numerical finite-difference methods to compute partial derivatives, $\partial^{|I_k|} f^\alpha / \partial x^{I_k}$, aligns naturally with such methodologies.

This approach enables our framework to learn the complementary submanifold $\mathcal{N}^{(s)^\perp}$ —a foliation of the flow generated by the non-trivial symmetry generators—without

disturbing the contact system (see Definition 38) associated with the DE system under consideration.

In the concluding part of this section, we employ iResNETs [41] to implement the exponential map $\exp : T_eG \to G$, as described in (3.10) (see Figure 4.1). Additionally, we propose a technique to address the deviation of the group action from the solution manifold $\Sigma^{(s)}$ (see Definition 24) as consecutive group actions are iteratively applied, beginning from an initial point $z_0^{(s)} \in \Sigma^{(s)}$.

4.1 Learning the Infinitesimal Lie Point Symmetries

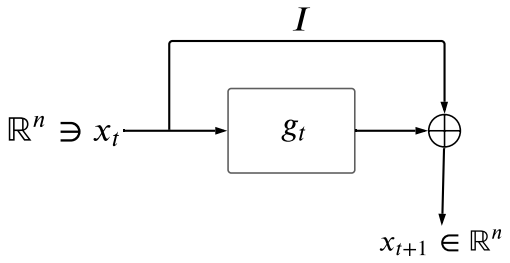

Figure 4.1: iResNET architecture with $\|g_t\| < 1$. Each iResNET block g_t is constructed using $\hat{v}_i$, representation of $v_i \in \mathfrak{g}$ in TM via (3.13), as a building block. See Section 4.4 for further details.

We begin by outlining the process of learning the finite-dimensional Lie algebra $\mathfrak{g}$. This sub-algorithm constitutes a key component of the iterative framework introduced in Section 4.3. Let $\mathfrak{g} = \{v_1, \ldots, v_{o_s}\}$ represent the vector fields spanning the Lie algebra of the symmetry group G associated with the DE system $\Delta^{(s_0)} = 0$, where s_0 is the original order of the system. The parameter $o_s = \dim(G)$ corresponds to the orbit dimension of $G^{(s)}$ acting on $J^{(s)}(M)$, with s denoting the stabilization order of the group orbit (see Definition 27). The parameters to be learned are denoted by $\mathbf{\Lambda} \coloneqq \{\lambda_1, \ldots, \lambda_{o_s}\}$, where each λ_i parametrizes the representation of $v_i \in \mathfrak{g}$. The corresponding approximation $\hat{v}_i \in TM$ is implemented using a fully connected neural network (FCN). In Sections 4 and 5, $\hat{v}_i$ and FCN_{λ_i} are used interchangeably to emphasize that the component functions of $\hat{v}_i$, as stated in (3.18), are determined by an FCN parameterized by the subset λ_i. Furthermore, the index t, denoted as

a subscript in g_t, specifies the number of times the corresponding group action is applied to transform an initial point $z_0^{(s)} \in J^{(s)}(M)$ into $z_t^{(s)} \in J^{(s)}(M)$. For further details, we refer the reader to Section 4.4.

To compute the infinitesimal group action of v_i on $J^{(s)}(M)$, we leverage the *back-propagation tools* provided by PyTorch [50]. The action of $\text{pr}^{(s)}(\hat{v}_i)$ on a scalar function $f \in C^\infty(J^{(s)}(M))$ is formulated as:

$$\text{pr}^{(s)}(\hat{v}_i)\,[f]\,|_{z^{(s)}} = \sum_j \text{FCN}_{\lambda_i}^{(s),j}|_{z^{(s)}} \underbrace{\frac{\partial}{\partial z^{(s),j}}(f)}_{(*)}\,[z(s)]\,, \tag{4.1}$$

where $(*)$ represents the derivative of f, with respect to jet space coordinates, computed using PyTorch's `grad` library, $|_{z^{(s)}}$ indicates evaluation at $z^{(s)}$, and $\text{pr}^{(s)}$ denotes the prolongation operator defined in Definition 16.

The remaining aspect of implementing (4.1) involves realizing the prolongation operator $\text{pr}^{(s)}$ as outlined in Theorem 3. To this end, we compute the characteristic of each $\hat{v}_i$ as given in (3.19) and implement the total derivative operator $D_{i_k} := {}^{d}/_{dx^{i_k}}$ in (3.17) using PyTorch's `grad` facility. While it is feasible to train the framework to learn the prolonged component functions $\hat{\varphi}^{\alpha}_{i,I_k}$ in (3.21), this approach would introduce additional parameters to optimize for each $\hat{v}_i$.

In this and the subsequent sections, let $\boldsymbol{u} = \boldsymbol{f}_{\text{tr}}(\boldsymbol{x})$ and $\boldsymbol{u} = \boldsymbol{f}_{\text{te}}(\boldsymbol{x})$ represent the training and testing solution datasets, respectively. The graphs of these functions, denoted as $\Gamma_{\boldsymbol{f}_{\text{tr}}}$ and $\Gamma_{\boldsymbol{f}_{\text{te}}}$, are further prolonged to $J^{(s)}(M) \to \mathcal{X}$ (see Remark 5) and represented as $\Gamma^{(s)}_{\boldsymbol{f}_{\text{tr}}}$ and $\Gamma^{(s)}_{\boldsymbol{f}_{\text{te}}}$, respectively.

To learn any FCN_{λ_i}, a set of conditions must be satisfied. First, each vector field $v_i \in \mathfrak{g}$ is required to satisfy the infinitesimal symmetry condition given in (3.35). Using (4.1), the first component of the final Lagrangian, $\mathcal{L}_{\text{Lie}}$, is defined as:

$$\mathcal{J}_{\text{sym}} = \frac{1}{2}\left\| \left(\text{pr}^{(s)}(\hat{v}_i)\,\left[\Delta^{(s)}\right]\right)_{|_{z^{(s)}}} \right\|^2, \tag{4.2}$$

where $\|\cdot\|$ denotes the ℓ_2 norm. Special care must be taken when implementing (4.2), as residual terms may arise after applying the infinitesimal action of $\hat{v}_i$, as defined in (3.35). These residual terms can involve $\Delta^{(s)}$ and $\left\{{}^{\partial\Delta^{(s),l_s}}/_{\partial z^{(s),j_s}}\right\}_{j_s}^{l_s}$ as their least common multiple (LCM), where $l_s = 1,\ldots,n_s$, $j_s = 1,\ldots,m_s = \dim(J^{(s)}(M))$, and n_s denotes the number of DEs in $\Delta^{(s_0)}$ prolonged to $J^{(s)}(M)$ (see Remark 35).

Such terms may lead to superfluous non-zero expressions, even though they are mathematically equal to zero. For instance, consider Example 2.41 in [8], where applying $\mathrm{pr}^{(2)}(v_3)$ to the linear *heat equation* yields $u_t - u_{xx} \neq 0$, despite being zero theoretically. To formalize this issue, let

$$\mathcal{R}_{\Delta^{(s)}} = \begin{cases} \mathrm{rest}^{(k),l_k} & = \Delta^{(k),l_k} = 0, \\ \mathrm{rest}^{(k),l_k}_{j_k} & = \frac{\partial \Delta^{(k),l_k}}{\partial z^{(k),j_k}} = 0 \end{cases} \tag{4.3}$$

represent the set of such restrictions on $\Sigma^{(k)}$, where $k = s_0, \ldots, s$, $l_k = 1, \ldots, n_k$, and $j_k = 1, \ldots, m_k$. Addressing $\mathcal{R}_{\Delta^{(s)}}$ typically requires symbolic computation tools such as SymPy [51] or REDUCE [46]. However, in our framework, these concerns are mitigated by the assumption that the sample datasets $\Gamma_{\boldsymbol{f}_{\mathrm{tr}}}$ and $\Gamma_{\boldsymbol{f}_{\mathrm{te}}}$ are the graphs of solutions to $\Delta^{(s_0)}$. This ensures that such residual terms are already evaluated to zero, thanks to the *computational graph* generated by PyTorch.

Remark 35 (Prolongation of $\Delta^{(s_0)}$)**.** *It is worth noting that* (4.2) *demonstrates that the infinitesimal symmetry condition can be imposed on the prolonged DE system* $\Delta^{(s)} = 0$*, where* $s \geq s_0$*, with* s_0 *denoting the order of the original system. As specified in Algorithm 2, any* $\hat{v}_i \in \mathcal{V}$ *must satisfy the resulting constraints* $\forall\, k = s_0, s_0 + 1, \ldots, s$*, which arise from the prolongation process. Therefore, the notation* $\Delta^{(s)}$ *encompasses DEs of all orders* $(s_0, \ldots, s)$*. This prolongation procedure is further detailed in Section 4.3.*

As we are focusing on learning Lie point transformations, which are a subset of contact transformations (see Section 3.3), the contact condition given in (3.33) (or equivalently in (3.34)) must also be satisfied:

$$\mathcal{J}_{\mathrm{cnt}} = \frac{1}{2} \left\| \left(\frac{\partial Q_i^\alpha}{\partial u_j^\beta} + \hat{\xi}_i^j \delta_\beta^\alpha \right)_{|_{z^{(1)}}} \right\|^2 . \tag{4.4}$$

Here, we prefer the formulation in (3.34) over (3.33), as the characteristic of the prolongation of $\hat{v}_i$ is already computed. The contact condition along with the prolongation formula in Theorem 3 can be interpreted as preserving the consistency of a solution $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$ with its partial derivatives $\partial^k \boldsymbol{f}^\alpha / \partial x^{I_k}$ under the action of $G^{(s)}$, where $\alpha = 1, \ldots, q$, $k = 1, \ldots, s$, and $|I_k| = k$ (see Definition 14). As noted in the discussion following Proposition 1, for Lie point transformations, the inclusion

of (4.4) in the ultimate cost function is not necessary. However, this cost function becomes essential when extending the framework to encompass Lie contact transformations, which are typically infinite-dimensional and necessitate a distinct implementation scheme.

The set $\text{pr}^{(s)}(\{\hat{v}_1, \ldots, \hat{v}_{o_s}\}) \subset TJ^{(s)}(M)$ must be linearly independent at any $z^{(s)} \in J^{(s)}(M)$. Given that Algorithm 3 learns each vector field v_i, which spans $\mathfrak{g}$, in a single optimization phase indexed by $i \in \{1, \ldots, o_s\}$, and under the assumption of a locally effective group action (see Definition12), we enforce the inclusion hierarchy:

$$\mathcal{V}_1^{(s)}|_{z^{(s)}} \subset \ldots \subset \mathcal{V}_i^{(s)}|_{z^{(s)}} \subset \ldots \subset \mathcal{V}_{o_s}^{(s)}|_{z^{(s)}} \subset T_{z^{(s)}} J^{(s)}(M) \tag{4.5}$$

for all $z^{(s)} \in J^{(s)}(M)$. Here, each $\mathcal{V}_i^{(s)} \coloneqq \text{pr}^{(s)}(\mathcal{V}_i) = \text{pr}^{(s)}(\{\hat{v}_1, \ldots, \hat{v}_i\})$ is a VFS of dimension i. For $i = 1$, linear independence is trivially satisfied. However, for $i \geq 2$, two cases arise: ***(i)*** $i = 2$ and ***(ii)*** $i \geq 2$. Define $\kappa_{12} = \angle(\hat{v}_1, \hat{v}_2)|_{z^{(s)}}$ as the angle between $\text{pr}^{(s)}(\hat{v}_1)$ and $\text{pr}^{(s)}(\hat{v}_2)$ at $z^{(s)}$. For $i = 2$, the objective is to minimize $|\cos(\kappa_{12})|^2$. For $i > 2$, we minimize the square of the cosine of the angle $\kappa_{(i-1)i} = \angle(\hat{v}_{i\|}, \hat{v}_i)|_{z^{(s)}}$, where $\hat{v}_{i\|}$ is the projection of $\text{pr}^{(s)}(\hat{v}_i)$ onto the subspace spanned by $\mathcal{V}_{i-1}^{(s)}$. To formalize this, let A_i represent the matrix:

$$A_i = \begin{bmatrix} | & & | \\ \text{pr}^{(s)}(\hat{v}_1) & \ldots & \text{pr}^{(s)}(\hat{v}_{i-1}) \\ | & & | \end{bmatrix}_{|_{z^{(s)}}}, \tag{4.6}$$

where $i = 3, \ldots, o_s$. The corresponding projection matrix $\mathbb{P}_i$ is given by $\mathbb{P}_i \coloneqq A_i \left(A_i^{\text{t}} A_i\right)^{-1} A_i^{\text{t}}$, where t denotes the matrix transposition operator [52]. Consequently, $\hat{v}_{i\|}$ becomes:

$$\hat{v}_{i\|}|_{z^{(s)}} = \mathbb{P}_i \cdot \text{pr}^{(s)}(\hat{v}_i)|_{z^{(s)}}. \tag{4.7}$$

It is important to note that the linear independence of $\mathcal{V}_i^{(s)}$ does not necessarily ensure that $|\cos(\kappa_{(i-1)i})|^2 \to 0$ as epoch $\to \infty$ in Algorithm 3. Hence, this condition must also be introduced as a constraint in $\mathcal{L}_{\text{Lie}}$:

$$\mathscr{C}_{\text{ind}} = |\cos(\kappa_{(i-1)i})|^2 - \text{tol}_{\text{ind}}, \tag{4.8}$$

where $\text{tol}_{\text{ind}} \in \mathbb{R}^+$ is a tolerance value. Alternatively, a simpler theoretical approach exists, albeit computationally more expensive than the previously described method.

Specifically, let the matrix A_i from (4.6) be extended by appending the column vector $\mathrm{pr}^{(s)}(\hat{v}_i)$, forming a new matrix denoted by $\tilde{A}_i$. Ensuring that $\tilde{A}_i$ is of maximal rank can be achieved by verifying the existence of at least one non-vanishing $i \times i$ minor of $\tilde{A}_i$. However, this method is not employed due to the high computational complexity associated with calculating such minors, which is given by:

$$O\left(\binom{m}{i} i^3\right), \tag{4.9}$$

where $m = \dim(J^{(s)}(M))$.

As discussed in Section 5.3, the ultimate cost function includes a crucial term to eliminate solution-dependent but non-trivial symmetry generators. The primary motivation for excluding such symmetries is to prevent the emergence of infinite-dimensional Lie algebras acting on $J^{(s)}(M)$. Furthermore, these solution-dependent symmetries are directly influenced by initial and/or boundary conditions (IBCs), which is particularly undesirable when focusing on *intrinsic* symmetry generators. For instance, in the case of linear DE systems, solution-dependent symmetry generators arise due to the superposition property of solutions and take the form:

$$\hat{v} = \sum_{\alpha=1}^{q} \hat{\varphi}^{\alpha}(\boldsymbol{x}) \frac{\partial}{\partial u^{\alpha}}, \tag{4.10}$$

where $\hat{\varphi}^{\alpha}$ depends only on the independent variables $\boldsymbol{x}$ and not on the dependent variables $\boldsymbol{u}$. However, for nonlinear PDEs, this assumption does not hold. Regardless of the linearity of $\Delta^{(s_0)} = 0$ or the dependence of $\hat{\varphi}^{\alpha}$ on $\boldsymbol{u}$, any solution-dependent symmetry generator must satisfy the following condition:

$$\Delta^{(s_0)}(\boldsymbol{x}, \hat{\varphi}^{\alpha}_J) = 0, \tag{4.11}$$

where $\alpha = 1, \ldots, q$, $0 \leq |J| \leq s_0$, and $\hat{\varphi}^{\alpha}_J$ denotes the prolongation of $\hat{\varphi}^{\alpha}$, as defined in Definition 16. Readers should note the distinction between $\hat{\varphi}^{\alpha}_J(\boldsymbol{x}, \boldsymbol{u})$ and $\hat{\varphi}^{\alpha}_I(z^{(|I|)})$ in (3.21). Specifically:

$$\hat{\varphi}^{\alpha}_J(\boldsymbol{x}, \boldsymbol{u}) = \frac{\partial^{\#J} \hat{\varphi}^{\alpha}(\boldsymbol{x}, \boldsymbol{u})}{\partial x^J} \tag{4.12}$$

where $\partial^{\#J}/\partial x^J$ is the partial derivative operator, of order $|J|$, with respect to x^J.

To exclude solution-dependent symmetry generators, the cost function must include a penalization term, defined as follows:

$$\mathscr{C}_{\text{dep}} = \frac{1}{\varepsilon_{\text{dep}} + \|\Delta^{(s)}(\boldsymbol{x}, \hat{\varphi}^{\alpha}_J)\|^2} - \text{tol}_{\text{dep}}, \tag{4.13}$$

where $\varepsilon_{\text{dep}} \in \mathbb{R}^+$ and $\text{tol}_{\text{dep}} \in \mathbb{R}^+$ are regularization and tolerance values, respectively, and $\|\cdot\|$ denotes the ℓ_2 norm. Here, $\alpha = 1, \ldots, q$, and $0 \leq |J| \leq s$ (see Remark 35).

Since the objective is to discover only intrinsic symmetries, any nonlinear dependence of $\hat{\xi}^i$ or $\hat{\varphi}^\alpha$, as defined in (3.18), on u^α leads to non-uniform transformations across different solutions of the given DE system. The issue is particularly evident for $\hat{\varphi}^\alpha$, as it explicitly determines the transformation of the dependent variables u^α in $\Sigma^{(s)}$. Although the effect of nonlinearity in $\hat{\xi}^i$ may appear more subtle, there are two fundamental reasons why intrinsic symmetry generators must also remain at most linearly dependent on u^α: *(i)* Non-uniform transformation of independent coordinates. If any $\hat{\xi}^i$ depends nonlinearly on u^α, then the transformation of the independent variables x^i will be solution-dependent. This results in a **non-uniform** deformation of the solution space $\Sigma^{(s)}$, much like the distortion caused by nonlinearity in $\hat{\varphi}^\alpha$. *(ii)* Potential violation of the Lie algebra closure property. Nonlinear dependence on u^α can obstruct the closure of the Lie algebra under the Lie bracket operation defined in (3.5). This closure is essential for integrating the action of the symmetry generators to construct a well-defined group transformation on $\Sigma^{(s)}$.

To illustrate these issues, consider the following first-order DE along with a nonlinear generator:

$$\begin{aligned} \Delta^{(1)}(x, u, u_x) &= \frac{du}{dx} - u = 0, \\ \hat{v} &= u^2 \frac{\partial}{\partial x}. \end{aligned} \tag{4.14}$$

The general solution to $\Delta^{(1)}(x, u, u_x)$ is given by

$$u(x) = Ce^x, \tag{4.15}$$

where $C \in \mathbb{R}$ is an arbitrary integration constant. Consider two particular solutions, $u_1(x) = e^x$ and $u_2(x) = 2e^x$. The integral curves of $\hat{v}$ in (4.14) yield the following transformations of x for these solutions, respectively:

$$\begin{aligned} x_1(t) &= -\frac{1}{2}\ln(C - 2t), \\ x_2(t) &= -\frac{1}{2}\ln(C - 8t). \end{aligned} \tag{4.16}$$

This result demonstrates that the rate of change in x differs across solutions, meaning the transformation induced by $\hat{v}$ is inherently solution-dependent. Consequently, such

a symmetry generator disrupts the uniform, global action of an intrinsic Lie symmetry group G. While solution-dependent symmetries are sometimes studied in applied settings, they are beyond the scope of this work.

To prevent the learned symmetry generators $\hat{v}_i$ from exhibiting nonlinear dependence on $\boldsymbol{u}$, we introduce the following regularization term:

$$\mathcal{J}_{\text{nlu}} = \sum_i \sum_\alpha \left\| \frac{\partial^2 \hat{\xi}^i}{\partial u^{j_1} \partial u^{j_2}} \right\|^2 + \left\| \frac{\partial^2 \hat{\varphi}^\alpha}{\partial u^{j_1} \partial u^{j_2}} \right\|^2, \tag{4.17}$$

where $\|\cdot\|$ denotes the *Frobenius* norm of each matrix indexed by i and α and $j_1, j_2 = 1, \ldots, q$. This term penalizes second-order derivatives with respect to u^α, ensuring that both $\hat{\xi}^i$ and $\hat{\varphi}^\alpha$ remain at most linearly dependent on u^α.

Incorporating (4.2), (4.8), (4.13), and (4.17) the final Lie Lagrangian $\mathcal{L}_{\text{Lie}}$ takes the form:

$$\mathcal{L}_{\text{Lie}} = \mathcal{J}_{\text{sym}} + \mathcal{J}_{nlu} + \mu_{\text{ind}} \mathscr{C}_{\text{ind}} + \mu_{\text{dep}} \mathscr{C}_{\text{dep}}, \tag{4.18}$$

where $\boldsymbol{\mu} \coloneqq [\mu_{\text{ind}}, \mu_{\text{dep}}]^{\text{t}}$ represents the dual optimal Karush-Kuhn-Tucker (KKT) multipliers (cf. λ_i^* variables in (5.49) of [53]). However, reader should keep in mind that (4.4) is included when transformation groups larger than Lie point transformations are considered, though such infinite dimensional groups require a modified approach as discussed in [48]. As noted earlier, the log-barrier extension method in [54] converts this constrained optimization problem into an unconstrained one.

Remark 36 (Zero Vector Field)**.** *The cost functions $\mathcal{J}_{sym}$, $\mathcal{J}_{nlu}$, and $\mathcal{J}_{cnt}$ may be trivially satisfied if any $\hat{v}_i$ is a zero vector field. However, normalizing any $\hat{v}_i$ to unity, $\hat{v}_i / \|\hat{v}_i\|$, prevents this triviality.*

The costs outlined thus far are consolidated in Algorithms 2 and 3.

Remark 37 (Jet Space Order s_c in Algorithms)**.** *It is important to note that the jet space order s_c, referred to as the current order of the jet space in the algorithms, does not necessarily coincide with s_0. Prolongation of $J^{(s_c)}(M)$ may be required until the stabilization order s is attained, when an involutive, non-trivial distribution is considered in the next section. Further elaboration is provided in Section 4.3.*

4.2 Learning Involutive Non-Trivial Distribution

This section builds upon the discussion in Section 3.5. As highlighted in the concluding paragraph of that section, unlike trivial symmetries, non-trivial symmetries actively influence the dynamics of DE systems and may compromise the integrity of the contact system characterizing the integral submanifold under consideration. Consequently, ensuring the involutivity of $\mathcal{W}_{\boldsymbol{f}}$, which corresponds to the representation of these non-trivial, contact-preserving symmetry generators, is essential in constructing a G-coframe. However, it is important to note that the symmetry generators learned in Section 4.1 are not necessarily non-trivial, as defined in Definition 26, when restricted to a particular solution $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$. Consequently, the non-triviality of these generators must be assessed on a per-solution basis. Since such symmetries $\hat{w}_j$ are already known to belong to $\mathcal{V}$, it follows that

$$\hat{w}_j|_z = \sum_i a_j^i(z)\hat{v}_i|_z \in \mathcal{W}_{\boldsymbol{f}}, \tag{4.19}$$

where $a_j^i \in C^\infty(M)$ are coefficient functions.

Observe that the distribution $\mathcal{V}$ obtained in Section 4.1 consists solely of intrinsic symmetry generators; that is, no vector field $\hat{v}_i \in \mathcal{V}$ depends on a particular solution, nor does any exhibit nonlinear dependence on the dependent variables $\boldsymbol{u}$. Nevertheless, several important questions arise—even when considering a single example from among Examples 2.41 through 2.45 in [8]. For instance, both the heat equation in Example 2.41 and the Burgers' equation in Example 2.42 possess solution-dependent symmetry generators, commonly denoted v_α. As indicated in these examples, such generators form infinite-dimensional Abelian subalgebras, whereas the intrinsic symmetry generators constitute a finite-dimensional Lie subalgebra $\mathfrak{g}$. Moreover, there exist differential equation systems, such as the Korteweg–de Vries (KdV) equation presented in Example 2.44 of [8], that do not admit any solution-dependent symmetry generators. In such cases, the consideration of solution-dependent symmetries becomes unnecessary.

In keeping with our notation, let $\hat{v}_\alpha$ denote any solution-dependent symmetry generator. This prompts the natural question: why does $\hat{v}_\alpha$ not appear in (4.19)? The first reason is that including this Abelian subalgebra would necessarily lead to an infinite-

dimensional representation for any $\hat{w}_j \in \mathcal{W}_{\boldsymbol{f}}$. Furthermore, generators of the form $\hat{v}_\alpha$ do not admit a global representation on the full jet space $J^{(s)}(M)$. To see this, consider the solution space

$$\mathcal{S}^{(s)} \coloneqq \left\{ \Gamma^{(s)}_{\boldsymbol{f}} \mid \Delta^{(s)}(\boldsymbol{x}, \boldsymbol{f}^{(s)}) = 0 \right\}, \tag{4.20}$$

which consists of all prolonged graphs corresponding to solutions of the DE system $\Delta^{(s)}$. Then, $\mathrm{pr}^{(s)}(\hat{v}_\alpha) \in T\mathcal{S}^{(s)}$. However, since $\mathcal{S}^{(s)} \subsetneq \Sigma^{(s)} \subset J^{(s)}(M)$, unless the system is trivial, it follows that $\hat{v}_\alpha$ is not defined globally on $J^{(s)}(M)$. Consequently, any $\hat{v}_\alpha$ fails to generate directions transverse to the orbits of the global symmetry group $G \subset \mathrm{Diff}(J^{(s)}(M))$, which is generated by the Lie algebra $\mathfrak{g} \cong \mathcal{V}^{(s)} \subset \mathfrak{X}(J^{(s)}(M))$.

Fortunately, identifying $\mathcal{W}_{\boldsymbol{f}}$ is computationally efficient due to the frozen parameter set $\boldsymbol{\Lambda}$ of $\mathcal{V}$, allowing only the coefficients $a^i_j(z)$ to be learned. In contrast, modifying $\boldsymbol{\Lambda}$ directly to extract the non-trivial, involutive subset of $\mathcal{V}$ for a given solution dataset would be computationally expensive, especially when a large number of symmetry generators have been discovered.

Remark 38 ($\mathcal{V}$ may not be a linearly independent set)**.** *We consistently denote any linearly independent set as $\{v_1, \ldots, v_n\}$. However, unless prolonged to a specific order s, a VFS $\mathcal{V} = \{\hat{v}_1, \ldots, \hat{v}_{o_s}\}$ may not constitute a linearly independent set. Hence, this non-prolonged $\mathcal{V}$ is treated as an overdetermined VFS.*

As discussed in Sections 1.2 and 3.4, our aim is to learn non-trivial symmetries (see Definition 26). To ensure this, the components of any vector field $\hat{v}_i$ must satisfy the following *non-zero* constraint:

$$Q^\alpha_i(z^{(1)})|_{\Gamma^{(1)}_{\boldsymbol{f}_{\mathrm{tr}}}} \neq 0, \ \forall\, \alpha = 1, \ldots, q. \tag{4.21}$$

Since (4.21) is expressed as a "$\neq 0$" condition, standard optimization methods, such as those used in (4.2) or (4.4), are not applicable. Instead, this constraint is incorporated into $\mathcal{L}_{\mathrm{inv}}$ in (4.27) as an additional term, denoted by $\mathscr{C}_{\mathrm{ntr}}$, and optimized using *log-barrier* method in [54]. Let $\mathrm{tol}_{\mathrm{ntr}} \in \mathbb{R}^+$ be a small tolerance value. The constraint is then expressed as:

$$\mathscr{C}_{\mathrm{ntr}} = -1 \left(\gamma(\boldsymbol{Q}_i(z^{(1)})) - \mathrm{tol}_{\mathrm{ntr}} \right), \tag{4.22}$$

where $\boldsymbol{Q}_i = [Q_i^1, \ldots, Q_i^q]^{\mathrm{t}}$ in (3.19), $z^{(1)} \in \Gamma^{(1)}_{\boldsymbol{f}_{\mathrm{tr}}}$, and γ is the "softmin" function:

$$\gamma(\boldsymbol{Q}) = \ln\left(\sum_j e^{Q^j}\right). \tag{4.23}$$

Observe that, (4.22) ensures that $\hat{v}_i$ has non-trivial effect on each dependent variable u^α.

According to Theorem 4, a VFS is involutive if and only if $\hat{v}_{jk} = [\hat{v}_j, \hat{v}_k] \in \mathcal{V} \coloneqq \mathcal{V}_{o_s}$, where $[\cdot,\cdot]$ denotes the Lie bracket as defined in (3.5), and $1 \leq j < k \leq o_s$. As discussed in the last paragraph of Section 3.1, transitivity of a group action, when restricted to a single orbit $\mathcal{O}_{z^{(s)}}$ of $G^{(s)}$, is crucial for involutivity as non-transitive actions might result in singularities or disconnected components in $\mathcal{O}_{z^{(s)}}$, potentially leading to a failure of the bracket closure property of $\mathcal{V}$. Under this assumption, the sub-algorithm for ensuring involutivity of $\mathcal{W}_{\boldsymbol{f}}$ begins with the following proposition:

Proposition 3. *Let $\mathcal{W}^{(s)}_{\boldsymbol{f}} \coloneqq pr^{(s)}(\mathcal{W}) = pr^{(s)}(\{\hat{w}_1, \ldots, \hat{w}_{o_s}\})$ be the VFS for a given DE system $\Delta^{(s_0)} = 0$, and the solution dataset $\Gamma^{(s)}_{\boldsymbol{f}}$, where each $\hat{w}_j$ satisfies* (3.34)*,* (3.35)*, and* (4.21) *for all $s_c = s_0, s_0+1, \ldots, s$. Then:*

$$pr^{(s)}(\hat{w}_{jk}) \coloneqq pr^{(s)}([\hat{w}_j, \hat{w}_k]) \in \mathcal{W}_{\boldsymbol{f}} \tag{4.24}$$

is sufficient for the vector field $pr^{(s)}(\hat{w}_{jk})$, $1 \leq j < k \leq o_s$, to satisfy the same conditions as any $pr^{(s)}(\hat{w}_j)$ spanning $\mathcal{W}^{(s)}_{\boldsymbol{f}}$ (see also (3.45)*).*

The proof of Proposition 3 is straightforward, when (3.46) is considered. This proposition implies that minimizing the Euclidean distance between $\mathrm{pr}^{(s)}(\hat{w}_{jk})$ and its projection onto $\mathcal{W}^{(s)}_{\boldsymbol{f}}$ is sufficient to satisfy the involutivity condition stated in Theorem 4:

$$\mathcal{J}_{\text{inv}} = \sum_{1\leq j<k\leq o_s} \|\mathrm{pr}^{(s)}(\hat{w}_{jk})|_{z^{(s)}} - \bar{\mathbb{P}} \cdot \mathrm{pr}^{(s)}(\hat{w}_{jk})|_{z^{(s)}}\|^2, \tag{4.25}$$

where $\bar{\mathbb{P}} \coloneqq \bar{A}\left(\bar{A}^{\mathrm{t}}\bar{A}\right)^{-1}\bar{A}^{\mathrm{t}}$ is the projection matrix, and

$$\bar{A} = \begin{bmatrix} | & & | \\ \mathrm{pr}^{(s)}(\hat{w}_1) & \ldots & \mathrm{pr}^{(s)}(\hat{w}_{o_s}) \\ | & & | \end{bmatrix}_{|_{z^{(s)}}}. \tag{4.26}$$

Since $\mathcal{W}_{\boldsymbol{f}}$ is required to form a linearly independent set, the constraint $\mathscr{C}_{\text{ind}}$ must be enforced for all pairs $(\hat{w}_j, \mathcal{W}_{\boldsymbol{f}} \setminus \hat{w}_j)$ simultaneously, where $j = 1, \ldots, \dim(\mathcal{W}_{\boldsymbol{f}})$ (see

Algorithm 4 for details). Incorporating the constraints (4.8), (4.25), and (4.22), the final Lagrangian including the involutivity condition is given by:

$$\mathcal{L}_{\text{inv}} = \mathcal{J}_{\text{inv}} + \mu_{\text{ind}}\mathscr{C}_{\text{ind}} + \mu_{\text{ntr}}\mathscr{C}_{\text{ntr}}, \tag{4.27}$$

where μ_{ind} and μ_{ntr} are the corresponding KKT multipliers associated with the linear independence and non-triviality constraints, respectively.

The concise yet profound algorithm is summarized in Algorithms 4 and 5. However, note that Algorithm 5 is invoked in Algorithm 6 only after reaching a specific order s_c, as detailed in Section 4.3.

One might propose, "Why not enforce the involutivity conditions during every i^{th} training phase alongside the optimization of $\mathcal{L}_{\text{Lie}}$?" While this observation is valid, enforcing involutivity at all training steps—except for the final step, when $i = o_s$—is redundant. Moreover, while the vector field system $\mathcal{V}$ learned in Section 4.1 is valid for any solution in $\Sigma^{(s)}$, constructing a non-trivial involutive distribution, as discussed at the beginning of this section, is necessary only for a particular solution $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$.

Another potential approach involves constructing a Lie algebra chain as follows:

$$0 \subset \{w_1\} = \mathfrak{g}_1 \subset \{w_1, w_2\} = \mathfrak{g}_2 \subset \ldots \subset \{w_1, \ldots, w_{o_s}\} = \mathfrak{g}, \tag{4.28}$$

where $[\mathfrak{g}_{i-1}, \mathfrak{g}_i] \subset \mathfrak{g}_i$. In this context, $\mathfrak{g}$ is referred to as a solvable Lie algebra (see Definition 2.63 in [8]). Equation (4.28) implies that each $\mathfrak{g}_{i-1}$ is not only a vector subspace but also an algebraic ideal in $\mathfrak{g}_i$. (Note: This is distinct from the algebraic ideal defined in Definition 32.) This chain creates a hierarchy of normal subgroups such that $\{I\} \subset G_1 \subset G_2 \subset \ldots \subset G_{o_s} = G$. However, as discussed in the Solvable Groups section of Chapter 2 in [8], this approach overly restricts the discovery of new symmetries due to the potential lack of sufficient number of normal subgroups. While such a chain is valuable for simplifying DE systems via normal subgroups G_i, without interference from higher-dimensional normal subgroups G_j (where $j > i$), it is not relevant to the context of this study or the construction of the G-coframe in future research.

Remark 39 (Arbitrary functions and Frobenius' Theorem)**.** *Learning $a^i_j(z)$'s in* (4.19) *via Algorithm 5 is equivalent to computing arbitrary (free) functions that satisfy in-*

tegrability conditions. These functions are derived as a consequence of the Cartan–Kähler theorem, a generalization of Frobenius' theorem (see, for example, the $\partial u^\alpha/\partial y$ *functions in (15.24) of [10]). However, in the case of Frobenius' theorem, such free functions are computed* ***indirectly****.*

4.3 Prolongation & Iterative Learning Algorithm

In this final subsection, we delve into the process of *prolongation*, which plays a key role in learning the involutive, finite-dimensional Lie algebra $\mathfrak{g}$. Note that this use of the term "prolongation" should not be confused with the prolongation of analytic objects, as defined in Definition 16 and Theorem 3. The culmination of this discussion is encapsulated in Algorithm 6, which constructs the involutive chain for $\mathcal{W}_{\boldsymbol{f}}$ similar to the one presented in (4.5).

Although Cartan's equivalence problem [15] is not explicitly addressed here, the problem at hand is inherently an equivalence problem. Any DE system can be expressed via a contact system that is specific to that DE system (see Section 3.5). Consequently, any contact form $\theta \in \mathcal{I}_{\text{diff}}$ must satisfy the following compatibility condition:

$$\phi^*(\bar{\theta}) = \theta, \tag{4.29}$$

where $\bar{\theta} \in \mathcal{I}_{\text{diff}}$ and ϕ is the transformation induced by the non-trivial, contact-preserving symmetry generators of $\mathfrak{g}$ (see the equivalence problem in Section 3.6). Analogous to Cartan's framework (see Chapter 12 in [10]), prolongation of the subspace underlying the DE system may become necessary to ensure the involutivity of $\mathcal{W}_{\boldsymbol{f}}^{(s)}$ in (4.5). In the context of DE systems, prolongation is intrinsically tied to the jet space framework, as DE systems are naturally defined in jet spaces. New constraints often emerge due to the relationships between higher-order derivatives of the solutions. For this reason, the term "prolongation" in this study refers specifically to the prolongation of the underlying jet space:

$$J^{(s_c)}(M) = \text{pr}^{(1)}(J^{(s_c-1)}(M)). \tag{4.30}$$

The need for prolongation arises from two key considerations. First, the symmetry group G must reach its stabilization order s. At this order, a *contact-invariant coframe*

of rank p exists (see Definition 9.11 and Theorem 9.13 in [13]). While this condition can be relaxed by forgoing *full-symmetry* reduction (which is computationally expensive), approximating the prolongation using finite differences [55] for:

$$\Gamma_{\boldsymbol{f}_{\text{tr}}}^{(s_c)} = \text{pr}^{(1)}(\Gamma_{\boldsymbol{f}_{\text{tr}}}^{(s_c-1)}) \tag{4.31}$$

may lead to inaccuracies in higher-order partial derivatives of the sample dataset (see also Remark 32). To construct a contact-invariant coframe, the following condition must be satisfied:

$$o_{s_c} \leq \dim(\Sigma^{(s_c)}) - p, \tag{4.32}$$

where o_{s_c} is the orbit dimension of $G^{(s_c)}$ in $J^{(s_c)}(M)$, and $\Sigma^{(s_c)} \subset J^{(s_c)}(M)$ is the regular submanifold representing the prolonged DE system $\Delta^{(s_c)} = 0$ as discussed below. This inequality ensures that the symmetry group action $G^{(s)}$ is compatible with the available degrees of freedom in the integral submanifold associated with $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$ at jet order s_c. However, this condition alone is insufficient to guarantee the involutivity of $\mathcal{W}_{\boldsymbol{f}}^{(s_c)}$ for $s_c < s$. Hence, further prolongation might be required to incorporate additional integrability conditions for involutivity.

In addition to prolonging the input dataset $\Gamma_{\boldsymbol{f}_{\text{tr}}}$ in (4.31), each vector field $\hat{w}_j \in \mathcal{W}_{\boldsymbol{f}}$ must also be prolonged. However, directly applying $\text{pr}^{(s_c)}(\hat{w}_j)$ to $\Delta^{(s_0)}$ is theoretically incorrect, as it neglects the higher-order relations on which $\text{pr}^{(s_c)}(\hat{v}_i)$ acts. Thus, the DE system $\Delta^{(s_0)}$ must also undergo prolongation. This is achieved through the total derivative operator D_j defined in (3.17) and is expressed by the following recursive formula (see Equation (2.54) in [23]):

$$\text{pr}^{(1)}(\Delta^{(k-1),l_{k-1}}) = \begin{cases} & \Delta^{(k-1),l_{k-1}} \\ \Delta_1^{(k),l_{k-1}} & \coloneqq D_1(\Delta^{(k-1),l_{k-1}}) \\ & \vdots \\ \Delta_p^{(k),l_{k-1}} & \coloneqq D_p(\Delta^{(k-1),l_{k-1}}) \end{cases}, \tag{4.33}$$

where $l_{k-1} = 1, \ldots, n_{k-1}$, $k = s_0 + 1, \ldots, s_c$, and n_{k-1} denotes the number of all constraints up to $(k-1)^{\text{th}}$ order of prolongation, with n_{s_0} being the number of DEs in $\Delta^{(s_0)}$. For convenience, we relabel $\Delta_j^{(k),l_{k-1}}$ as $\Delta^{(k),l_k}$, where $j = 1, \ldots, p$ and $l_k = 1, \ldots, n_k$. Consequently, the prolongation defined in (4.33) introduces new

constraints, which must be satisfied via the following conditions:

$$\mathrm{pr}^{(k)}(\hat{w}_j)\left[\Delta^{(k)}\right]_{|_{\Delta^{(k)}=0}} = 0, \tag{4.34}$$

for $j = 1, \ldots, o_{s_c}$.

Remark 40 (Polynomial DE Systems and Prolongation)**.** *In this study, we assume that all DE systems are expressed as polynomials in jet space coordinates. Consequently, employing a computational algebra tool rather than PyTorch's computational graph to implement* (4.33) *offers an advantage in terms of computational efficiency. Although the gain is not substantial for smaller systems, it can become valuable when dealing with large-scale DE systems. Moreover, since SymPy provides functionality for bi-directional conversion of expressions,* (4.33) *can be implemented without disrupting PyTorch's computational graph. This interoperability enables seamless integration of symbolic computation with machine learning frameworks, enhancing flexibility and efficiency in handling complex DE systems.*

Not all such restrictions, however, contribute to the geometric structure of $\Sigma^{(k)}$, though they remain part of the differential ideal generated by the prolonged DE system via (4.33). This leads to the following definition:

Definition 59. *Let $\Delta^{(k),l_k} \in \Delta^{(k)} \setminus \Delta^{(k-1)}$, as given by* (4.33)*, for a particular pair (k, l_k), where $k \in \{s_0 + 1, \ldots, s_c\}$ and $l_k \in \{n_{k-1} + 1, \ldots, n_k\}$. If $\Delta^{(k),l_k}$ satisfies:*

$$0 \equiv \Delta^{(k),l_k} mod \left(\Delta^{(k-1)}\right), \tag{4.35}$$

then $\Delta^{(k),l_k}$ is trivially satisfied and is termed a redundant differential consequence, where $\Delta^{(k-1)}$ forms a differential ideal (see Remark 41).

Remark 41 ($\mathcal{I}_{\text{diff}}$ in Differential Geometry vs the One in Differential Algebra)**.** *The term differential ideal, as used in Definition 59, slightly differs from $\mathcal{I}_{diff}$ in Definition 33. In differential geometry, $\mathcal{I}_{diff}$ characterizes the geometric properties of DE systems via differential forms and exterior derivative (see Definitions 18 and 19). Conversely, in differential algebra, a differential ideal models DE systems using polynomials in jet space coordinates and their derivatives (see Definition B.1.24 in [23]). Despite these differences, both notions encode the same principle of closure under differentiation and are often interconvertible.*

According to Definition 59, redundant differential consequences should not be included in the total count of non-trivial constraints c_k for all $k = s_0 + 1, \ldots, s_c$. In computational algebra, tools such as *Gröbner* or *Pommaret* bases are employed to identify and eliminate these trivial differential consequences, as discussed extensively in [23]. These algebraic techniques determine the involutivity conditions and the prolongation step s_c at which involutivity is achieved.

Fortunately, in our case, such computational tools are unnecessary to evaluate the condition in (4.32), as our objective is not to identify the maximal integral submanifold (see Definition 37) defined by the involutive subsystem of (4.33). Instead, our focus lies on the submanifold $\mathcal{N}^{(s_c)^\perp} \subset \Sigma^{(s_c)}$, which is transverse to that maximal integral submanifold $\boldsymbol{S}_{\boldsymbol{f}}^{(s_c)}$ and is foliated by the non-trivial, contact-preserving vector fields $\mathcal{W}_{\boldsymbol{f}}^{(s_c)}$ (see Section 3.5). Assuming regularity of $\Sigma^{(k)}$ for all $k = s_0, \ldots, s_c$, we avoid computing $\dim(\Sigma^{(k)})$ at every point $z^{(k)} \in \Sigma^{(k)}$, as it retains constant rank. Thanks to the following proposition, the redundant differential consequences do not have to be detected:

Proposition 4. *Let $\Delta^{(k),l_k}$ be a redundant differential consequence as defined in* (4.35)*. Then:*

$$\textit{rank}\left[\textit{Jac}(\Delta^{(k)})\right] = \textit{rank}\left[\textit{Jac}(\Delta^{(k)} \setminus \{\Delta^{(k),l_k}\})\right], \tag{4.36}$$

where Jac$(\cdot)$ is the Jacobian matrix operator with respect to the jet coordinates of $J^{(k)}(M)$.

A proof of Proposition 4 is provided in Appendix A.3 and is inspired by the result in Theorem 7.1.6 in [23]. In contrast to Proposition 4, Theorem 7.1.6 in Seiler's work excludes the columns of Jac$(\Delta^{(k)})$ corresponding to ${}^{\partial\Delta^{(k),l_k}}/_{\partial x^i}$. This exclusion arises because Seiler's focus is not on the **geometric** differential consequences, but rather on those induced by relations among the fiber coordinates. Here, a geometric differential consequence refers to the natural tangency to the base manifold M. As such, it does not represent a non-trivial differential consequence within the emerging fiber $T\Sigma^{(k)} \cap \ker(d\pi_{k-1}^k)$, relative to the fibration π_{k-1}^k (see (3.22)). Specifically, $\ker(d\pi_{k-1}^k)$ denotes the kernel of the differential map associated with π_{k-1}^k (see (3.24)). However, it does not mean that such geometric consequences do not influence the rank of Jac$(\Delta^{(k)})$ and hence, we include them in our implementation.

Consequently, by Proposition 4, redundant differential consequences introduced at the k^{th} prolongation step do not influence $\dim(\Sigma^{(k)})$, allowing us to compute:

$$\dim(\Sigma^{(k)}) = m_k - \text{rank}\left[\text{Jac}(\Delta^{(k)})\right], \tag{4.37}$$

where $m_k = \dim(J^{(k)}(M))$ (see (3.37)).

In summary, redundant differential consequences do not need to be eliminated to evaluate (4.32) in our algorithms. Nonetheless, for generalized symmetries and/or DE systems that define non-trivial contact structures, the number of constraints introduced via (4.33) can become computationally prohibitive, necessitating algebraic methods to discard such differential consequences. However, this aspect is reserved for future research, as the prolongation process is seldom required for Lie point transformations.

We also want reader to recall that, in (3.22), the canonical projection for a given jet bundle is defined. As elaborated in the corresponding section, this operator can be reliably employed in conjunction with Lie point transformations. Notably, different jet space orders are required for the various cost functions discussed thus far. Consequently, the projection π_s^t is utilized across various algorithms in Appendix D and implicitly implemented by discarding the prolonged components of order $k = t, t-1, \ldots, s+1$ from the relevant analytic objects, where $s < t$.

Remark 42 (Discovering no new Symmetry upon Prolongation)**.** *The purpose of the inner while-loop in Algorithm 6 is to prolong the analytical objects of interest and search for additional symmetries that may arise from the constraints introduced by* (4.33)*. However, prolonging the system as described in* (4.33) *to address non-involutivity and establish a contact-invariant coframe does not guarantee the discovery of new symmetries. Consequently, the absence of newly identified symmetries upon prolongation does not indicate a failure of the algorithm. Thus, uncovering new symmetries is not a prerequisite for satisfying the conditions in* (4.32) *and* (4.25)*.*

4.4 Constructing the Group Action from the Learned Lie Algebra

Generating group actions is crucial for determining a G-coframe, as fundamental lifted differential invariants, defined in (3.93), are constructed using these actions.

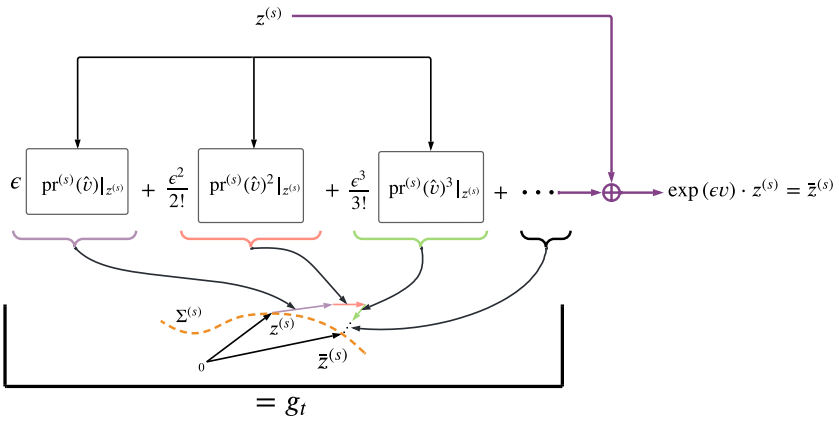

Figure 4.2: The implementation of $\exp : T_eG \to G$ in (3.10) using iResNET. The iResNET block g_t represents the action of $\{\exp(\epsilon w) - e\}$ in $J^{(s)}(M)$ (see Figure 4.1) and ensures all the group axioms in Definition 3.

Having obtained an approximation of the representation of $\mathfrak{g} \in J^{(s)}(M)$, we now turn to the construction of any $g_\epsilon \in G$ in the neighborhood of $e \in G$. For this purpose, we revisit the exponential map $\exp : T_eG \to G$, introduced in (3.10), in the context of the group action $\Psi : G \times J^{(s)}(M) \to J^{(s)}(M)$ defined in (3.1). Using (3.13) and considering $z^{(s)}$ as a coordinate function, we write:

$$\begin{aligned}\Psi(\exp(\epsilon w), z^{(s)}) &= z^{(s)} + \underbrace{\epsilon\, \mathrm{pr}^{(s)}(\hat{w})|_{z^{(s)}} + \frac{\epsilon^2}{2!}\mathrm{pr}^{(s)}(\hat{w})\left[\mathrm{pr}^{(s)}(\hat{w})\right]|_{z^{(s)}} + \dots}_{=g_t} \\ &= \bar{z}^{(s)},\end{aligned} \tag{4.38}$$

where $\epsilon \in \mathbb{R}$ and $z^{(s)},\ \bar{z}^{(s)} \in J^{(s)}(M)$ (see Figure 4.2).

The group axioms, listed in Definition 3, must be satisfied by our implementation. We employ iResNET, as discussed in Section 4.1, to fulfill these requirements. The identity element $e \in G$ is trivially satisfied when $\epsilon = 0$, as it nullifies the g_t term in (4.38). Associativity follows from the compatibility rule of the group action Ψ, stated in (3.1), and the composition law of iResNET, given as:

$$(\mathbb{1} + g_{t+1}) \cdot [(\mathbb{1} + g_t) \cdot z_{t-1}^{(s)}] = [(\mathbb{1} + g_{t+1}) \cdot (\mathbb{1} + g_t)] \cdot z_{t-1}^{(s)}, \tag{4.39}$$

where $\mathbb{1}$ is the identity operator on $J^{(s)}(M)$. Ensuring invertibility is more challenging but can be addressed using Theorem 3.42 in [1] and the following result:

Proposition 5. *Let $\epsilon \in \mathbb{R}$, $\hat{v} \in \mathcal{V}_{o_s}$, and*

$$\|\boldsymbol{pr}^{(s)}(\hat{v})\|_{op} \coloneqq \sup_{TJ^{(s)}(M)\ni\hat{w}\neq 0} \frac{\|\boldsymbol{pr}^{(s)}(\hat{v})[\hat{w}]\|}{\|\boldsymbol{pr}^{(s)}(\hat{v})\|\|\hat{w}\|} = K, \tag{4.40}$$

where $\|\cdot\|_{op}$ is the operator norm on $TJ^{(s)}(M)$ with respect to a metric defined on $J^{(s)}(M)$. If $K < {\ln(2)}/{|\epsilon|}$, then, by Theorem 1 in [41], the transformation specified in (4.38)*, implemented via iResNET, is invertible.*

Remark 43 (Why $\mathcal{V}_{o_s}$ in Proposition 5?)**.** *Observe that $\mathcal{W}_{\boldsymbol{f}}$ is not involved in Proposition 5, but $\mathcal{V}_{o_s}$. The reason is that the operator norm of any $\hat{w}_j \in \mathcal{W}_{\boldsymbol{f}}$ could be controlled by $\hat{v}_i$'s spanning $\mathcal{V}_{o_s}$.*

The proof of Proposition 5 is provided in Appendix A.4. As its name suggests, iResNET ensures invertibility by controlling the spectral norms $\{\sigma_i^1, \ldots, \sigma_i^{N_i}\}$ of the linear layers in FCN_{λ_i} associated with $\hat{v}_i$, and by employing contractive non-linear activation functions. Here, N_i denotes the number of linear layers in FCN_{λ_i}. The following inequality illustrates why bounding the spectral norms of $\{W_i^j\}_{j=1}^{N_i-1}$ constrains the operator norm as stated in Proposition 5:

$$\|W_i^j\|_{\text{op}}^2 \geq \frac{\hat{v}^{\text{t}}(W_i^j)^{\text{t}}W_i^j\hat{v}}{\|\hat{v}\|^2}, \tag{4.41}$$

where $\hat{v} \in TJ^{(s)}(M)$. The maximum value of the *Rayleigh quotient* (the right-hand side of the inequality) is achieved when $\hat{v}$ is the eigenvector of W_i^j corresponding to σ_i^j. Therefore,

$$\|W_i^j\|_{\text{op}} = \sigma_i^j, \tag{4.42}$$

where $\|\cdot\|_{\text{op}}$ denotes the ℓ_2-induced operator norm. Moreover, σ_i^j may provide an upper bound on $\|W_i^j\|_{\text{op}}$ when the operator norm is induced by a norm other than ℓ_2. Consequently, iResNET must ensure $\sigma_i^j < {\ln(2)}/{|\epsilon|}$, where $i = 1, \ldots, o_s$ and $j = 1, \ldots, N_i$. Although contractive activation functions naturally reduce K below this threshold, minor adjustments can further optimize it.

An accuracy issue arises due to truncating the expansion in (4.38), which becomes significant as consecutive actions are applied to $z_0^{(s)} \in \Sigma^{(s)}$. To mitigate this deviation from the submanifold $\Sigma^{(s)}$ of interest, a retraction map $r : U \to \Sigma^{(s)}$, where $U \subset J^{(s)}(M)$, is employed. By Proposition 6.25 in [39], r is a smooth submersion

computed via:

$$\min_{z^{(s)},\mu} \|\bar{z}^{(s)} - z^{(s)}\|^2 + \mu^{\mathrm{t}} \Delta^{(s)}(z^{(s)}), \tag{4.43}$$

where $\mu \in \mathbb{R}^n$ are Lagrange multipliers, n is the number of equations in $\Delta^{(s)} = 0$, and $\bar{z}^{(s)}$ is the truncated group action output from (4.38). Typically, we expand (4.38) to second or third order for computational efficiency.

The retraction map r can be simplified when linear DE systems are considered, as such systems are embedded into $J^{(s)}(M)$ as affine submanifolds—for instance, a hyperplane. This observation is utilized in Section 5.4, where the sample DE system is linear and homogeneous, corresponding to a hyperplane passing through the origin in $J^{(s_0)}(M)$. Consequently, the retraction map reduces to the canonical projection from the normal bundle $N_{\Sigma^{(s)}} = TJ^{(s)}(M)|_{\Sigma^{(s)}} \setminus T\Sigma^{(s)}$ (see the definition preceding Theorem 6.23 in [39]) onto $\Sigma^{(s)}(M)$. Explicitly, it is given by:

$$r(\bar{z}^{(s)}) = \bar{z}^{(s)} - \underbrace{\tilde{A} \cdot (A \cdot \bar{z}^{(s)} - b)}_{\in N_{\Sigma^{(s)}}}, \tag{4.44}$$

where A represents the matrix form of $\Delta^{(s)} = 0$, $\tilde{A} = A^{\mathrm{t}}(A \cdot A^{\mathrm{t}})^{-1}$ is the *Moore–Penrose inverse* of A, and $b \in \mathbb{R}$ is the inhomogeneous component of $\Delta^{(s)} = 0$ (see the prolongation of $\Delta^{(s_0)}$ to order $s > s_0$ in (4.33)).

Remark 44 (Activation of retraction map)**.** *The retraction map $r : U \to \Sigma^{(s)}$, implemented via* (4.43)*, is only activated when $\|\Delta^{(s)}(\bar{z}^{(s)})\|$, measured in the ℓ_2 norm, exceeds a predefined threshold.*

CHAPTER 5

EXPERIMENTS

This section presents our experiments to validate and assess the generalization capabilities of our neural architecture, implemented in PyTorch [50]. First, we introduce and analyze a selected DE system with **known** symmetry generators, obtained by solving the *determining equations*. Second, these known symmetries are classified into trivial and non-trivial categories using the computational algebra tool SymPy [51]. For one of these non-trivial, contact-preserving, canonical symmetry generators, we verify whether it remains within the span of the learned involutive VFS $\mathcal{V}_{os}^{(s)}$. Finally, we alter the domain and initial conditions (ICs) to generate entirely new datasets, distinct from those used during training and validation, to evaluate the framework's generalization under varying conditions. Additionally, we demonstrate how the group actions, constructed in Section 4.4, begin to diverge from the solution manifold $\Sigma^{(s)}$ and highlight how the retraction map in (4.44) can mitigate the adverse effects caused by truncating the $\exp$ in (4.38).

5.1 The Differential Equation System & The Input Data

We consider a simple, linear, and homogeneous DE system whose symmetry generators, spanning $\mathfrak{g}$, are known. This setup enables direct comparison between the learned $\bar{\mathfrak{g}}$ and the original symmetries, as well as the dimension of its representation in $J^{(s)}(M)$. For instance, in Section 5.3, the representation of a non-trivial symmetry generator $v \in \mathfrak{g}$ is used to demonstrate that it lies within $\mathrm{Span}(\mathcal{V}_{os}^{(s)})$, which corresponds to the representation of $\bar{\mathfrak{g}}$ in $TJ^{(s)}(M)$.

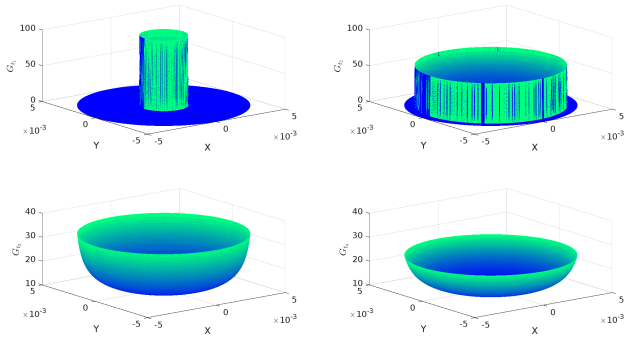

Figure 5.1: The Green's function for the wave equation in two spatial dimensions at different time steps. Time runs forward from left to right and from top to bottom. The point source is located at $[x_1\ y_1]^{\mathrm{t}} = [0\ 0]^{\mathrm{t}}$, its time $t_1 = 0.0$, and $c = 1.0$. Observe that, the wave front is still within the data boundary in the figures located at the top; while it goes out of that boundary in the figures at the bottom.

To validate these principles, we employ the wave equation in two spatial dimensions:

$$\Delta^{(2)}(z^{(2)}) = u_{tt} - c^2\left(u_{xx} + u_{yy}\right) = 0, \tag{5.1}$$

where (x, y, t) are the independent variables, u is the dependent variable, and $c = 1$ is the wave propagation speed used throughout all experiments in this study. As discussed in Section 3.5, the fact that a DE system is a Pfaffian does not **explicitly** influence the construction of our framework. Nevertheless, we provide the proof that (5.1) constitutes a Pfaffian DE in Appendix B, as it **implicitly** reflects the complexity of the system and, consequently, the difficulty of the learning process. To construct a simple solution $u = f(x, y, t)$, we utilize an elementary setup.

The Green's function for (5.1) is given by (see Exercise 10.2.12-(a) in [56]):

$$G\left(x, y, t, x_1, y_1, t_1\right) = \left\{\frac{1}{2\pi c\sqrt{c^2(t_1 - t)^2 - r^2}}\right\} H\left(c\left(t_1 - t\right) - r\right), \tag{5.2}$$

where $H(\cdot)$ denotes the Heaviside function in two dimensions, $r = \sqrt{\left(x_1 - x\right)^2 + \left(y_1 - y\right)^2}$, and the coordinate pairs (x_1, y_1, t_1) and (x, y, t) represent source and response locations, respectively (see Figure 5.1).

To simplify the numerical integration further, the following ICs are imposed:

$$u_0 = 0,$$
$$u_{t_0} = \begin{cases} 1, & -0.05 \leq x, y \leq 0.05, \\ 0, & \text{otherwise}, \end{cases} \tag{5.3}$$

where u_{t_0} represents the IC for u_t. The ICs in (5.3) states that the field disturbance (initial velocity) is spread across the specified rectangular region. The fact that $u_0 = 0$ implies the medium starts at rest with no initial displacement, but the nonzero initial velocity implies that the medium is "kicked" in this small rectangular region, initiating wave propagation. Furthermore, no boundary condition (BC) is imposed as boundaries are assumed at infinity and the wave dies down due to its finite energy. The closed-form expression for the numerical integration is then given by (see Chapter 10 in [56]):

$$u(x, y, t) = \int_{y_1} \int_{x_1} u_{t_0} G(x, y, t, x_1, y_1, t_1), dx_1, dy_1. \tag{5.4}$$

For the training and validation purposes, we define a rectangular spatial domain $[-0.03, 0.03] \times [-0.03, 0.03]$ (response locations) and a temporal domain $[0.2, 0.206]$, denoted by:

$$\boldsymbol{\mathcal{D}}_{\text{tr}} = \{(x_i, y_i, t_i) \mid t_i \in (0.2,\ 0.206),$$
$$(x_i, y_i) \in [-0.03,\ 0.03] \times [-0.03,\ 0.03]\}. \tag{5.5}$$

The spatial domain is discretized into a 101×101 uniform grid, while the temporal domain is divided into 10 equal subintervals, yielding 11 equally spaced time points. This results in a total of 112211 grid points in $\Gamma_{\boldsymbol{f}_{\text{tr}}}$.

Remark 45 (Maximal Rank Condition and Sample Points)**.** *Let* $\partial\Delta^{(s),i}/\partial z^{(s),j}$ *denote the Jacobian matrix of* $\Delta^{(s)} : J^{(s)}(M) \to \mathbb{R}^n$*. Any irregular sampling point* $z^{(s)} \in \Gamma_f^{(s)}$ *such that* $\partial\Delta^{(s),l_s}/\partial z^{(s),j_s}\big|_{z^{(s)}}$ *is not of maximal rank should be excluded. Otherwise, such samples* ***may*** *compromise the local effectiveness and/or transitivity of the group action* $G^{(s)}$ *(see Definitions 12 and 13).*

In Figure 5.2, the graphs of u_{xx}, u_{yy}, and u_{tt}, constructed using the *findiff* library [55], are displayed sequentially. The lower-right subfigure demonstrates the numerical accuracy of evaluating these derivatives within (5.1). The computed numerical

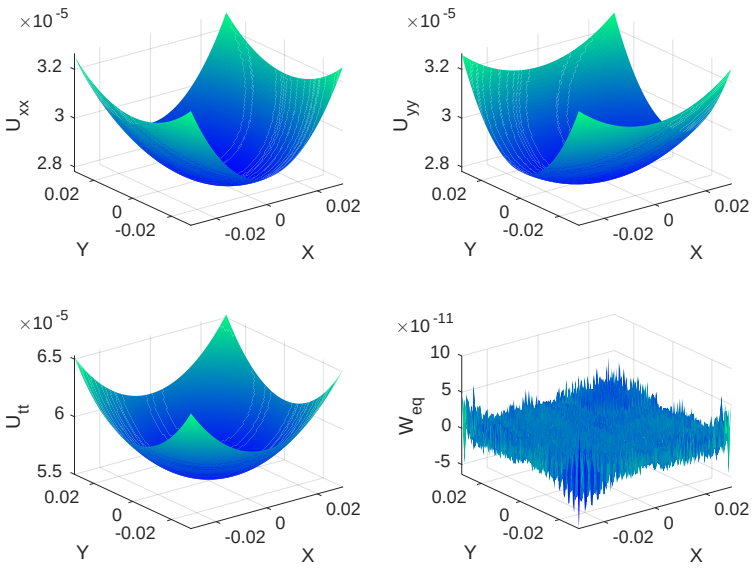

Figure 5.2: The graphs of u_{xx}, u_{yy}, and u_{tt} at $t_0 = 0.216$, and the accuracy of (5.1). The accuracy is illustrated in the lower-right graph and is in the order of $\approx 7e-11$.

accuracy is approximately 7×10^{-11}, which serves as a relative benchmark for accepting a model.

A valid question arises: "Does utilizing a solution obtained with imposing IBCs restrict the process of learning a **full set** of non-trivial symmetries, since the solution submanifold $\Sigma^{(s)}|_{\text{IBC}}$ is a subset of $\Sigma^{(s)}$?" [57]. The concise answer is: "No, it does not, provided that the symmetries are non-trivial and their flows are non-degenerate". Consider the cost functions discussed in Section 4. None of those functions depend on whether a sample solution lies within a submanifold of $\Sigma^{(s)}$ or not. This is because the flow of a non-trivial symmetry generator maps a solution to an entirely new solution, independent of how the original solution was obtained. The only limitation is that the mapped solution may no longer belong to $\Sigma^{(s)}|_{\text{IBC}}$, although it will still reside in $\Sigma^{(s)}$. Thus, we conclude that imposing IBCs does not hinder the learning process or reduce the number of discoverable symmetry generators. However, for future studies aimed at constructing invariant G-coframes to simplify $\Delta^{(s)} = 0$, the coframe should be established on $\Sigma^{(s)}|_{\text{IBC}}$ rather than on the broader $\Sigma^{(s)}$.

Remark 46 (Why not using a known solution?)**.** *A closed-form solution, such as the one specified in* (5.8)*, could have been utilized without imposing any IBCs. This approach would certainly simplify data generation, especially in practical applications.*

However, our decision to generate an approximated solution serves three primary purposes. First, it simulates a dataset as if captured from a dynamical system, which in real-world scenarios is always constrained by a set of IBCs. Second, there might be DE systems that have no closed form definition and learned/approximated by another machine learning framework by analyzing a given dataset [19]. Lastly, it provides a means to evaluate the findiff library's capability in prolonging a solution to $J^{(s)}(M)$. This is particularly important since the output of a dynamical system typically lacks partial derivatives of the associated sample solution.

5.2 Training Results

In this section, we analyze the training performance corresponding to the dataset introduced in Section 5.1. As computed in Example 2.43 of [8], the dimension of the solution manifold for the wave equation in two spatial dimensions is given by $\dim(\Sigma^{(2)}) = 13 - 1 = 12$, based on the assumption that $\Delta^{(2)}$ is of maximal rank (see Definition 15). Furthermore, since the wave equation constitutes a linear differential system, the submanifold $\Sigma^{(2)}$ forms a hyperplane within the jet space $J^{(2)}(M)$.

The goal of the training procedure is to recover eleven intrinsic symmetry generators using the methodology developed in Section 4.1, despite the fact that $\Sigma^{(2)}$ is a 12-D submanifold invariant under the symmetry group. This discrepancy arises because one of the symmetry generators, denoted v_α in Example 2.43 of [8], is solution-dependent. To avoid generating an infinite-dimensional Lie algebra, this generator is excluded from $\mathcal{V}$ by imposing the constraint $\mathscr{C}_{\text{dep}}$ in (4.13). Indeed, it is a well-established fact that any linear differential equation system admits a solution-dependent symmetry generator as described in (4.10). Consequently, the final learned vector field system is denoted $\mathcal{V}_{11}^{(2)}$.

The results of the training process used to learn $\mathcal{V}_{11}^{(2)}$ are presented in Table 5.1. For each $\hat{v}_i \in \mathcal{V}_{11}^{(2)}$, only the model corresponding to the lowest validation cost is reported. A primary observation is that the cost function tends to increase as additional vector fields are incorporated into $\mathcal{V}_{11}^{(2)}$. This trend is expected, as the simultaneous satisfaction of all imposed constraints—particularly the linear independence condi-

tion—becomes increasingly difficult for the optimizer.

Given that the quality of the input dataset in satisfying $\Delta^{(2)} = 0$ from (5.1) is approximately 7×10^{-11}, and that $\mathscr{J}_{\text{sym}}$ remains well below this threshold, we may confidently conclude that each $\hat{v}_i \in \mathcal{V}_{11}$ is tangent to $\Sigma^{(2)}$. While the constraint $\mathscr{C}_{\text{dep}}$ in (4.13) is theoretically justified, it becomes numerically steep as $\|\Delta^{(s)}\| \to 0$, leading to a loss of smoothness in practical implementation. This, in turn, causes abrupt discontinuities in the cost function, which may destabilize the training process at the epochs where such discontinuities occur.

To mitigate this issue, a smoother formulation of the constraint is introduced as follows:

$$\mathscr{C}_{\text{dep}} = h(e)\text{log1p}(\max(\text{tol}_{\text{dep}} - \|\Delta^{(s)}(\boldsymbol{x}, \hat{\varphi}^{\alpha}_{J})\|, 0)/\text{tol}_{\text{dep}}), \tag{5.6}$$

where $h(e)$ is a ramp factor, depending on the current epoch number e, that increases up to a predefined epoch and then remains constant. The function $\text{log1p}(x) = \ln(1 + x)$, and tol_{dep} is typically chosen to match the step size of the input dataset, which is 6×10^{-4} in this experimental setup.

Both $\mathscr{J}_{\text{nlu}}$ and the smoothed constraint $\mathscr{C}_{\text{dep}}$ are observed to be satisfied at an early stage of training. Once satisfied, these terms remain inactive throughout the remainder of the training process, as they only become active if $J_{\text{nlu}} > 6 \times 10^{-4}$ or $\mathscr{C}_{\text{dep}} > 6 \times 10^{-4}$.

In the rest of this section, we examine the properties of the Lie subalgebra comprising solely **intrinsic** symmetry generators and justify why $\mathcal{W}_{\boldsymbol{f}}^{(2)}$ must form a subalgebra within the subspace spanned by $\mathcal{V}_{11}^{(2)}$, as outlined in Section 4.2. Moreover, no further prolongation beyond $s_0 = 2$ is required, as expected. This result aligns with the discussion in Section 4.3, where it is noted that prolongation is unnecessary for Lie point symmetries (see also Remark 33).

In accordance with the condition stated in (4.32), the dimension of any orbit generated by non-trivial, contact-preserving symmetry generators in $\mathcal{W}_{\boldsymbol{f}}$ must satisfy:

$$o_{s_2} \leq \dim(\Sigma^{(2)}) - p = 12 - 3 = 9. \tag{5.7}$$

Since all symmetries under consideration are Lie point symmetries, they are inherently contact-preserving, as discussed in the paragraph following Proposition 1. The

Table 5.1: Quality of validation of costs and constraints discussed in Section 4.1.

Idx	$\mathscr{J}_{\text{sym}}$	$\mathscr{J}_{\text{nlu}}$	$\mathscr{C}_{\text{ind}}$ $\text{tol}_{\text{ind}} = 0.846$	$\mathscr{C}_{\text{dep}}$ $\text{tol}_{\text{dep}} = 6e-4$
1	3.907e-13	0.0	**N/A**	0.0
2	8.122e-13	0.0	-0.842	0.0
3	6.972e-13	0.0	-0.801	0.0
4	8.658e-13	0.0	-0.714	0.0
5	7.580e-13	0.0	-0.658	0.0
6	9.011e-13	0.0	-0.529	0.0
7	6.001e-13	0.0	-0.405	0.0
8	8.843e-13	0.0	-0.328	0.0
9	9.240e-13	0.0	-0.219	0.0
10	9.341e-13	0.0	-0.116	0.0
11	1.086e-12	0.0	-0.001	0.0

eleven symmetry generators obtained by solving the determining equations are listed in Example 2.43 of [8]. However, not all of these generators are necessarily non-trivial. Their classification is carried out via the following procedure:

(i) Select a closed-form solution $u = f(x, y, t)$ to (5.1).

(ii) Compute the characteristics Q_i of each symmetry generator $\hat{v}_i$ from (2.65) and the subsequent equation in [8] using a symbolic computation library, specifically *Sympy* [51].

(iii) Evaluate Q_i on $\Gamma^{(1)}_{f_{\text{te}}}$: Symbolically test whether Q_i vanishes for $i = 1, \ldots, 12$.

To this end, we select the following elementary solution:

$$u(x, y, t) = \cos(kx)\cos(ly)\cos(\omega t), \tag{5.8}$$

where $\omega^2 = c^2(k^2 + l^2)$ represents the dispersion relation, and $c = 1$ as specified in Section 5.1. Steps (ii) and (ii) are then executed sequentially.

Remark 47 ($c = 1$ in Example 2.43 in [8])**.** *Assigning any constant value to c does not affect the conclusions drawn in this thesis. Furthermore, the general algebraic expressions of the symmetry generators presented in (2.65) and the subsequent equation in [8] remain structurally intact. However, c-dependent multiplicative factors*

may *appear in the expressions. Accordingly, the symmetry generators described in Example 2.43 of [8] remain valid.*

We find that all eleven symmetry generators listed in Example 2.43 of [8] are non-trivial, as $Q_i|_{\Gamma^{(1)}_{\boldsymbol{f}}} \neq 0$ for all $i = 1, \ldots, 11$. It is important, however, not to conflate the dimension of an involutive distribution with the number of linearly independent symmetry generators that act non-trivially at each point of the surface $\Gamma^{(2)}_{\boldsymbol{f}_{\text{te}}}$, defined by (5.8). To illustrate, consider a 3-D surface embedded in a 12-D space: although there may exist eleven vector fields that deform the surface in various directions, the normal bundle of the surface is limited to nine dimensions, as dictated by (5.7). Hence, only nine of these directions can be linearly independent, while the remaining two must be expressible as linear combinations of the others, modulo the tangent space of the surface.

As outlined in Algorithm 5, we initialize with the maximum number of generators that can be linearly independent and act non-trivially on $\Gamma^{(2)}_{\boldsymbol{f}_{\text{te}}}$, namely $\dim(\mathcal{W}^{(2)}_f) = 9$. Subsequently, $\dim(\mathcal{W}^{(2)}_f)$ is iteratively reduced until either $\mathcal{W}^{(2)}_f$ becomes involutive or its dimension reaches unity. For the specific function $\boldsymbol{f}(\boldsymbol{x})$ given in (5.8), involutivity is achieved precisely when $\dim(\mathcal{W}^{(2)}_{\boldsymbol{f}}) = 9$. The corresponding validation results for the optimal model, representing the coefficients a^i_j in (4.19), are summarized in Table 5.2. The constraint $\mathscr{C}_{\text{ind}}$ is not applied to individual generators $\hat{w}_j \in \mathcal{W}_{\boldsymbol{f}}$ in isolation, as is the case for each $\hat{v}_i \in \mathcal{V}$. Instead, it is designed to ensure linear independence across the entire set $\mathcal{W}_{\boldsymbol{f}}$ simultaneously. For the best-performing model, the average value is observed to be $\mathscr{C}_{\text{ind}} \approx -0.7$. Regarding $\mathscr{C}_{\text{ntr}}$, it is deactivated in the early stages of training, similar to the behavior of $\mathscr{C}_{\text{dep}}$ and $\mathcal{J}_{\text{nlu}}$ as discussed above. Furthermore, the value of $\mathcal{J}_{\text{inv}}$ satisfies $\mathcal{J}_{\text{inv}} \ll 6e - 11$, where $6e - 11$ corresponds to the average residual $\|\Delta^{(s)}\|$ of the input dataset. This result verifies that a full symmetry reduction via $\mathcal{W}^{(2)}_{\boldsymbol{f}}$ is indeed achievable.

5.3 Verifying the Learned Algebra against a Known Canonical Symmetry

The objective of this experiment is to determine whether a known non-trivial, contact-preserving symmetry of (5.1) lies within $\text{Span}(\mathcal{V}^{(s_c)}_{o_{s_c}})$. This validation seeks to ensure

Table 5.2: Quality of validation of costs and constraints discussed in Section 4.2.

Idx	$\mathcal{J}_{\text{inv}}$	$\mathscr{C}_{\text{ntr}}$ $\text{tol}_{\text{ntr}} = 6e-4$
1	2.014e-13	0.0
2	8.994e-13	0.0
3	3.467e-12	0.0
4	7.146e-13	0.0
5	3.582e-12	0.0
6	4.088e-13	0.0
7	6.114e-13	0.0
8	9.056e-13	0.0
9	2.348e-12	0.0

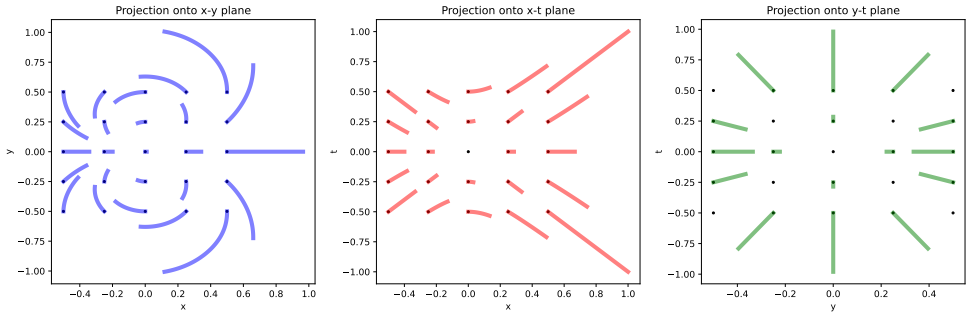

Figure 5.3: Some integral curves of $\hat{v}_{\text{can}}$ in (5.9), that are projected onto $x-y$, $x-t$, and $y-t$ planes respectively. Black dots show the initial points of the flows.

that $\mathcal{V}_{o_{sc}}^{(s_c)}$, associated with $\bar{\mathfrak{g}}$, represents the correct subspace.

We begin by identifying the canonical symmetry generator, $\hat{v}_{\text{can}} \in \mathfrak{g}$. In light of none of $Q_i|_{\Gamma^{(2)}_{\boldsymbol{f}_{\text{te}}}} = 0$ for the solution in (5.8), we select the following vector field as the canonical symmetry generator (see (3.7) and (3.20) for notation):

$$\hat{v}_{\text{can}} := \hat{v}_8 = x^2 - y^2 + t^2\frac{\partial}{\partial x} + 2xy\frac{\partial}{\partial y} + 2xt\frac{\partial}{\partial t} - xu\frac{\partial}{\partial u} \tag{5.9}$$

The projections of the flows generated by $\hat{v}_{\text{can}}$ onto various planes in the base manifold M are depicted in Figure 5.3.

To determine whether $\hat{v}_{\text{can}} \in \text{Span}(\mathcal{V}_{11})$, a model must be proposed to approximate $\hat{v}_{\text{can}}$. Definition 39 suggests the appropriate model for this approximation. Specifi-

cally, any vector field within $\mathcal{V}_{11}$ takes the following form:

$$\hat{v}|_z = \sum_{i=1}^{11} b^j(z)\hat{v}_j|_z, \tag{5.10}$$

where $z = (\boldsymbol{x}, \boldsymbol{u})$. Consequently, the goal is to learn smooth coefficient functions $b^j_{\text{can}}(z)$, where $j = 1, \ldots, 4$. If $\hat{v}_{\text{can}} \notin \mathcal{V}_{11}$, the algorithm described below will fail to identify such functions $b^j_{\text{can}}(z)$. Assuming that $\boldsymbol{b}_{\text{can}}(z) = [b^1_{\text{can}}, \ldots, b^{11}_{\text{can}}]^{\text{t}}$ is modeled by FCN_θ, where θ represents the parameter set, the cost function to be minimized is given by:

$$\mathcal{J}_{\text{can}} = \frac{1}{2}\left\| \sum_{i=1}^{11} b^j_{\text{can}}(z)\hat{v}_j|_z - \hat{v}_{\text{can}|_z}. \right\|^2 \tag{5.11}$$

If $\mathcal{J}_{\text{can}}$ cannot be reduced below a predefined threshold tol_{can}, it can be concluded that $\mathcal{V}^{(2)}_4$ fails to span the correct subspace of non-trivial, contact-preserving symmetry generators. The corresponding learning algorithm is presented as Algorithm 7.

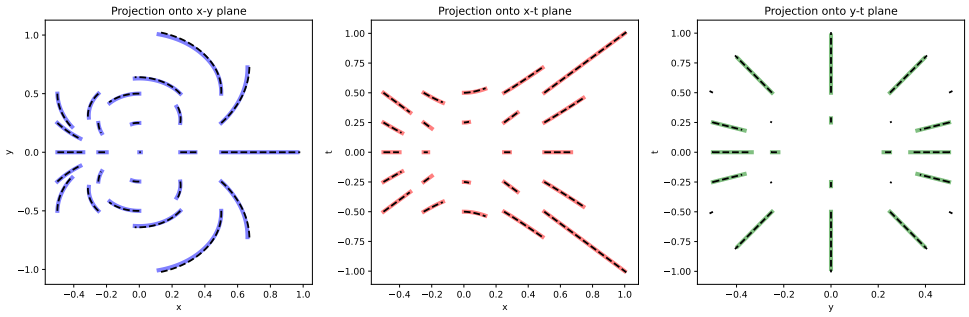

Figure 5.4: Some integral curves of $\hat{\tilde{v}}_{\text{can}}$, an approximation to $\hat{v}_{\text{can}}$ in (5.9) via the learned $\mathcal{V}_{11}$, that overlay the projections of $\hat{v}_{\text{can}}$ onto $x - y$, $x - t$, and $y - t$ planes, respectively, illustrated in Figure 5.3. They are drawn in dashed, black, and thinner lines.

We now focus on verifying whether $\hat{v}_{\text{can}} \in \text{Span}(\mathcal{V}_{11})$. Although, as stated in Remark 38, $\mathcal{V}_{11}$ may not form a linearly independent set until it is prolonged to a certain jet space of order s, $\hat{v}_{\text{can}}$ is still expected to lie in the span of $\mathcal{V}_{11}$. Thus, employing its prolonged version, $\mathcal{V}^{(2)}_{11}$, is necessary.

Next, we implement Algorithm 7 to evaluate whether the cost function (5.11) can be reduced below the specified threshold $\text{tol}_{\text{can}} = 6e - 4$. This threshold value was chosen to match the dataset step size. Remarkably, $\mathcal{J}_{\text{can}}$ drops below $5e - 5$ within

approximately 200 training epochs, and the entire computation completes in under four minutes on a grid platform equipped with **NVIDIA A100** graphical processing units (GPUs).

In conclusion, this experiment validates that $\mathcal{V}_{11}$ accurately spans the subspace defined by $\{\hat{v}_1, \ldots, \hat{v}_{11}\}$ from Example 2.43 in [8]. Additionally, we provide visualizations of the integral curves of the approximated vector field, denoted $\hat{\hat{v}}_{\text{can}}$, in Figure 5.4.

Remark 48 (Querying Datasets for Known Symmetry Generators)**.** *The straightforward yet effective method discussed in this subsection, namely "approximating a canonical symmetry generator", can also be applied to query datasets for* ***known*** *symmetries. This capability is especially valuable for interpreting datasets originating from physical and engineering problems, offering insights into underlying symmetries present in such systems.*

5.4 Generalization of the Learned Algebra to Other Domains and The Effect of Retraction

In this final subsection, we perform two types of experiments: *(i)* evaluating the generalization capabilities of the proposed neural architecture under different conditions, and *(ii)* analyzing the cumulative effect of repeatedly applying a small group action g_ϵ, as well as how the accuracy of the symmetries can be recovered.

To assess whether the learned symmetries are influenced by changes in the input domain, we generate a new dataset, $\Gamma^{(2)}_{\boldsymbol{f}_{\text{te}}}$, by modifying the domain defined in Section 5.1 as follows:

$$\boldsymbol{\mathcal{D}}_{\text{tr}} = \{(x_i, y_i, t_i) \mid t_i \in (0.3,\, 0.306),\\ (x_i, y_i) \in [-0.04,\, 0.02] \times [-0.04,\, 0.02]\}. \quad (5.12)$$

It is important to emphasize that the initial conditions (ICs) defined in (5.3) remain unchanged. Thus, only the sampling points at which the response of the propagating wave is measured are modified. The results obtained for all symmetry generators in $\mathcal{V}_{11}$ are summarized in the upper section of Table 5.3. Notably, altering the input

domain does not result in any discernible impact on the quality of the cost function. The last column indicates the measure of divergence from the solution submanifold $\Sigma^{(2)}$:

$$\mathcal{J}_{\text{div}} = |\Delta^{(2)}(g_\epsilon^n \cdot z^{(2)})|, \tag{5.13}$$

where $z^{(2)} \in \Gamma^{(2)}_{\boldsymbol{f}_{\text{te}}}$, $\Delta^{(2)}$ is the DE system in (5.2), and g_ϵ^n denotes the continued application of $g_\epsilon \in G$ for $n \in \mathbb{N}$ times.

Table 5.3: The quality of testing in case $\boldsymbol{\mathcal{D}}_{\text{te}}$ is set to (5.12). $\#g_{\epsilon^i}$ stands for the number of repeated application of the group action g_{ϵ^i}, which is generated by $\hat{v}_i \in \mathcal{V}_{11}$ from Section 5.2

	Idx	$\mathcal{J}_{\text{sym}}$	$\mathcal{J}_{\text{div}}$
$\#g_{\epsilon^i} = 1$			
	1	6.110e-13	8.414e-13
	2	2.446e-12	3.113e-12
	3	9.119e-13	5.821e-12
	4	1.366e-12	2.886e-12
	5	1.002e-12	1.922e-12
	6	3.241e-12	5.286e-12
	7	5.794e-13	8.011e-13
	8	7.587e-13	9.046e-13
	9	4.371e-12	6.478e-12
	10	2.244e-12	4.919e-12
	11	7.866e-11	1.009e-10
$\#g_{\epsilon^i} = 1000$			
	1	4.188e-11	6.744e-11
	2	7.211e-10	1.016e-9
	3	5.183-11	7.621e-11
	4	9.666e-11	2.042e-10
	5	6.137e-11	8.414e-11
	6	9.011e-11	3.874e-10
	7	1.866e-11	8.554e-10
	8	5.843e-11	8.117e-11
	9	9.240e-10	1.811e-9
	10	9.341e-11	2.018e-10
	11	1.332e-9	4.019e-9

In the second experiment aimed at testing the generalization of our model, the training domain $\mathcal{D}_{\text{tr}}$ defined in (5.5) remains unchanged, while the ICs specified in (5.3) are

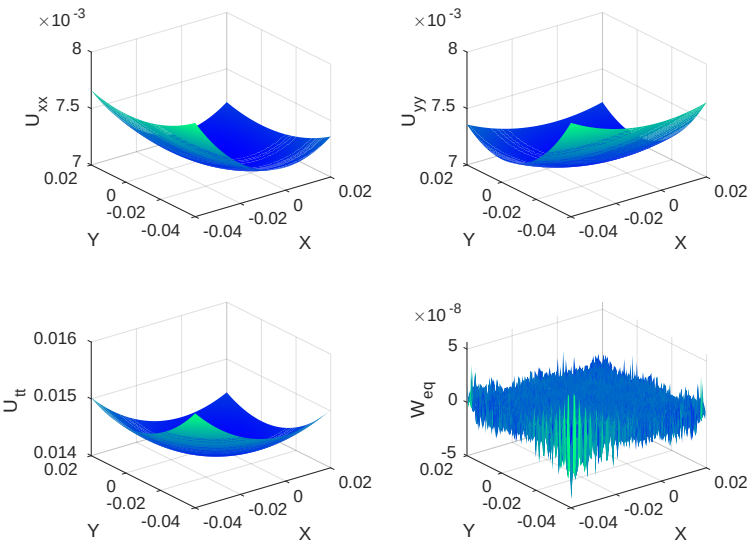

Figure 5.5: The graphs of u_{xx}, u_{yy}, and u_{tt} at $t_0 = 0.06$, and the accuracy of (5.1) when ICs are set as in (5.14). The accuracy is illustrated in the lower-right graph and is $\approx 6e-8$.

modified as follows:

$$u_0 = 0,$$
$$u_{t_0} = \begin{cases} \frac{\mathrm{e}^{\left(-\frac{\mathbf{s}^{\mathrm{t}}\cdot\mathbf{V}^{-1}\cdot\mathbf{s}}{2}\right)}}{\sqrt{(2\pi)^2|\mathbf{V}|}}, & -0.05 \leq x, y \leq 0.05, \\ 0, & \text{otherwise} \end{cases} \tag{5.14}$$

Here, $\mathbf{s} = (x, y)^{\mathrm{t}}$ denotes the spatial sample point, $\mathbf{V}$ is the covariance matrix, and $|\mathbf{V}|$ is its determinant. In other words, the initial velocity field is *normally* distributed over $\boldsymbol{\mathcal{D}}_{\text{tr}}$. In this experiment, we assume that x and y are uncorrelated, so $\mathbf{V}$ is a diagonal matrix with entries equal to 0.0255.

As illustrated in Figure 5.5, similar to Figure 5.2, the quality of the updated dataset is evaluated with respect to (5.1). The results are summarized in the upper section of Table 5.4. Changing the initial conditions does not result in any significant degradation in quality. However, compared to the cost function values reported in Table 5.3, the values in this case are on the order of $1e-8$, which is attributed to the dataset's accuracy being restricted to $\Sigma^{(2)}$. Consequently, we conclude that intrinsic **noise** plays a critical role. In real-world applications, *denoising* methods for collected signal data—applied prior to both training and testing—are essential for improving the

robustness of the system.

Table 5.4: The quality of testing in case $\boldsymbol{\mathcal{D}}_{\text{te}}$ and ICs are set to (5.5) and (5.14), respectively.

	Idx	$\mathcal{J}_{\text{sym}}$	$\mathcal{J}_{\text{div}}$
$\#g_{\epsilon^i} = 1$			
	1	6.110e-11	8.084e-11
	2	7.117e-10	9.912e-10
	3	4.831e-11	7.015e-11
	4	3.199e-10	5.001e-10
	5	9.111e-11	8.190e-10
	6	3.066e-10	4.656e-10
	7	3.217e-11	6.818e-11
	8	5.103e-11	8.999e-11
	9	2.377e-10	3.717e-10
	10	8.200e-10	1.655e-9
	11	4.778e-9	9.000e-9
$\#g_{\epsilon^i} = 1000$			
	1	2.101e-9	6.146e-9
	2	2.100e-9	7.002e-9
	3	1.818e-9	2.664e-9
	4	1.002e-8	2.015e-8
	5	4.142e-9	9.378e-9
	6	1.676e-8	3.211e-8
	7	6.606e-9	9.898e-9
	8	5.111e-9	7.379e-9
	9	6.140e-8	8.011e-8
	10	9.341e-8	2.099e-7
	11	1.002e-7	3.077e-7

In the final set of experiments, we investigate the effect of the retraction map defined in (4.43). Since (5.1) is a linear DE system, we employ the simplified retraction map given in (4.44). The group action $g_{\epsilon^i} \in G$ is constructed separately for each $\hat{v}_i \in \mathcal{V}_{11}$, $i = 1, \ldots, 11$, and the exponential map in (4.38) is truncated to third order.

The lower sections of Tables 5.3 and 5.4 report the average accuracy of the cost functions after transforming a set of initial points $z_0^{(2)} \in \Gamma^{(2)}_{\boldsymbol{f}_{\text{te}}}$ by repeatedly applying the group action $g_{\epsilon^i} \cdot z^{(2)} = \exp(\epsilon^i v_i) \cdot z_0^{(2)}$ $1,000$ times $(\#g_{\epsilon^i})$. As observed, the accumulating errors lead to a progressive decline in accuracy. However, incorporating the retraction map (4.44) when the error exceeds a predefined threshold prevents any

transformed point $\bar{z}^{(2)}$ from escaping the tubular neighborhood of $\Sigma^{(2)}$. The results demonstrating this effect are presented in Tables 5.5 and 5.6 (cf. Tables 5.3 and 5.4, respectively). For this experiment, we set $\epsilon^i = 0.6$, $i = 1, \ldots, 11$, and activate the retraction map when $\mathcal{J}_{\text{div}} > 7e{-}11$ and $\mathcal{J}_{\text{div}} > 6e{-}8$, respectively. It should be noted that, given the dataset's step size of $6e{-}4$, the retraction map would ideally be activated at $\mathcal{J}_{\text{div}} > 6e{-}4$. While the error would inevitably surpass the $6e{-}4$ threshold with continued application of g_{ϵ^i}, this experiment primarily demonstrates the necessity of the retraction map when the truncated expansion in (4.38) is employed.

Table 5.5: The quality of testing the retraction map in case $\boldsymbol{\mathcal{D}}_{\text{te}}$ and ICs are set to (5.12) and (5.3), respectively (cf. Table 5.3).

Idx	$\mathcal{J}_{\text{sym}}$	$\mathcal{J}_{\text{div}}$
1	4.815e-13	7.413e-14
2	1.017e-12	4.289e-13
3	4.088e-13	8.009e-14
4	9.067e-13	2.001e-13
5	8.414e-13	5.386e-13
6	9.017e-13	6.588e-13
7	2.696e-13	5.002e-14
8	7.587e-13	3.087e-14
9	1.0917e-12	1.314e-13
10	7.364e-13	3.145e-13
11	6.009e-12	9.877e-13

Table 5.6: The quality of testing the retraction map in case $\mathcal{D}_{\text{te}}$ and ICs are set to (5.5) and (5.14), respectively (cf. Table 5.4).

Idx	$\mathcal{J}_{\text{sym}}$	$\mathcal{J}_{\text{div}}$
1	4.833e-11	7.211e-12
2	3.677e-10	6.084e-11
3	1.011e-11	3.269e-12
4	3.199e-10	5.568e-11
5	3.077e-11	4.996e-11
6	8.099e-11	3.069e-11
7	1.119e-11	8.917e-12
8	1.737e-11	6.556e-12
9	9.316e-11	2.709e-11
10	5.333e-11	9.419e-11
11	8.118e-10	1.116e-10

CHAPTER 6

CONCLUSION AND REMARKS

This study aimed to develop a neural framework capable of learning non-trivial, contact-preserving symmetries of a DE system, while laying the groundwork for constructing G-invariant coframes in future research. Unlike computational algebraic methods, our approach does not seek to approximate solutions of determining equations using neural networks. First, algebraic methods fail to distinguish non-trivial symmetries from trivial ones. Second, the proposed framework extracts symmetries of interest directly from input datasets. Third, algebraic approaches require the explicit closed form of the dynamics and are thus incompatible with analytical models, such as the one introduced in [19]. The framework in [19] is capable of approximating governing equations (i.e., the DE system $\Delta^{(s)} = 0$) in an analytic form by training on data collected from the dynamical system of interest. Therefore, neural frameworks like ours could integrate with such architectures to simultaneously discover the DE system and its non-trivial, contact-preserving symmetries.

A significant limitation of computational algebraic methods is their dependency on the polynomial structure of DE systems, where these systems are viewed as elements of differential modules over rings of polynomials in jet space coordinates. Unlike these methods, our framework avoids producing an **intractable** number of **non-linear** PDEs requiring symbolic manipulation and resolution. Furthermore, in the absence of a polynomial structure, it is unclear whether the system satisfies necessary closure conditions for Lie symmetry analysis when algebraic methods are employed. In contrast, our analytical approach leverages the computational graph of neural networks, making it capable of handling non-polynomial DE systems when cost functions, as outlined in Section 4, are considered.

Algebraic methods traditionally focus on solving the involutivity problem (see Section 3.5) by constructing integral submanifolds, which represent solutions to the PDE system. However, prior knowledge of non-trivial symmetries—those that preserve contact structure and map one solution to a fundamentally distinct one—can provide significant advantages. Such symmetries often can be studied independently of the full contact system, with their involutivity conditions governed by Frobenius' theorem (Theorem 4), which is **computationally** simpler than the constraints imposed by the Cartan–Kähler theorem (see Chapter 15 of [10]).

Another objective of this study was to systematically introduce the symmetries of DE systems to the machine learning community. In Sections 1 and 3, we provided the necessary tools from differential geometry and algebra to facilitate this. As discussed in Section 2, recent studies have explored these symmetries; however, their goals were not to discover symmetries through training on datasets derived from dynamical systems, but rather to apply existing symmetries to specific problems. For example, [36] is among the first studies proposing a framework to learn symmetries within a subset of $GL(n, \mathbb{R})$. However, our focus on Lie point symmetries encompasses $GL(n, \mathbb{R})$, making our scope broader. Furthermore, unlike our study, [36] does not present a learning-based architecture, focusing solely on theoretical development. Their framework targets DE systems defined in linear subspaces and does not operate within the jet space framework, which might restrict their ability to learn more general classes of symmetries, such as Lie contact symmetries and Lie pseudogroups, in future studies.

Our framework also facilitates the construction of the associated G-coframe (as in [13]), which consists of invariant and covariant directions with respect to G in a given jet space. The invariant subspace spanned by these directions plays a crucial role in simplifying DE systems. For instance, such simplifications may lead to dimensionality reduction or decoupling (see the example in Section 1.1) or to order reduction (as in Example 2.65 of [8]). From a statistical perspective, this process is akin to principal component analysis (PCA), as it identifies directions irrelevant to particular solutions. For example, PINNs [27] could benefit from our framework by improving the accuracy of analytically approximated solutions and reducing computational complexity.

Additionally, as noted in Remark 48, physicists and engineers may query their datasets for specific symmetries of interest. Since our framework learns symmetries in the Lie algebra $\bar{\mathfrak{g}}$ without assigning intrinsic geometric meaning (e.g., rotation, translation, inversion), users can test for specific geometric symmetries relevant to their theoretical models and simplify these models accordingly.

A compelling application of our framework lies within generative models. As demonstrated in [58], the Liouville PDE serves as the *mathematical dual* of the log-likelihood function in normalizing flow and diffusion approaches [59, 60]. Discovering non-trivial symmetries of this highly non-linear PDE could facilitate searching for flows in lower-dimensional submanifolds, tailored to specific applications.

This work thus lays the foundation for future exploration of symmetries, bridging the gap between traditional differential geometry and modern machine learning.

REFERENCES

[1] B. Hall, *Lie groups, Lie algebras, and representations: an elementary introduction*, vol. 222. Springer, 2015.

[2] P. Seshu, *Textbook of finite element analysis*. PHI Learning Pvt. Ltd., 2003.

[3] J. W. Thomas, *Numerical partial differential equations: finite difference methods*, vol. 22. Springer Science & Business Media, 2013.

[4] M. T. Heath, "Scientific computing: An introductory survey," *McGraw-Hill*, 1997.

[5] S. Mallat, "Understanding deep convolutional networks," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150203, 2016.

[6] S. Sinha, S. P. Nandanoori, and E. Yeung, "Koopman operator methods for global phase space exploration of equivariant dynamical systems," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 1150–1155, 2020.

[7] D. J. Rezende, S. Racanière, I. Higgins, and P. Toth, "Equivariant hamiltonian flows," *arXiv preprint arXiv:1909.13739*, 2019.

[8] P. J. Olver, *Applications of Lie groups to differential equations*, vol. 107. Springer Science & Business Media, 2000.

[9] G. Bluman and S. Anco, *Symmetry and integration methods for differential equations*, vol. 154. Springer Science & Business Media, 2008.

[10] P. J. Olver, *Equivalence, invariants and symmetry*. Cambridge University Press, 1995.

[11] J. Starrett, "Solving differential equations by symmetry groups," *The American Mathematical Monthly*, vol. 114, no. 9, pp. 778–792, 2007.

[12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[13] M. Fels and P. J. Olver, "Moving coframes: Ii. regularization and theoretical foundations," *Acta Applicandae Mathematica*, vol. 55, pp. 127–208, 1999.

[14] T. A. Ivey and J. M. Landsberg, *Cartan for beginners: differential geometry via moving frames and exterior differential systems*, vol. 61. American Mathematical Society Providence, RI, 2003.

[15] R. B. Gardner, *The method of equivalence and its applications*. SIAM, 1989.

[16] W. E. Boyce and R. C. DiPrima, *Elementary differential equations and boundary value problems*. John Wiley & Sons, 2009.

[17] G. J. Reid, "Algorithms for reducing a system of pdes to standard form, determining the dimension of its solution space and calculating its taylor series solution," *European Journal of Applied Mathematics*, vol. 2, no. 4, pp. 293–318, 1991.

[18] MATLAB, *9.13.0.2145394 (R2022b) Update 3*. Natick, Massachusetts: The MathWorks Inc., 2022.

[19] R. T. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, "Neural ordinary differential equations," *arXiv preprint arXiv:1806.07366*, 2018.

[20] E. Quinlan-Gallego, "Rings of differential operators,"

[21] F. Schwarz, "An algorithm for determining the size of symmetry groups," *Computing*, vol. 49, no. 2, pp. 95–115, 1992.

[22] K. Iohara and P. Malbos, "Maurice janet's algorithms on systems of linear partial differential equations," *Archive for History of Exact Sciences*, vol. 75, pp. 43–81, 2021.

[23] W. M. Seiler, *Involution: The Formal Theory of Differential Equations and Its Applications in Computer Algebra*, vol. 24 of *Algorithms and Computation in Mathematics*. Berlin, Heidelberg: Springer, 2010.

[24] D. Hartley and R. W. Tucker, "A constructive implementation of the cartan-kähler theory of exterior differential systems," *J. Symb. Comput.*, vol. 12, no. 6, pp. 655–668, 1991.

[25] D. Hartley, R. W. Tucker, and P. A. Tuckey, "Constrained dynamics and exterior differential systems," *Journal of Physics A: Mathematical and General*, vol. 24, no. 22, p. 5253, 1991.

[26] N. H. Ibragimov, *CRC handbook of Lie group analysis of differential equations*, vol. 3. CRC press, 1995.

[27] 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," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.

[28] B. O. Koopman, "Hamiltonian systems and transformation in hilbert space," *Proceedings of the national academy of sciences of the united states of America*, vol. 17, no. 5, p. 315, 1931.

[29] P. Kidger and T. Lyons, "Universal approximation with deep narrow networks," in *Conference on learning theory*, pp. 2306–2327, PMLR, 2020.

[30] S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz, "Modern koopman theory for dynamical systems," *arXiv preprint arXiv:2102.12086*, 2021.

[31] M. Budišić, R. Mohr, and I. Mezić, "Applied koopmanism," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 4, p. 047510, 2012.

[32] D. Floryan and M. D. Graham, "Data-driven discovery of intrinsic dynamics," *Nature Machine Intelligence*, pp. 1–8, 2022.

[33] G. Mialon, Q. Garrido, H. Lawrence, D. Rehman, Y. LeCun, and B. Kiani, "Self-supervised learning with lie symmetries for partial differential equations," *Advances in Neural Information Processing Systems*, vol. 36, pp. 28973–29004, 2023.

[34] T. Akhound-Sadegh, L. Perreault-Levasseur, J. Brandstetter, M. Welling, and

S. Ravanbakhsh, "Lie point symmetry and physics-informed networks," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[35] M. Finzi, M. Welling, and A. G. Wilson, "A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups," in *International conference on machine learning*, pp. 3318–3328, PMLR, 2021.

[36] S. E. Otto, N. Zolman, J. N. Kutz, and S. L. Brunton, "A unified framework to enforce, discover, and promote symmetry in machine learning," *arXiv preprint arXiv:2311.00212*, 2023.

[37] S. Hassani, *Mathematical physics: a modern introduction to its foundations*. Springer Science & Business Media, 2013.

[38] P. Renteln, *Manifolds, tensors, and forms: an introduction for mathematicians and physicists*. Cambridge University Press, 2013.

[39] J. M. Lee, "Smooth manifolds," in *Introduction to smooth manifolds*, pp. 1–31, Springer, 2013.

[40] A. J. Parzygnat, "Math 3151 analysis ii, spring 2017,"

[41] J. Behrmann, W. Grathwohl, R. T. Chen, D. Duvenaud, and J.-H. Jacobsen, "Invertible residual networks," in *International Conference on Machine Learning*, pp. 573–582, PMLR, 2019.

[42] V. S. Varadarajan, *Lie groups, Lie algebras, and their representations*, vol. 102. Springer Science & Business Media, 2013.

[43] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*. Academic press, 1979.

[44] H. Liu, "Hopf fibration," *Department of Mathematics, Cornell University*, p. 3, 2011.

[45] M. Fels and P. J. Olver, "Moving coframes: I. a practical algorithm," *Acta Applicandae Mathematica*, vol. 51, pp. 161–213, 1998.

[46] A. C. Hearn and R. Schöpf, *Reduce Users' Guide, Free Manual*. Sourceforge, March 29, 2024 ed., 2024.

[47] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry, Volume I*. No. 1. c. in Foundations of Differential Geometry [by] Shoshichi Kobayashi and Katsumi Nomizu, Interscience Publishers, 1963.

[48] P. J. Olver and J. Pohjanpelto, "Moving frames for Lie pseudo–groups," *Canadian Journal of Mathematics*, vol. 60, no. 6, p. 1336–1386, 2008.

[49] P. J. Olver and J. Pohjanpelto, "Maurer–cartan forms and the structure of lie pseudo-groups," *Selecta Mathematica*, vol. 11, pp. 99–126, 2005.

[50] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), pp. 8024–8035, Curran Associates, Inc., 2019.

[51] A. Meurer, C. P. Smith, M. Paprocki, O. Čertík, S. B. Kirpichev, M. Rocklin, A. Kumar, S. Ivanov, J. K. Moore, S. Singh, T. Rathnayake, S. Vig, B. E. Granger, R. P. Muller, F. Bonazzi, H. Gupta, S. Vats, F. Johansson, F. Pedregosa, M. J. Curry, A. R. Terrel, Š. Roučka, A. Saboo, I. Fernando, S. Kulal, R. Cimrman, and A. Scopatz, "Sympy: Symbolic computing in python," *PeerJ Computer Science*, vol. 3, p. e103, 2017.

[52] H. Anton and R. C. Busby, *Contemporary linear algebra*. Wiley, 2003.

[53] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[54] H. Kervadec, J. Dolz, J. Yuan, C. Desrosiers, E. Granger, and I. B. Ayed, "Log-barrier constrained cnns," *Computing Research Repository (CoRR)*, 2019.

[55] M. Baer, "findiff software package," 2018.

[56] R. Haberman, *Elementary applied partial differential equations*, vol. 987. Prentice Hall Englewood Cliffs, NJ, 1983.

[57] A. T. Jassim, “Symmetry for initial boundary value problems of pdes,” *Journal of the College of Basic Education*, vol. 25, no. 105, pp. 178–194, 2019.

[58] E. G. Tabak and E. Vanden-Eijnden, “Density estimation by dual ascent of the log-likelihood,” 2010.

[59] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, “Normalizing flows for probabilistic modeling and inference,” *Journal of Machine Learning Research*, vol. 22, no. 57, pp. 1–64, 2021.

[60] Z. Chang, G. A. Koulieris, and H. P. Shum, “On the design fundamentals of diffusion models: A survey,” *arXiv preprint arXiv:2306.04542*, 2023.

APPENDICES

A Proofs

A.1 Theorem 1

The normal coordinate map $\mathbf{\Phi}$ in (3.11) clearly satisfies:

$$\mathbf{\Phi}(\epsilon^j v_j) = \exp(\epsilon^j v_j), \tag{A.1}$$

where $\epsilon^j \in \mathbb{R}$. Similar to (3.13):

$$\frac{d}{d\epsilon} \exp(\epsilon v)|_{\epsilon=0} = v \in \mathfrak{g}. \tag{A.2}$$

Therefore, using the definition of differential:

$$d\mathbf{\Phi}[v_j]\,|_{\boldsymbol{\epsilon}=0} = v_j \tag{A.3}$$

for all $j = 1, \ldots, r$, so that $d\mathbf{\Phi}|_{\boldsymbol{\epsilon}=0} = \mathbb{1}_{\mathfrak{g}}$.

Consequently, by the Implicit Function Theorem (see Theorem 10.26 in [40]), $\mathbf{\Phi} : V \to U$ defines a local diffeomorphism from an open neighborhood $0 \in V \subset \mathfrak{g}$ to a neighborhood $e \in U \subset G$, so that every $g_{\boldsymbol{\epsilon}} \in U$ can be uniquely defined written as in (3.11).

One now uses Proposition 1.24 in [8] to prove (3.12) (see Corollary 3.47 in [1] for its proof).

□

A.2 Proof of Proposition 2

We aim to show that the connection one-forms $\boldsymbol{\eta}_U$ together with the lifted one-forms $\boldsymbol{\omega}$ provide a basis for the cotangent bundle of the localized principal bundle, i.e.

$T^*\mathscr{P} = T^*U \oplus T^*G$. This follows from the properties of the M-C forms and the connection one-forms, as detailed below.

First, recall the definition of the lifted one-forms $\boldsymbol{\omega}$ given in (3.59), which expresses their dependence on the local trivialization:

$$\omega^i|_{(z,\tau_U)} = \left(\tau_U(z)^{-1}\right)^i_j \, \omega^j_U|_z. \tag{A.4}$$

Since $\tau_U(z)^{-1} \in G$, the set of forms $\boldsymbol{\omega} = \{\omega_1, \ldots, \omega_m\}$ constitutes a basis for T^*U. Consequently, the key remaining step is to establish that the connection one-forms $\boldsymbol{\eta}_U$ span T^*G. By Definition 50, the components ${\eta_U}^i_k$ define a linear map $\boldsymbol{\eta}_U : T\mathscr{P} = TU \oplus TG \to T_eG$. Furthermore, by Definition 51, $\boldsymbol{\eta}_U$ annihilates all vectors $\hat{v} \in TU$, implying that it acts nontrivially only on elements of TG.

Now, recall from Definition 47 that the M-C forms $\boldsymbol{\mu}$ satisfy

$$\boldsymbol{\mu}|_g : T_gG \to T_eG. \tag{A.5}$$

Comparing this with the action of $\boldsymbol{\eta}_U$ on $T\mathscr{P}$, we observe that $\boldsymbol{\eta}_U$ behaves analogously to $\boldsymbol{\mu}$ when restricted to TG (see also Definition 8.3.1 in [14]). Since $\boldsymbol{\mu}$ is known to form a basis for T^*G, it follows that $\boldsymbol{\eta}_U$ spans T^*G within $T^*\mathscr{P}$.

Thus, combining the basis $\boldsymbol{\omega}$ for T^*U with the basis $\boldsymbol{\eta}_U$ for T^*G, we conclude that $\{\boldsymbol{\omega}, \boldsymbol{\eta}_U\}$ provides a basis for the cotangent bundle of the principal bundle $\mathscr{P}$, yielding the desired decomposition:

$$T^*\mathscr{P} = T^*U \oplus T^*G. \tag{A.6}$$

This result is consistent with Definition 8.3.1 in [14], which establishes a natural splitting of the cotangent bundle in terms of a principal bundle structure.

□

A.3 Proposition 4

Recall the **regularity** condition imposed on the map $\Delta^{(k)} : J^{(k)}(M) \to \mathbb{R}^{n_k}$, as discussed in Section 3, where $n_k \leq m_k = \dim(J^{(k)}(M))$, and n_k represents the total number of DEs generated by the prolongation procedure in (4.33) across all

orders $s_0, \ldots, k$, with s_0 being the original order of the system. Consequently, the submanifold $\Sigma^{(k)}$, defined by the zero set of $\Delta^{(k)}$, is a regular submanifold of constant rank $m_k - n_k + r_k$, embedded (or immersed, though the distinction is irrelevant for our discussion) into $J^{(k)}(M)$, where r_k denotes the total number of redundant differential consequences as defined in Definition 59.

The regularity condition has significant implications for the higher-order derivatives of $\Delta^{(k)}$ when restricted to $\Sigma^{(k)}$. This effect, utilized in the proof, is demonstrated constructively as follows. Let $\mathrm{Jac}(\Delta^{(k)})$ denote the Jacobian matrix of $\Delta^{(k)}$. Then, the tangent space to $\Sigma^{(k)}$ at a point $z^{(k)} \in \Sigma^{(k)}$ is given by $T_{z^{(k)}}\Sigma^{(k)} = \ker\left(\mathrm{Jac}(\Delta^{(k)})|_{z^{(k)}}\right)$.

Since $\Delta^{(k)}$ is regular and $\Delta^{(k)}(z^{(k)}) = 0$, the following constraints hold:

$$\begin{aligned} \frac{\partial \Delta^{(k),l_k}}{\partial z^{(k),i}}(z^{(k)}) &= 0 \\ \frac{\partial^2 \Delta^{(k),l_k}}{\partial z^{(k),i_2} \partial z^{(k),i_1}}(z^{(k)}) &= 0, \end{aligned} \tag{A.7}$$

where $z^{(k)} \in \Sigma^{(k)}$, $l_k = 1, \ldots, n_k$, and $i, i_1, i_2 = 1, \ldots, m_k$. Fixing indices l_k and i_1, define the vector

$$\hat{v}_{i_1}^{l_k} := \sum_{i_2} \frac{\partial^2 \Delta^{(k),l_k}}{\partial z^{(k),i_2} \partial z^{(k),i_1}} \frac{\partial}{\partial z^{(k),i_2}} \in TJ^{(k)}(M). \tag{A.8}$$

By (A.7), we find that $\hat{v}_{i_1}^{l_k}|_{\Sigma^{(k)}} = 0$, meaning $\hat{v}_{i_1}^{l_k} \in \mathrm{Span}(\mathrm{Jac}(\Delta^{(k)})^{\mathrm{t}}|_{\Sigma^{(k)}})$. Therefore, appending $\hat{v}_{i_1}^{l_k}$ as a row to $\mathrm{Jac}(\Delta^{(k)})$ does not alter its rank.

Having established this critical property of $\mathrm{Jac}(\Delta^{(k)})$, we now proceed to the proof of the proposition. By Definition 59, any redundant $\Delta^{(k),l_k}$ is expressed as

$$\Delta^{(k),l_k} = \sum_{l_{k-1}} \sum_{J} f_J^{l_{k-1}} \frac{\partial^{|J|} \Delta^{(k-1),l_{k-1}}}{\partial_J z^{(k-1)}}, \tag{A.9}$$

where $\partial^{|J|}/\partial_J z^{(k-1)}$ denotes $\partial^{|J|}/\partial z^{(k-1),j_1} \ldots \partial z^{(k-1),j_{|J|}}$, $l_k \in \{n_{k-1}+1, \ldots, n_k\}$ (i.e., a newly emerged differential consequence due to the prolongation in (4.33)), $f_J^{l_{k-1}} \in C^\infty(J^{(k-1)}(M))$ are coefficient functions, $l_{k-1} = 1, \ldots, n_{k-1}$, $z^{(k-1)}$ are jet coordinates on $J^{(k-1)}(M)$, J is a multi-index on jet coordinates, and $|J|$ is bounded by some finite $N \in \mathbb{N}$.

The entries in the row of $\mathrm{Jac}(\Delta^{(k)})$ corresponding to $\Delta^{(k),l_k}$ take the form

$$\sum_{l_{k-1}} \sum_{J} f_J^{l_{k-1}} \frac{\partial^{|J|+1} \Delta^{(k-1),l_{k-1}}}{\partial z^{(k),i} \partial_J z^{(k-1)}} + \frac{\partial f_J^{l_{k-1}}}{\partial z^{(k),i}} \frac{\partial^{|J|} \Delta^{(k-1),l_{k-1}}}{\partial_J z^{(k-1)}} \tag{A.10}$$

where $i = 1, \ldots, m_k$. Note that the partial derivatives with respect to the fiber coordinates $(z^{(k),m_{k-1}+1}, \ldots, z^{(k),m_k})$ are all zero, as $\Delta^{(k),l_k}$ does not depend on the fibers of $J^{(k)}(M)$ over $J^{(k-1)}(M)$.

Since all lower-order differential consequences $\Delta^{(k-1)}$ are already represented in the relevant rows of $\mathrm{Jac}(\Delta^{(k)})$, specifically as

$$\left[\frac{\partial \Delta^{(k-1),l_{k-1}}}{\partial z^{(k),1}}, \ldots, \frac{\partial \Delta^{(k-1),l_{k-1}}}{\partial z^{(k),m_k}}\right] \tag{A.11}$$

for all $l_{k-1} = 1, \ldots, n_{k-1}$, the row containing the entries in (A.10), when evaluated at $z^{(k)} \in \Sigma^{(k)}$, remains in $\mathrm{Span}(\mathrm{Jac}(\Delta^{(k-1)})^{\mathrm{t}}|_{z^{(k)}})$, thanks to the discussion about $\mathrm{Jac}(\Delta^{(k)})$ in the beginning of the proof.

□

A.4 Proposition 5

Theorem 1 in [41] states that if the operator norm of the iResNET block g_t satisfies $\|g_t\|_{\mathrm{op}} < 1$, then the transformation given in (4.38) is invertible by *Banach fixed-point iteration* theorem.

Let $z_1^{(s)}$ and $z_2^{(s)}$ be two sufficiently close points in $J^{(s)}(M)$. Then:

$$\begin{aligned}\left\| \exp(\epsilon v) \cdot z_2^{(s)} - \exp(\epsilon v) \cdot z_1^{(s)} \right\| \leq \\ \left\| z_2^{(s)} - z_1^{(s)} \right\| + \epsilon \left\| \hat{v}^{(s)}|_{z_2^{(s)}} - \hat{v}^{(s)}|_{z_1^{(s)}} \right\| + \\ \frac{\epsilon^2}{2} \left\| (\hat{v}^{(s)})^2|_{z_2^{(s)}} - (\hat{v}^{(s)})^2|_{z_1^{(s)}} \right\| + O(\epsilon^3),\end{aligned} \tag{A.12}$$

where $\hat{v}^{(s)} \coloneqq \mathrm{pr}^{(s)}(\hat{v})$, $\|\cdot\|$ is ℓ_2 norm, and, without loss of generality, $\epsilon \in \mathbb{R}^+$. We may treat $\hat{v}^{(s)} : J^{(s)}(M) \to TJ^{(s)}(M)$ as an operator. Therefore, defining *Lipschitz constant* for $\hat{v}^{(s)}$ makes sense and it signifies how "smoothly" $\hat{v}^{(s)}$ varies over $J^{(s)}(M)$:

$$\left\| \hat{v}^{(s)}|_{z_2^{(s)}} - \hat{v}^{(s)}|_{z_1^{(s)}} \right\| \leq K \left\| z_2^{(s)} - z_1^{(s)} \right\|, \tag{A.13}$$

where $K \in \mathbb{R}^+$ is the Lipschitz constant for $\hat{v}^{(s)}$. Therefore:

$$\begin{aligned}\Big\|(\hat{v}^{(s)})^n\big|_{z_2^{(s)}} - (\hat{v}^{(s)})^2\big|_{z_1^{(s)}}\Big\| \leq \\ K\Big\|(\hat{v}^{(s)})^{n-1}\big|_{z_2^{(s)}} - (\hat{v}^{(s)})^{n-1}\big|_{z_1^{(s)}}\Big\| \leq \ldots \leq \\ K^n\Big\|z_2^{(s)} - z_1^{(s)}\Big\|, \quad \text{(A.14)}\end{aligned}$$

for all $n \geq 1$. Substitute (A.14) into (A.12):

$$\begin{aligned}\Big\| \exp(\epsilon v) \cdot z_2^{(s)} - \exp(\epsilon v) \cdot z_1^{(s)}\Big\| \leq \\ \Big(1 + \epsilon K + \frac{1}{2}(\epsilon K)^2 + \ldots\Big)\Big\|z_2^{(s)} - z_1^{(s)}\Big\| = \\ e^{\epsilon K}\Big\|z_2^{(s)} - z_1^{(s)}\Big\|. \quad \text{(A.15)}\end{aligned}$$

Therefore, $L = e^{\epsilon K}$ is the Lipschitz constant for $\exp$ map. Since Lipschitz constant could be considered as an upper bound on $\|\cdot\|_{\text{op}}$, the cap on $\|\exp\|_{\text{op}}$ depends on the Lipschitz constant of $\hat{v}$. The iResNET block g_t represents the transformation $[\exp(\epsilon v) - e]$. Thus, $\|g_t\|_{\text{op}} \leq L - 1$. In accordance with Theorem 1 in [41]:

$$\|g_t\|_{\text{op}} \leq L - 1 < 1 \quad \text{(A.16)}$$

and hence, $L < 2$ implies $K < {}^{\ln 2}/_{\epsilon}$.

□

B Pfaffian Characterization of the Wave Equation in Two Spatial Dimensions

This section establishes that the DE system, utilized throughout this study, is of Pfaffian type (see Section 3.5). The significance of this classification lies in its implication that $\Delta^{(2)} = 0$, as given in (5.1), possesses a simpler contact structure compared to non-Pfaffian systems, particularly regarding the satisfaction of involutivity conditions. Furthermore, for the interested reader, we demonstrate how a DE system can be reformulated as a contact system.

The two-dimensional wave equation (5.1) in its standard form is not directly suitable for verifying the inclusion $d\theta_i \in \mathcal{I}_{\text{diff}}$. Without loss of generality, let $c = 1$. To render the equation more tractable, we introduce the characteristic coordinates $\xi = x + t$ and $\eta = x - t$. The corresponding first- and second-order partial derivatives in terms of (ξ, η) are:

$$\begin{aligned}
\frac{\partial}{\partial x} &= \frac{\partial}{\partial \xi} + \frac{\partial}{\partial \eta}, \\
\frac{\partial}{\partial t} &= \frac{\partial}{\partial \xi} - \frac{\partial}{\partial \eta}, \\
\frac{\partial^2}{\partial x^2} &= \frac{\partial^2}{\partial \xi^2} + 2\frac{\partial^2}{\partial \xi \partial \eta} + \frac{\partial^2}{\partial \eta^2}, \\
\frac{\partial^2}{\partial t^2} &= \frac{\partial^2}{\partial \xi^2} - 2\frac{\partial^2}{\partial \xi \partial \eta} + \frac{\partial^2}{\partial \eta^2}.
\end{aligned} \tag{B.1}$$

Using (B.1) alongside $\partial/\partial y$ and $\partial^2/\partial y^2$, the wave equation (5.1) transforms into:

$$\Delta^{(2)} = \frac{\partial^2 u}{\partial \xi \partial \eta} + \frac{1}{4}\frac{\partial^2 u}{\partial y^2} = 0, \tag{B.2}$$

where $u = \bar{u}(\xi, \eta, y)$.

The associated EDS for (B.2) consists of the following one-forms:

(i) First-order consistency condition:

$$\theta_1 = du - \sum_{i=1}^{3} p_i \, dx^i. \tag{B.3}$$

where $p_i = \frac{\partial u}{\partial x^i}$ and $(x^1, x^2, x^3) \coloneqq (\xi, \eta, y)$.

(ii) Second-order consistency conditions:

$$\theta_{i+1} = dp_i - \sum_{j=1}^{3} q_{ij}\, dx^j, \tag{B.4}$$

for $i = 1, 2, 3$, where $q_{ij} = \frac{\partial p_i}{\partial x^j}$.

(iii) Constraint imposed by the DE system:

From (B.2), we find $q_{12} = -\frac{1}{4}q_{33}$. Thus, the additional condition is:

$$\theta_5 = dq_{12} + \frac{1}{4}dq_{33}. \tag{B.5}$$

Consequently, the EDS for (B.2) is given by:

$$\begin{aligned} \Delta^{(2)}(x^1, x^2, x^3, u^{(2)}) &= 0, \\ \mathcal{I}_{\text{diff}} &= \{\theta_1, \ldots, \theta_5\}, \\ \alpha &= dx^1 \wedge dx^2 \wedge dx^3. \end{aligned} \tag{B.6}$$

First, we verify the independence condition $\alpha \neq 0$:

$$\begin{aligned} dx^1 \wedge dx^2 \wedge dx^3 &= d(x+t) \wedge d(x-t) \wedge dy \\ &= dx \wedge dy \wedge dt + dt \wedge dx \wedge dy \\ &= 2dx \wedge dy \wedge dt \neq 0. \end{aligned} \tag{B.7}$$

Thus, α ensures independence on the base manifold M.

To verify the Pfaffian criterion for every $\theta_i \in \mathcal{I}_{\text{diff}}$, we do not rely on the *geometric* condition in (3.42). Instead, we employ the *algebraic* condition $d\theta_i \in \mathcal{I}_{\text{diff}}$, which is both sufficient and computationally simpler for the case of (B.2). Specifically, we note that $d\theta_5 = 0 \in \mathcal{I}_{\text{diff}}$, while for $d\theta_1$, we have $d\theta_1 = -\sum_j dp_j \wedge dx^j$. Observe that:

$$-\sum_{j=1}^{3}(\theta_{j+1}) \wedge dx^j = \underbrace{-\sum_{j=1}^{3} dp_j \wedge dx^j}_{d\theta_1} + \underbrace{\sum_{\substack{i_j=1,\\ j=1}}^{3} q_{ji_j} dx^{i_j} \wedge dx^j}_{=0}. \tag{B.8}$$

The second term on the right-hand side vanishes because, for each pair $(q_{ji_j}\, dx^{i_j} \wedge dx^j, q_{i_jj}\, dx^j \wedge dx^{i_j})$, the terms cancel each other. Consequently, $d\theta_1 \in \mathcal{I}_{\text{diff}}$.

Computing $d\theta_{i+1} \in \mathcal{I}_{\text{diff}}$ for $i = 1, 2, 3$ is more intricate compared to the case of θ_1. Let

$$d\theta_{i+1} = -\sum_{j=1}^{3} dq_{ij} \wedge dx^j, \tag{B.9}$$

where $i = 1, 2, 3$. Note that $dq_{ij} = d\big({}^{\partial p_i}/_{\partial x^j}\big) = {}^{\partial}/_{\partial x^j}(dp_i)$. Since $p_i \in J^{(1)}(M)$, we have:

$$dp_i \coloneqq \sum_{k=1}^{3} \frac{\partial p_i}{\partial x^k}\, dx^k + \frac{\partial p_i}{\partial u}\, du, \tag{B.10}$$

where terms involving partial derivatives with respect to p_l (for $l = 1, 2, 3$) are omitted because they vanish when ${}^{\partial}/_{\partial x^j}$ is applied. Substituting this into (B.9), we obtain:

$$\begin{aligned} d\theta_{i+1} &= \underbrace{-\sum_{j=1}^{3}\sum_{k=1}^{3} \frac{\partial^2 p_i}{\partial x^j \partial x^k}\, dx^k \wedge dx^j}_{=0} \qquad -\sum_{j=1}^{3} \frac{\partial^2 p_i}{\partial x^j \partial u}\, du \wedge dx^j \\ &= -\sum_{j=1}^{3} \frac{\partial^2 p_i}{\partial x^j \partial u}\, du \wedge dx^j. \end{aligned} \tag{B.11}$$

The first term vanishes due to the same reasoning as in (B.8). Next, we introduce the expression:

$$\frac{\partial^2 p_i}{\partial x^j \partial u} \theta_1 \wedge dx^j \in \mathcal{I}_{\text{diff}} \tag{B.12}$$

which holds because $f\theta \wedge \omega \in \mathcal{I}_{\text{diff}}$ for all $f \in C^\infty(J^{(s)}(M))$ and $\omega \in \Omega^{(s),k}$, where $\theta \in \mathcal{I}_{\text{diff}}$. This leads to:

$$d\theta_{i+1} = \underbrace{-\sum_{j=1}^{3} \frac{\partial^2 p_i}{\partial x^j \partial u} \theta_1 \wedge dx^j}_{\in \mathcal{I}_{\text{diff}}} \underbrace{-\sum_{j=1}^{3}\sum_{\substack{k=1 \\ k\neq j}}^{3} \frac{\partial^2 p_i}{\partial x^j \partial u} p_k dx^k \wedge dx^j}_{(*)}. \tag{B.13}$$

Thus, the final step is to prove whether $(*) \in \mathcal{I}_{\text{diff}}$. Using the substitution:

$$p_k dx^k = du - \theta_1 - \sum_{\substack{l=1 \\ l\neq k}}^{3} p_l dx^l \tag{B.14}$$

and applying it to $(*)$ in (B.13), we have:

$$\begin{aligned} \sum_{j=1}^{3}\sum_{\substack{k=1 \\ k\neq j}}^{3} \frac{\partial^2 p_i}{\partial x^j \partial u}\left(du - \theta_1 - \sum_{\substack{l=1 \\ l\neq k}}^{3} p_l dx^l\right) \wedge dx^j = \quad & \sum_{j=1}^{3} \frac{\partial^2 p_i}{\partial x^j \partial u} \theta_1 \wedge dx^j + \\ & \sum_{j=1}^{3}\sum_{\substack{k=1 \\ k\neq j}}^{3} \big(du - \sum_{\substack{l=1 \\ l\neq k}}^{3} p_l dx^l\big) \wedge dx^j. \end{aligned} \tag{B.15}$$

The second term on the right-hand side can be safely replaced with $\sum_{j=1}^{3} \theta_1 \wedge dx^j$, since terms such as $dx^l \wedge dx^j = 0$ when $j = l$. Thus, $(*)$ in (B.13) simplifies to:

$$\sum_{j=1}^{3} \left(\frac{\partial^2 p_i}{\partial x^j \partial u} + 1 \right) \theta_1 \wedge dx^j \in \mathcal{I}_{\text{diff}}. \tag{B.16}$$

Combining (B.13) and (B.16), we conclude that $d\theta_{i+1} \in \mathcal{I}_{\text{diff}}$ for all $i = 1, 2, 3$.

□

C Table of Symbols

Symbol	Description
G, H	Lie group and its subgroup
$GL(r, \mathbb{R})$	General linear group of degree r over $\mathbb{R}$
M, N	Manifold (or submanifold) on which G acts
$J^{(s)}(M)$	Jet space of order s over the base manifold M
$\boldsymbol{S}_{\boldsymbol{f}}^{(s)}$	Integral submanifold in $J^{(s)}(M)$, associated with solution $\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x})$
$\mathcal{V}$	System of vector fields
$\mathfrak{g}$, $\mathfrak{h}$	Lie algebra and its subalgebra
TG, TM	Tangent bundles of G and M, respectively
T^*G, T^*M	Cotangent bundles of G and M, respectively
$\mathfrak{X}(M)$	Space of vector fields on M
$\Omega^l(M)$	Space of l-forms on M
$\Omega^{(s),l}(M)$	Space of l-forms on $J^{(s)}(M)$
$\mathcal{C}^{(s)}(M)$	Space of contact forms of order s in $T^*J^{(s)}(M)$
$\mathcal{H}(M)$	Space of horizontal one-forms in $T^*J^{(s)}(M)$
$\mathcal{I}_{\text{alg}}$	Algebraic ideal of l-forms
$\mathcal{I}_{\text{diff}}$	Differential ideal of l-forms
x^i	i^{th} local independent coordinate on M
I, I_k	Multi-index notation for independent coordinates, e.g., $I \equiv I_k = (x^{i_1}, \ldots, x^{i_k})$ of length k
$\lvert I \rvert$	Length (order) of the multi-index I
u^α	α^{th} local dependent coordinate on M

Continued on next page

Continued from previous page

Symbol	Description
$u^\alpha_{x^{I_k}}$, u^α_I	k^{th}-order partial derivative of u^α with respect to x^{I_k}
p	Number of independent coordinates $\boldsymbol{x}$
q	Number of independent coordinates $\boldsymbol{u}$
z	Coordinate in the base manifold, assuming $M = \mathcal{X} \times \mathcal{U}$
$z^{(s)}$	Coordinate in $J^{(s)}(M)$
$C^\infty(M)$	Space of smooth functions on M
$\mathcal{X}, \mathcal{U}$	Spaces spanned by independent and dependent coordinates, respectively
$\mathcal{U}^s$	Space spanned **only** by the s^{th}-order derivatives of u^α
$\mathcal{U}^{(s)}$	Space spanned by all derivatives of u^α up to order s
$\boldsymbol{x}$	Coordinate vector in $\mathcal{X}$, i.e., $\boldsymbol{x} = (x^1, \dots, x^p)$
$\boldsymbol{u}$	Coordinate vector in $\mathcal{U}$, i.e., $\boldsymbol{u} = (u^1, \dots, u^q)$
$\boldsymbol{u}^{(s)}$	Coordinate vector in $\mathcal{U}^{(s)}$
$\frac{\partial}{\partial x^i}$	Basis vector in the coordinate frame of M corresponding to x^i
$\frac{\partial}{\partial u^\alpha_I}$	Basis vector in the coordinate frame of $J^{(s)}(M)$ corresponding to u^α_I
dx^i	Basis one-form in the coordinate coframe of M corresponding to x^i
du^α_I	Basis one-form in the coordinate coframe of $J^{(s)}(M)$ corresponding to u^α_I
θ^α_I	Contact form of order $\#I$ given by $du^\alpha_I - \sum_{i=1}^p u^\alpha_{x^I x^i}(\boldsymbol{x}) dx^i$
g, h	Elements of the Lie group G
$g^{(s)}$	Prolongation of g acting on $J^{(s)}(M)$

Continued on next page

Continued from previous page

Symbol	Description
e	Identity element of G
v, w	Arbitrary vector fields in TG
$\hat{v}$, $\hat{w}$	Arbitrary vector fields in TM
ω	Arbitrary l-form in $\Omega^l(M)$
$\Upsilon^{(s)}$	Coframe on $J^{(s)}(M)$
$\Upsilon_G^{(s)}$	G-Coframe on $J^{(s)}(M)$
$\hat{\xi}^i$	Component function of $\hat{v}$ in the $\frac{\partial}{\partial x^i}$ direction
$\hat{\varphi}_I^\alpha$	Component function of $\hat{v}$ in the $\frac{\partial}{\partial u_I^\alpha}$ direction
$\boldsymbol{Q}$	Characteristic of $\hat{v}$, given by $\boldsymbol{Q} = (Q^1, \ldots, Q^q)$
Ψ	Group action $\Psi : G \times M \to M$
$\exp(\cdot)$	Exponential map
$\boldsymbol{\Phi}$	Normal coordinate map
$g \cdot x$	Shorthand for $\Psi(g, x)$
$\mathbb{1}$	Identity operator
$[\cdot, \cdot]$	Lie bracket operator
$v(\cdot)$, $\hat{v}(\cdot)$	Infinitesimal action of v (or $\hat{v}$) on an analytical object in G (or M)
$\mathrm{pr}^{(s)}(\cdot)$	Prolongation operator of order s
$\pi_s^t(\cdot)$	Canonical projection from $J^{(t)}(M)$ to $J^{(s)}(M)$, where $s < t$
D_i	Total derivative operator in the x^i direction
d	Exterior derivative operator
$\wedge$	Wedge product operator

Continued on next page

Continued from previous page

Symbol	Description
ι	Immersion map $\iota : N \to M$
r	Retraction map $r : U \to M$ for a tubular neighborhood $U \subset M$
$\text{Jac}(\cdot)$	Jacobian matrix operator
$\text{rank}(\cdot)$	Rank operator for a given matrix
$\sigma^{(s)}$	Moving frame section of order s, $\sigma^{(s)} : J^{(s)}(M) \to G \times J^{(s)}(M)$
Γ_f	Graph of the function $u = f(\boldsymbol{x})$
$\Gamma_f^{(s)}$	Prolonged graph of $u = f(\boldsymbol{x})$ in $J^{(s)}(M)$
$\Delta^{(s)}(\boldsymbol{x}, \boldsymbol{u}^{(s)})$	System of differential equations of order s
$\Sigma^{(s)}$	Submanifold of $J^{(s)}(M)$ defined by $\Delta^{(s)}(\boldsymbol{x}, \boldsymbol{u}^{(s)}) = 0$
$\boldsymbol{\Lambda}$	Parameter set of a neural model
$\mathscr{J}_*$	Cost function notation
$\mathscr{C}_*$	Constraint function notation
μ	Vector of Lagrange multipliers
ℓ_p	p-norm space

D Algorithms

All algorithms (except the first one) presented in this appendix are discussed sequentially in Sections 4 and 5. They illustrate the core concepts underlying the implementation but do not account for any instruction-level or data-level optimizations provided by PyTorch [50]. A flow-chart depicting the control flow throughout all algorithms is included in the end of this section.

Algorithm 1: Cartan's Algorithm

input: (ϕ, $\boldsymbol{\omega}_U$, $\boldsymbol{\Omega}_V$, G)

$f \leftarrow 0$

Formulate the lifted coframes ((3.59)) and equivalence map ((3.61)):

$(\boldsymbol{\omega}, \boldsymbol{\Omega}, \Phi)$

while $G \neq \{e\}$ **do**

Generate structure equations (3.81) and (3.88):

$d\,\omega^i = \eta_{U\,k}^{\,i} \wedge \omega^k + (T_{\eta_U})^i_{jk}\,(z,g)\,\omega^j \wedge \omega^k$

$d\,\Omega^i = A^i_{k\kappa}\bar{\pi}^\kappa \wedge \Omega^k + (T_V)^i_{jk}(\bar{z},\bar{g})\,\Omega^j \wedge \Omega^k$

Construct and solve absorption equations (3.85) for ν^κ_j:

$\left(T_{\boldsymbol{\eta}_U}\right)^i_{jk}(z,g) = -\left(A^i_{k\kappa}\nu^\kappa_j - A^i_{j\kappa}\nu^\kappa_j\right),\ j<k$

Extract essential torsion coefficients ((3.89)):

$B^i_{jk}\,[\boldsymbol{\nu}] \leftarrow A^i_{k\kappa}\nu^\kappa_j - A^i_{j\kappa}\nu^\kappa_k,\ j<k$

forall i, j, k; *s.t.* $j<k$ **do**

if $B^i_{jk} = 0$ **then**

$T_E \leftarrow T_E \cup (T_{\eta_U})^i_{jk}$

if $T_E = \varnothing$ **then**

return $(\boldsymbol{\omega}, \boldsymbol{\Omega})$

Normalize group parameters $\boldsymbol{g}$ of $G_{(f)}$:

forall $(T_E)^{i_l}_{j_l k_l} \in T_E$ **do**

if $(T_E)^{i_l}_{j_l k_l}$ ***explicitly*** *depends on any* g **then**

$g(z,\mathbf{h}) \leftarrow$ Solve (3.91) for g

Substitute $g(z,\mathbf{h})$ into T_E

Substitute all normalized params $g(z,\mathbf{h})$ into lifted coframes and hence, reduce the structure group G:

$(\widetilde{\boldsymbol{\omega}}, \widetilde{\boldsymbol{\Omega}}, G_{(f+1)})$

Set the reduced structure group and coframes for the next iteration:

$(\boldsymbol{\omega}, \boldsymbol{\Omega}, G) \leftarrow (\widetilde{\boldsymbol{\omega}}, \widetilde{\boldsymbol{\Omega}}, G_{(f+1)})$

$f \leftarrow f+1$

return $(\boldsymbol{\omega}, \boldsymbol{\Omega})$

Algorithm 2: Cost for discovering a single $v \in \mathfrak{g}$

input: (isSltnInd, $\mathcal{V}$, $\hat{v}$, $\mathbf{\Delta}^{(s_c)}$, $z^{(s_c)}$, tol_{ind}, tol_{dep}, ε_{dep})

$\mathcal{J}_{\text{sym}} \leftarrow$ evalEqn-(4.2)$(\Delta^{(s_c)},\ \hat{v}, z^{(s_c)})$

if *isSltnInd* **then**

  $\mathcal{J}_{\text{nlu}} \leftarrow$ evalEqn-(4.17)$(\hat{v}, z^{(s_c)})$

else

  $\mathcal{J}_{\text{nlu}} \leftarrow 0.0$

if *len*$(\mathcal{V}) > 1$ **then**

  $\mathscr{C}_{\text{ind}} \leftarrow$ evalEqn-(4.8)$(\mathcal{V}, z^{(s_c)}, \hat{v}, \text{tol}_{\text{ind}})$

else

  $\mathscr{C}_{\text{ind}} \leftarrow 0.0$

if *isSltnInd* **then**

  $\mathscr{C}_{\text{dep}} \leftarrow$ evalEqn-(4.13)$(\varepsilon_{\text{dep}}, \mathbf{\Delta}^{(s_0)}, \hat{v}, z^{(s_c)}, \text{tol}_{\text{dep}})$

else

  $\mathscr{C}_{\text{dep}} \leftarrow 0.0$

$\mathcal{L}_{\text{Lie}} \leftarrow$ evalEqn-(4.18)$(\mathcal{J}_{\text{sym}}, \mathcal{J}_{\text{nlu}}, \mathscr{C}_{\text{ind}}, \mathscr{C}_{\text{dep}})$

return $\mathcal{L}_{Lie}$

Algorithm 3: Discovering symmetries up to order s_c

input: (isSltnInd, s_c, $\Gamma^{(s_c)}_{\boldsymbol{f}_{\text{tr}}}$, $\boldsymbol{\Delta}^{(s_c)}$, tol_{halt}, tol_{ind}, tol_{dep}, ε_{dep}, $\dim\left(\Sigma^{(s_c)}\right)$,
maxNumEp, $o_{(s_c-1)}$, $\mathcal{V}$, $\boldsymbol{\Lambda}$)

$i \leftarrow o_{s_c-1}$

numSmpl $\leftarrow$ numSamples$(\Gamma^{(s_c)}_{\boldsymbol{f}_{\text{tr}}})$

failed $\leftarrow$ **False**

while not *failed* **and** $i \leq \dim\left(\Sigma^{(s_c)}\right)$ **do**

  // Initialize FCN_{λ_i}

  cost $\leftarrow \infty$

  epoch $\leftarrow 0$

  $\lambda_i \leftarrow$ createAndInitParams$(\dots)$

  $\hat{v}_{\lambda_i} \leftarrow \text{FCN}(\lambda_i)$

  // Learn $v_i \in \mathfrak{g}$

  while *cost* $\geq$ *tol*$_{halt}$ **and** *epoch* $<$ *maxNumEp* **do**

    dataset $\leftarrow$ randomize$(\Gamma^{(s_c)}_{\boldsymbol{f}_{\text{tr}}})$

    epochCost $\leftarrow 0.$

    forall $z^{(s_c)} \leftarrow$ *next*(*dataset*) **do**

      $\mathcal{L}_{\text{Lie}} \leftarrow$ callAlgorithm-2(
        isSltnInd, s_c, $z^{(s_c)}$, $\mathcal{V}$, $\hat{v}_{\lambda_i}$, $\boldsymbol{\Delta}^{(s_c)}$,
        tol_{ind}, tol_{dep}, ε_{dep}
)

      epochCost $\leftarrow$ epochCost $+ \min_{\lambda_i,\boldsymbol{\mu}} \mathcal{L}_{\text{Lie}}$

    cost $\leftarrow$ epochCost/numSmpl

    epoch $\leftarrow$ epoch $+ 1$

  if *cost* $\geq$ *tol*$_{halt}$ **then**

    failed $\leftarrow$ **True**

  else

    $\mathcal{V} \leftarrow \mathcal{V} + [\hat{v}_{\lambda_i}]$

    $\boldsymbol{\Lambda} \leftarrow \boldsymbol{\Lambda} + [\lambda_i]$

    $i \leftarrow i + 1$

$o_{s_c} \leftarrow i$

return *(*$\mathcal{V}$, $\boldsymbol{\Lambda}$, o_{s_c}*)*

Algorithm 4: Cost for learning $\mathcal{W}_{\boldsymbol{f}}$

input: ($z^{(s_c)}$, $\mathcal{W}$, tol_{ntr}, tol_{ind})

$\bar{\mathcal{W}} \leftarrow \text{pr}^{(s_c)}(\mathcal{W})|_{z^{(sc)}}$

`// +Cost for involutivity`

$\mathscr{J}_{\text{inv}} \leftarrow 0.0$

if *len*$(\mathcal{W}) > 1$ **then**

  $\bar{A} \leftarrow$ evalEqn-(4.26)$(\mathcal{W}, z^{(s_c)})$

  $\bar{\mathbb{P}} \leftarrow \bar{A}\left(\bar{A}^{\text{t}}\bar{A}\right)^{-1}\bar{A}^{\text{t}}$

  `// Lie bracketing` $\forall\, 1 \leq j < k \leq o_s$

  $\bar{\mathcal{W}}_{jk} \leftarrow \left[\bar{\mathcal{W}}\left[j\right], \bar{\mathcal{W}}\left[k\right]\right]$

  $\mathscr{J}_{\text{inv}} \leftarrow$ evalEqn-(4.25)$(\bar{\mathbb{P}}, \bar{\mathcal{W}}_{jk})$

`// +Constraint for linear-indep`

$\mathscr{C}_{\text{ind}} \leftarrow 0.0$

if *len*$(\mathcal{W}) > 1$ **then**

  foreach $\hat{w}_j \in \bar{\mathcal{W}}$ **do**

    $\mathscr{C}_{\text{ind}} \leftarrow \mathscr{C}_{\text{ind}} +$ evalEqn-(4.8)$(\bar{\mathcal{W}}, z^{(s_c)}, \hat{w}_j, \text{tol}_{\text{ind}})$

`// +Constraint for non-triviality`

$\mathscr{C}_{\text{ntr}} \leftarrow$ evalEqn-(4.22)$(\hat{v}, z^{(1)}, \text{tol}_{\text{ntr}})$

$\mathcal{L}_{\text{inv}} \leftarrow$ evalEqn-(4.27)$(\mathscr{J}_{\text{inv}}, \mathscr{C}_{\text{ind}}, \mathscr{C}_{\text{ntr}})$

return $\mathcal{L}_{inv}$

Algorithm 5: Learning $\mathcal{W}_{\boldsymbol{f}}$

input: (o_{s_c}, dof, $\Gamma^{(s_c)}_{\boldsymbol{f}_{\text{tr}}}$, $\mathcal{V}$, tol_{halt}, tol_{ntr}, tol_{ind}, maxNumEp)

// $o_{s_c} = 1$ taken care of by ODE's

// existence/uniqueness theorem

if $o_{s_c} < 2$ **then**

  return True

$\mathcal{V} \leftarrow \text{pr}^{(s_c)}(\mathcal{V})$

cost $\leftarrow \infty$

epoch $\leftarrow 0$

// See (4.19) and (4.32)

$\boldsymbol{\Lambda}_{\mathcal{W}} \leftarrow \text{createAndInitParams}(\text{dof}, o_{s_c})$

$\mathcal{W} \leftarrow \text{FCN}(\boldsymbol{\Lambda}_{\mathcal{W}}, \mathcal{V})$

success $\leftarrow$ **False**

while not *success* **and** *len*$(\mathcal{W}) > 1$ **do**

  while *cost* $>$ *tol*$_{halt}$ **and** *epoch* $<$ *maxNumEp* **do**

    dataset $\leftarrow \text{randomize}(\Gamma^{(s_c)}_{\boldsymbol{f}_{\text{tr}}})$

    epochCost $\leftarrow 0.$

    forall $z^{(s_c)} \leftarrow$ *next*(*dataset*) **do**

      $\mathcal{L}_{\text{inv}} \leftarrow$ callAlgorithm-4(

        $z^{(s_c)}, \mathcal{W}, \text{tol}_{\text{ntr}}, \text{tol}_{\text{ind}}$

)

      epochCost $\leftarrow$ epochCost $+ \min_{\boldsymbol{\Lambda}_{\mathcal{W}}, \boldsymbol{\mu}} \mathcal{L}_{\text{inv}}$

    cost $\leftarrow$ epochCost/numSmpl

    epoch $\leftarrow$ epoch $+ 1$

  if *cost* $<$ *tol*$_{halt}$ **then**

    success $\leftarrow$ **True**

  else

    // Remove the last generator

    $\mathcal{W} \leftarrow \mathcal{W} \setminus \mathcal{W}[-1]$

    $\boldsymbol{\Lambda}_{\mathcal{W}} \leftarrow \boldsymbol{\Lambda}_{\mathcal{W}} \setminus \boldsymbol{\Lambda}_{\mathcal{W}}[-1]$

    reinitializeParams$(\boldsymbol{\Lambda}_{\mathcal{W}})$

return $\mathcal{W}$

Algorithm 6: The main algorithm with prolongation

input: ($\Gamma^{(s_0)}_{\boldsymbol{f}_{\text{tr}}}$, $\boldsymbol{\Delta}^{(s_0)}$, $\dim\left(\Sigma^{(s_0)}\right)$, tol_{halt}, tol_{ntr}, tol_{ind}, tol_{dep}, ε_{dep}, p, q, s_0, maxNumEp, maxProlong)

$\mathcal{V} \leftarrow []$

$\boldsymbol{\Lambda} \leftarrow []$

$s_c \leftarrow s_0$

isStabOrder $\leftarrow$ **False**

prCnt $\leftarrow 0$

isInvolutive $\leftarrow$ **False**

```
// Discover symmetries independent
// of a particuler solution
```

$(\mathcal{V}, \boldsymbol{\Lambda}, o_{s_c}) \leftarrow$ callAlgorithm-3(

True, s_c, $\Gamma^{(s_c)}_{\boldsymbol{f}_{\text{tr}}}$, $\boldsymbol{\Delta}^{(s_0)}$, tol_{ind}, tol_{halt}, tol_{dep}, ε_{dep}, $\dim\left(\Sigma^{(s_c)}\right)$, maxNumEp, 0, $\mathcal{V}$, $\boldsymbol{\Lambda}$

)

```
// Failed to find any symmetry...
```

if $o_{s_c} < 1$ **then**
  return None

Continues on the next page...

Algorithm 6: The main algorithm with prolongation (cnt.)

```
// Prolong until (4.32) and
// involutivity conditions are met
```
$\boldsymbol{\Delta}^{(s_c)} \leftarrow \boldsymbol{\Delta}^{(s_0)}$

while *prCnt* $<$ *maxProlong* **do**

```
    // Construct involutive 𝒲_f
```
$\mathcal{W} \leftarrow$ callAlgorithm-5(
$o_{s_c}, \dim\left(\Sigma^{(s_c)}\right) - p, \Gamma^{(s_c)}_{\boldsymbol{f}_{\text{tr}}}, \mathcal{V},$
$\text{tol}_{\text{halt}}, \text{tol}_{\text{ntr}}, \text{tol}_{\text{ind}}, \text{maxNumEp}$
)

isInvolutive $\leftarrow \text{len}(\mathcal{W}) > 0$

if *isInvolutive* **then**
    break

```
    // Necessary prolongation and
    // further symmetry discovery
```
$s_c \leftarrow s_c + 1$

$\Delta^{(s_c)} \leftarrow$ evalEqn-(4.33)$(\Delta^{(s_c-1)})$

$\dim\left(\Sigma^{(s_c)}\right) \leftarrow$ evalEqn-(4.37)$(\boldsymbol{\Delta}^{(s_c)}, s_c, p, q)$

$\Gamma^{(s_c)}_{\boldsymbol{f}_{\text{tr}}} \leftarrow$ evalEqn-(4.31)$(\Gamma^{(s_c-1)}_{\boldsymbol{f}_{\text{tr}}})$

$(\mathcal{V}, \boldsymbol{\Lambda}, o_{s_c}) \leftarrow$ callAlgorithm-3(
False, $s_c, \Gamma^{(s_c)}_{\boldsymbol{f}_{\text{tr}}}, \boldsymbol{\Delta}^{(s_c)}, \text{tol}_{\text{halt}}, \text{tol}_{\text{ind}}, \text{tol}_{\text{dep}}, \varepsilon_{\text{dep}},$
$\dim\left(\Sigma^{(s_c)}\right), \text{maxNumEp}, o_{s_c-1}, \mathcal{V}, \boldsymbol{\Lambda}$
)

prCnt $\leftarrow$ prCnt $+ 1$

```
// Not sufficient prolongation
```
if not *isInvolutive* **then**
    return None

```
// Is s_c = s?
```
if $s_c = s_0$ **or** $o_{s_c-1} = o_{s_c} \leq \dim\left(\Sigma^{(s_c-1)}\right) - p$ **then**
    isStabOrder $\leftarrow$ **True**

return *($\mathcal{V}$, $\boldsymbol{\Lambda}$, s_c, isStabOrder)*

Algorithm 7: Learning $\hat{v}_{\text{can}}$ in Section 5.3

input: ($\Gamma_{\boldsymbol{f}_{\text{tr}}}$, $\mathcal{V}_{o_s}$, tol_{can}, $\hat{v}_{\text{can}}$, maxNumEp)

cost $\leftarrow \infty$

epoch $\leftarrow 0$

numSmpl $\leftarrow$ numSamples($\Gamma_{\boldsymbol{f}_{\text{tr}}}$)

$\theta \leftarrow$ createAndInitParams(...)

$\boldsymbol{b}_{\text{can}} \leftarrow \text{FCN}(\theta)$

while *cost* $\geq$ *tol*$_{can}$ **and** *epoch* $<$ *maxNumEp* **do**

  dataset $\leftarrow$ randomize($\Gamma_{\boldsymbol{f}_{\text{tr}}}$)

  epochCost $\leftarrow 0.$

  forall $z \leftarrow$ *next*(*dataset*) **do**

    $\mathcal{J}_{\text{can}} \leftarrow$ evalEqn $-$ (5.11)($\mathcal{V}_{o_s}, \hat{v}_{\text{can}}, \boldsymbol{b}_{\text{can}}$)

    epochCost $\leftarrow$ epochCost $+ \min_\theta \mathcal{J}_{\text{can}}$

  cost $\leftarrow$ epochCost/numSmpl

  epoch $\leftarrow$ epoch $+ 1$

if *cost* $\geq$ *tol*$_{can}$ **then**

  return None

return *(*$\boldsymbol{b}_{can}$*, cost, epoch)*

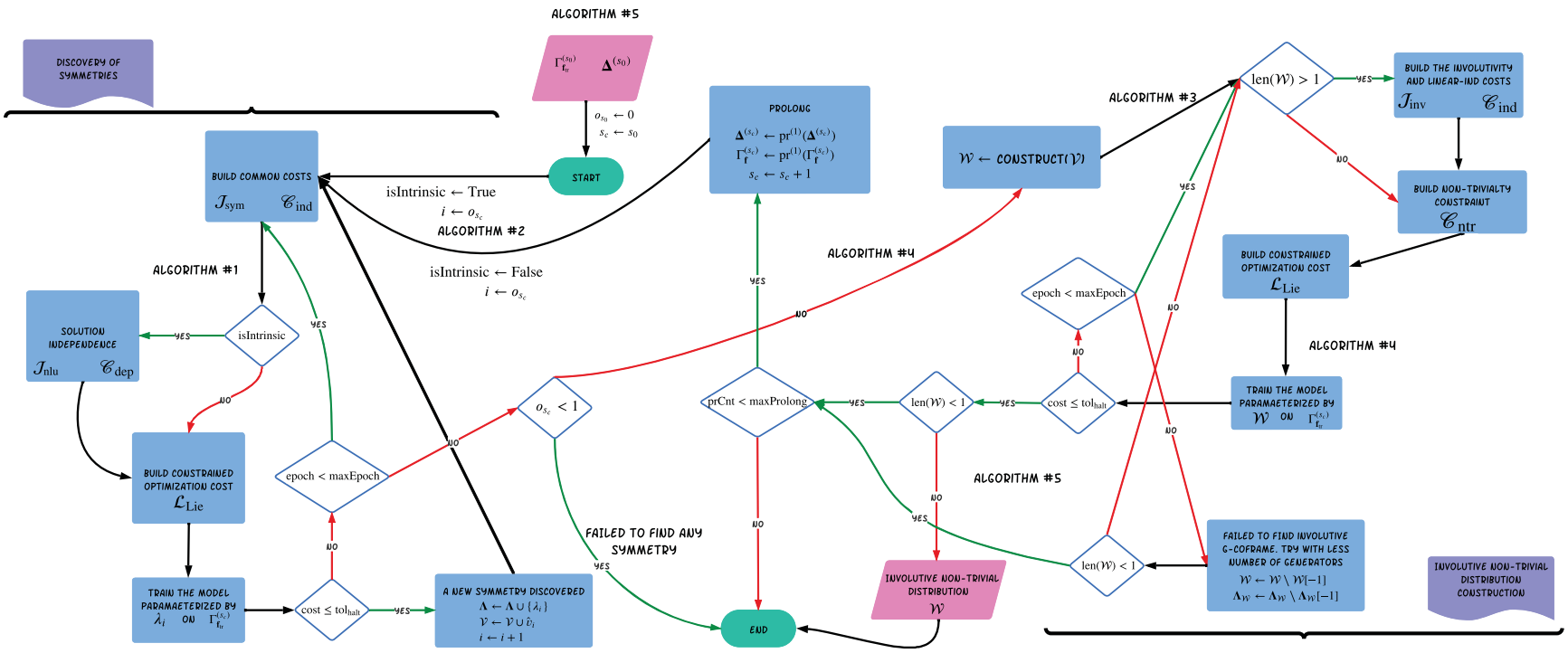

Figure D.1: The labelings “ALGORITHM #X” are locations where control is transferred to the specified algorithm

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Gürcan, İlker
Nationality: Turkish (TC)

EDUCATION

Degree	Institution	Year of Graduation
M.S.	METU	2013
B.S.	Başkent University	2009
High School	Kırıkkale Science High School	2004

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2019-2020	Ayaslı Research Centre	Senior Researcher
2017-2018	University of Houston	Teaching & Research Assistant
2015-2017	TÜBİTAK	Researcher
2010-2015	Başarsoft Company	Software Engineer
2009-2010	DIFA Bilişim	Cofounder

PUBLICATIONS

1. Eldjon Kepucka, Ilker Gurcan, and Alptekin Temizel. *Fast Omnidirectional Image Unwrapping on GPU*. In *Work-in-Progress Proceedings of the Euromicro International Conference on Parallel, Distributed and Network-Based Computing*, 2012.

2. Ilker Gurcan and Alptekin Temizel. *Heterogeneous CPU–GPU Tracking–Learning–Detection (H-TLD) for Real-Time Object Tracking*. *Journal of Real-Time Image Processing*, 16:339–353, Springer, 2015.

3. Ilker Gurcan, Mustafa Teke, and Ugur Murat Leloglu. *Land Use/Land Cover Classification for Göktürk-2 Satellite*. In *2016 24th Signal Processing and Communication Application Conference (SIU)*, pages 2097–2100, IEEE, 2016.

4. Sebastian Berisha, Mahsa Lotfollahi, Jahandar Jahanipour, Ilker Gurcan, Michael Walsh, Rohit Bhargava, Hien Van Nguyen, and David Mayerich. *Deep Learning for FTIR Histology: Leveraging Spatial and Spectral Features with Convolutional Neural Networks*. *Analyst*, 144(5):1642–1653, Royal Society of Chemistry, 2019.

5. Ilker Gurcan and Hien Van Nguyen. *Surgical Activities Recognition Using Multi-Scale Recurrent Networks*. In *ICASSP 2019 – IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2887–2891, IEEE, 2019.