Maier, Markus:

Towards the genuinely ghost–free massive vector Horndeski theory
On the Lagrangian characterization of first–order gravity

Masterarbeit, Wintersemester 2020

Gutachter*in: Weller, Jochen und Schmidt-May, Angnis ; Méndez–Zavaleta, Julio A. und Errasti Díez, Verónica

Fakultät für Physik
Universitäts-Sternwarte München
MSc Astrophysik

Ludwig-Maximilians-Universität München

https://doi.org/10.5282/ubm/epub.104078

Towards the genuinely ghost–free massive vector Horndeski theory

On the Lagrangian characterization of first–order gravity

Beitrag zur wirklich geistfreien massiven Vektor–Horndeski Theorie

Über die Lagrange Charakterisierung der Gravitation erster Ordnung

Thesis written by

Markus Maier

under the supervision of
Prof. Dr. Jochen Weller, **Dr. Angnis Schmidt–May** and
Dr. Julio A. Méndez–Zavaleta
at the Max Planck Institute for Physics

Submitted in partial fulfillment of the requirements for the degree of

M.Sc. in Astrophysics

at the Ludwig Maximilian University of Munich

Munich, 30.03.2020

Abstract

We put forward an extension to already existing Lagrangian constraint algorithms, which is readily applicable to (almost all) first–order classical field theories. Our algorithm is optimized to obtain the explicit constraints and thus count the number of propagating degrees of freedom in said theories. This is the main result of the thesis. We employ both the renowned Dirac–Bergmann procedure and our own formalism to obtain the constraint structure of the $\mathcal{H}(\mathcal{P})$–formulation of non–linear electrodynamics and two–dimensional Palatini gravity. Both approaches yield the same results. We observe that our proposed method is an algebraically simpler and conceptually clearer way to calculate the number of physical modes.

The relevance and usefulness of our novel Lagrangian iterative procedure are twofold. On the one hand, it simplifies the determination of the constraint structure of these theories. This is particularly pertinent for effective theories of multiple interacting fields of different spins, whose analysis is in general cumbersome and which are prone to the presence of additional unphysical modes — ghosts. On the other hand, it constitutes an essential first step towards establishing a Lagrangian building principle for genuinely ghost–free theories. Indeed, given a first–order Lagrangian, our method yields its associated constraint structure. It is then possible to reverse the logic and find out the conditions a Lagrangian must satisfy in order to possess a certain constraint structure. This natural follow–up is work in progress.

Wir schlagen eine Erweiterung zu bereits existierenden Lagrange Zwangsbedingungsalgorithmen vor, welche ohne weiteres auf (fast alle) klassischen Feldtheorien erster Ordnung anwendbar ist. Unser Algorithmus ist optimiert, um die explizite Form der Zwangsbedingungen zu erhalten und ermöglicht damit das Zählen der propagierenden Freiheitsgrade in diesen Theorien. Das ist das Hauptergebnis dieser Arbeit. Wir wenden sowohl das renommierte Dirac–Bergmann Verfahren, als auch unseren eigenen Formalismus an, um die Zwangsbedingungsstruktur der nicht–linearen Elektrodynamik in ihrer $\mathcal{H}(\mathcal{P})$–Formulierung und der zweidimensionalen Palatini Gravitation zu bestimmen. Beide Ansätze liefern das gleiche Ergebnis. Wir beobachten, dass unsere Methode der algebraisch einfachere und konzeptuell klarere Weg ist, um die Anzahl der physikalischen Moden zu berechnen.

Relevanz und Nutzen unseres neuen iterativen Lagrange Verfahrens ist zweifältig. Einerseits vereinfacht es die Bestimmung der Zwangsbedingungsstruktur besagter Theorien. Das ist besonders relevant für effektive Feldtheorien mehrerer wechselwirkender Felder mit verschiedenen Spins. Deren Analyse ist im Allgemeinen umständlich und sie sind anfällig für das Auftauchen zusätzlicher, unphysikalischer Moden — Geister. Andererseits stellt es den ersten essenziellen Schritt in Richtung eines Konstruktionsprinzips für wirklich geistfreie Theorien dar. Unter Voraussetzung einer Lagrangedichte erster Ordnung kann mit unserer Methode die zugehörige Struktur der Zwangsbedingungen bestimmt werden. Es ist dann möglich die Logik umzukehren und Bedingungen zu finden, die eine Lagrangedichte erfüllen muss, um eine bestimmte Zwangsbedingungsstruktur aufzuweisen. Dies ist Teil einer Follow–up–Arbeit in Vorbereitung.

Acknowledgements

I would like to express my gratitude to Prof. Dr. Jochen Weller for being the official supervisor of this work. I thank Dr. Angnis Schmidt–May for the chance to be part of her research group during my thesis as well as the Max Planck Institute for Physics for employing me during the last year.
I owe special thanks to my day–to–day supervisor Dr. Julio A. Méndez–Zavaleta and to Dr. Verónica Errasti Díez, on whose support I could always rely. Without them this thesis would not have been possible.
Last but not least I want to thank Dr. Mojtaba Taslimi Tehrani for his valuable input, in particular regarding the canonical structure of two–dimensional Palatini.

Motivation

Today, dark energy and dark matter are thought to make up about 95% of the energy density of the universe. Type Ia Supernovae, the cosmic microwave background and baryonic acoustic oscillations are convincing evidence for an accelerated expansion of the universe driven by dark energy. Without dark matter, we cannot explain the rotational curves of galaxies, gravitational lensing and structure formation in the universe. Even though the existence of the dark sector is strongly implied by observations, we are far from settling the question of its true nature. The simplest dark energy candidate — a cosmological constant — suffers from a huge discrepancy between the quantum field theoretical explanation of its origin and experimental data [1]. With a difference of around 120 orders of magnitude, this is sometimes called the worst prediction of theoretical physics.

Evidently, not all observations can be explained by a plain constant on its own. Therefore, one must entertain additional degrees of freedom in order to account for the plethora of cosmological data. The aspiration is not to develop a fundamental theory to this aim, but to model the still unexplained observations in the context of effective classical field theories. In the following, we review, in increasing order of algebraic complexity, the pertinent different gravitational models theorists have put forward so far.

Scalar fields

The simplest possibility is to couple a scalar, i.e. spin–0, field to the metric. As it happens, not any arbitrary real scalar–tensor coupling will introduce one and only one additional degree of freedom. When assembling new models, one has to be careful not to do more harm than good: it can happen that a naively proposed theory contains unphysical propagating modes — so called *ghosts*. Notably, the exhaustive and consistent way to couple a scalar field to the metric was derived already over 40 years ago by G.W. Horndeski [2]. Now known as *(scalar) Horndeski theory*[1], it represents the most general scalar–tensor theory that keeps all physical requirements. Namely, it has a positive energy density, propagates the appropriate number of degrees of freedom — two modes of the metric and one mode of the scalar field — and yields at most second–order equations of motion, thus avoiding Ostrogradsky instabilities [5].

Already this simple scenario of an additional scalar field exhibits a rich phenomenology. The corresponding one degree of freedom (two in the complex case) has proven to be fruitful for cosmological applications, including the modeling of dark matter [6], inflation [7] and the accelerated late–time expansion of the universe [8].

Vector fields

The logical continuation of our previous line of though is to consistently couple vector fields to gravity. We thus introduce two (in the massless case) or three (in the massive case) additional degrees of freedom to the system. There is also a strong physical motivation to consider spin–1 fields. It is well–known that the appearance of additional symmetries enhances the protection of the natural values of effective parameters. Indeed, endowing gravity with symmetries beyond diffeomorphism invariance has proven to be able to raise the classically predicted value of the cosmological constant [9]. Further, vector models allow for the simultaneous realization of two fundamental symmetries of nature. It is thus pertinent to look at the coupling between spin–2 and spin–1 fields.

[1] Recently discussed Galileon interactions [3] represent a subclass of scalar Horndeski interactions [4].

Maxwell

Interestingly, also G.W. Horndeski came to this logical conclusion several decades ago — shortly after he developed the most general scalar–tensor theory. His idea was to couple a massless spin–1 field to the metric, under the requirement that the vector field's equations of motion reduce to Maxwell's equations. He thus derived the most general Lagrangian that can be constructed from the metric and a vector (gauge) field in four–dimensional spacetime, which yields at most second–order equations of motion [10]. He did so without worrying about any canonical analysis, but rather pursuing an on–shell study of the field equations. Despite being overlooked for a long time by the physics community, the aforementioned *vector Horndeski* theories are now applied regularly in astrophysics [11] and cosmology [12].

Proca

The textbook example of a massive vector field A_μ is the standard Proca action:

$$S_P = \int d^4x \Big[-\frac{1}{2} F_{\mu\nu} F^{\mu\nu} + m^2 A_\mu A^\mu \Big],$$

where the (non–derivative) self–interaction[2] of the vector field explicitly breaks the $U(1)$ gauge invariance. As a consequence, the longitudinal mode of the vector field propagates and contributes an additional degree of freedom to the system. This is in stark contrast to Maxwell's electrodynamics, where the gauge symmetry eliminates the longitudinal polarization mode and only the two transversal helicities propagate. Indeed, it is noted in [13] that the additional "degree of freedom [...] might not be so undesirable after all" and can have interesting cosmological applications. Investigations of the coupling between a standard Proca field and gravity include dark matter [14] and novel black holes [15, 16]. Provided that only the mass m entered as a new parameter, options are quite limited.

Generalized Proca

An extension of the notion of mass in classical field theory allows for more complex realizations of the Proca field. The difference between a massive and a massless vector field is reflected in the presence of an additional physical degree of freedom. Thus, one might try to explicitly break the gauge symmetry of the vector field by different self–interactions and ask what happens if, besides the hard mass term of the form $A_\mu A^\mu$, one includes derivative self–interactions of the vector field. Do these necessarily introduce further, unphysical, degrees of freedom? This question was independently posed for the first time in [4] and [13] and lead to the construction of the so–called *generalized Proca* (or *vector Galileon*) theory. Schematically, its action takes the form

$$S_{gP} = \int d^4x \Big[-\frac{1}{2} F_{\mu\nu} F^{\mu\nu} + \mathcal{L}_{int} \Big],$$

where the self–interaction part $\mathcal{L}_{int}$ now also contains first–order derivatives of the vector field and certain products thereof. These models are widely used for their rich phenomenology in astrophysics, including dark energy [17], self-accelerating cosmologies [18] and inflation [19].

[2] This is sometimes called the hard mass term.

The need for taking one step back

As should be apparent by now, possibilities to alter the dynamical content of a gravitational theory are vast and one can possibly lose track of all the different attempts. Furthermore, by no means is there any proof that all the proposed and applied models are actually free of ghosts. As an example, consider two generalized Proca fields $A^{(1)}$ and $A^{(2)}$ in flat spacetime, interacting via

$$\mathcal{L}_{int} = \alpha\ \partial \cdot A^{(1)} + \beta\ \partial \cdot A^{(2)},$$

where α and β are smooth functions of $A^{(1)}$ and $A^{(2)}$. This model does indeed propagate a ghost–like degree of freedom, if the non–trivial condition

$$\frac{\partial \alpha}{\partial A_0^{(2)}} - \frac{\partial \beta}{\partial A_0^{(1)}} = 0,$$

is not satisfied [20]. The moral is clear: ghosts are to be expected in generalized Proca theories, unless they are purposely avoided.

Of course, it is interesting to explore the cosmological implications of any generalized Proca theory in the presence of gravity. However, if nobody takes care of the foundations, at some point the whole building is bound to collapse. Realizing the need for taking one step back, in this thesis we begin the rigorous analysis of consistent models of modified gravity. By virtue of simplicity, we focus on the easiest possibility mentioned above: couplings of (massive) vector fields to gravity. These theories are often claimed to be free of unphysical pathologies — ghosts — but a complete proof of this assertion is still missing. To guarantee ghost–freedom, an explicit canonical analysis is inevitable. For instances consider [20, 21], where it was shown that multiple interacting generalized Proca fields in flat spacetime generically propagate ghosts unless stringent conditions are enforced. Nonetheless, the constraint analysis of any extended gravity action is a highly non–trivial task, which gets especially involved when fields of spin $s > 0$ are coupled. Such an analysis is far from being made systematic as of yet. This thesis contributes to fill in this fundamental gap.

Organization of the thesis

This thesis is divided into four parts. Part I provides a self–contained introduction to the mathematical foundations of constrained systems. Part II starts with an introduction to singular systems in the form of Lagrangian and Hamiltonian mechanics. Here, we lay the basis for their treatment under constraint algorithms. Subsequently, we discuss the Dirac–Bergmann theory of constraints, which is a Hamiltonian approach, and we put forward a novel Lagrangian constraint algorithm. In part III, we analyze in detail the concrete examples of non–linear electrodynamics and $d = 2$ Palatini gravity. We do so using the Dirac–Bergmann prescription and our own Lagrangian approach and verify that the latter is much more efficient in both cases. The thesis concludes with part IV, where the main results are summarized and discussed. In the end, we give an outlook to a manuscript under preparation, which shall be a natural follow–up to this thesis.

Main results

The main, and novel, results in this thesis are:

1. An extension to already existing constraint algorithms, which allows for a straightforward count of physical degrees of freedom in (almost) any first–order field theory, given its Lagrangian. This algorithm is presented in Sec. 5.2 and its relevance and usefulness are explained in Sec. 8.

2. The complete characterization of non–linear electrodynamics in purely Lagrangian as well as Hamiltonian terms in Secs. 6.2 and 6.3, respectively. The Lagrangian approach can be readily seen to be much more convenient and conceptually clearer.

3. The complete characterization of $d = 2$ Palatini in purely Lagrangian as well as Hamiltonian terms in Secs. 7.3 and 7.4, respectively. These analyses unequivocally show that the Lagrangian approach is computationally and conceptually simpler.

Quick read

For the comfort of the reader, this thesis admits the following quick read. In the beginning of every section, a brief summary of the most important ideas, concepts and results can be found. These overviews are easy to spot, as they are highlighted by boxes. Together with the present motivation section and the concluding section, the boxed text constitutes a synopsis of the thesis.

Contents

List of Figures

Part I
Mathematical framework

1 From sets to manifolds

In this section, we build up the notion of smooth manifolds, starting from set theory. Sets, as collections of distinct objects, carry little structure on their own. Still, already at this elementary stage, we can introduce important operations and make the notion of a mapping precise. It is insightful to equip sets only stepwise with more and more structure. By doing so, we understand which structure is necessary and sufficient to introduce each pertinent concept. We thus arrive at topological manifolds, i.e. spaces that locally resemble $\mathbb{R}^d$, and construct natural coordinate systems in them, which we refer to as charts. These coordinates allow us to talk about differentiability of manifolds in terms of the usual concept of differentiation in $\mathbb{R}^d$. Crucially, a differentiable manifold has an associated tangent space at each point, which will be essential for our geometrical formulation of classical mechanics. fiber bundles and sections are introduced as purely mathematical concepts at this stage and only later will be used to define important physical concepts.

1.1 Sets

A set is one of the most basic mathematical objects one can even think of and essentially every concept in modern mathematics can be defined in terms of sets [22]. Especially, in the context of differential geometry the relevance and ubiquity of ideas from set theory will be clear and hopefully prevent us from getting more confused than necessary. Even though there is much to say about the theory of sets, for the intents and purposes of this work it is enough to stick to the naive intuition of a set:

A set is a collection of distinct objects, which can be considered as an object in its own right[3].

We call such a distinct object an *element* of the set and write $a \in A$ to state that a is an element belonging to the set A. Similarly we write $a \notin A$ if a is not an element of A. When explicitly writing down a set, we usually enclose its elements with curly brackets. For example, $A = \{2, 3, 5, 7\}$ is the set that contains all prime numbers smaller than 10 as its elements.

Set terminology There is some basic notation regarding the relationship between two sets A and B.

1. $A = B$ means that both sets have exactly the same elements.
2. $A \subseteq B$ denotes that A is a *subset* of B, i.e. every element of A is also an element of B.
3. A is a *proper subset* of B, denoted by $A \subset B$, if $A \subseteq B$ and $A \neq B$.
4. Two sets are called *disjoint*, if they have no elements on common.

On the size of sets The *cardinality* of a set is a measure of how many elements are in the set. The set $A = \{2, 3, 5, 7\}$ has 4 elements, hence its cardinality is 4 and we write $|A| = 4$. The set that has no elements is called the *empty set* and is denoted by $\varnothing$. We can therefore write $|\varnothing| = 0$ [22].

[3]Notice that in modern, so–called axiomatic, set theory there is no definition of a set itself, but only of axioms concerning the sets and the relation between sets.

Set operations For a pair of sets A and B there are three basic operations we can perform to build new sets.

1. $A \cup B = \{x : x \in A \text{ or } x \in B\}$ is the *union* of A and B.
2. $A \cap B = \{x : x \in A \text{ and } x \in B\}$ is the *intersection* of A and B.
3. $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$ is the *set difference* of A and B.

Cartesian product Let A and B be two sets. The *Cartesian product* $A \times B$ is defined to be the set of pairs

$$A \times B = \{(x, y) : x \in A \text{ and } y \in B\}.$$

Intuitively the Cartesian product is very clear, namely we pair up the elements of the set A with the elements of B and put them together as an ordered pair (a, b) [22].

1.1.1 Maps between sets

A *map* $\phi : A \to B$ is a relation[4] between two sets. For every $a \in A$ there exists exactly one $b \in B$ such that $\phi(a, b)$[5]. The usual notation is

$$\begin{aligned} \phi : A &\to B, \\ a &\mapsto \phi(a) \coloneqq b, \end{aligned}$$

which is strictly speaking an abuse of notation since ϕ should have two arguments. Anyhow, since for every $a \in A$ there exists exactly one $b \in B$ such that $\phi(a, b)$ is true, for each a we can define $\phi(a)$ to be that unique b. Even though we have no formal understanding of what a relation is, the naive intuition of a map should be clear: It acts on elements of A and relates them to elements of B [23].

Terminology for sets Let $\phi : A \to B$ be a map and let $V \subseteq B$. We refer to

1. A as the *domain* of ϕ,
2. B as the *target* of ϕ,
3. $\mathrm{im}_\phi(A) \coloneqq \{\phi(a) \mid a \in A\}$ as the *image*[6] of A under ϕ, and
4. $\mathrm{preim}_\phi(V) \coloneqq \{a \in A \mid \phi(a) \in V\}$ as the *preimage* of V under ϕ.

While giving names to objects and concepts doesn't change their nature, it is very helpful in order to specify what we talk about and to communicate statements to others in an unambiguous way. Similarly, we can also encompass properties of a map itself in an extra label.

[4] A relation is a set of ordered pairs. We can e.g. fully represent the squaring function $x \mapsto x^2$ by ordered pairs, one for each real number.

[5] We say it is a single–valued relation.

[6] If we talk about the image of the whole domain this is often abbreviated as $\mathrm{im}(\phi)$.

Terminology for maps Let $\phi : A \to B$ be a map. We call the map

1. *surjective*, if $\phi(A) = B$,
2. *injective*, if $\phi(a_1) = \phi(a_2) \Rightarrow a_1 = a_2$,
3. *bijective*, if it is injective and surjective.

If there exists a bijection $\phi : A \to B$, the two sets A and B are called (set–theoretically) *isomorphic* and we write $A \cong B$. Even though $A \neq B$ (i.e. the elements are not identical) the map is structure preserving and can be reversed by an inverse mapping. Classifying sets (and later spaces) by means of their properties and the structure–preserving maps between them is a recurrent theme in mathematics.

Composition of maps Given two maps $\phi : A \to B$ and $\psi : B \to C$ we can construct a map

$$\begin{aligned} \psi \circ \phi : A &\to C, \\ a &\mapsto \psi(\phi(a)), \end{aligned}$$

called *composition*, which first applies the map ϕ and afterwards the map ψ. Representing the situation pictorially as in Fig. 1 seems unnecessary at this point but will turn out to be very helpful once things get more involved. Using the notion of composition we can make the idea of an inverse

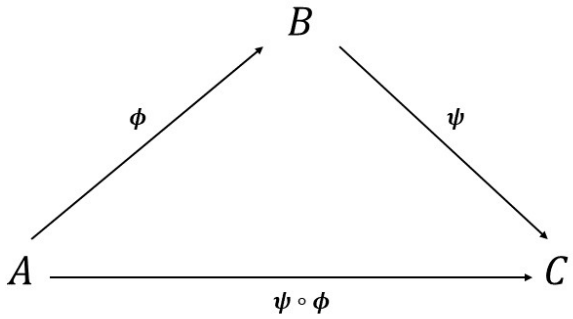

Figure 1: Illustration of a composition of maps.

map precise. Let $\phi : A \to B$ be a bijection. Then the *inverse* of ϕ is the map $\phi^{-1} : B \to A$, defined uniquely by

$$\begin{aligned} \phi^{-1} \circ \phi &= id_A, \\ \phi \circ \phi^{-1} &= id_B, \end{aligned}$$

where id_M is the *identity map* which maps every element $m \in M$ to itself

$$\begin{aligned} id_M : M &\to M, \\ m &\mapsto m. \end{aligned}$$

1.2 Manifolds

Obviously, bare sets don't carry enough structure by themselves to be of direct use for our applications. We could just start our discussion at a more advanced stage but it is interesting and insightful to see how it is — in principle — possible to build the spaces we work with in physics from the theory of sets.

1.2.1 Topological spaces

The most general notion of a mathematical space[7] is a *topological space* and it allows for the definition of concepts such as continuity, connectedness, and convergence. Let M be a non–empty set. Then a choice $\mathcal{O}_M \subseteq \mathcal{P}(M)$[8] is called a *topology* on M if [24]

1. $\varnothing, M \in \mathcal{O}_M$,
2. the intersection of two sets in $\mathcal{O}_M$ remains in $\mathcal{O}_M$, and
3. the arbitrary union of sets in $\mathcal{O}_M$ is again in $\mathcal{O}_M$.

The pair $(M, \mathcal{O}_M)$ is then called a topological space and we call the sets in $\mathcal{O}_M$ open sets. Notice that there are, in general, different topologies we can choose on one and the same set.

Neighborhood Let $(M, \mathcal{O}_M)$ be a topological space and let $m \in M$. A *neighborhood* of m in M is any open set $U \in \mathcal{O}_M$ which contains m [24].

Continuity Let $(M, \mathcal{O}_M)$ and $(N, \mathcal{O}_N)$ be topological spaces and let $\phi : M \to N$ be a map. Then ϕ is called *continuous*, if

$$\forall V \in \mathcal{O}_N : \mathrm{preim}_\phi(V) \in \mathcal{O}_M.$$

Homeomorphism Let $\phi : M \to N$ be a bijection between topological spaces $(M, \mathcal{O}_M)$ and $(N, \mathcal{O}_N)$. We call ϕ a *homeomorphism* if [25]

1. $\phi : M \to N$ is continuous and
2. $\phi^{-1} : N \to M$ is continuous.

This means that ϕ provides a one–to–one pairing of the open sets of M and N and we write $(M, \mathcal{O}_M) \cong (N, \mathcal{O}_N)$. Homeomorphisms are the structure preserving maps in topology.

Hausdorff condition A topological space $(M, \mathcal{O}_M)$ is called *Hausdorff* (or *separated*) if given any pair of distinct points $q_1, q_2 \in M$, there exist neighborhoods U_1 of q_1 and U_2 of q_2 with $U_1 \cap U_2 = \varnothing$ [26]. Sometimes this is simplified as *open sets separate points*. Notably, the Hausdorff condition implies the uniqueness of limits of sequences on the topological space.

1.2.2 Topological manifolds

The topological spaces that physicists use to represent spacetime have a special property. Namely, it is possible to uniquely label any particular spacetime point by specifying the values of a finite set of real numbers. The size of this finite set is identified with the *dimension* of the space [27]. This leads us to the strikingly simple idea behind a topological manifold: It is a topological space that *locally* looks like $\mathbb{R}^d$.

[7] Space here means some set equipped with some structure.
[8] $\mathcal{P}(M)$ is the power set of M. Its elements are precisely the subsets of M.

Definition A Hausdorff topological space $(M, \mathcal{O}_M)$ is called a *d–dimensional (topological) manifold* if for every $p \in M$ there exists an open neighborhood $U \in \mathcal{O}_M$ and a homeomorphism $\phi : U \to \phi(U) \subseteq \mathbb{R}^d$.

The above definition is the lowest common denominator among different authors. Usually one puts additional requirements on topological manifolds at some point, like *paracompactness*[9] [28] and/or *second-countability*[10] [26]. Luckily, we will not deal with such pathological cases in the course of this work and are happy to stick to our intuition of a topological space that locally looks like Euclidean space.

1.2.3 Charts and Atlases

Let $(M, \mathcal{O}_M)$ be a topological manifold of dimension d. The pair (U, ϕ), where

$$U \in \mathcal{O}_M \quad \text{and} \quad \phi : U \to \phi(U) \subseteq \mathbb{R}^d,$$

is called a *chart* of the manifold. By defining a chart, one labels each point $p \in U$ by d real numbers

$$\phi(p) = (\phi^1(p), \phi^2(p), ..., \phi^d(p)).$$

The d functions $\phi^1, ..., \phi^d$ are called the coordinates associated with the chart [29]. This means that coordinates are to be thought of as (local) functions on the manifold, see Fig. 2.

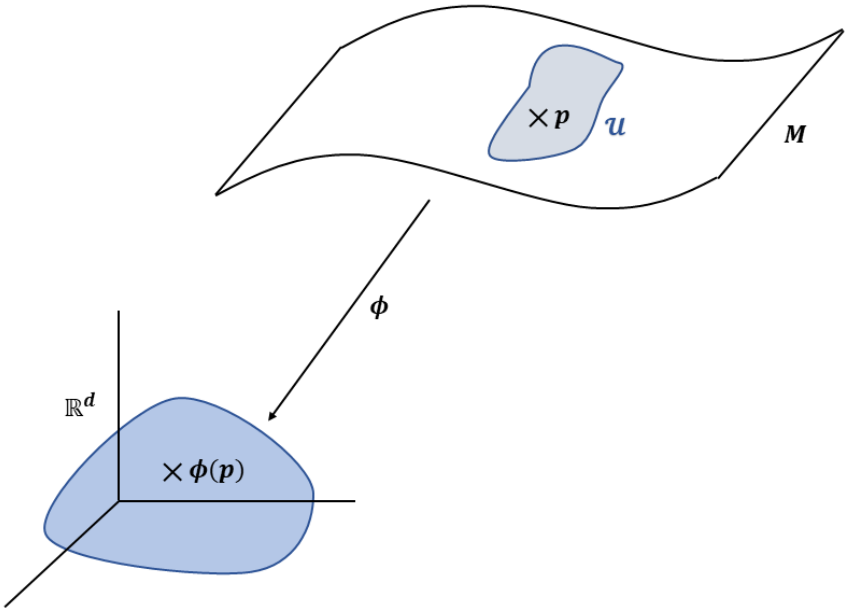

Figure 2: A local coordinate chart.

[9] The weakest topological notion which implies the existence of a partition of unity which in turn is used to define integration on the manifold.

[10] Ensures that there are "not too many" open sets.

Chart transition functions Crucially, charts are not unique and often a single chart is not enough to cover the whole domain of interest. Let (U_1, ϕ_1) and (U_2, ϕ_2) be two coordinate charts on M that have a non–vanishing intersection $U_1 \cap U_2 \neq \varnothing$. The *chart transition function* is the map $\phi_2 \circ \phi_1^{-1}$ from the open subset $\phi_1(U_1 \cap U_2) \subseteq \mathbb{R}^d$ to the open subset $\phi_2(U_1 \cap U_2) \subseteq \mathbb{R}^d$ as illustrated in Fig. 3. The two charts are said to be C^k–compatible if their chart transition functions (in both directions) are *differentiable of class* C^k, i.e. if they are k–times continuously differentiable as maps between open subsets of $\mathbb{R}^d$.

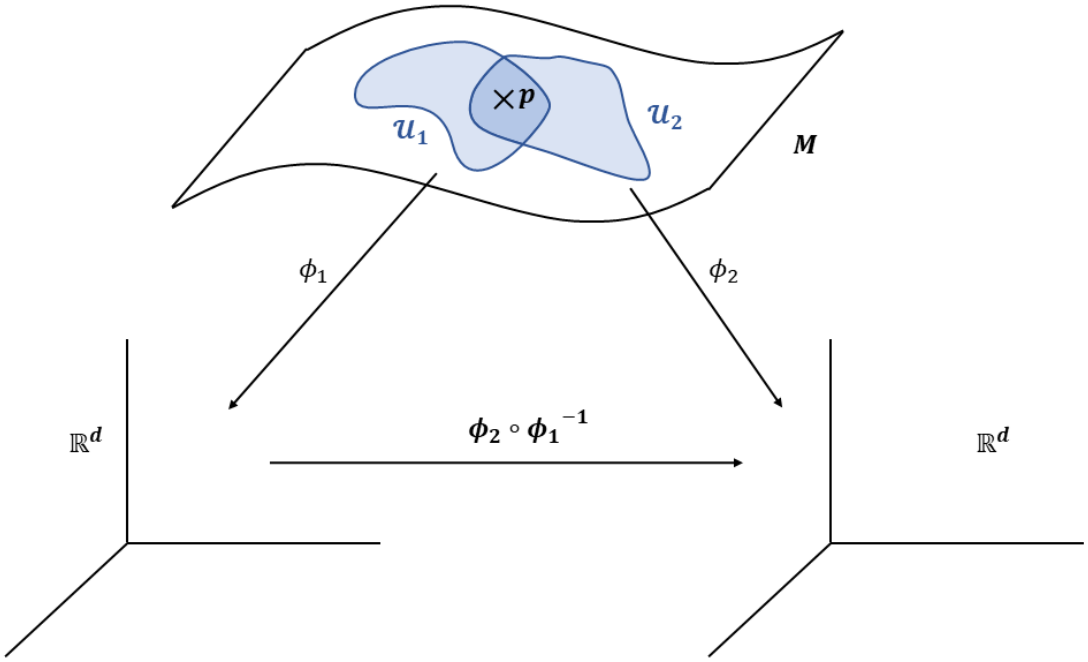

Figure 3: Coordinate transition function of two charts.

Atlases An atlas $\mathcal{A}$ on M is a family of coordinate charts $(U_i, \phi_i)_{i \in I}$ such that their union covers all of M:

$$\bigcup_{i \in I} U_i = M.$$

If its charts are pairwise C^k–compatible it is called a C^k–atlas. A C^k–atlas is called complete if it is maximal, i.e. if any chart (U, ϕ) that is C^k–compatible with any other chart $(V, \psi) \in \mathcal{A}$ is already contained in $\mathcal{A}$. For a complete atlas, the family $(U_i, \phi_i)_{i \in I}$ is called a *differentiable structure* on M [27].

1.2.4 Fiber bundles

A *bundle (of topological manifolds)* is a triple (E, π, M), where the topological manifolds E and M are called *total space* and *base space*, respectively, and the continuous map

$$\pi : E \to M,$$

is called *projection*. Let $p \in M$, then

$$F_p \coloneqq \mathrm{preim}_\pi(\{p\}),$$

is called *fiber over* p. The above definition of a bundle is very general and in all applications in physics the bundles that arise have a special property: the fibers are all homeomorphic to a common space F. Therefore, let (E, π, M) be a bundle, such that

$$\forall p \in M : \mathrm{preim}_\pi(\{p\}) \cong F,$$

for some topological manifold F. Then (E, π, M) is called a *fiber bundle* with (typical) fiber F [27].

Sections

Let (E, π, M) be a fiber bundle. A map

$$\sigma : M \to E,$$

is called a *section* of the bundle, if

$$\pi \circ \sigma = id_M.$$

Notice that a map from M to the total space E could go anywhere, but only if it "goes back to the fiber" it is a section, as depicted in Fig. 4.

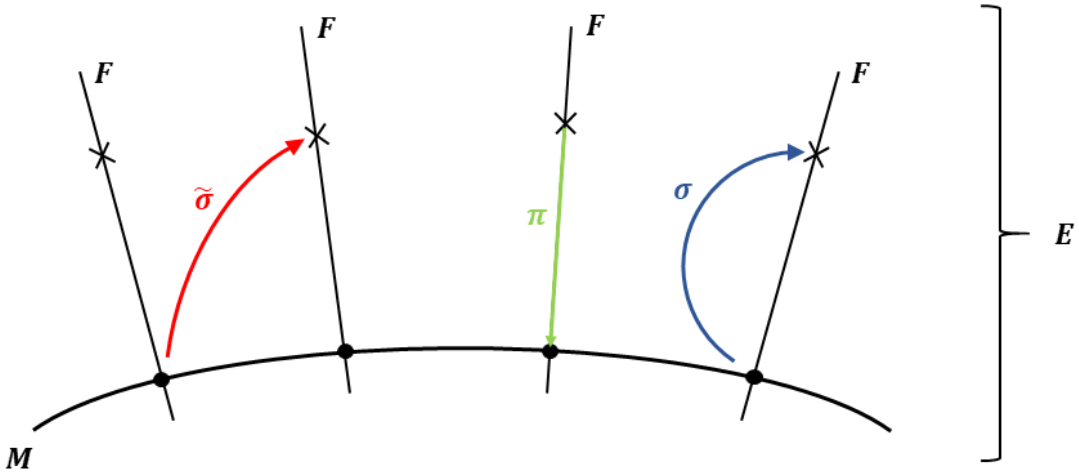

Figure 4: The map σ constitutes a section of the bundle, whereas the $\tilde{\sigma}$ fails to do so.

1.2.5 Differentiable manifolds

A C^k–manifold is a triple $(M, \mathcal{O}_M, \mathcal{A}_M)$ where $(M, \mathcal{O}_M)$ is a topological manifold and $\mathcal{A}_M$ is a maximal C^k–atlas. That means the C^k.-manifold is equipped with a differentiable structure of class C^k. If $k \geq 1$ it is called a *differentiable manifold*. The case of a C^∞–manifold is also referred to as a *smooth manifold*. Without loss of (physically relevant) generality all manifolds that are to appear in this work are assumed to be smooth. Notice that smoothness is a condition on the differentiability of transition functions between different charts covering the manifold.

Tangent space A key feature of a differentiable manifold is that there exists a *tangent space* at each point of the manifold, which is essential in order to formulate physics in geometrical terms[11]. In the case of a 2–sphere ($\mathcal{S}^2$) we can imagine the tangent space as a plane at each point of the sphere. If we think about this tangent plane naively as in Fig. 5, we assume an ambient space (in this case $\mathbb{R}^3$) in which the 2–sphere is embedded. However, it is possible to define this concept

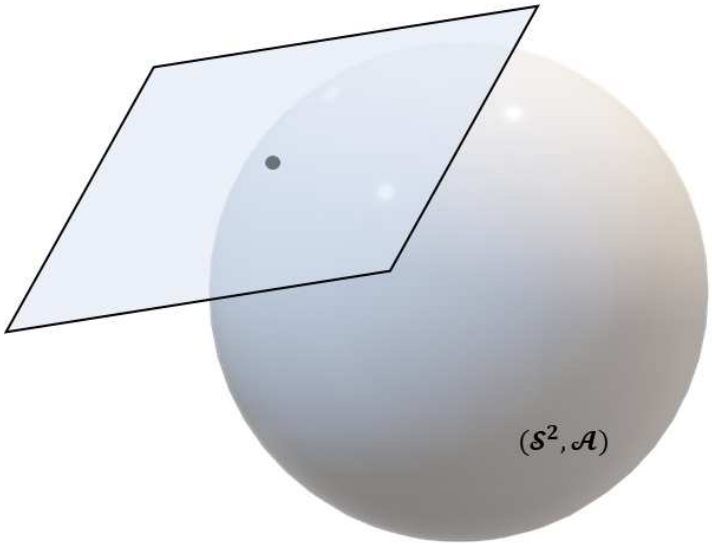

Figure 5: How we think about the tangent space of $\mathcal{S}^2$ at a certain point.

intrinsically in two dimensions without any embedding. To do so, we first have to introduce some basic concepts from linear algebra to complement the machinery we built so far. Notice that there is no mathematical reason for us to choose the intrinsic definition over the embedded one — rather a physical one. Later on, when we understand a certain manifold as spacetime itself, it is conceptually alleviating not to think about any embedding.

Diffeomorphisms Let $\phi : M \to N$ be a bijective map and let both ϕ and ϕ^{-1} be of class C^k, then ϕ is called a *(C^k–)diffeomorphism*. Diffeomorphisms are the structure preserving maps between differentiable manifolds. Consequently $(M, \mathcal{O}_M, \mathcal{A}_M)$ and $(N, \mathcal{O}_N, \mathcal{A}_N)$ are called *diffeomorphic* if there exists a diffeomorphism between them. In that case we write $M \cong N$. A diffeomorphism $\phi : M \to M$ is said to be a *transformation* of M [30].

[11] In particular, tangent spaces will provide the structure for the notion of a derivative.

2 Vector spaces and tangent bundles

We begin this section by introducing essential ideas from linear algebra, such as (dual) vector spaces and their corresponding (dual) basis, together with linear transformations. These concepts are of immediate interest for physics. Rank and nullity of linear transformations will be of utmost importance in our discussion of constrained systems.

Based on our prior discussion, we equip the tangent space with a vector space structure. This allows not only for a generalized idea of the directional derivative in terms of tangent vectors, but also for the construction of a basis on the tangent and cotangent spaces. This is crucial for our purposes, since most of our physical discussion will take place in these spaces. Thereafter, push–forward and pull–back, which will be used frequently throughout this thesis, are introduced.

By putting together all the tangent spaces of a manifold we construct its tangent bundle. Among other things, it provides a domain and range for the derivative of a smooth function. Vector fields and their generalization — tensor fields — can be intrinsically defined in terms of tangent bundles. We further investigate special kinds of tensor fields, namely differential forms, and establish important operations between them. Additionally, integral curves and second–order differential equations and their formal definition are discussed. This provides a strong link to physical applications, in particular to the Hamiltonian and Lagrangian formulations of classical mechanics.

2.1 Linear algebra

Fields A *field* is a triple $(\mathbb{K}, +, \cdot)$, where $\mathbb{K}$ is a set and $+$ and $\cdot$ are two maps

$$+ : \mathbb{K} \times \mathbb{K} \to \mathbb{K},$$
$$\cdot : \mathbb{K} \times \mathbb{K} \to \mathbb{K},$$

that satisfy the so–called field axioms[12]. A well known example of a field are the real or the complex numbers. As a rule of thumb, one can do math in a field as one would do in $\mathbb{R}$ [31]. If the map $\cdot$ fails to satisfy commutativity and does not have an inverse, we usually call the triple a *ring*. For example, $(\mathbb{Z}, +, \cdot)$ is a commutative ring. It only lacks the multiplicative inverse for a field.

2.1.1 Vector spaces and algebras

Let $\mathbb{K}$ be a field. If we have two maps (or *operations*)

$$\oplus : V \times V \to V,$$
$$\odot : \mathbb{K} \times V \to V,$$

which satisfy the vector space axioms[13], we call $(V, \oplus, \odot)$ a $\mathbb{K}$*–vector space* [31].

Subspaces A *subspace* of a vector space V is a subset $S \subseteq V$ that is a vector space in its own right under the operations obtained by restricting the operations of V to S. Notice that any subspace of a vector space has a *complement*. If $S \subseteq V$, then there exists a subspace T for which $V = S \oplus T$. We then call T a complement of S in V [32].

[12] Each map has to satisfy commutativity, associativity and possess a neutral element as well as an inverse.
[13] $\oplus$ follows the field axioms, $\odot$ has to be distributive instead of commutative.

Linear maps Let $f : V \to W$ be a map between the two $\mathbb{K}$–vector spaces $(V, \oplus, \odot)$ and $(W, \boxplus, \boxdot)$. If f satisfies the properties

1. $\forall v_1, v_2 \in V : f(v_1 \oplus v_2) = f(v_1) \boxplus f(v_2)$,
2. $\forall \lambda \in \mathbb{K}, v \in V : f(\lambda \odot v) = \lambda \boxdot f(v)$,

we call it a *linear map* and write $V \xrightarrow{\sim} W$. A bijective linear map is called a *vector space isomorphism*. We denote by $\mathrm{Lin}(V, W)$ the set of all linear maps from V to W, which is a vector space in its own right [32].

Dual vector space Let V be a $\mathbb{K}$–vector space. The *dual* vector space to V is

$$V^* \coloneqq \{f : V \xrightarrow{\sim} \mathbb{K}\},$$

where $\mathbb{K}$ is considered as a vector space over itself. The dual vector space to V is the vector space of linear maps from V to the underlying field $\mathbb{K}$, which are called *linear functionals*, *covectors*, or *one-forms* on V.

Algebras A vector space $(V, +, \cdot)$ equipped with a *product*, i.e. a bilinear map $\bullet : V \times V \xrightarrow{\sim} V$, is called an *algebra* $(V, +, \cdot, \bullet)$. For example, the set of all linear operators $\mathrm{Lin}(V, V)$ can be made into an algebra [32].

2.1.2 Tensors

Bilinear maps Let V, W, Z be vector spaces over some field $\mathbb{K}$. A map $f : V \times W \to Z$ is said to be bilinear if

1. $\forall w \in W : \forall v_1, v_2 \in V : \forall \lambda \in \mathbb{K} : f(\lambda v_1 + v_2, w) = \lambda f(v_1, w) + f(v_2, w)$,
2. $\forall v \in V : \forall w_1, w_2 \in W : \forall \lambda \in \mathbb{K} : f(v, \lambda w_1 + w_2) = \lambda f(v, w_1) + f(v, w2)$.

In other words, when we hold the first entry (resp. second entry) of the bilinear map fixed, while letting the second entry (resp. first entry) vary, the result is a linear operator. The generalization of this concept are *multilinear maps*.

Definition Let V be a vector space over some field $\mathbb{K}$. A (p, q)–*tensor* T on V is a multilinear map

$$T : \underbrace{V^* \times \cdots \times V^*}_{p \text{ copies}} \times \underbrace{V \times \cdots \times V}_{q \text{ copies}} \to \mathbb{K},$$

i.e. it is linear in each slot. We write

$$T^p_q V \coloneqq \underbrace{V \otimes \cdots \otimes V}_{p} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{q} \coloneqq \{T \mid T \text{ is a } (p, q)\text{-tensor}\},$$

and equip it with the usual point–wise addition and S–multiplication, such that $(T^p_q, \oplus, \odot)$ is a vector space.

Tensor product Let $T \in T^p_q V$ and $S \in T^r_s V$. The *tensor product* of T and S is the tensor $T \otimes S \in T^{p+r}_{q+s} V$ defined by

$$\begin{aligned}&(T \otimes S)(w_1, \ldots, w_p, w_{p+1}, \ldots, w_{p+r}, v_1, \ldots, v_q, v_{q+1}, \ldots, v_{q+s}),\\ &\coloneqq T(w_1, \ldots w_p, v_1, \ldots v_q) \cdot_{\mathbb{K}} S(w_{p+1}, \ldots, w_{p+r}, v_{q+1}, \ldots, v_{q+s}).\end{aligned}$$

2.1.3 (Dual) Basis

Let $(V, +, \cdot)$ be a vector space over $\mathbb{K}$. Then a subset $B \subseteq V$ is called a (Hamel) basis for V if [31]

1. every finite subset $\{b_1, \dots, b_N\}$ of B is linearly independent, i.e.

$$\sum_{i=1}^{N} \lambda^i b_i = 0 \Rightarrow \lambda^1 = \cdots = \lambda^N = 0.$$

2. B is a spanning set of V, i.e.

$$\forall v \in V : \exists v^1, \dots, v^M \in \mathbb{K} : \exists b_1, \dots, b_M \in B : v = \sum_{i=1}^{M} v^i b_i.$$

We define the *dimension* of V as $dim(V) \coloneqq |B|$. Furthermore, we can construct a basis on V^* once we have a basis B on V. This induced basis is called *dual basis* and is defined by

$$\epsilon^a(e_b) \overset{!}{=} \delta^a_b,$$

where basis vectors of V are labeled by a downstairs index $(e_1, \dots, e_d \in V)$ and basis vectors of V^* are labeled by upstairs indices $(\epsilon^1, \dots, \epsilon^d \in V^*)$.

From now on we will make use of the *Einstein summation convention*, which means that we suppress the summation sign when the indices to be summed over each appear once as a subscript and once as a superscript in the same term. Indices that are summed over are called *dummy indices*.

2.1.4 Rank and nullity of a linear transformation

Let V and W be $\mathbb{K}$-vector spaces and let $f \in \mathrm{Lin}(V, W)$. We define the *kernel* of f as the subspace of V which gets mapped to 0 by application of f, i.e.

$$\ker(f) \coloneqq \{v \in V \mid f(v) = 0\}.$$

The dimension of $\ker(f)$ is called *nullity* of f and we denote it by $\mathrm{null}(f)$. Furthermore, we define the *rank* of f as the dimension of its image, i.e. $\mathrm{rank}(f) \coloneqq \dim(\mathrm{im}(f))$.

Rank-nullity theorem The *rank-nullity theorem* relates the dimensions of a linear map's kernel and image with the dimension of its domain. It states that

$$\mathrm{rank}(f) + \mathrm{null}(f) = \dim(V),$$

and only holds true for finite-dimensional vector spaces [32].

2.2 Tangent spaces

2.2.1 Tangent vectors

Let M be a smooth manifold. Then we can construct the $\mathbb{R}$–vector space $(C^\infty(M), \oplus, \odot)$ with point–wise defined operations:

$$(f \oplus g)(p) \coloneqq f(p) +_{\mathbb{R}} g(p),$$
$$(\lambda \odot f)(p) \coloneqq \lambda \cdot_{\mathbb{R}} f(p),$$

where $f, g \in C^\infty(M)$, $\lambda \in \mathbb{R}$ and $p \in M$.

Directional derivative operator Let $\gamma : \mathbb{R} \to M$ be a smooth curve through a point $p \in M$. Without loss of generality we choose $p = \gamma(0)$. Then the *directional derivative operator* (at p along γ) is the linear map

$$\begin{aligned} X_{\gamma,p} : C^\infty(M) &\xrightarrow{\sim} \mathbb{R}, \\ f &\mapsto (f \circ \gamma)'(0). \end{aligned}$$

Note that $(f \circ \gamma)$ is a map $\mathbb{R} \to \mathbb{R}$, hence we can use our usual notion of a derivative, denoted by an apostrophe ('). We often call $X_{\gamma,p}$ the *tangent vector*[14] to the curve γ at $p \in M$. Notice that the tangent vector formally is an operator that acts on smooth functions [27] but we often think of it as depicted in Fig. 6.

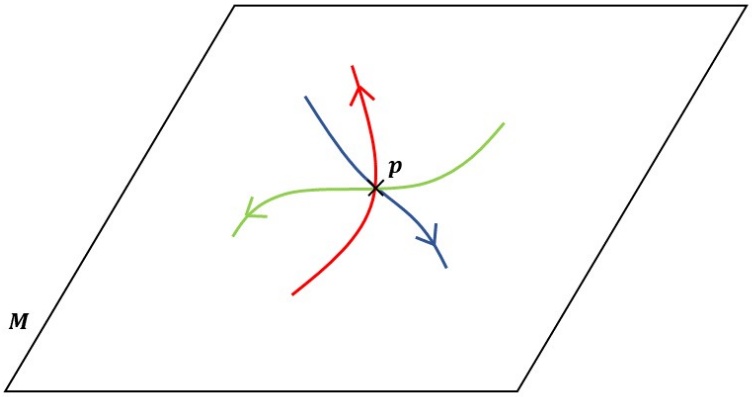

Figure 6: Intuition for the tangent vectors at $p \in M$. On has to run through all curves with all possible velocities to get every $X_{\gamma,p}$.

2.2.2 Tangent vector space

The *tangent vector space* T_pM is defined as the set of all tangent vectors to the smooth curve γ at $p \in M$.

$$T_pM \coloneqq \{X_{\gamma,p} \mid \gamma \text{ smooth curve through } p\},$$

equipped with addition and S–multiplication

$$\begin{aligned} \oplus &: T_pM \times T_pM \to T_pM, \\ \odot &: \mathbb{R} \times T_pM \to T_pM, \end{aligned}$$

defined pair–wise as

$$\begin{aligned} (X_{\gamma,p} \oplus X_{\delta,p})(f) &\coloneqq X_{\gamma,p}(f) +_{\mathbb{R}} X_{\delta,p}(f), \\ (\lambda \odot X_{\gamma,p})(f) &\coloneqq \lambda \cdot_{\mathbb{R}} X_{\gamma,p}(f). \end{aligned}$$

[14] There are many conceptually different, but equivalent, definitions of a tangent vector in the literature. For an extensive overview see e.g. Alan Kenningtons stupendous work in progress "Differential geometry reconstructed".

Tangent space basis Let M be a smooth manifold and let (U, x) be a local chart, centered at $p \in U$. Consider $(\dim\text{–}M)$ many curves $\gamma_{(a)} : \mathbb{R} \to U$ through p. The action of a tangent vector $e_a \coloneqq X_{\gamma_{(a)},p}$ on an arbitrary function $f \in C^\infty(U)$ is given by

$$e_a f = X_{\gamma_{(a)},p} f = (f \circ \gamma_{(a)})'(0) = \partial_a (f \circ x^{-1}) \coloneqq (\frac{\partial}{\partial x^a} f)_p.$$

As a matter of fact, any $X \in T_pM$ can be written as $X = X^a (\frac{\partial}{\partial x^a})_p$, i.e. the $(\frac{\partial}{\partial x^1})_p, \ldots, (\frac{\partial}{\partial x^{\dim(M)}})_p$ are a generating system for T_pM. Additionally they are linearly independent and thus constitute a basis of T_pM. The real numbers $X^1, \ldots, X^{\dim(M)}$ are called the *components* of the vector X with respect to the *tangent space basis* induced from the chart (U, x), $p \in U$ [29]. Notice that this implies $\dim(T_pM) = \dim(M)$, i.e. the vector space dimension of the tangent space at each point of the manifold coincides with the dimension of M as a topological manifold.

2.2.3 Cotangent vector space

Let M be a smooth manifold, we define the *cotangent space* as the dual of the tangent vector space: $T_p^*M \coloneqq (T_pM)^*$ [33].

Gradient Let $f \in C^\infty(M)$ and $X \in T_pM$. At every point $p \in M$ we have a linear map

$$\begin{aligned} \mathrm{d}_p : C^\infty(M) &\xrightarrow{\sim} T_p^*M, \\ f &\mapsto d_p f, \end{aligned}$$

defined by $(\mathrm{d}_p f)(X) \coloneqq Xf$, called the *gradient operator* at $p \in M$. We call $\mathrm{d}_p f$ the gradient of the function f at p. Notice that $\mathrm{d}_p f$ is a covector, not a vector.

Cotangent space basis We can use the gradient operator to construct a basis of T_p^*M: Let $p \in U \subseteq M$ and (U, x) be a chart. Then $\mathrm{d}_p x^1, \mathrm{d}_p x^2, \ldots, \mathrm{d}_p x^{\dim(M)}$ is called the *chart–induced covector basis at* p. In fact this is even the dual basis, as can be seen easily:

$$\mathrm{d}_p x^a ((\frac{\partial}{\partial x^b})_p) = (\frac{\partial}{\partial x^b})_p x^a = \partial_b (\underbrace{x^a \circ x^{-1}}_{proj_a})(x(p)) = \delta^a_b, \qquad a, b = 1, \ldots, \dim(M).$$

2.2.4 Push–forward and pull–back

Push–forward Let $\phi : M \to N$ be a smooth map between smooth manifolds. Furthermore let $f : N \to \mathbb{R}$, $X \in T_pM$, and $p \in M$. Then the *Push–forward* of the map ϕ at the point p is the linear map

$$\begin{aligned} (\phi_*)_p : T_pM &\xrightarrow{\sim} T_{\phi(p)}N, \\ X &\mapsto (\phi_*)_p(X), \end{aligned}$$

defined by $(\phi_*)_p(X) f \coloneqq X(f \circ \phi)$[15]. Fig. 7 illustrates how the tangent vector $X_{\gamma,p}$ of a smooth curve γ at p is pushed forward to the tangent vector of the smooth curve $(\phi \circ \gamma)$ at $\phi(p)$, i.e. $(\phi_*)_p(X_{\gamma,p}) = X_{\phi\circ\gamma,\phi(p)}$.

[15] Notice that this is the derivative of ϕ at p and agrees with our definition of a tangent vector.

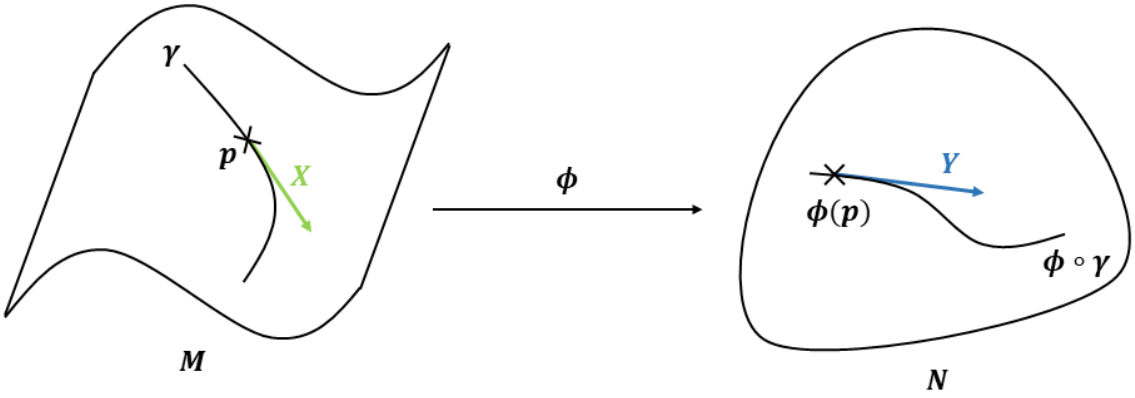

Figure 7: The idea of a push–forward: $Y = \phi_*(X)$.

pull–back Let $\phi : M \to N$ be a smooth map, let $p \in M$, $X \in T_pM$, and $\omega \in T^*_{\phi(p)}N$. Then the *pull–back* of ϕ at $\phi(p) \in N$ is the linear map

$$\begin{aligned}(\phi^*)_p : T^*_{\phi(p)}N &\xrightarrow{\sim} T^*_pM,\\ \omega &\mapsto (\phi^*)_p(w),\end{aligned}$$

where $(\phi^*)_p(\omega)$ is defined as

$$\begin{aligned}(\phi^*)_p(\omega) : T_pM &\xrightarrow{\sim} \mathbb{R},\\ X &\mapsto \omega((\phi^*)_p(X)),\end{aligned}$$

hence $(\phi^*)_p(\omega)(X) \coloneqq \omega((\phi_*)_p(X))$. So, if ω is a covector on N, its pull–back $(\phi^*)_p(\omega)$ is a covector on M. It acts on tangent vectors on M by first pushing them forward to tangent vectors on N, and then applying ω to get a real number.

Remarks In general, one should keep in mind that vectors are pushed forward and covectors are pulled back. If the map $\phi : M \to N$ is a diffeomorphism we can also pull a vector $Y \in T_{\phi(p)}N$ back to a vector $(\phi^*)_p(Y) \in T_pM$, and push a covector $\eta \in T^*_pM$ forward to a covector $(\phi_*)_p(\eta) \in T^*_{\phi(p)}N$. We do this by using ϕ^{-1} in the following way:

$$\begin{aligned}(\phi^*)_p(Y) &\coloneqq ((\phi^{-1})_*)_{\phi(p)}(Y),\\ (\phi_*)_p(\eta) &\coloneqq ((\phi^{-1})^*)_{\phi(p)}(\eta).\end{aligned}$$

2.3 Tangent bundles

Idea Let M be a smooth manifold and $p \in M$, let $I \subset \mathbb{R}$ be an open interval and $\gamma : \mathbb{R} \to M$ be a smooth curve. For every $t \in I$ the tangent vector $X_{\gamma,t}$ is an element of $T_{\gamma(t)}M$ and therefore, as t runs through I, $X_{\gamma,t}$ runs through the tangent spaces along γ. To follow the tangent vector along γ it is convenient to consider the totality of all tangent spaces of M [34]. This idea is sketched in Fig. 8.

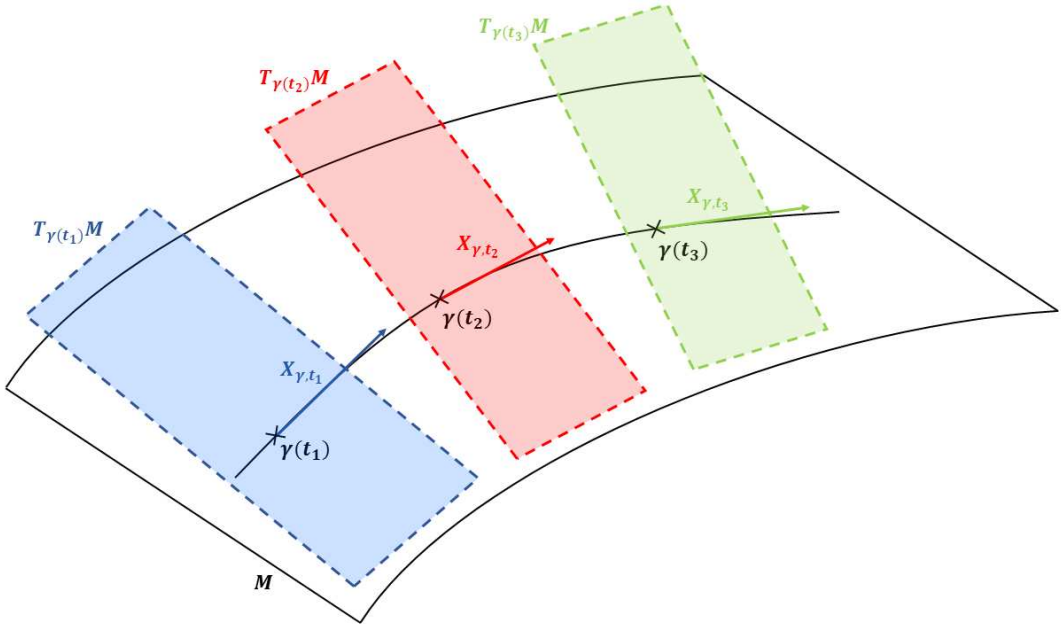

Figure 8: Tangent vectors along a curve.

Tangent bundle The *tangent bundle* TM is defined as the disjoint union of the tangent spaces at all points of the manifold:

$$TM \coloneqq \bigcup_{p \in M} T_p M,$$

equipped with the canonical projection map

$$\begin{aligned} \pi : TM &\to M, \\ X &\mapsto p, \end{aligned}$$

where p is the unique $p \in M$ such that $X \in T_p M$. The map π naturally associates with each tangent vector the point at which it is tangent [27]. In fact, for any smooth manifold M of dimension d, the set TM is a smooth $2d$–dimensional manifold. This can be shown by constructing a smooth atlas on TM from a given smooth atlas on M [33]. The tangent bundle (TM, π, M) is a so–called *vector bundle*, i.e. a fiber bundle whose fibers are vector spaces. In an abuse of notation TM alone is often called the tangent bundle.

Cotangent Bundle Completely analogous we define the *cotangent bundle* as the vector bundle (T^*M, π, M), where

$$T^*M \coloneqq \bigcup_{p \in M} T_p^* M,$$

is the disjoint union of all cotangent spaces of M. The bundle projection map π is defined accordingly.

2.3.1 Vector fields

Consider (TM, π, M) as above. We define a *vector field* as a smooth section σ of the tangent bundle, i.e. it satisfies $\pi \circ \sigma = id_M$. A vector field is a smooth assignment of a tangent vector at

each point of M. We call the set of all vector fields $\Gamma(TM) \coloneqq \{\sigma : M \to TM \mid \pi \circ \sigma = id_M\}$ (sometimes also $\mathfrak{X}(M)$) and equip it with two operations[16].

1.
$$\begin{aligned} \oplus : \Gamma(TM) \times \Gamma(TM) &\to \Gamma(TM), \\ (\sigma, \tau) &\mapsto \sigma \oplus \tau, \end{aligned}$$

where

$$\begin{aligned} \sigma \oplus \tau : M &\to \Gamma(TM), \\ p &\mapsto (\sigma \oplus \tau)(p) \coloneqq \sigma(p) + \tau(p), \end{aligned}$$

2.
$$\begin{aligned} \odot : C^\infty(M) \times \Gamma(TM) &\to \Gamma(TM), \\ (f, \sigma) &\mapsto f \odot \sigma, \end{aligned}$$

where

$$\begin{aligned} f \odot \sigma : M &\to \Gamma(TM), \\ p &\mapsto (f \odot \sigma)(p) \coloneqq f(p) \cdot \sigma(p), \end{aligned}$$

Apparently, we can give the set of all vector fields a "vector space–like" structure itself. However, notice that $(C^\infty, +, \bullet)$ is a ring, not a field, and therefore $\Gamma(TM)$ is not a vector space. Instead we call it a *module*, and we call $(\Gamma, \oplus, \odot)$ a C^∞–module. In contrast to a vector space, a module needn't have a basis in general[17].

2.3.2 Integral curves and Flows

Let M be a n-dimensional smooth manifold and let X be a vector field on M, i.e. $X \in \Gamma(TM)$. A smooth curve $\gamma : I \to M$ on M is an *integral curve* of X if

$$\dot{\gamma}(t) = X_{\gamma(t)}, \quad \forall t \in I. \tag{1}$$

In other words, an integral curve of a vector field is a smooth curve whose tangent vectors are given by the values of this vector field along the curve [34], as illustrated in Fig. 9. We choose local coordinates, such that

$$\gamma(t) = (\gamma^1(t), \dots, \gamma^n(t)),$$

and $X^i(\gamma(t))$ are the coefficients of X at $\gamma(t)$. We can infer from (1) that γ is an integral curve of X iff it locally satisfies the system of first–order ordinary differential equations [35]

$$\dot{\gamma}^i(t) = X^i(\gamma(t)), \quad i = 1, \dots, n. \tag{2}$$

If $X^i(\gamma(t))$ is smooth, the system (2) is guaranteed to have a unique solution for each set of initial data. This in turn implies the existence of a unique *maximal* integral curve $\gamma_{max} : I \to M$ passing through a given point $\gamma_{max}(0) \in M$. Maximal means that γ_{max} is not contained in any longer integral curve, i.e. if $\tilde{\gamma} : \tilde{I} \to M$ is any other integral curve with the same initial value $\tilde{\gamma}(0) = \gamma_{max}(0)$, then $\tilde{I} \subset I$ and $\tilde{\gamma}(t) = \gamma_{max}(t)$ for $t \in \tilde{I}$ [36].

[16] $+$ and $\cdot$ are the addition and scalar multiplication on T_pM.

[17] However, if the underlying ring is a division ring, i.e. a ring in which division is possible, the module is guaranteed to have a basis. Therefore, every vector space has a basis, since any field is a division ring.

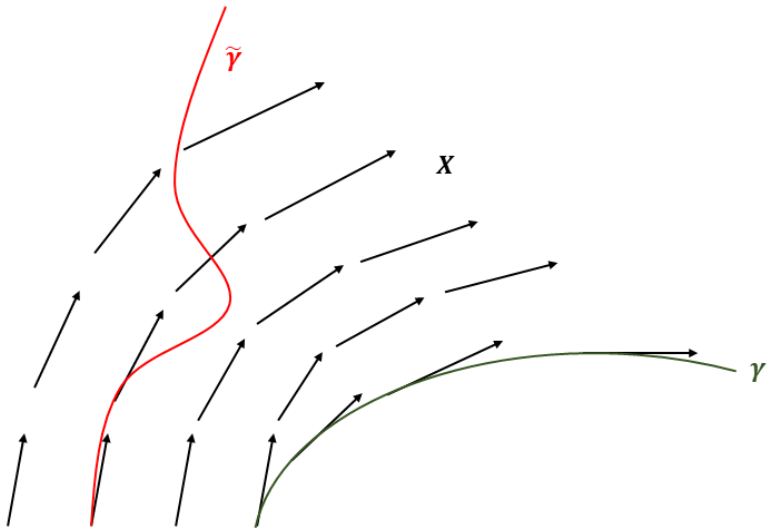

Figure 9: The curve γ is an integral curve of X, whereas $\tilde{\gamma}$ is not.

Flows Denote the maximal integral curve passing through a point p in M by $\Psi_t(p)$ and call Ψ the *flow* generated by X. For each $p \in M$ and $t \in I \subseteq \mathbb{R}$ containing 0, $\Psi_t(p)$ is a point on the integral curve passing through p. The flow of a vector field can be viewed as a smooth map

$$\Psi : I \times M \to M,$$

satisfying the basic properties [36, 37]

1. $\Psi_s(\Psi_t(p)) = \Psi_{s+t}(p), \quad \forall s, t \in \mathbb{R}, \quad \text{and } p \in M,$
2. $\Psi_0(p) = p,$
3. $\frac{d}{dt}\Psi_t(p) = X|_{\Psi_t(p)}.$

The reason for the terminology "flow" can be understood in the following way. Picture X as a velocity vector field of some steady state fluid. The integral curves of X are the stream lines followed by the fluid particles. If a particle started at position p at time $t = 0$, then the flow $\Psi_t(p)$ is the position of the particle after "flowing" for the amount of time t on its integral curve.

Complete vector fields A vector field X is said to be *complete* if all of its integral curves have the domain $I = \mathbb{R}$ [37], i.e. every one of its integral curves exists for all time. In this case,

$$\Psi_t(\Psi_{-t}(p)) = \Psi_0(p) = p, \quad \forall t \in \mathbb{R},$$

hence the Ψ_t are diffeomorphisms of M and we call the family $\{\Psi_t\}_{t\in\mathbb{R}}$ *one–parameter group of diffeomorphisms* [37].

2.3.3 Second–order differential equations

The geometrical description of integral curves and flows will find applications in the context of Hamiltonian mechanics, where the equations of motion are first–order differential equations. As

we will see later, the Lagrangian formulation of classical mechanics involves less variables, but as a consequence one has to deal with second–order differential equations. In the following they will be introduced from a geometrical standpoint. Let M be a smooth manifold and consider its tangent bundle TM as a smooth manifold in its own right. The tangent bundle $T(TM)$ has two different vector bundle structures over the base TM. One is given by the natural projection

$$\tau_{TM} : T(TM) \to TM,$$

while the other is given by the bundle map

$$\tau_* : T(TM) \to TM,$$

which is the the Push–forward of

$$\tau : TM \to M.$$

The vector fields X in TM are sections with respect to the natural projection, i.e. $\tau_{TM} \circ X = id_{TM}$. Those vector fields, that are also a section with respect to τ_* are a special type of vector field: They are the geometric version of second–order differential equation systems. We define a *second–order differential equation field* as a vector field $X \in \mathfrak{X}(TM)$, such that [38]

$$\tau_* \circ X = id_{TM}.$$

The situation is illustrated in Fig 10.

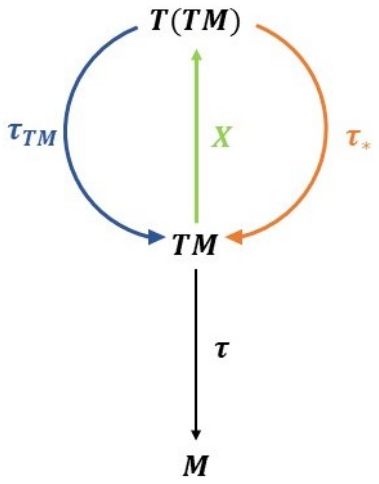

Figure 10: A second–order differential equation field X consists of simultaneous sections with respect to τ_{TM} and τ_*.

2.3.4 Lie algebra

Let M be a smooth manifold and let $X, Y \in \Gamma(TM)$. The *Lie bracket* (or *commutator*) of X and Y is defined as

$$\begin{aligned}[X,Y] : C^\infty(M) &\xrightarrow{\sim} C^\infty(M) \\ f &\mapsto [X,Y](f) \coloneqq X(Y(f)) - Y(X(f)).\end{aligned}$$

As a map

$$[\cdot,\cdot]:\Gamma(TM)\times\Gamma(TM)\to\Gamma(TM),$$

it satisfies the properties

1. $\mathbb{R}$–bilinearity,
2. Antisymmetry,
3. Jacobi identity.

Hence, the pair $(\Gamma(TM),[\cdot,\cdot])$[18] has the structure of an algebra, and we call it *Lie algebra* of vector fields [34].

Lie subalgebras Let $X_1,\dots,X_s$ be vector fields on M, such that

$$[X_i,X_j]=C_{ij}^k X_k,\quad i,j=1,\dots,s.$$

The coefficients $C_{ij}^k\in\mathbb{R}$ are called *structure constants* of the *Lie subalgebra* $L\coloneqq(\mathrm{span}_{\mathbb{R}}\{X_1,\dots,X_s\},[\cdot,\cdot])$ [39].

2.3.5 Tensor fields

We can extend the concept of a vector field to include the idea of a (r,s)–*tensor field* on M. This is, in complete analogy, a smooth assignment of a tensor of type (r,s) to each point $p\in M$ [27]. We could therefore understand a tensor field as a section of the tensor bundle of type (r,s). Equivalently, a tensor field τ on M is as a $C^\infty(M)$–multilinear map

$$\tau:\underbrace{\Gamma(T^*M)\times\cdots\times\Gamma(T^*M)}_{r\text{ copies}}\times\underbrace{\Gamma(TM)\times\cdots\times\Gamma(TM)}_{s\text{ copies}}\to C^\infty(M).$$

Notice, that we could have defined a vector field (being a special kind of tensor field) as a derivation on the algebra $C^\infty(M)$, i.e. an $\mathbb{R}$–linear map $\sigma:C^\infty\xrightarrow{\sim}C^\infty(M)$. Due to the point–wise nature of tensors both definitions are equivalent, hence we can pick whichever suits our purpose the most.

2.3.6 Differential forms

Our discussion of differential forms is mainly based on [27] and [34]. If necessary, we give further references throughout the text. Let M be a smooth manifold and $X_1,\dots,X_k\in\Gamma(TM)$. A *(differential) k–form* is a $(0,k)$–tensor field ω that is antisymmetric under the exchange of any pair of indices, i.e.

$$\omega(X_1,\dots,X_k)=\mathrm{sgn}(\pi)\omega(X_{\pi(1)},\dots,X_{\pi(k)}).$$

We call $\pi\in\mathrm{Perm}(k)$ a permutation of the integers $\{1,\dots,k\}$ and $\mathrm{sgn}(\pi)$ is the signature of the permutation, defined as $+1$ if π is even and -1 if π is odd. The set of all k–forms on M is denoted by $\Omega^k(M)$ and can be made into a $C^\infty(M)$-module by defining the addition and multiplication operations point–wise. Notice that $\Omega^0(M)\equiv C^\infty(M)$ and $\Omega^1(M)\equiv\Gamma(T_1^0M)\equiv\Gamma(T^*M)$.

[18] Regard $\Gamma(TM)$ as an $\mathbb{R}$–vector space.

Pull–back of forms In Fig. (11) we show the action of the pull–back of a covector (or 1–form), which we have introduced earlier. We can generalize this idea to forms of arbitrary degree[19]. Let $\phi : M \to N$ be a smooth map and let $\omega \in \Omega^k(N)$. We define the pull–back $\phi^*(\omega) \in \Omega^k(M)$ of ω as

$$\begin{aligned} \phi^*(\omega) : M &\to T^*M, \\ p &\mapsto \phi^*(\omega)(p), \end{aligned}$$

where $\phi^*(\omega)(p)(X_1, \ldots, X_k) \coloneqq w(\phi(p))(\phi_*(X_1), \ldots, \phi_*(X_k))$, for $X_i \in T_pM$.

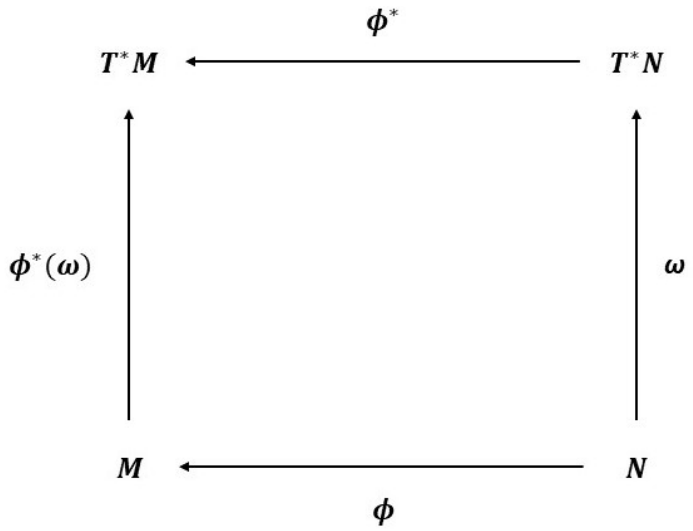

Figure 11: Action of the pull–back of $\omega \in \Gamma(T^*N)$

Wedge product The tensor product $\otimes$ does not possess all the desired properties for a product of forms, e.g. the tensor product of two forms is not necessarily a form. We therefore define the *wedge* (or *exterior*) *product* of forms as the map

$$\begin{aligned} \wedge : \Omega^k(M) \times \Omega^l(M) &\to \Omega^{k+l}(M), \\ (\omega, \sigma) &\mapsto w \wedge \sigma, \end{aligned}$$

where

$$(w \wedge \sigma)(X_1, \ldots, X_{k+l}) \coloneqq \frac{1}{k!l!} \sum_{\pi \in \mathrm{Perm}(k+l)} \mathrm{sgn}(w \otimes \sigma)(X_{\pi(1)}, \ldots, X_{\pi(k+l)})$$

and $X_1, \ldots, X_{k+l} \in \Gamma(TM)$. As an example, suppose that $\omega, \sigma \in \Omega^1(M)$. Then, for any $X, Y \in \Gamma(TM)$ we find

$$(\omega \wedge \sigma)(X, Y) = (w \otimes \sigma - \sigma \otimes \omega)(X, Y),$$

hence we write $\omega \wedge \sigma = (w \otimes \sigma - \sigma \otimes \omega)$.

Let $\phi : M \to N$ be smooth, $\omega \in \Omega^k(M)$, and $\sigma \in \Omega^l(N)$. Two, particularly convenient, properties of the wedge product are

1. The pull–back distributes over the wedge product, such that $\Phi^*(\omega \wedge \sigma) = \Phi^*(\omega) \wedge \Phi^*(\sigma)$, and
2. $(\omega \wedge \sigma) = (-1)^{k \cdot l}(\sigma \wedge \omega)$.

[19] In a similar way one can define the pull–back of any smooth (p, q) tensor field.

Exterior derivative We can extend the definition of the gradient operator at a point $p \in M$ and understand it as an operator that takes in 0–forms and produces 1–forms

$$\mathrm{d} : \Omega^0(M) \xrightarrow{\sim} \Omega^1(M).$$

This, in turn, can be extended to an operator which acts on any k–form. The *exterior derivative* on M is the $\mathbb{R}$-linear operator (formally a *derivation*)

$$\begin{aligned} \mathrm{d} : \Omega^k(M) &\xrightarrow{\sim} \Omega^{k+1}(M), \\ \omega &\mapsto \mathrm{d}\omega, \end{aligned}$$

where

$$\begin{aligned} \mathrm{d}\omega(X_1, \ldots, X_{k+1}) \coloneqq &\sum_{i=1}^{k+1} (-1)^{i+1} X_i(\omega(X_1, \ldots, \cancel{X_i}, \ldots, X_{k+1})) \\ &+ \sum_{i<j} (-1)^{i+j} \omega([X_i, X_j], X_1, \ldots, \cancel{X_i}, \ldots \cancel{X_j}, \ldots, X_{k+1}), \end{aligned}$$

and $X_i \in \Gamma(TM)$. In the case $k = 1$, the form $\mathrm{d}\omega \in \Omega^2(M)$ is given by

$$\mathrm{d}\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y]).$$

The exterior derivative satisfies $\mathrm{d} \circ \mathrm{d} = 0$ as well as a graded version of the Leibniz rule with respect to the wedge product. Let $\omega \in \Omega^k(M)$ and $\sigma \in \Omega^l(M)$ [40]. Then

$$\mathrm{d}(\omega \wedge \sigma) = \mathrm{d}\omega \wedge \sigma + (-1)^k \omega \wedge \mathrm{d}\sigma.$$

Furthermore, if $\phi : M \to N$ is a smooth map and $\omega \in \Omega^k(N)$, one can show that

$$\Phi^*(\mathrm{d}\omega) = \mathrm{d}(\Phi^*(\omega)).$$

Informally, one sometimes says that the exterior derivative "commutes" with the pull–back.

Interior product Let $\omega \in \Omega^k(M)$ and $X \in \Gamma(TM)$. Another basic operation we can perform is the *contraction* of ω with X and intuitively we just insert the vector field X into the first slot of ω. We can make our intuition precise by understanding this operation as a map

$$\begin{aligned} \iota_X : \Omega^k(M) &\to \Omega^{k-1}(M), \\ \omega &\mapsto \iota_X \omega, \end{aligned}$$

defined as $\iota_X \omega(X_1, \ldots, X_{k-1}) \coloneqq \omega(X, X_1, \ldots, X_{k-1})$ [41]. Like the exterior derivative it is a graded derivation on $\Omega^n(M)$, i.e. it satisfies

$$\iota_X(\omega \wedge \sigma) = (\iota_X \omega) \wedge \beta + (-1)^k \omega \wedge (\iota_X \sigma),$$

$\forall \omega \in \Omega^k(M)$ and σ a form of arbitrary degree.

Closed and exact forms We say that a differential form $\omega \in \Omega^k(M)$ is *closed* if $\mathrm{d}\omega = 0$, and *exact* if there exists a $(k-1)$-form η on M such that $\omega = \mathrm{d}\eta$. The fact that $\mathrm{d} \circ \mathrm{d} = 0$ implies that every exact form is closed, but the converse may not be true [40].

2.3.7 Lie derivative

The *Lie derivative* $\pounds$ with respect to a smooth vector field X on a smooth manifold M extends the directional derivative along the integral curves of X from functions to tensor fields T on M. The idea is to compare the value of $T \in \Gamma(T^p_q M)$ at a given point m with its values along the integral curve through m, transported back to m by the flow of X [34]. We define

$$(\pounds_X T)_m \coloneqq \left.\frac{d}{dt}\right|_{t=0} ((\Psi_{-t})_* T)_m, \tag{3}$$

where Ψ denotes the flow generated by X. The Lie derivative satisfies the following properties.

1. $\pounds_X f = X(f)$,
2. $\pounds_X Y = [X, Y]$,
3. $\pounds_X (T + S) = \pounds_X T + \pounds_X S$,
4. $\pounds_X (T(\omega, Y)) = (\pounds_X T)(\omega, Y) + T(\pounds_X \omega, Y) + T(\omega, \pounds_X Y)$,
5. $\pounds_{X+Y} = \pounds_X T + \pounds_Y T$,

where $f \in C^\infty(M)$, $X, Y \in \Gamma(TM)$, $\omega \in \Omega^1(M)$ and T a $(1,1)$–tensor field. Property 4 can easily be extended to arbitrary (p, q)–tensor fields.

3 Geometry: additional structure

We add the last puzzle pieces to our formal construction and define (pseudo–)Riemannian and symplectic geometries. We call a smooth manifold, equipped with a specific $(0, 2)$ tensor field, a (pseudo–)Riemannian manifold. This additional structure allows to define intuitive notions like distances and the length of a curve on the manifold. Even though it is beyond the scope of this thesis to define them in their full glory, we briefly introduce connections and relate them to the covariant derivative of tensor fields. We discuss symmetries of the metric, known as isometries, and their infinitesimal generators – Killing fields.

If a smooth manifold is equipped with a non–degenerate, closed 2–form, we call it a symplectic manifold. One of the most fundamental theorems in the context of symplectic geometry is Darboux's theorem. It allows to choose canonical local coordinates in every open neighborhood around a point of the symplectic manifold. Similar to how we can always represent a metric at every point in its diagonal form with entries ± 1, it is possible to always find the so–called normal form of a symplectic form. The development of symplectic geometry is closely related to the geometrical formulation of Hamiltonian mechanics, which will already be obvious from the terminology at hand. We introduce Hamiltonian vector fields and the Poisson bracket in purely geometrical terms and see how the cotangent bundle admits a natural symplectic structure.

At the end of this section we have, by systematically equipping sets with more and more structure, arrived at the very language in which most of modern theoretical physics is formulated. The pinnacle of our construction are Riemannian manifolds and the cotangent bundle of a configuration space, that naturally admits a symplectic structure. The link between symplectic geometry and Hamiltonian mechanics at this point becomes self–evident. On these spaces, we can perform all the operations we shall naively require and we understand their origin and limitation. Thus, we finally have all the machinery at hand to discuss singular systems in a thorough manner.

3.1 Riemannian geometry

Despite its clear intuition, we did not introduce the notion of a metric up to this point. This is due to the simple fact that it was not necessary yet. As we stressed before, we try to establish the basic concepts of differential geometry in a systematic manner, in particular not assume more structure than needed at any given point. But when it comes to physical systems, our manifolds will, of course, always be equipped with a metric. We therefore need to introduce this concept in the following.

3.1.1 Aside: metric spaces

In some sense a *metric space* is a special kind of topological space, since every metric space determines a unique canonical topology. A metric function on the other hand is not, in general, consistent with the topology on a topological space[20] [25]. In Fig. 12 we try to put the concept of a metric space into perspective. Notice how we took the left path to arrive at differentiable manifolds without the need for a metric. Usually the defining properties of a topological manifold are strong enough to guarantee its metrizability[21].

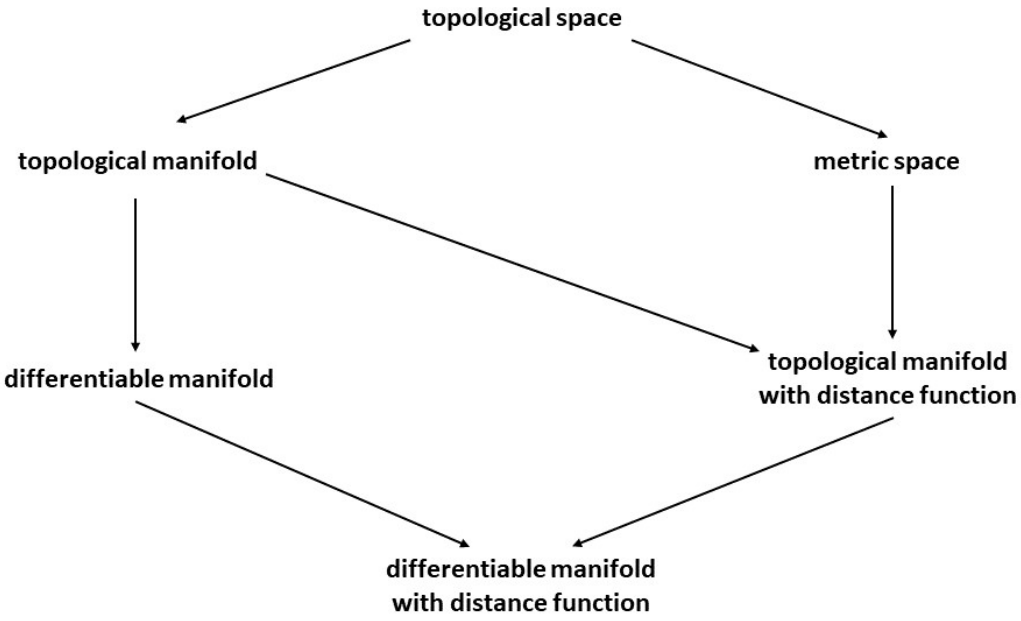

Figure 12: The role of metric spaces

Metric A metric space is a set A together with a function

$$\text{dist} : A \times A \to \mathbb{R},$$

called a *metric*. It satisfies the following three properties:

[20] In other words: Not every topological space is metrizable.
[21] See e.g. the Urysohn Metrization Theorem.

1. $\text{dist}(x, y) \geq 0$ and $\text{dist}(x, y) = 0 \Leftrightarrow x = y$ (Positivity)
2. $\text{dist}(x, y) = \text{dist}(y, x)$ (Symmetry)
3. $\text{dist}(x, z) \leq \text{dist}(x, y) + \text{dist}(y, z)$ (Triangle inequality)

3.1.2 Riemannian manifolds

The metric tensor Let M be a smooth manifold. A *Riemannian metric* on M is a tensor field $g \in \Gamma(T_2^0 M)$ such that

1. $\forall p \in M, \forall V_1, V_2 \in T_p M : \; g_p(V_1, V_2) = g(p)(V_2, V_1)$ and
2. $\forall p \in M, \forall V \in T_p M \setminus \{0\} : \; g_p(V, V) > 0.$

In other words, at each point of the manifold g_p is a symmetric positive definite bilinear form. One usually omits the lowercase p unless it may cause confusion. If we relax the requirement of positive-definiteness and the metric instead satisfies for any $V_1 \in T_p M$

$$g(V_1, V_2) = 0 \Leftrightarrow V_2 = 0,$$

we call it *pseudo–Riemannian* metric. By making use of the spectral theorem we can always diagonalize a symmetric matrix and Sylvester's law allows us to choose a basis such that the entries of the metric are ± 1. We can use this to categorize a metric by its *signature*, i.e. a pair (i, j), where i is the number of positive eigenvalues and j the number of negative eigenvalues. The case of $j = 1$ is called *Lorentz metric* [42].

Definition A *(smooth) Riemannian manifold* is a pair (M, g), where M is a smooth manifold and g a Riemannian metric. A Riemannian manifold is a differentiable manifold with a metric-space distance function which is required to have a particular kind of relation to the differentiable structure.

The notion of distance On a Riemannian manifold (M, g) the metric assigns, in a differentiable fashion, a positive definite inner product $\langle \cdot, \cdot \rangle$ in each tangent space $T_p M$ [43]. The *norm* (or *length*) of a tangent vector $v_p \in T_p M$ is defined by $\|v_p\| \coloneqq \sqrt{g_p(v_p, v_p)}$ and the length of a curve $\gamma : [a, b] \to M$ is defined by [29]

$$L_\gamma \coloneqq \int_a^b \|\gamma'(t)\| dt.$$

Local representation Let (U, ϕ) be a chart on M with coordinates $\{x^\mu\}$. Since g is a $(0, 2)$ tensor field we can express it in terms of the canonical basis as

$$g_p = g_{\mu\nu}(p) dx^\mu \otimes dx^\nu,$$

and it is common practice to regard $(g_{\mu\nu})$ as a matrix whose $(\mu, \nu)^{\text{th}}$ entry is $g_{\mu\nu}$. Furthermore, since $(g_{\mu\nu})$ has maximal rank, it has an inverse[22] denoted by $(g^{\mu\nu})$, which satisfies

$$g_{\mu\nu} g^{\nu\lambda} = \delta_\mu^\lambda.$$

[22] Actually, since the metric tensor is a bilinear map $g_p : T_p M \times T_p M \to \mathbb{R}$, it cannot be inverted in the way a linear map can. This issue can be resolved with musical isomorphisms [33], where one holds one of the inputs of g_p constant and ends up with a linear map.

Notice that a Riemannian metric gives rise to an isomorphism between T_pM and T_p^*M and with the help of the inverse metric we can write [42]

$$\omega_\mu = g_{\mu\nu}U^\nu \quad \text{and} \quad U^\mu = g^{\mu\nu}\omega_\nu,$$

for $\omega \in T_p^*M$ and $U \in T_pM$. This is often called *lowering and raising indices* of tensors

3.1.3 Connections

A *connection* is related to the concepts of *parallel transport* and *covariant differentiation* on a differentiable manifold. In fact, the terms connection and covariant derivative are often used as synonyms at this stage. We already know that a vector $X_p \in T_pM$ is a directional derivative acting on a function $f \in C^\infty(M)$ as $X_p : f \mapsto X_p(f)$, but there is no notion of a directional derivative acting on a vector field (let alone a general (p, q) tensor field). By introducing a connection we will be able to transport tensors along a curve and hence specify a directional derivative of tensors.

Affine connection An *affine* (or *linear*) *connection* ∇ is a map

$$\begin{aligned}\nabla : \Gamma(TM) \times \Gamma(TM) &\to \Gamma(TM),\\ (X, Y) &\mapsto \nabla_X Y,\end{aligned}$$

which satisfies the following conditions [42]:

1. $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$,
2. $\nabla_{(X+Y)}Z = \nabla_X Z + \nabla_Y Z$,
3. $\nabla_{(fX)}Y = f\nabla_X Y$,
4. $\nabla_X(fY) = X(f)Y + f\nabla_X Y$,

where $f \in C^\infty(M)$ and $X, Y, Z \in \Gamma(TM)$.

Connection coefficients Let (U, ϕ) be a chart with the coordinate $x = \phi(p)$ on M. Let $\{e_\nu\} = \{\frac{\partial}{\partial x^\nu}\}$ be the coordinate basis in T_pM. We define the *connection coefficients* $\Gamma^\lambda_{\mu\nu}$ as the functions that specify how the basis vectors change from point to point, such that

$$\nabla_\mu e_\nu = e_\lambda \Gamma^\lambda_{\mu\nu}.$$

Once we know how ∇ acts on the basis vectors, we can calculate its action on any vector. For $V, W \in T_pM$ we find

$$\begin{aligned}\nabla_V W = V^\mu \nabla_\mu (W^\nu e_\nu) &= V^\mu (e_\mu(W^\nu)e_\nu + W^\nu \nabla_\mu e_\nu)\\ &= V^\mu \left(\frac{\partial W^\lambda}{\partial x^\mu} + W^\nu \Gamma^\lambda_{\mu\nu}\right) e_\lambda.\end{aligned}$$

That means, ∇ maps the two vectors V and W to a new vector, whose λ^{th} component is $V^\mu \nabla_\mu W^\lambda$ and we make the identification

$$(\nabla_\mu W)^\lambda \equiv \frac{\partial W^\lambda}{\partial x^\mu} + \Gamma^\lambda_{\mu\nu} W^\nu.$$

It is common to omit the brackets and simply write $\nabla_\mu W^\lambda$ but one should not mistake it for a covariant derivative of a component W^λ [29, 42].

Covariant derivative of tensor fields We define, in a natural way, the covariant derivative of a function $f \in C^\infty(M)$ by the usual directional derivative, such that

$$\nabla_X f \equiv X(f).$$

By doing so, we find that the affine connection satisfies the Leibniz rule $\nabla_X(fY) = (\nabla_X f)Y + f\nabla_X Y$ and we generalize this idea to the product of any two tensor fields T_1 and T_2 of arbitrary type:

$$\nabla_X(T_1 \otimes T_2) = (\nabla_X T_1) \otimes T_2 + T_1 \otimes (\nabla_X T_2).$$

We can first compute the covariant derivative of a 1–form $\omega \in \Omega(M)$ and then straightforwardly generalize to arbitrary tensor fields. For $Y \in \Gamma(TM)$ we know that $\iota_Y \omega = \langle \omega, Y \rangle \in C^\infty(M)$, such that [42]

$$X(\langle \omega, Y \rangle) = \nabla_X(\langle \omega, Y \rangle) = \langle \nabla_X \omega, Y \rangle + \langle \omega, \nabla_X Y \rangle.$$

Analogous to the vector case above we can go to a local chart, in terms of components we find

$$(\nabla_\mu \omega)_\nu = \partial_\mu \omega_\nu - \Gamma^\lambda_{\mu\nu} \omega_\lambda,$$

i.e. the connection coefficients pick up a minus sign for covariant objects. The covariant derivative of general (p, q) tensor T therefore reads

$$\begin{aligned} \nabla_\nu T^{\lambda_1 \ldots \lambda_p}_{\mu_1 \ldots \mu_q} = \partial_\nu T^{\lambda_1 \ldots \lambda_p}_{\mu_1 \ldots \mu_q} + \Gamma^{\lambda_1}_{\nu\kappa} T^{\kappa \lambda_2 \ldots \lambda_p}_{\mu_1 \ldots \mu_q} + \cdots + \Gamma^{\lambda_p}_{\nu\kappa} T^{\lambda_1 \ldots \lambda_{p-1} \kappa}_{\mu_1 \ldots \mu_q} \\ - \Gamma^\kappa_{\nu\mu_1} T^{\lambda_1 \ldots \lambda_p}_{\kappa \mu_2 \ldots \mu_q} - \cdots - \Gamma^\kappa_{\nu\mu_q} T^{\lambda_1 \ldots \lambda_p}_{\mu_1 \ldots \mu_{q-1} \kappa}. \end{aligned}$$

Remarks Obviously, our quick introduction to connections was less rigorous than what we did before. We gave a minimum working definition of a connection as one often finds in undergraduate text books. That does not mean that one cannot define a connection in more sophisticated (and less coordinate dependent) ways — it is simply beyond the scope of this work to do so. One would have to introduce *principle fiber bundles* and *Lie groups* to discuss connections in their full geometrical glory, see e.g. [27, 33]. The basic idea is to compare the points in "neighboring" fibers in a way that is not dependent on any particular local bundle trivialization[23]. In this context it becomes clear that all the concepts we usually use synonymous are not equivalent; the hierarchy goes something like this:

Choose a connection on a principal G–bundle
↳parallel transport on the principal bundle
↳parallel transport on any associated fiber bundle
↳covariant derivative on a vector bundle

In order to define a covariant derivative we need the linear structure in the fiber of a vector bundle. So it is possible to talk about parallel transport in any of the associated fiber bundles, but only when an associated bundle is a vector bundle it is possible to define a covariant differentiation [44].

[23] A fiber bundle (E, π, M) is trivial if it is M–isomorphic to the product bundle $((M \times F), pr_1, M)$ for some space F. It is locally trivial if its is locally isomorphic to the product bundle.

3.1.4 Isometries

Isometries are the structure preserving maps on the level of (pseudo)–Riemannian manifolds. In other words, they are the distance preserving transformations between metric spaces. Let $R = (M, g)$ and $R' = (M', g')$ be two (pseudo)–Riemannian manifolds and let $\phi : R \to R'$ be a (local) diffeomorphism. The map ϕ is called a (local) isometry if

$$\phi^* g' = g, \tag{4}$$

and the manifolds are then called *(locally) isometric* [33]. An isometry is a map, which pulls back the metric tensor on the second manifold to the metric tensor on the first. Such "symmetries of the metric" are important tools to find solutions to the field equations of General Relativity [45].

Killing fields It would be convenient to have a systematic procedure in order to find and characterize isometries. We thus search for the *generators* of isometries, which will take us back to our discussion on vector fields. Given a vector field X on a (pseudo)–Riemannian manifold (M, g), what are the necessary and sufficient conditions for the one–parameter group generated by X (i.e. its flow) to consist of local isometries? Compare the defining property of isometries (4) with the definition of the Lie derivative (3): by setting

$$\pounds_X g = \frac{d}{dt}\bigg|_{t=0} (\Psi_t^* g) = 0, \tag{5}$$

we obtain $\Psi_t^* g = g$ for all $t \in I$, since Eq. (5) holds for every point of M [46]. A vector field X on M satisfying $\pounds_X g = 0$ is called a *Killing field*. Killing fields are the infinitesimal generators of isometries and constitute a Lie algebra, i.e. the space of Killing fields is closed under the Lie bracket $[\cdot, \cdot]$ [46].

Killing equation Let X be a vector field on the (pseudo)–Riemannian manifold (M, g). The Lie derivative of a $(0, 2)$–tensor field T along X takes the form

$$\pounds_X T_{\mu\nu} = X^\sigma \nabla_\sigma T_{\mu\nu} + (\nabla_\mu X^\sigma) T_{\sigma\nu} + (\nabla_\nu X^\sigma) T_{\mu\sigma},$$

where ∇ represents any symmetric covariant derivative. If we now set $T = g$, we find that

$$\pounds_X g_{\mu\nu} = X^\sigma \nabla_\sigma T_{\mu\nu} + (\nabla_\mu X^\sigma) T_{\sigma\nu} + (\nabla_\nu X^\sigma) T_{\mu\sigma} = \nabla_\mu X_\nu + \nabla_\nu X_\mu,$$

assuming that $\nabla_\sigma g_{\mu\nu} = 0$[24]. In this case, X is a Killing field if it satisfies the *Killing equation*

$$\nabla_\mu X_\nu + \nabla_\nu X_\mu = 0.$$

[24] This is called metric–compatibility and we will discuss it in more detail in Sec. 7 on General Relativity.

3.2 Symplectic geometry

3.2.1 Symplectic manifolds

A *symplectic structure* on a smooth manifold M is defined by a 2–form $\Omega \in \Omega^2(M)$, which satisfies the following properties:

1. At every point $p \in M$ of the manifold $\Omega(p)$ is non–degenerate, and
2. The 2–form Ω is closed, i.e. $\mathrm{d}\Omega = 0$.

We call the pair (M, Ω) *symplectic manifold* and Ω *symplectic form* [41]. Notice that a symplectic manifold is of even dimension, and its tangent bundle is a symplectic vector bundle[25]. If the 2–form on M does not satisfy the non-degeneracy condition, we call it *presymplectic form*.

Symplectomorphisms A smooth mapping $\phi : M \to N$ between symplectic manifolds (M, Ω) and (N, ρ) is called symplectic, if $\phi^*\rho = \Omega$. If ϕ is additionally a (local) diffeomorphism, it is called a (local) *symplectomorphism* [34].

3.2.2 Intuition on symplectic geometry

To get a first intuition on symplectic geometry it can be helpful to compare it to the familiar Riemannian geometry in two dimensions.

Riemannian geometry For a Riemannian manifold (M, g) all the information is encoded in the metric g. In the case of $M = \mathbb{R}^2$ the metric just associates to two vectors in the plane, $v = (v_1, v_2)$ and $w = (w_1, w_2)$, the number $g(v, w) = v_1w_1 + v_2w_2$. One then defines the length of a vector v to be $\|v\| = \sqrt{g(v, v)}$ and similarly the angle θ between two vectors v and w as $\cos\theta = \frac{g(v,w)}{\|v\|\|w\|}$. Notice that v is at right angles to w iff $g(v, w) = 0$. The symmetry of g guarantees that the angle between v and w is the same as that between w and v, the non–degeneracy ensures that no non-zero vector can be perpendicular to every other vector.

Symplectic geometry The standard symplectic form Ω on $\mathbb{R}^2$ also associates to every two vectors v and w a number, in this case

$$\Omega(v, w) = v_1w_2 - v_2w_1.$$

As a result of the antisymmetry of Ω the symplectic form does not give rise to a notion of length or angle. The "symplectic lenght" $\|v\| = \sqrt{\Omega(v, v)}$ of a vector v always vanishes and every vector is perpendicular to itself! Instead, it exactly gives the area of the parallelogram that is formed by two vectors v and w as its sides. Thus the symplectic form makes precise the concept of *oriented area*. Non–degeneracy of Ω implies that only the collapsed parallelogram (when v and w are parallel) has zero area.

Symplectic geometry in higher dimensions Symplectic geometry is a purely *areal* type of geometry and area is essentially a two–dimensional construct. It is, in fact, not possible to consistently define an oriented area for odd–dimensional spaces without introducing degeneracies. But it is still possible to build a symplectic form on the four–dimensional space $\mathbb{R}^4$ in the following

[25] The total space is equipped with a symplectic structure.

way: View $\mathbb{R}^4$ as the sum of two planes $\mathbb{R}^2 \oplus \mathbb{R}^2$. Then the oriented area of a parallelogram in this space is the sum of the oriented areas of its shadows on these planes. This idea can be extended to $\mathbb{R}^6, \mathbb{R}^8, \ldots$ etc. [47].

3.2.3 Darboux's Theorem

The Darboux Theorem is fundamental for the theory of symplectic manifolds and its application to classical mechanics. Let (M, Ω) be a symplectic manifold of dimension $2n$. Darboux's Theorem states [41] that every point $m \in M$ has an open neighborhood U, which is the domain of a chart (U, ϕ) with local coordinates $q^1, \ldots q^n, p_1, \ldots, p_n$, such that the 2–form Ω has the local expression

$$\Omega = \mathrm{d}p_i \wedge \mathrm{d}q^i,$$

on U. Such a chart, and its associated local coordinates, are said to be *canonical*. This means, in particular, that two symplectic manifolds of the same dimension are always locally symplectomorphic and hence there are no local invariants[26] and the two symplectic manifolds can at most differ globally [34].

3.2.4 Natural isomorphism

Provided with the structure of a symplectic manifold (M, ω) we can construct the *natural isomorphism* between the tangent spaces of M. We introduce it as the map

$$\begin{aligned} \flat : TM &\to T^*M, \\ X &\mapsto X^\flat, \end{aligned}$$

defined by $X^\flat := \iota_X w$ and referring to musical notation[27] it is called *flat*. Its inverse is the map

$$\sharp : T^*M \to TM,$$

called *sharp*. These are exactly the *musical isomorphisms* one uses to specify the notion of an inverse metric on a (pseudo–)Riemannian manifold where they are used to lower and raise indices [40].

Hamiltonian vector field With the help of the natural isomorphism we can assign a vector field to every smooth function $f \in C^\infty(M)$. We call

$$X_f := -(\mathrm{d}f)^\sharp,$$

the *Hamiltonian vector field* generated by f and in canonical coordinates it has the form [34]

$$X_f = \left(\frac{\partial f}{\partial p_i}\right)\frac{\partial}{\partial q^i} - \left(\frac{\partial f}{\partial q^i}\right)\frac{\partial}{\partial p_i}.$$

Poisson bracket Let (M, Ω) and $f, g \in C^\infty(M)$. We define the *Poisson bracket* of f and g as

$$\{f, g\} := \Omega(X_f, X_g) = X_f(g),$$

and in a Darboux chart we recover the well known expression from classical mechanics

$$\{f, g\} = \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i} - \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i}.$$

[26] In sharp contrast to a Riemannian manifold where curvature is a local invariant.
[27] In the context of music $\flat$ and $\sharp$ denote a lower and higher pitch, respectively.

3.2.5 Symplectic structure on the cotangent bundle

Canonical 1–form Let M be a n–dimensional smooth manifold and $\xi \in T^*M$ a point in its cotangent bundle. The *canonical* (also *tautological* or *Liouville*) 1–form is defined by [48, 49]

$$\theta_\xi(v) \coloneqq \xi(\pi_*(v)),$$

where $v \in T_\xi(T^*M)$ and $\pi : T^*M \to M$ is the bundle projection and π_* its Push–forward, i.e. $\pi_* : T(T^*M) \to TM$. In other words the value of the canonical 1–form at $\xi \in T^*M$ on a tangent vector v is obtained by projecting the vector to a tangent vector on M and evaluating the form ξ. In Darboux coordinates the canonical 1–form is given by

$$\theta = p_i \mathrm{d}q^i.$$

The canonical 1–form θ is the unique 1–form on T^*M which satisfies $\mu^*\theta = \mu$ for every 1–form $\mu \in \Omega^1(M)$.

Canonical symplectic form The exterior differential of the canonical 1–form defines a 2–form on T^*M, called the *canonical symplectic form*, as

$$\omega = -\mathrm{d}\theta.$$

Locally, it can be written as $\omega = \mathrm{d}x^i \wedge \mathrm{d}p_i$ [48].

Part II
Dynamics of constrained systems

4 Classical mechanics

There are two main frameworks in which one usually addresses mechanics: the Lagrangian and Hamiltonian formalisms. As a first step, we introduce both of them and discuss their underlying geometric structures. Mechanics deals not only with the dynamics of particles, but also with field theories such as electromagnetism and gravity [50]. We discuss the transition from particle systems to classical field theories in detail at the end of this section.

In the regular case, the Lagrangian and Hamiltonian formulations are equivalent, i.e. they generate the same equations of motions for a given physical system. Their mathematical content is, of course, different and there are various opinions on the question as to which one is more fundamental or natural.

Subsequently, we investigate the consequences of a singular Lagrangian. In this case, it is not possible to express all velocities in terms of the generalized momenta. This will lead us naturally to the notion of a constrained system. As it turns out, we will have to restrict the dynamics of a singular system in a certain way to guarantee that the equations of motion possess solutions.

Last but not least, we investigate the relation between the Lagrangian and Hamiltonian formulations and understand the well–know Legendre transformation as a special case of a fiber derivative. In the context of constrained systems, this can be used to classify objects as projectable and non–projectable.

4.1 Lagrangian formulation

4.1.1 Principle of stationary action

Lagrangian mechanics is based on the *principle of stationary action*, which plays a role of utmost importance in physics. It states that the dynamics of a physical system is determined by a variational problem for a functional based on a single function, the Lagrangian $\mathcal{L}$, which contains all physical information concerning the system. Even though it was originally formulated for classical mechanics it also applies to classical field theory and can be traced back to Feynman's path integral formulation of quantum mechanics [51]. Let Q be a differentiable manifold, called the *configuration space*. A Lagrangian $\mathcal{L} : TQ \to \mathbb{R}$ is a function on the tangent bundle TQ (sometimes called the *velocity phase space*) to the manifold [41, 50]. The action S of a system is defined as the functional[28]

$$S[q^i] = \int_{t_1}^{t_2} dt\ \mathcal{L}(q^i, \dot{q}^i)$$

where $q^i, \dot{q}^i \in TQ$ are the generalized positions and the generalized velocities, respectively. Consider some variation on the paths $q + \delta q$. According to the principle of stationary action the dynamics of the system between two fixed points is given by the family of curves that satisfies

$$\delta S[q^i] \overset{!}{=} 0.$$

[28] Formally, a functional is a mapping from a vector space V over some field F onto that field F.

Euler–Lagrange equations The variation δS is

$$\delta S = \int_{t_1}^{t_2} dt \, \big[\mathcal{L}(q+\delta q, \dot{q}+\delta\dot{q}) - \mathcal{L}(q,\dot{q})\big], \tag{6}$$

and for infinitesimal δq^i the first–order approximation

$$\mathcal{L}(q+\delta q, \dot{q}+\delta\dot{q}) = \mathcal{L}(q,\dot{q}) + \left(\frac{\partial\mathcal{L}}{\partial q^i}\delta q^i + \frac{\partial\mathcal{L}}{\partial\dot{q}^i}\delta\dot{q}^i\right) + \ldots,$$

holds. Thus, we find that (6) takes the form

$$\begin{aligned}
\delta S &= \int_{t_1}^{t_2} dt \left[\frac{\partial\mathcal{L}}{\partial q^i}\delta q^i + \frac{\partial\mathcal{L}}{\partial\dot{q}^i}\delta\dot{q}^i\right], \\
&= \int_{t_1}^{t_2} dt \left[\frac{\partial\mathcal{L}}{\partial q^i}\delta q^i + \frac{d}{dt}\big(\frac{\partial\mathcal{L}}{\partial\dot{q}^i}\delta q^i\big) - \big(\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{q}^i}\big)\delta q^i\right], \\
&= \int_{t_1}^{t_2} dt \left[\frac{\partial\mathcal{L}}{\partial q^i} - \frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{q}^i}\right]\delta q^i + \textit{surface term},
\end{aligned}$$

which must be valid for arbitrary variations δq^i. Furthermore, assuming that the variations at the boundary vanish, i.e $\delta q^i(t_1) = 0 = \delta q^i(t_2)$, the surface term vanishes. Then, the requirement $S[q^i] \stackrel{!}{=} 0$ yields the well-known Euler–Lagrange equations

$$\frac{d}{dt}\frac{\partial\mathcal{L}}{\partial\dot{q}^i} - \frac{\partial\mathcal{L}}{\partial q^i} = 0, \quad \forall i, \tag{7}$$

a system of $\dim(Q)$ second–order differential equations [50–53]. For later convenience, one can explicitly carry out the time derivative and write equation (7) as

$$W_{ij}\ddot{q}^j + \alpha_i \coloneqq \frac{\partial\mathcal{L}}{\partial q^i} - \frac{\partial^2\mathcal{L}}{\partial\dot{q}^i\partial q^j}\dot{q}^j - \frac{\partial^2\mathcal{L}}{\partial\dot{q}^i\partial\dot{q}^j}\ddot{q}^j = 0,$$

where

$$W_{ij} \equiv \frac{\partial^2\mathcal{L}}{\partial\dot{q}^i\partial\dot{q}^j},$$

is called the Hessian matrix. Notice at this point, that if the Hessian has an inverse, say $\widetilde{W}^{ij}$, the Euler–Lagrange equations can be solved for the accelerations as

$$\ddot{q}^i = \widetilde{W}^{ij}\alpha_j.$$

The inverse of the Hessian is guaranteed to exist if $\det W \neq 0$. A Lagrangian is said to be regular if this condition holds, otherwise it is called singular [51].

4.1.2 Geometrical approach

In the following q^i are coordinates on the smooth manifold M, and correspondingly (q^i, u^i) are the coordinates on TM $(i = 1, \ldots, \dim(M))$. For convenience, we write $\mathfrak{X}(M) \equiv \Gamma(TM)$ for the set of all vector fields on M.

Vertical lift Let M be a smooth manifold with tangent bundle TM and canonical bundle projection $\pi : TM \to M$. An element v of $T_{(q,u)}(TM)$ satisfies $\pi_* v = 0$ if and only if it is tangent to the fiber $\pi^{-1}(q)$. Such a vector is said to be *vertical* and a vector field V on TM is said to be vertical if $V_{(q,u)}$ is vertical at each point (q, u). The $\dim(M)$–dimensional subspace of $T_{(q,u)}TM$ consisting of vertical vectors is called the *vertical subspace*. Any element $\xi \in T_qM$ determines a vertical vector at any point (q, u) in the fiber over q, called its *vertical lift* to (q, u), denoted by $\xi^V_{(q,u)}$ [38, 54]. In local coordinates, the vertical lift of ξ (whose components are ξ^i) is $\xi^V = \xi^i \frac{\partial}{\partial u^i}$ [55].

Vertical endomorphism We construct a tensor field J of type $(1, 1)$, called *vertical endomorphism*, whose action is to map each tangent vector $v \in T_{(q,u)}TM$ to a vertical tangent vector, by first projecting it into T_qM and then lifting the result vertically to (q, u). We define the linear map [56]

$$J : \mathfrak{X}(TM) \to \mathfrak{X}^V(TM),$$

via $J_{(q,u)}(v) \coloneqq (\pi_* v)^V_{(q,u)}$, where $\mathfrak{X}^V(TM)$ denotes the subspace of vertical vector fields. Notice that J is the vertical lift of the identity tensor field, i.e. the Kronecker δ, from M to TM [57]. Locally, the vertical endomorphism is given by

$$J = \frac{\partial}{\partial u^i} \otimes \mathrm{d}q^i.$$

Liouville vector field The *Liouville vector field* (sometimes also *dilation field*) Δ is a vertical vector field generating the one–parameter group of dilations $\delta_t : (q, u) \mapsto (q, e^t u)$ [57]. In particular, any vector field $\Gamma \in \mathfrak{X}(TM)$ is a second-order differential equation field[29] if and only if $J(\Gamma) = \Delta$. In coordinates we can write it as

$$\Delta = u^i \frac{\partial}{\partial u^i}.$$

Lagrangian forms With the help of the objects we introduced above, we can now make the essential definitions for Lagrangian mechanics in purely geometrical terms. Notice how this intrinsic construction doesn't make use of any concepts from the Hamiltonian formulation, i.e. from the cotangent bundle. Let $\mathcal{L} \in C^\infty(TM)$, we define the 1–form $\theta_\mathcal{L} \in \Omega^1(TM)$, called *Euler–Poincaré 1–form*, as $\theta_\mathcal{L} \coloneqq \mathrm{d}\mathcal{L} \circ J$ [56]. In natural coordinates it reads

$$\theta_\mathcal{L} = \frac{\partial \mathcal{L}}{\partial u^i} \mathrm{d}q^i.$$

Furthermore, the exterior derivative of the Euler–Poincaré 1–form gives rise to a natural 2–form, called the *Lagrange 2–form*, as $\Omega_\mathcal{L} \coloneqq -\mathrm{d}\theta_\mathcal{L}$ [51, 58]. In fact, this natural 2–form is even a symplectic form and $(TM, \omega_\mathcal{L})$ is a symplectic manifold [59]. In coordinates we find

$$\Omega_\mathcal{L} = \frac{\partial^2 \mathcal{L}}{\partial q^j \partial u^i} dq^i \wedge dq^j + \frac{\partial^2 \mathcal{L}}{\partial u^j \partial u^i} dq^i \wedge du^j.$$

The only condition for $\mathcal{L}$ to be an adequate function to describe the dynamics of the system is that the rank of $\Omega_\mathcal{L}$ has to be constant [60]. If $\Omega_\mathcal{L}$ is of constant rank, then $\mathcal{L}$ is called the Lagrangian and if $\Omega_\mathcal{L}$ is of maximal rank, then the Lagrangian is regular [58].

[29] Locally, the projections of its integral curves onto M are the solutions of a system of second–order ordinary differential equations.

Lagrangian energy and equations of motion The *energy function* of the system, determined by the Lagrangian function $\mathcal{L}$, is defined as $E_{\mathcal{L}} \coloneqq \Delta(\mathcal{L}) - \mathcal{L}$ [38]. Its local coordinate expression is

$$E_{\mathcal{L}} = u^i \frac{\partial \mathcal{L}}{\partial u^i} - \mathcal{L}.$$

The possible dynamics of the system are determined by the solutions to the Lagrangian equations of motion,

$$\iota_\Gamma \Omega_{\mathcal{L}} = \mathrm{d}E_{\mathcal{L}}, \tag{8}$$

and in the regular case the projection of the integral curves of Γ onto M will satisfy the Euler–Lagrange equations [61]. Notice, that these solutions arise from the determination of the integral curves of the dynamical vector field and do not stem from any variational principle [38].

4.1.3 Singular case

At this point, we can understand the singularity of a Lagrangian in two equivalent ways. Locally, it means that the Hessian matrix W is non–invertible, i.e. $\det W = 0$, and therefore not all Euler–Lagrange equations are second-order differential equations[30] [62]. In fact, there can be initial conditions for which there exists no solution at all, or only some initial positions and velocities are compatible with the equations. And even if there exists a solution for particular initial values of q^i and $\dot{q}^i$ it might not be unique. Obviously, this is in conflict with the foundations of mechanics and hints at the existence of gauge degrees of freedom [38]. In geometrical terms, the consequence of a singular Lagrangian is that the Lagrangian 2–form $\Omega_{\mathcal{L}}$ does not have maximal rank, i.e. it is merely a presymplectic form. Therefore, the Lagrangian equations of motion have in general no solution everywhere in TM. In particular, equation (8) has only solutions in the points of the set

$$P = \{x \in TM \mid \iota_\Gamma \mathrm{d}E_{\mathcal{L}}(x) = 0,\ \forall \Gamma \in \ker \Omega_{\mathcal{L}}\},$$

and in P the solutions are only determined up to contributions from $\ker\Omega_{\mathcal{L}}$ [56].

4.2 Hamiltonian formulation

4.2.1 Hamilton equations

The formulation in terms of the Lagrangian presupposes that the state of a system is described by specifying its generalized coordinates q^i and velocities $\dot{q}^i$. This is not the only possibility, however. An equally valid option is to treat the system in terms of generalized coordinates and momenta of the system and we can pass from one set of independent variables to the other [63].

Generalized momenta Define the *generalized momenta* for a Lagrangian system $\mathcal{L}$ as

$$p_i(q, \dot{q}) = \frac{\partial \mathcal{L}}{\partial \dot{q}^i},$$

and, in principle, we can solve these relations to find functional expressions for the $\dot{q}^i$ in terms of q^i and p_i. In terms of these new variables the Euler–Lagrange equations become

$$\dot{p}_i = \frac{\partial \mathcal{L}}{\partial q^i}.$$

[30] Such that some of them are non–dynamical.

Hamiltonian Consider the total differential of the Lagrangian as a function of coordinates and velocities,

$$\mathrm{d}\mathcal{L} = \frac{\partial \mathcal{L}}{\partial q^i}\mathrm{d}q^i + \frac{\partial \mathcal{L}}{\partial \dot{q}^i}\mathrm{d}\dot{q}^i = \dot{p}_i \mathrm{d}q^i + p_i \mathrm{d}\dot{q}^i, \tag{9}$$

where we used the definition of the generalized momenta and the Euler–Lagrange equations. We can rewrite the second term on the right hand side as $p_i \mathrm{d}\dot{q}^i = \mathrm{d}(p_i\dot{q}^i) - \dot{q}^i \mathrm{d}p_i$ and therefore obtain from (9)

$$\mathrm{d}(p_i\dot{q}^i - \mathcal{L}) = -\dot{p}_i \mathrm{d}q^i + \dot{q}^i \mathrm{d}p_i. \tag{10}$$

The argument of the differential on the left hand side is the energy of the system expressed in terms of generalized coordinates and momenta [63]. We call

$$H(p,q) = p_i\dot{q}^i - \mathcal{L}, \tag{11}$$

the *Hamiltonian* of the system.

Hamilton equations From

$$\mathrm{d}H = -\dot{p}_i \mathrm{d}q^i + \dot{q}^i \mathrm{d}p_i,$$

we can read off the *Hamilton equations*

$$\dot{q}^i = \frac{\partial H}{\partial p_i} \qquad \text{and} \qquad \dot{p}_i = -\frac{\partial H}{\partial q^i},$$

the equations of motion in the Hamiltonian formulation. Alternatively, one can also obtain them from a variational principle [51].

4.2.2 Geometrical approach

Let (M, ω) be a symplectic manifold, called the *phase space* and $H : M \to \mathbb{R}$ a given smooth function. Every point $m \in M$ describes a state of the system and smooth functions on M are observables. The result of a measurement of a observable $f \in C^\infty(M)$ at m is given by $f(m)$ [34]. As introduced before, the vector field X_H on M determined by the condition

$$\iota_{X_H}\omega = \mathrm{d}H, \tag{12}$$

is called Hamiltonian vector field. We call the function H *energy function* [35]. The dynamics of the system are determined by the integral curves of the Hamiltonian vector field X_H. If ω is non–degenerate, the linear map $\flat : TM \to T^*M$ is an isomorphism, thus for any H we can solve (12) uniquely by $X_H = (\mathrm{d}H)^\sharp$. In the context of classical mechanics we usually take $M = T^*Q$, i.e. our phase space is the cotangent bundle of the configuration space Q. As discussed earlier, any cotangent bundle is locally a symplectic manifold [51] and comes equipped with a canonical symplectic structure. In terms of canonical coordinates (12) is simply

$$\frac{dq^i}{dt} \equiv X_H(q^i) = \frac{\partial H}{\partial p_i} \qquad \text{and} \qquad \frac{dp_i}{dt} \equiv X_H(p_i) = -\frac{\partial H}{\partial q_i},$$

and we recover the Hamilton equations [49]. Making use of the Poisson bracket, the time development of any function in phase space $f \in C^\infty(T^*Q)$ can be written as

$$\frac{df}{dt} = \{f, H\},$$

and the Hamilton equations take the form

$$\dot{q}^i = \{q^i, H\} = \frac{\partial H}{\partial p_i} \qquad \text{and} \qquad \dot{p}_i = \{p_i, H\} = -\frac{\partial H}{\partial q^i}.$$

Canonical transformations A vector field X_H on the symplectic manifold (T^*Q, ω) is said to be (globally) Hamiltonian if the 1–form associated with the vector field, $\iota_{X_H}\omega$, is exact [41]. A vector field X_H is called locally Hamiltonian if $\iota_{X_H}\omega$ is closed. Notice that every exact form is closed, therefore every globally Hamiltonian vector field is also locally Hamiltonian. The inverse is generically not the case. The flow Ψ_t of a locally Hamiltonian vector field satisfies

$$\Psi_t^*\omega = \omega,$$

i.e. the flow preserves ω. Thus, if X_H is complete, Ψ_t determines a one–parameter group of symplectomorphisms on the manifold [41, 50]. In physics a symplectomorphism of the phase space is called a canonical transformation, hence the dynamics of a Hamiltonian system can be seen as a time–dependent canonical transformation [34].

4.2.3 Singular case

The transformation from Lagrangian to Hamiltonian requires to invert the generalized coordinates for the velocities, i.e. $\dot{q}^i = \dot{q}^i(q, p)$, which is possible if and only if the matrix

$$\frac{\partial p_i}{\partial \dot{q}^j} = \frac{\partial^2 \mathcal{L}}{\partial \dot{q}^i \partial \dot{q}^j} \equiv W_{ij},$$

has non–vanishing determinant. From this one can clearly see the implications of a singular Lagrangian on the Hamiltonian formulation: The transformation is no longer well–defined since we are unable to completely eliminate all the velocity variables [64]. In geometrical terms, the singularity has the consequence that the canonical 2–form ω on T^*M is only presymplectic, i.e. it is not of maximal rank. Hence, the mapping $\flat : TM \to T^*M$ is no longer bijective and in general we cannot uniquely solve Eq. (12).

4.3 Relation between the Lagrangian and Hamiltonian formulations

4.3.1 Legendre transformation

It sometimes appears as if the transition from the Lagrangian to the Hamiltonian is merely a change of variables, since the prescription of the canonical momenta

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}^i},$$

is usually invertible and one can always calculate the Hamiltonian from a given Lagrangian via

$$H(q, p) = p_i \dot{q}^i(p, q) - \mathcal{L}(q, \dot{q}(p, q, t)).$$

This prescription is formally a *Legendre transformation* and it is only well–defined for regular Lagrangians [51, 65]. We make this idea more precise and insightful in the following.

fiber derivative Given a Lagrangian $\mathcal{L} : TM \to \mathbb{R}$, define the map

$$F\mathcal{L} : TM \to T^*M,$$

called the *fiber derivative*, by

$$F\mathcal{L}(v) \cdot w \coloneqq \frac{d}{ds}\Big|_{s=0} \mathcal{L}(v + sw),$$

where $v, w \in T_qM$. Thus, $F\mathcal{L}(v) \cdot w$ is the derivative of $\mathcal{L}$ at v along the fiber T_qM in the direction w. The fiber derivative is fiber preserving, i.e. it maps the fiber T_qM to the fiber T_q^*M. For finite dimensional manifolds with induced coordinates $(q^i,\dot{q}^i)$ on TM the fiber derivative takes the form

$$F\mathcal{L}(q,\dot{q}^i) = (q^i, \frac{\partial \mathcal{L}}{\partial \dot{q}^i}),$$

and we can make the identification

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}^i},$$

i.e. we find the origin of the canonical momenta. $F\mathcal{L}$ is also called *Legendre map* [50] and when restricted to a particular point the fiber derivative is a Legendre transformation. If $F\mathcal{L}$ is a diffeomorphism the Lagrangian is called regular and we can obtain an equivalent Hamiltonian formulation by defining the Hamiltonian as [58]

$$H = E_{\mathcal{L}} \circ (F\mathcal{L})^{-1}. \tag{13}$$

Pullback under $F\mathcal{L}$ With the geometrical understanding of the Legendre map at hand, one might ask if it is possible to relate the canonical symplectic structure on T^*Q with the one we intrinsically constructed on TQ. In fact [35], the pullback of the canonical 1–form $\theta \in \Omega^1(T^*Q)$ under $F\mathcal{L}$ independently yields the Euler–Poincaré 1–form, i.e.

$$(F\mathcal{L})^*\theta = \theta_{\mathcal{L}},$$

and consequently we can also pull back the canonical symplectic form as

$$(F\mathcal{L})^*\omega = \Omega_{\mathcal{L}}.$$

4.3.2 Singular case

As mentioned before, there are several ways to understand the singularity of a Lagrangian. Among other things, it implies that the Legendre map is not a (local) diffeomorphism and therefore the prescription (13) is not well–defined. It is not possible to isolate all the coordinates (u^i) of a local chart in TQ as functions of the natural symplectic coordinates (q^i, p_i) of T^*Q. Hence, any object in TQ with explicit dependence on these coordinates does not have a canonical expression in T^*Q.

$F\mathcal{L}$ projectability Based on this straight forward observation, we make the following definitions.

1. A differential p–form $l \in \Omega^p(TQ)$ is said to be $F\mathcal{L}$ projectable iff $\exists h \in \Omega^p(T^*Q)$ such that $F\mathcal{L}^*h = l$.

2. A vector field $X_l \in \mathfrak{X}(TQ)$ is said to be $F\mathcal{L}$–projectable iff $\exists X_h \in \mathfrak{X}(T^*Q)$ such that $F\mathcal{L}_*X_l = X_h$.

The Lagrange 2–form $\Omega_{\mathcal{L}} = F\mathcal{L}^*\omega$ and the energy function $E_{\mathcal{L}} = F\mathcal{L}^*H$ are important examples of $F\mathcal{L}$–projectable objects [60].

Equivalence theorem While the equivalence of Lagrangian and Hamiltonian formulation is well established in the regular case, one might wonder if both descriptions are still equivalent in the singular case. Since the Legendre map is in general not invertible in the latter case, we don't have a naive way to relate TQ and T^*Q. Still, the dynamical equivalence between Lagrangian and Hamiltonian formalism for singular systems can be proven [60]. Remarkably though, we can only have solutions in a submanifold of the respective space. In fact, this is one of the core ideas of the theory of constrained systems — to find the final submanifold on which dynamically equivalent solutions exist. Note at this point, that we always assume our Lagrangian to be *almost regular* in the course of this work. Physically, these are very reasonable assumptions and we refer the interested reader to [59] for more information and the exact definition. Only in the almost regular case every system has a Hamiltonian counterpart.

4.4 Transition to field theory

As a matter of fact, the systems of interest in this work are not point particle systems but classical field theories. Our above discussion is still valid for field theories and can mostly be applied unchanged to these cases. Yet, it is important to bear in mind some of the subtleties that arise when transitioning from point particles to fields. The fundamental difference is that we no longer have coordinates q^i labelling points in the manifold M, but rather field variables that in turn depend on the points of the manifold we operate on. Symbolically, we make the extension

$$q^i(t) \to q^i(t, \vec{x}),$$

which comes with a number of consequences. One should now distinguish between a Lagrangian L and the *Lagrangian density* $\mathcal{L}(q, \partial_\mu q)$, which can also depend on spatial derivatives of the field variables. Say, we describe the q–fields on a d–dimensional Lorentzian manifold with local coordinates $(x^0, \vec{x})$. The Lagrangian is the integral over space of the Lagrangian density, i.e.

$$L = \int d^{d-1}\mathcal{L}(\phi, \partial_\mu \phi),$$

and the action is defined as before, leading to

$$S[q^i] = \int dt L = \int d^d x \mathcal{L}.$$

Thus, the equations of motion read

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu q)} - \frac{\partial \mathcal{L}}{\partial q} = 0,$$

and we see that in comparison with Eq. (7) we substituted ∂_0 by ∂_μ.

Poisson bracket The second modification we have to be aware of in this work concerns the definition of the Poisson bracket. In the context of particle mechanics, the fundamental Poisson brackets read

$$\{q^i, p_i\} = \delta^i_j, \qquad \{q^i, q^j\} = 0, \qquad \{p_i, p_j\} = 0. \tag{14}$$

A naive extension to field theory leads us to the notion of the (equal–time) Poisson bracket [66]

$$\{A(q(x), p(x)), B(q(x'), p(x'))\} \coloneqq \int d^{d-1}y \left[\frac{\partial A}{\partial q(y)} \frac{\partial B}{\partial p(y)} - \frac{\partial A}{\partial p(y)} \frac{\partial B}{\partial q(y)} \right],$$

such that Eqs. (14) turn into

$$\{q^i(x), p_j(x')\} = \delta^i_j \delta(x - x'), \qquad \{q^i(x), q^j(x')\} = 0, \qquad \{p_i(x), p_j(x')\} = 0,$$

which is the appropriate version in most cases [67]. However, if derivatives of fields appear inside the Poisson bracket additional complications arise. The easiest way to deal with this issue is by introducing *test–* or *smearing functions*[31]. In the following, test functions are generically called f and g and their Poisson brackets with phase space variables vanish [68]. The most general version of the Poisson bracket we will use in this work thus reads

$$\int dx \int dy f(x)\{A(x), B(y)\}g(y) \equiv f(x)\{A(x), B(y)\}g(y),$$

which we will usually abbreviate as $f\{A, B\}g$. If the explicit calculation of a Poisson bracket yields a result like

$$f\{A, B\}g = fCg,$$

i.e. one can simply factor out the test functions, we simply write $\{A, B\} = C$ [69]. This will always be possible, except if A or B depend on the partial derivative of a relevant generalized coordinate. In this case the result will, after partial integration, depend on a spatial derivative of a test function and we cannot simply factor the test functions out. We call such Poisson brackets *non–local* [70].

5 On singular systems

By the end of the last section, we obtained a solid understanding of what it means for a system to be singular. The relation between a singular Lagrangian and its Hamiltonian became clear, thanks to our geometric insights on the fiber derivative. In the following, we want to examine how to treat singular systems, in particular how we can restrict the dynamics of the system to a subspace in which the solutions are well–defined. This is the main idea behind constraint algorithms on both the Hamiltonian and Lagrangian sides.

The starting point of the Dirac–Bergmann theory of constraints are the canonical momenta, in particular the ones that do not depend on the generalized velocities. Exactly these momenta will be the so–called primary constraints of the theory. Step by step, the constraints have to be enforced and stabilized, further and further restricting the admissible phase space. After a finite iteration (for physically sensible theories), the algorithm terminates. By classifying the full set of constraints, it is possible to count the physical degrees of freedom.

The starting point of the Lagrangian algorithm is a degenerate primary Hessian. Its null vectors determine the primary Lagrangian constraints, which in turn have to be stabilized. This procedure can unravel new, secondary, Lagrangian constraints. This logic has to be repeated until no further constraints arise.

Singularity of a Lagrangian is deeply connected with internal constraints and symmetries of the system. For local symmetries, Noether identities imply that the canonical conjugate momenta are not independent anymore — they are *constrained* [51]. As a matter of fact, fundamental physics is full of theories based on singular Lagrangians and symmetries play a role of utmost importance. Despite being more involved, it is possible to pass to the Hamiltonian description of a system starting from a singular Lagrangian. The proper treatment of such systems, called *constraint analysis*,

[31] Well–behaved functions that vanish at spatial infinity.

can be used to count the number of physical degrees of freedom of a theory. In fact, it is *the* mathematically rigorous way to determine the propagating degrees of freedom and should therefore be an inevitable tool when constructing new theories. A Lagrangian can exhibit beautiful symmetries and seem like a perfect candidate for certain applications. But if it does not meet the requirements for a consistent theory, it cannot describe the physical world. Propagating the correct number of degrees of freedom is not a sufficient, but a *necessary* condition. Even in view of this, purely classical, discussion one cannot overstate the importance of the theory of constrained systems. Not to mention its transcendence towards quantization schemes [71].

Historical context The systematic procedure to deal with constrained systems goes back to work in the late 1940s and early 1950s, mainly by Dirac [72], Bergmann and colleagues [73, 74], for the most part motivated by the quest for quantization of such systems. Despite not being the main motivation back then, their work marks the beginning of the systematic study of the canonical fomarlism for what we today call *gauge theories* [75]. Recently, there have been efforts [76, 77] to pay tribute to Léon Rosenfeld for his pioneering works on this subject, that appeared nearly 20 years earlier. In their original work, Anderson and Bergmann do indeed cite his work, stating that their "examination [...] is in some respects similar to the results obtained by L. Rosenfeld" [78]. Efforts to put the Dirac-Bergmann theory of constraints on a solid mathematical footing were made by many authors during the following decades. Gotay et al. developed a geometric generalization and globalization of the Dirac-Bergmann algorithm in terms of presymplectic manifolds [49] and consequently applied this framework to the Lagrangian side [59, 79]. It is impossible to give an exhaustive overview of the work and research that has been done on the topic of constrained systems since its beginnings 70 years ago, but one has to mention N. Mukundas influence [64, 80]. K. Kamimura was one of the first to address the Lagrangian side and the relation between singular Lagrangian and Hamiltonian systems locally [62]. Subsequently, a lot of work has been devoted to establish the equivalence between and relation of Lagrangian and Hamiltonian formalisms for constrained systems [56, 81]. More recently, noteworthy developments include the extension of the geometric formulation to field theories [82, 83] and the approach to count the physical degrees of freedom in a geometrical, and purely Lagrangian, setting [84, 85].

5.1 Dirac–Bergmann theory of constraints

The *Dirac–Bergmann algorithm* is the original procedure to transit from the Lagrangian to the Hamiltonian formulation in the singular case. Let $\mathcal{L}(q, \dot{q})$ be a time–independent first–order Lagrangian, defined in the tangent bundle TQ of some N–dimensional configuration manifold Q. The defining feature of a singular theory is that its primary Hessian

$$W_{kl} \coloneqq \frac{\partial^2 \mathcal{L}}{\partial \dot{q}^k \partial \dot{q}^l},$$

has a vanishing determinant, or in other words the Legendre map

$$F\mathcal{L} : TQ \to T^*Q,$$

is not invertible. This makes already clear that there is an issue about the projectability of structures from the tangent bundle to phase space [75]. The momenta are defined in the canonical way as

$$p_n(q, \dot{q}) = \frac{\partial \mathcal{L}}{\partial \dot{q}^n}, \quad n = 1, \ldots, N,$$

and it becomes apparent that only in the regular case the $p_n(q, \dot{q})$ can be solved for all the velocities in the form $\dot{q}^n(q, p)$ [51].

5.1.1 Primary stage

Primary constraints The p's involve only $\text{rank}(W) = N - M$ independent functions of the $\dot{q}$'s, hence there will be M relations

$$\Psi_m(q, p) = 0, \quad m = 1, 2, \ldots, M, \tag{15}$$

connecting the momentum variables [86]. The relations (15) are called *primary constraints*. Notice that the Ψ_m must actually be independent functions of the p's, otherwise we could eliminate momenta from (15) to obtain constraints that only involve the q's alone [64].

Primary constraint submanifold The relations Ψ_m restrict the dynamics to a subspace $M_p \subset T^*Q$ of the full phase space [51], called the *primary constraint submanifold*. This submanifold coincides with the image of TQ under the Legendre map, i.e. [81]

$$F\mathcal{L}(TQ) = M_p,$$

as depicted in Fig. 13. We use the *weak equality* "$\overset{1}{\approx}$" to denote that a relation only holds in the

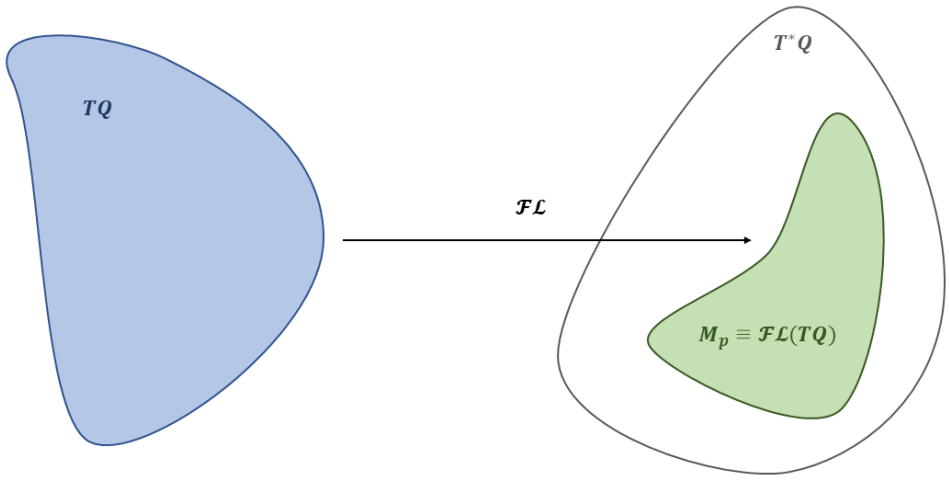

Figure 13: The primary constraint submanifold M_p

primary constraint submanifold and consequently we write

$$\Psi_m :\overset{1}{\approx} 0,$$

to make clear that the primary constraints *define* the primary constraint submanifold.

Canonical Hamiltonian The initial definition of the Hamiltonian does not change, compared to the regular case (11). Due to reasons that will become apparent soon, we want to stress this fact and call

$$H_{can} = p_n \dot{q}^n - \mathcal{L}(q, \dot{q}),$$

the *canonical Hamiltonian*. There is no explicit dependence on any velocity variable in the canonical Hamiltonian, as can already be seen from (10). This is still true in the singular case, despite the fact that the Legendre map is non–invertible[32] [51].

[32] This is due to the $F\mathcal{L}$ projectability of the energy function $E_{\mathcal{L}}$.

Primary Hamiltonian However, the canonical Hamiltonian is not uniquely determined as a function of the p's and the q's. This is because the p's are not all independent, but rather restricted to satisfy the primary constraints $\Psi_m \overset{1}{\approx} 0$. The aim is to obtain equations of motion in the Hamiltonian form, while paying attention to the constraints [87]. In order to do so, we add the primary constraints to the canonical Hamiltonian via Lagrange multipliers. We define the *primary Hamiltonian* as

$$H_p \coloneqq H_{can} + \lambda^m \Psi_m,$$

with arbitrary multiplier functions λ^m. This ambiguity in the definition of the Hamiltonian has profound consequences: it sets the stage for gauge freedom [75]. Varying the primary Hamiltonian with respect to (λ, q, p), we find

$$\frac{\partial H_{can}}{\partial p_n} + \lambda^m \frac{\partial \Psi_m}{\partial p_n} = \dot{q}^n, \tag{16}$$

$$\frac{\partial H_{can}}{\partial q^n} + \lambda^m \frac{\partial \Psi_m}{\partial q^n} = -p_n, \tag{17}$$

which are sometimes referred to as *Hamilton–Dirac equations*. They contain extra terms depending on the primary constraints compared to the Hamilton equations. Consider the time evolution of any function $f(p,q) \in C^\infty(T^*Q)$ in terms of the Poisson bracket

$$\{f, H_p\} = \{f, H_{can} + \lambda^m \Psi_m\} = \{f, H_{can}\} + \lambda^m \{f, \Psi_m\} + \{f, \lambda^m\}\Psi_m. \tag{18}$$

Since the Lagrange multipliers λ^m are not phase–space functions, the Poisson brackets $\{f, \lambda^m\}$ are not defined [51]. Notice however, that on the primary constraint submanifold the last term in (18) vanishes and we can write the dynamical evolution for any phase space function as [51, 71]

$$\dot{f}(p,q) = \{f, H_p\} \overset{1}{\approx} \{f, H_{can}\} + \lambda^m \{f, \Psi_m\}. \tag{19}$$

5.1.2 Secondary stage

In order to consistently respect the primary constraints, we have to require their preservation in time. $\Psi_m \overset{1}{\approx} 0$ should not only hold at an instant, but throughout all time. We therefore require $\dot{\Psi}_m$ to be zero.

The geometrical point of view can contribute some intuition here. In the primary stage we search for solutions to

$$(\iota_X \omega - \mathrm{d}H_p)|_{M_p},$$

where ω is the canonical symplectic form on T^*Q. The non–degeneracy of ω guarantees that solutions X exist. Preservation of the primary constraints is expressed as $X(\Psi_m)|_{M_p} = 0$. This can be understood as a tangency condition: consistent solutions X must be tangent to M_p, i.e. the dynamics are restricted to stay inside the primary constraint submanifold [78].

Stability of constraints If we take f in (19) to be one of the Ψ_r, we can express the time evolution of the primary constraints as

$$\dot{\Psi}_r \overset{1}{\approx} \{\Psi_r, H_{can}\} + \lambda^m \{\Psi_r, \Psi_m\}, \quad r, m = 1, \dots M. \tag{20}$$

The vanishing of (20) can be divided into three distinct cases [86]:

1. A constraint is already stable inside M_p, i.e. $\dot{\Psi}_r \overset{1}{\approx} 0$.

2. $\{\Psi_r, \Psi_m\} \overset{1}{\not\approx} 0$, such that it imposes a restriction on the λ's on M_p.

3. $\dot{\Psi}_m$ may not be identically zero on M_p, but $\{\Psi_r, \Psi_m\} \overset{1}{\approx} 0$ (and hence no λ's remain).

Case 1 is trivially satisfied in M_p and for case 2 the λ's get fixed in such a way that the constraint is stable.

Secondary constraints Case 3 on the other hand is different, since we can not satisfy the stability of the primary constraints on the whole primary constraint submanifold. In order to guarantee their stability, we further restrict the dynamics of the system to the *secondary constraint submanifold* $M_s \subset M_p$. It is defined by the *secondary constraints*

$$\chi_m \equiv \dot{\Psi}_m :\overset{2}{\approx} 0.$$

Secondary Hamiltonian As before, we add the new constraints to our Hamiltonian via Lagrange multipliers. Say we found M primary constraints and R secondary constraints. We write the *secondary Hamiltonian* as

$$H_s \coloneqq H_{can} + \lambda^m \Psi_m + \lambda^r \chi_r,$$

where $m = 1, \dots M$ and $r = 1, \dots, R$. Some of the λ^m's might already be determined from the stability conditions on the primary constraints. In addition to the stability conditions for the secondary constraints we also have to check the *consistency conditions* on the primary constraints with respect to the secondary Hamiltonian:

$$\{\Psi_m, H_s\} \overset{2}{\approx} 0.$$

This consistency check cannot yield new constraints but it might fix some of the remaining Lagrange multipliers.

5.1.3 Tertiary stage and beyond

By ensuring the stability of the secondary constraints we might enter the tertiary stage of the constraint algorithm. This happens if, for some r,

$$\{\chi_r, H_s\} \overset{2}{\not\approx} 0.$$

Then, we have to restrict ourselves to the tertiary constraint submanifold $M_t \subset M_s \subset M_p \subset T^*Q$, defined by the demand for stability of the secondary constraints

$$\tau_r \equiv \dot{\chi}_r :\overset{3}{\approx} 0.$$

Tertiary Hamiltonian The *tertiary Hamiltonian* is accordingly defined as

$$H_t \coloneqq H_{can} + \lambda^m \Psi_m + \lambda^r \chi_r + \lambda^k \tau_k,$$

where k labels the number of tertiary constraints. The procedure should be clear at this point. After we find the primary constraints we have to ensure their stability, which might lead to new

constraints, and which in turn have to be stabilized and so on. The algorithm terminates if we find a *final constraint submanifold* $M_f \subset \cdots \subset M_p \subset T^*Q$ on which no further constraints arise. In other words, the submanifold to which all solutions of the dynamical equations are already tangent. The Hamiltonian pertaining to M_f is usually referred to as the *extended* or *total* Hamiltonian. For most physically relevant cases, the algorithm terminates already with the secondary constraints, i.e. no tertiary constraints arise [51].

5.1.4 First– and second–class constraints

Denote by Φ_l the set of all constraints defining the final constraint submanifold M_f, independent of the stage in which they appeared. A function $f(p,q)$ (and in particular any constraint) is said to be *first–class*, if its Poisson bracket with every constraint vanishes weakly,

$$\{f, \Phi_l\} \overset{f}{\approx} 0.$$

Any function of the canonical variables that is not first class is called *second–class*. Notice, that the first–class property is preserved under the Poisson bracket operation, i.e. the Poisson bracket of two first–class functions is first–class [71]. The motivation for this definition becomes clear when we look back at (20): On the primary constraint submanifold the stability of a first–class constraint will generate a secondary constraint, whereas stability of a second–class constraint fixes a Lagrange multiplier. In principle, we can categorize the constraints at any stage of the constraint algorithm, but it is possible for a first–class constraint on the n^{th} stage to become second–class on the $(n+1)^{\text{th}}$ stage. Only on the final constraint submanifold one can reliably classify all constraints.

Gauge freedom The existence of arbitrary functions in the final Hamiltonian H_f is connected to *gauge freedom* of the system. There will exist *gauge transformations* connecting different solutions of the dynamical equations that share the same initial conditions. Although the physical state is uniquely defined once a set of q's and p's is given, there is more than one set of values of the canonical variables representing a physical state [71]. Noteworthy, not every constrained system has to exhibit gauge freedom. It might happen that all constraints eventually become second class on the final constraint submanifold and hence there is no arbitrariness left [75]. Even though we will not go into the details of constructing the gauge transformation on the Hamiltonian side, a word of caution regarding the concrete relation between first–class constraints and gauge is necessary. The view that any primary first–class constraint generates a gauge transformation is usually referred to as *Dirac's conjecture*. This conjecture has been criticized before [75] but surprisingly enough, it was not until 2014 that it was actually refuted by a straight forward check of Maxwell's electromagnetism. Rather than a one–to–one correspondence between primary first–class constraints and gauge transformations, one can construct a tuned sum of first–class constraints that will serve as gauge generator [88].

5.1.5 Physical degrees of freedom

Gauge fixation It is possible to eliminate the ambiguity of first–class constraints and the associated gauge freedom by imposing further restrictions on the canonical variables. The well–known procedure of *gauge fixation* leads to a one–to–one correspondence between physical states and values of those canonical variables that are left independent after imposing the additional conditions. Say we have N_1 first–class constraints on the final constraint submanifold. By fixing the gauge we introduce N_1 *gauge conditions*, demoting the former first–class constraints to second–class constraints. These new conditions are not an intrinsic feature of the theory, but rather brought in from

the outside to cope with redundant degrees of freedom. They remove the unobservable arbitrary elements of the theory[33] and we therefore avoid a multiple counting of states [71].

Counting The formula to count the physical degrees of freedom can be deduced from the above considerations. Since it was already implied in their original works [72, 78, 89], the counting procedure is usually attributed to Dirac and Bergmann, but it was not *explicitly* stated there. Consider a theory, which only contains second–class constraints. In that case no arbitrary functions appear in the Hamiltonian and a set of canonical variables that satisfies the constraint equations determines the only physical state. In that case we count for the number of physical degrees of freedom n_{dof}

$$2 \cdot n_{dof} = P_0 - N_2,$$

where P_0 is the total number of canonical variables and N_2 the number of second–class constraints. The 2 in front of n_{dof} stems from the fact that we usually count the degrees of freedom in configuration space, which is half the number of degrees of freedom in phase space. Now consider the case where we also have N_1 first–class constraints. In order to eliminate the redundant degrees of freedom we introduce as many gauge conditions G as there are first–class constraints and we find $2 \cdot n_{dof} = P_0 - N_2 - N_1 - G = P_0 - N_2 - 2 \cdot N_1$ [71]. Denoting by N the number of field variables, this is commonly written as [21]

$$n_{dof} = N - N_1 - \frac{N_2}{2}, \tag{21}$$

and constitutes the well–defined and unambiguous way to count the physical degrees of freedom on the Hamiltonian side.

5.2 A purely Lagrangian approach

On–shell, our analysis follows closely the iterative algorithm presented in [60, 62] and employed in Appendix A of [21]. This is the suitable generalization to field theory of the coordinate–dependent method used in [84] for particle systems, which is in turn based on [90, 91]. The non–trivial geometric extension to field theory of [84] was done in [85]. Thus, one can view our discussion as complementary to both these references. The main outcome of this algorithm shall be the total number of on–shell (functionally independent) constraints l.

Off–shell, we limit ourselves to obtaining, in the most streamlined manner for the theories of our interest, g and e, the number of gauge identities and effective gauge parameters, respectively. By effective gauge parameters we mean the number of independent gauge parameters plus their successive time derivatives that explicitly appear in the gauge transformations. Once we find (l, g, e), the physical degrees of freedom $n_{\rm dof}$ in a first–order classical field theory can be counted, employing the formula derived in [84]:

$$n_{\rm dof} = N - \frac{1}{2}(l + g + e), \tag{22}$$

with M_0 the dimension of the theory's configuration space, as in the Hamiltonian method. The remarkable feature about this counting is that it is purely Lagrangian.

[33] Whereas the observable properties are gauge invariant.

5.2.1 On–shell constraints

Let Q be the configuration space of a classical field theory. As usual, we take Q to be a smooth manifold whose points are labeled by N field variables Q^A:

$$Q = \text{span}\{Q^A\}, \qquad A = 1, 2, \ldots, N.$$

We stress that A comprises all possible indices that the field variables have. For instance, if one considers Yang–Mills theory, A consists of both spacetime indices and color indices. Notice that Q^A are functions of spacetime: $Q^A = Q^A(x^\mu)$. Then, TQ is the tangent space of Q, which is spanned by $\{Q^A, \dot{Q}^A\}$. We refer to $(Q^A, \dot{Q}^A, \ddot{Q}^A)$ as the generalized coordinates, velocities and accelerations of the theory, respectively.

As already stated and common to most field theories, we assume that the dynamics are derivable from a principle of least action. In other words, the equations of motion E_A for the field theory follow from the requirement that the action functional

$$S = S[Q^A] = \int_{\mathcal{M}} d^d x\, \mathcal{L} = \int_{t_1}^{t_2} dx^0 \int_\Sigma d^{d-1}x\, \mathcal{L} =: \int_{t_1}^{t_2} dx^0 L \tag{23}$$

remains stationary under arbitrary functional variations $\delta Q^A = \delta Q^A(x^0, x^i)$ that vanish at t_1 and t_2:

$$\delta S = \frac{\delta S}{\delta Q^A}\delta Q^A \equiv \int_{\mathcal{M}} d^d x\, E_A \delta Q^A \overset{!}{=} 0, \tag{24}$$

with $\delta Q^A(t_1, x^i) = 0 = \delta Q^A(t_2, x^i)$. The above variational derivative is defined as

$$E_A := \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu Q^A)} \right) - \frac{\partial \mathcal{L}}{\partial Q^A} \overset{!}{=} 0, \tag{25}$$

where the latter equality is the on–shell demand. Here, $\mathcal{L} = \mathcal{L}[Q^A]$ is the Lagrangian density and, accordingly, L is the Lagrangian. Observe that we have already restricted attention to first order field theories, i.e. we consider $\mathcal{L}$ depends only on Q^A and its first derivatives $\partial_\mu Q^A$, up to integration by parts. The study of higher order field theories lies beyond the scope of our present investigations. We omit the possible dependence of $\mathcal{L}$ on non–dynamical field variables, such as the spacetime metric in any special relativistic theory. The said dependence can be easily incorporated to the following analysis, but it does not arise in the theories we discuss in this work.

It is convenient to recast the Euler–Lagrange equations of motion (25) in the form

$$E_B = \ddot{Q}^A W_{AB} + \alpha_B \overset{!}{=} 0, \tag{26}$$

where we have defined the primary Hessian $W_{AB} := \partial_{\dot{A}}\partial_{\dot{B}}\mathcal{L}$, as well as

$$\alpha_B := (\partial_{\dot{B}}\partial^i_A\mathcal{L} + \partial^i_B\partial_{\dot{A}}\mathcal{L})\partial_i\dot{Q}^A + (\partial^i_B\partial^j_A\mathcal{L})\partial_i\partial_j Q^A + (\partial^i_B\partial_A\mathcal{L})\partial_i Q^A + (\partial_{\dot{B}}\partial_A\mathcal{L})\dot{Q}^A - \partial_B\mathcal{L}. \tag{27}$$

To alleviate notation, we have introduced the following short–hand notation:

$$\partial_{\dot{A}} := \frac{\partial}{\partial \dot{Q}^A}, \qquad \partial^i_A := \frac{\partial}{\partial(\partial_i Q^A)}, \qquad \partial_A := \frac{\partial}{\partial Q^A},$$

which we shall extensively employ henceforth.

It is well–known that any field theory of the kind here considered is described by a singular Lagrangian density; that is, a Lagrangian density whose primary Hessian has a vanishing determinant $\det(W_{AB}) = 0$. This means that $\text{rank}(W_{AB}) < \dim(W_{AB}) = N$. Further, for a first–order field theory, the rank of the primary Hessian can only be reduced by setting to zero entire rows (or, equivalently, columns). This is because requiring linear combinations of rows to vanish amounts to a redefinition of the field variables Q^A.

It follows that not all equations in (26) are equations of motion, in the sense that not all of them depend on the generalized accelerations $\ddot{Q}^A$. Instead, $M_1 := \dim(W_{AB}) - \text{rank}(W_{AB})$ of them are (primary) constraints; that is, relations between the generalized coordinates Q^A and velocities $\dot{Q}^A$. In the following, we obtain and ensure the stability of these constraints.

Let γ_I be the set of M_1 null vectors of the primary Hessian:

$$(\gamma_I)^A W_{AB} = 0, \qquad I = 1, 2, \ldots, M_1. \tag{28}$$

We require that these form a basis of the kernel of W_{AB}. This amounts to imposing the normalization condition

$$(\gamma_I)^A (\gamma^J)_A = \delta_I^J, \qquad \text{where } \gamma^I := (\gamma_I)^T, \tag{29}$$

with T standing for the transpose operation. Then, the M_1 primary constraints are given by the contraction of the equations of motion (26) with the above null vectors:

$$\varphi_I \equiv (\gamma_I)^A E_A = (\gamma_I)^A \alpha_A \overset{!}{=} 0. \tag{30}$$

These primary constraints need not be functionally independent from each other. When they are, the field theory is said to be irreducible at the primary stage. Otherwise, the theory is reducible at the primary stage. Before we carry on, we must restrict attention to the functionally independent primary constraints $\varphi_{I'} \overset{!}{=} 0$, where $I' = 1, 2, \ldots M_1' \leq M_1$. Their number is given by $M_1' = \text{rank}(J_{I\Lambda})$, where the Jacobian matrix $J_{I\Lambda}$ is defined as

$$J_{I\Lambda} := \frac{\partial \varphi_I}{\partial X^\Lambda}, \qquad X^\Lambda = \{Q^A, \dot{Q}^A\}. \tag{31}$$

The above is the pullback of the phase space regularity conditions in [92, 93]. For the theories we are concerned with in this work, we verify $M_1' = M_1$. Hence, all of the primary constraints in (30) must be considered in the following[34].

The vanishing of all (the functionally independent) on–shell primary constraints defines the so–called primary constraint surface $TQ_1 \subset TQ$. We write

$$\varphi_I \overset{1}{:\approx} 0. \tag{32}$$

Equalities that hold true in TQ_1 (and not in the entire of TQ) shall be denoted $\overset{1}{\approx}$.

Consistency requires us to not only enforce the primary constraints (32), but also to ensure that

[34] If $M_1' < M_1$, more work is required. Indeed, there exists an iterative algorithm to extract the functionally independent subset of constraints from (30). This is explained in section IID of [85] and subsequently exemplified. When the said algorithm requires a(n) finite (infinite) number of iterations, we face a(n) finitely (infinitely) reducible theory. As already pointed out, the procedure here described requires, at the very least, the closure of the reducibility algorithm to proceed. Thus, infinitely reducible theories cannot be studied with the present formalism.

these are preserved under time evolution. Explicitly, $\widetilde{E}_J := \dot{\varphi}_J \overset{!}{\underset{1}{\approx}} 0$. $\widetilde{E}_J$[35] are known as the secondary Euler–Lagrange equations of motion and it is convenient to write them as

$$\widetilde{E}_J = \ddot{Q}^A (\gamma^I)_A \widetilde{W}_{IJ} + \widetilde{\alpha}_J \overset{!}{\underset{1}{\approx}} 0, \tag{33}$$

where $\widetilde{W}_{IJ} := (\gamma_I)^A \partial_{\dot{A}} \varphi_J$ is the so–called secondary Hessian and we have defined

$$\widetilde{\alpha}_J := -\alpha_A M^{AB} \partial_{\dot{B}} \varphi_J + \dot{Q}^A \partial_A \varphi_J + (\partial_i \dot{Q}^A) \partial^i_A \varphi_J. \tag{34}$$

Here, the auxiliary matrix M^{AB} is the Moore–Penrose pseudoinverse (as defined in [94]) of the primary Hessian. This is ensured to always exist and be unique. Its defining relations are

$$W_{AB} M^{BC} - \delta^C_A + (\gamma^I)_A (\gamma_I)^C = 0, \qquad M^{AB} (\gamma^I)_B = 0. \tag{35}$$

In [62], the first relation is referred to as completeness relation. There, both equations in (35) are further used to obtain the subset of functionally independent secondary Euler–Lagrange equations.

If $\mathrm{rank}(\widetilde{W}_{IJ}) = \dim(\widetilde{W}_{IJ}) = M_1$, then the stability of the primary constraints is ensured by a set of secondary Euler–Lagrange equations $\widetilde{E}_J$. In such a case, all of the $\widetilde{E}_J$'s are dynamical, i.e. they involve generalized accelerations $\ddot{Q}^A$, and thus no secondary constraints arise. As a result, the total number of on–shell constraints present in such field theory is $l = M_1'$. However, this is not what generically happens in the theories of our interest.

In most theories, the rank of the secondary Hessian is smaller than its dimension. As a result, $M_2 := \dim(\widetilde{W}_{IJ}) - \mathrm{rank}(\widetilde{W}_{IJ})$ of the equations in (33) are (secondary) constraints, whose stability must be ensured. This is done exactly as in the primary stage before. In other words, the analysis from equation (28) onwards is to be repeated.

In details, first the null vectors $\widetilde{\gamma}_R$ of the secondary Hessian must be obtained:

$$(\widetilde{\gamma}_R)^I \widetilde{W}_{IJ} = 0, \qquad R = 1, 2, \ldots, M_2,$$

and chosen so that the normalization condition

$$(\widetilde{\gamma}_R)^I (\widetilde{\gamma}^S)_I = \delta^S_R, \qquad \text{with } \widetilde{\gamma}^S := (\widetilde{\gamma}_S)^T$$

is satisfied. Then, these must be contracted with the secondary Euler–Lagrange equations to yield the secondary constraints of the theory,

$$\widetilde{\varphi}_R \equiv (\widetilde{\gamma}_R)^I \widetilde{E}_I \overset{!}{\underset{1}{\approx}} 0. \tag{36}$$

If the secondary constraints identically vanish in the first constraint surface $\widetilde{\varphi}_R \overset{1}{\approx} 0$, then the number of on–shell constraints is $l = M_1'$. Again, this is not what happens in the theories of our interest.

As a consequence, we must proceed with the algorithm. First, we need to obtain the $M_2' \leq M_2$ of $\widetilde{\varphi}_R$'s which are functionally independent among themselves when evaluated in the first constraint surface. This number is given by

$$M_2' = \mathrm{rank}(\widetilde{J}_{R\Lambda}), \qquad \text{where} \quad \widetilde{J}_{R\Lambda} := \frac{\partial}{\partial X^\Lambda} \left(\widetilde{\varphi}_R \Big|_{TQ_1} \right) \tag{37}$$

[35] For clarity, we will use a notation where tilde quantities belong to the secondary stage and hat quantities pertain to the tertiary stage.

and X^Λ was introduced in (31). We verify $M_2' = M_2$ for the theories we shall consider. Thus, all secondary constraints in (36) must be considered subsequently[36].

The vanishing of the (functionally independent) secondary constraints will define the second constraint surface $TQ_2 \subset TQ_1$; which we write as $\widetilde{\varphi}_R \overset{2}{:\approx} 0$. Equalities holding true in TQ_2 shall be denoted $\overset{2}{\approx}$.

Let $\widehat{W}_{RS} := (\widetilde{\gamma}_R)^I (\gamma_I)^A \partial_{\dot{A}} \widetilde{\varphi}_S$ be the tertiary Hessian. When the tertiary Hessian's rank does not match its dimension, the stability of $M_3 := \dim(\widehat{W}_{RS}) - \mathrm{rank}(\widehat{W}_{RS})$ of the secondary constraints in TQ_2 is not guaranteed. Instead, it must be enforced through a third iteration of the above described procedure.

In full generality, this algorithm stops when all constraints have been stabilized. This can happen in either of the following three different manners:

1. Firstly, it may happen when $M_n := \dim(W^{(n)}) - \mathrm{rank}(W^{(n)}) = 0$ for some n^{th} stage Hessian $W^{(n)}$. This implies no constraints arise at the n^{th} stage.

2. Secondly, it may happen when the n^{th} stage constraints vanish identically in the $(n-1)^{th}$ constraint surface: $\varphi^{(n)} \overset{n-1}{\approx\approx} 0$. Such $\varphi^{(n)}$'s are known as Lagrangian identities.

3. Thirdly, it may happen when all the n^{th} stage constraints are functionally dependent on the $(n-1)^{th}$ stage constraints: $\varphi^{(n)} = (f_1 + f_2 \partial_i)\varphi^{(n-1)}$, where (f_1, f_2) are arbitrary smooth functions of the generalized coordinates and velocities $(Q^A, \dot{Q}^A)$, such that (f_1, f_2) are naturally defined in TQ_{n-1}.

Here, $\varphi^{(n)}$ and $\varphi^{(n-1)}$ refer to functionally independent constraints at the n^{th} and $(n-1)^{th}$ stage, respectively. Observe that in the cases 2 and 3, one finds $M_n' = 0$. This implies that such $\varphi^{(n)}$'s do not define a new constraint surface, so that $TQ_n \equiv TQ_{n-1}$.

In all three cases, the total number of on–shell constraints is given by

$$l = \sum_{a=1}^{n-1} M_a',$$

where M_a' counts the number of functionally independent a^{th} stage constraints. We are not aware of any physically relevant example of a field theory where n is infinite.

5.2.2 General expressions for the on–shell constraints

For completeness, we now provide the explicit expressions for all relevant quantities at some arbitrary iteration of the algorithm. These have not appeared in the literature, as far as we know.

Let $\varphi_{A_a} \overset{!}{=} 0$ be a set of M_a many functionally independent constraints at the a^{th} stage, with $A_a = 1, 2, \ldots, M_a$. These constraints are relations between the generalized coordinates Q^A and velocities $\dot{Q}^A$ of the field theory under consideration. They define the so–called a^{th} constraint surface $TQ_a \subset TQ = \mathrm{span}\{Q^A, \dot{Q}^A\}$, which we express as $\varphi_{A_a} \overset{a}{:\approx} 0$. In order to ensure the preservation of the said constraints under time evolution, we require

$$E_{A_a} := \dot{\varphi}_{A_a} \overset{!}{\underset{a}{\approx}} 0. \tag{38}$$

[36] When $M_2' < M_2$, the iterative algorithm referenced in footnote 34 must be employed to extract the functionally independent secondary constraints from (36).

We refer to E_{A_a} as the $(a+1)^{th}$ stage Euler–Lagrange equations. Next, we will explicitly write E_{A_a}. But to do so, we must first define the following objects. Let $W_{A_aA_b}$ denote the $(a+1)^{th}$ stage Hessian. This is a square matrix of dimension M_a that allows us to define $M_{a+1} := \dim(W_{A_aA_b}) - \mathrm{rank}(W_{A_aA_b})$. We refer to the M_{a+1} null vectors associated to $W_{A_aA_b}$ as $\gamma_{A_{a+1}}$. Explicitly, $(\gamma_{A_{a+1}})^{A_a} W_{A_aA_b} = 0$, with $A_{a+1} = 1, 2, \ldots, M_{a+1}$. We require them to fulfill the normalization condition

$$(\gamma_{A_{a+1}})^{A_a}(\gamma^{A_{b+1}})_{A_a} = \delta^{A_{b+1}}_{A_{a+1}}, \qquad \text{with } \gamma^{A_{a+1}} := (\gamma_{A_{a+1}})^T,$$

where T stands for the transpose operation. With the help of the above null vectors, the $(a+1)^{th}$ stage Hessian can be expressed in terms of the the a^{th} stage (functionally independent) constraints as follows:

$$W_{A_aA_b} = (\gamma_{A_a})^{A_{a-1}}(\gamma_{A_{a-1}})^{A_{a-2}} \ldots (\gamma_{A_1})^A \partial_{\dot{A}} \varphi_{A_b}, \qquad \text{with } A = 1, 2, \ldots, N = \dim(Q). \tag{39}$$

Finally, we introduce the auxiliary matrix $M^{A_aA_b}$ which always exists and is uniquely determined from the relations

$$W_{A_aA_b} M^{A_bA_c} - \delta^{A_c}_{A_a} + (\gamma^{A_{a+1}})_{A_a}(\gamma_{A_{a+1}})^{A_c} = 0, \qquad M^{A_aA_b}(\gamma^{A_{a+1}})_{A_b} = 0. \tag{40}$$

Using the above, the $(a+1)^{th}$ stage Euler–Lagrange equations in (38) can be written as

$$E_{A_b} = \ddot{Q}^A (\gamma^{A_1})_A (\gamma^{A_2})_{A_1} \ldots (\gamma^{A_a})_{A_{a-1}} W_{A_aA_b} + \alpha_{A_b} \overset{!}{\underset{a}{\approx}} 0, \tag{41}$$

where the expression (39) is to be employed for $W_{A_aA_b}$ and where we have (recursively) defined

$$\begin{aligned} \alpha_{A_b} := \Big[& -\alpha_{A_{a-1}} M^{A_{a-1}A_{b-1}} (\gamma_{A_{b-1}})^{A_{a-2}} (\gamma_{A_{a-2}})^{A_{a-3}} \ldots (\gamma_{A_1})^A \partial_{\dot{A}} \\ & -\alpha_{A_{a-2}} M^{A_{a-2}A_{b-2}} (\gamma_{A_{b-2}})^{A_{a-3}} (\gamma_{A_{a-3}})^{A_{a-4}} \ldots (\gamma_{A_1})^A \partial_{\dot{A}} \\ & - \ldots - \alpha_A M^{AB} \partial_{\dot{B}} + \dot{Q}^A \partial_A + (\partial_i \dot{Q}^A) \partial^i_A \Big] \varphi_{A_b}, \end{aligned}$$

with α_A as given in (27).

In order to reproduce the results in section 5.2.1 from the above discussion, the reader need only do the index replacements $(A_0 \equiv A, B, \ldots)$, $(A_1 \to I, J, \ldots)$, $(A_2 \to R, S, \ldots)$, etc., as well as take footnote 35 into account.

5.2.3 Off–shell identities

In the principle of least action (24), we have so far only considered that $\delta S \overset{!}{=} 0$ follows from the on–shell demand $E_A \overset{!}{=} 0$. In the following, we briefly review how the vanishing of δS may also be a consequence of off–shell identities stemming from the arbitrary functional variations δQ^A. There are different methods to obtain such off–shell identities, but it is not our goal to provide an overview of them here. Our subsequent discussion summarizes and employs the approach put forward in [95] and later on adapted to exhibit manifest covariance in [69]. This adaptation makes it straightforward to apply [95] to any first–order classical field theory, which is our interest.

Let δQ^A be of the form

$$\delta Q^A = \sum_{s=0}^{n} (-1)^s \left(\frac{\partial^s \theta_\beta}{\partial x^{\mu_1} \partial x^{\mu_2} \ldots \partial x^{\mu_s}} \right) (\Omega^{A\beta})^{\mu_1 \mu_2 \ldots \mu_s}, \tag{42}$$

where $n, \beta \in \mathbb{N} \cup \{0\}$ and the θ_β's and $\Omega^{A\beta}$'s are smooth functions of the spacetime coordinates x^μ, known as the gauge parameters and gauge generators of the transformation, respectively. The former are unspecified, while the latter are to be determined. Introducing the above in (24) and operating, we find that

$$\delta S = \int_{\mathcal{M}} d^d x \, \theta_\beta \varrho^\beta, \qquad \varrho^\beta := \sum_{s=0}^{n} \frac{\partial^s \left[E_A (\Omega^{A\beta})^{\mu_1 \mu_2 \ldots \mu_s} \right]}{\partial x^{\mu_1} \partial x^{\mu_2} \ldots \partial x^{\mu_s}}. \tag{43}$$

If, under the field variations (42) for some $\Omega^{A\beta}$, the action remains invariant $\delta S = 0$, then we have that

$$\varrho^\beta = 0. \tag{44}$$

holds true off–shell (i.e. without making use of $E_A \overset{!}{=} 0$). In such a case, (42) and (44) are known as the gauge transformations and gauge identities in the theory, respectively.

The recursive construction of the gauge generators $\Omega^{A\beta}$ has been a subject of vivid interest for decades. The approach in [96] is perhaps the most befitting to our own exposition, requiring only a suitable adaptation from particle systems to field theories that is devoid of subtleties. We shall not present the corresponding discussion here because, for the theories at hand, the explicit form of the gauge transformations is already known. This a priori knowledge allows to straightforwardly infer the generators $\Omega^{A\beta}$.

Further, given (42), g is equal to the number of distinct θ parameters there present. Namely, $\beta = 1, 2, \ldots, g$. On the other hand, e is equal to the number of distinct $(\theta, \dot{\theta}, \ddot{\theta}, \ldots)$ that appear in (42). Obviously, $e \geq g$.

We have now achieved our goal: we have determined g and e in theories for which the gauge transformations are known beforehand. Together with the number of on–shell constraints l determined in the previous section 5.2.1, we can now count degrees of freedom in (almost) all first–order field theories by means of (22).

We stress that the determination of g and e is possible and has been made systematic in theories for which the gauge transformations are unknown from the onset. The calculations in such theories are more involved, but there is no conceptual obstacle that has to be overcome. The reader can consult [95] for the explicit derivation of the gauge generators in Yang–Mills theory and both the metric and Palatini formulations of General Relativity, by means of the formalism put forward in [96].

We conclude this section with an important remark. The off–shell identities (44) here obtained need not always coincide with the Lagrangian identities defined in 2. Maxwell's theory is an example where this coincidence does happen. Indeed, the Bianchi identities are the projection onto the kernel of the primary Hessian of the theory's $U(1)$ gauge identity. On the other hand, the diffeomorphism invariance of relativistic theories does not lead to Lagrangian identities. For theories of this second type, our above off–shell analysis is indispensable.

Part III
Applications

6 Non–linear electrodynamics

Having explained in detail constrained systems, we proceed by applying them to physical theories. We will explicitly employ the algorithms previously introduced, thereby demonstrating their usefulness and exposing the mathematical structure of the theories.

In this section, we introduce non–linear electrodynamics in arbitrary dimensions and analyze it in the Hamiltonian, as well as in the Lagrangian, framework. We make use of Mathematica and, in particular, the xAct package [97] to complement and double–check lengthy calculations.

The Hamiltonian analysis terminates after the secondary stage and we find 2 first–class and $d(d-1)$ second–class constraints, resulting in $(d-2)$ physical degrees of freedom.

The Lagrangian analysis terminates at the secondary level and we find a total of d^2-d+1 constraints. We determine the number of gauge identities g and effective gauge parameters e in two complementary ways. In both cases, we find $g=1$ and $e=2$, which results in $d-2$ physical degrees of freedom.

Historically, non–linear electrodynamics (NLED) was an attempt to overcome conceptual difficulties when trying to quantize electromagnetism [98]. In fact, the development of quantum electrodynamics (QED) showed that non–linear effects arise in the interaction of electromagnetic fields with the vacuum [99]. It is possible to obtain NLED from QED, such that it can be considered as a classical model of the vacuum polarization [100]. It was, among other things, used to construct the first regular exact black hole solution in General Relativity [101]. Diverse applications of NLED to cosmology and high–energy astrophysics are recently investigated (see [99] and references therein).

6.1 $\mathcal{H}(\mathcal{P})$–formulation of NLED

The naive way to introduce non–linear effects in electrodynamics is to include general functions of the field strength invariant $\mathcal{F}=\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$. Alternatively, one can describe the system rewriting the Lagrangian $\mathcal{L}(\mathcal{F})$ in terms of its Legendre transform $\mathcal{H}$ [101]. The advantage of this formulation is that many results can be achieved more easily [102], e.g. the variation of the Lagrangian with respect to the electromagnetic field A_μ now gives the non–linear version of the Maxwell equations $\nabla_\mu P^{\mu\nu}=0$ [103]. We will study NLED in flat space with Minkowski metric $\eta_{\mu\nu}$, but the analysis does not change if carried out over a fixed curved background. The Lagrangian of non–linear electrodynamics can be written as

$$\mathcal{L}=-\frac{1}{2}F_{\mu\nu}P^{\mu\nu}+\mathcal{H}(\mathcal{P}), \tag{45}$$

where $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu$ corresponds to the intensity of the electric field and magnetic induction vectors and $P^{\mu\nu}=-P^{\nu\mu}$ corresponds to the electric induction and the intensity of the magnetic field vectors. The *structural function* $\mathcal{H}$ can in general depend on the two invariants of the tensor $P_{\mu\nu}$, i.e.

$$\mathcal{P}\coloneqq\frac{1}{4}P^{\mu\nu}P_{\mu\nu}\quad\text{and}\quad\mathcal{P}^*\coloneqq\frac{1}{4}P^{*\mu\nu}P_{\mu\nu},$$

but we will only consider $\mathcal{H} = \mathcal{H}(\mathcal{P})$ here [100]. By means of a Legendre transformation we find

$$\mathcal{H} = 2\mathcal{F}\frac{\partial \mathcal{L}}{\partial \mathcal{F}} - \mathcal{L}.$$

The variation of the action with respect to the conjugate antisymmetric tensor $P^{\mu\nu}$ gives the so–called *constitutive relations*

$$F_{\mu\nu} = \frac{\partial \mathcal{H}}{\partial \mathcal{P}} P_{\mu\nu},$$

and one can express the Lagrangian (45) in terms of $\mathcal{H}$ and $\mathcal{P}$ as

$$\mathcal{L} = -2\mathcal{P}\frac{\partial \mathcal{H}}{\partial \mathcal{P}} + \mathcal{H}.$$

The theory has to transition into Maxwell's electrodynamics in the weak field limit, demanding that $\mathcal{H} \approx \mathcal{P}$ holds for $\mathcal{P} \ll 1$. Furthermore, $\mathcal{H}$ is required to satisfy the weak energy condition, such that $\mathcal{H} < 0$ and $\frac{\partial \mathcal{H}}{\partial \mathcal{P}} > 0$ [101]. In the following, we write

$$\mathcal{H}' \equiv \frac{d\mathcal{H}}{d\mathcal{P}},$$

for derivatives of $\mathcal{H}(\mathcal{P})$ with respect to $\mathcal{P}$ and note that

$$\frac{\partial \mathcal{H}}{\partial P^i} = \frac{d\mathcal{H}}{d\mathcal{P}}\frac{\partial P}{\partial P^i} = P_i \mathcal{H}', \qquad \frac{\partial \mathcal{H}}{\partial P^{ij}} = \frac{d\mathcal{H}}{dP}\frac{\partial \mathcal{P}}{\partial P^{ij}} = \frac{1}{2} P_{ij} \mathcal{H}'.$$

6.2 Hamiltonian analysis

As a starting point, we take the Lagrangian from Eq. (45) and explicitly carry out the summation over repeated indizes. In other words, we decompose it into time and space components and arrive at

$$\mathcal{L} = -\dot{A}_i P^i + P^i \partial_i A + P^{ij} \partial_j A_i + \mathcal{H}. \tag{46}$$

Our set of generalized coordinates is

$$Q = \{A, A_i, P^i, P^{ij}\},$$

where A is the zero component of A_μ and $P^i := P^{0i}$. The corresponding set of generalized momenta reads

$$P = \{\Pi, \Pi^i, \mathbb{P}_i, \mathbb{P}_{ij}\}.$$

Notice that there are $N = d(d+1)/2$ configuration space variables. It is easy to see that the conjugate momenta of A, P^i and P^{ij} vanish identically, since

$$\frac{\partial \mathcal{L}}{\partial \dot{A}} = \frac{\partial \mathcal{L}}{\partial \dot{P}^i} = \frac{\partial \mathcal{L}}{\partial \dot{P}^{ij}} = 0. \tag{47}$$

Only the contribution from A_i,

$$\Pi^i = \frac{\partial \mathcal{L}}{\partial \dot{A}_i} = -P^i, \tag{48}$$

is non-trivial. The canonical Hamiltonian $H_{can} = \sum_n \frac{\partial L}{\partial \dot{q}_n}\dot{q}_n - \mathcal{L}$ reads

$$H_{can} = -P^i \partial_i A - P^{ij} \partial_j A_i - \mathcal{H}.$$

6.2.1 Primary stage

Primary constraint submanifold The primary constraints can be read off from Eqs. (47) and (48), such that

$$\Psi_m = \{\Pi, \Pi^i + P^i, \mathbb{P}_i, \mathbb{P}_{ij}\} \overset{1}{:\approx} 0, \quad m = 1, \dots, 4 \ ,$$

defines the primary constraint submanifold.

Primary Hamiltonian We have to include the primary constraints (the vanishing momenta as well as the constraint from relation (48)) via Lagrange multipliers. We obtain the primary Hamiltonian

$$H_p = H_{can} + \lambda_i \overbrace{(\Pi^i + P^i)}^{(\Psi_2)^i} + \lambda\Pi + \Lambda^i \mathbb{P}_i + \Lambda^{ij}\mathbb{P}_{ij},$$

where λ and Λ are arbitrary and only denoted differently for convenience.

Stability conditions We want to check if the constraints and relations on the momenta are stable, i.e. constant in time. To do so we use Eq. (19), such that

$$\dot{\Psi}_m = \{H_p, \Psi_m\},$$

for our primary constraints. We find the following stability conditions

1. $\{H_p, \Pi\} = \partial_i P^i$,
2. $\{H_p, \Pi^i + P^i\} = \partial_j P^{ij} - \Lambda^i \quad \rightarrow \Lambda^i \overset{1}{\approx} \partial_j P^{ij}$,
3. $\{H_p, \mathbb{P}_i\} = -P_i\mathcal{H}' + \lambda_i - \partial_i A \quad \rightarrow \lambda_i \overset{1}{\approx} P_i\mathcal{H}' + \partial_i A$,
4. $\{H_p, \mathbb{P}_{ij}\} = -\frac{1}{2}P_{ij}\mathcal{H}' + \partial_{[i}A_{j]}$.

Condition 2 and 3 determine two of the Lagrange multipliers. They can be fixed on the primary constraint submanifold. Condition 1 and 4 on the other hand do not include arbitrary functions and thus are true constraints that need to be enforced.

6.2.2 Secondary stage

Secondary constraint submanifold Ensuring the stability of the primary constraints generates two secondary constraints

$$\chi_r = \{\partial_i P^i, -\frac{1}{2}P_{ij}\mathcal{H}' + \partial_{[i}A_{j]}\} \overset{2}{:\approx} 0, \quad r = 1, 2 \ ,$$

which define the secondary constraint submanifold.

Secondary Hamiltonian Now, we add the new constraints to the primary Hamiltonian and fix the Lagrange multipliers that were determined in the first stage. We obtain the secondary Hamiltonian

$$\begin{aligned} H_s = H_{can} &+ (P_i\mathcal{H}' + \partial_i A)(\Pi^i + P^i) + \partial_j P^{ij}\mathbb{P}_i \\ &+ \lambda\Pi + \Lambda^{ij}\mathbb{P}_{ij} + \tilde{\lambda}(\partial_i P^i) + \tilde{\lambda}^{ij}(-\frac{1}{2}P_{ij}\mathcal{H}' + \partial_{[i}A_{j]}), \end{aligned}$$

where $\tilde{\lambda}$ are the Lagrange multipliers for the new constraints.

Consistency relations The consistency conditions $\{\Psi_m, H_s\}$ for the primary constraints read

1. $\{H_s, \Pi\} = -\partial_i \Pi^i \overset{2}{\approx} 0$,
2. $\{H_s, \Pi^i + P^i\} = \partial_j \tilde{\lambda}^{ij}$,
3. $\{H_s, \mathbb{P}_i\} = -\partial_i \tilde{\lambda} + \frac{1}{2}\tilde{\lambda}^{ab} P_{ab} P_i \mathcal{H}''$,
4. $\{H_s, \mathbb{P}_{ij}\} = \frac{1}{2}\tilde{\lambda}^{kl}\left[\eta_{i[k}\eta_{l]j}\mathcal{H}' + \frac{1}{2}P_{kl}P_{ij}\mathcal{H}''\right] \equiv \tilde{\lambda}^{kl}\tau_{klij}$.

As expected, relation 1 is not functionally independent from our prior constraints, since

$$\partial_i \Pi^i = \partial_i (\Psi_2)^i - \chi_1 \overset{2}{\approx} 0,$$

and therefore weakly vanishes in the secondary constraint submanifold. Conditions 2 to 4 fix one or more Lagrange multipliers, the exact solution is not important for our purposes. For convenience, we introduced the object

$$\tau_{klij} = \frac{1}{2}\left[\eta_{i[k}\eta_{l]j}\mathcal{H}' + \frac{1}{2}P_{kl}P_{ij}\mathcal{H}''\right],$$

since it will appear several times in the following analysis.

Stability conditions As for the primaries, the secondary constraints need to be conserved in time. We calculate

1. $\{H_s, \partial_i P^i\} = 0$,
2. $\{H_s, (\partial_{[i}A_{j]} - \frac{1}{2}P_{ij}\mathcal{H}')\} = \frac{1}{2}P_{ij}\partial_l P^{kl} P_k \mathcal{H}'' + \Lambda^{kl}\tau_{klij}$.

Condition 1 tells us that the constraint is automatically stable and does not evolve in time. Condition 2 determines a Lagrange multiplier. Since no new constraints arise at the secondary stage the algorithm terminates here. As long as we are only interested in counting the physical degrees of freedom there is no need to explicitly solve for the Lagrange multipliers.

Final Hamiltonian The form of the final Hamiltonian is not explicitly worked out here. To do so one has to solve for the Lagrange multipliers that got determined at the secondary stage and plug them into H_s. This is a technical problem to be solved and not necessary for the counting of degrees of freedom and the canonical structure, as already pointed out before. Strictly speaking, one should determine the final Hamiltonian and check the two secondary constraints for consistency — this might yield another relation on Lagrange multipliers but cannot generate a new constraint. The form of the final Hamiltonian can be of interest though, since the number of arbitrary Lagrange multipliers should be related to the gauge symmetry of the theory. For (non–linear) electrodynamics we would expect two λ's to remain arbitrary.

6.2.3 Degrees of freedom

To count the physical degrees of freedom using Eq. (21) we still need to classify the full set of constraints into first– and second–class.

First and second class constraints Our full set of constraints is

$$\{\Psi,\chi\}=\{\Pi,\Pi^i+P^i,\mathbb{P}_i,\mathbb{P}_{ij},\partial_i P^i,\partial_{[i}A_{j]}-\frac{1}{2}P_{ij}\mathcal{H}'\}.$$

Per definition a first class constraint has a vanishing Poisson bracket with every other constraint. If a constraint does not meet that criterion it is called second–class. Our set of constraints is fairly easy and we find the following situation:

$\{\cdot,\cdot\}$	Π	Π^i+P^i	$\mathbb{P}_i$	$\mathbb{P}_{ij}$	$\partial_i P^i$	$\partial_{[i}A_{j]}-\frac{1}{2}P_{ij}\mathcal{H}'$
Π	0	0	0	0	0	0
Π^k+P^k	0	0	δ^k_i	0	0	0
$\mathbb{P}_k$	0	δ^i_k	0	0	0	0
$\mathbb{P}_{kl}$	0	0	0	0	0	τ_{klij}
$\partial_i P^i$	0	0	0	0	0	0
$\partial_{[k}A_{l]}-\frac{1}{2}P_{kl}\mathcal{H}'$	0	0	0	τ_{ijkl}	0	0

One might wonder about the fact that τ_{ijkl} appears in the constraint algebra. Indeed, it seems as if there are two branches of the theory, depending on the specific choice of $\mathcal{H}$. However, for generic field configurations, where all Q's are functionally independent, there is no solution that satisfies $\tau_{ijkl}=0$. Therefore, we do not have any branching of the theory and the constraint structure is unambiguous. Interestingly, one could also reason from the above situation that $\tau_{ijkl}=0$ cannot be possible for a general theory of the form (45). Otherwise, it could not be considered dynamically equivalent to standard electromagnetism.

Counting We have the first–class constraints

$$\mathrm{FC}=\{\Pi,\partial_i P^i\},$$

and the second–class constraints

$$\mathrm{SC}=\{\Pi^i+P^i,\mathbb{P}_i,\mathbb{P}_{ij},\partial_{[i}A_{j]}-\frac{1}{2}P_{ij}\mathcal{H}'\}.$$

That amounts to a total number of $N_1=2$ first–class constraints and $N_2=d(d-1)$ second–class constraints. Together with $N=\frac{1}{2}d(d+1)$ we find for the total number of physical degrees of freedom in d dimensions

$$n_{dof}=\frac{1}{2}d(d+1)-2-\frac{d(d-1)}{2}=d-2.$$

As expected, non–linear electrodynamics propagates the same number of physical degrees of freedom as Maxwell's theory. In particular, for $d=4$ we recover the 2 degrees of freedom corresponding to the independent polarization directions of electromagnetic waves. Of course NLED respects the constraint structure of the Maxwell–Proca theory [20], since it constitutes a sub–theory thereof.

6.3 Lagrangian analysis

In the following we apply our algorithm from Sec. 5.2 to the Lagrangian of non–linear electrodynamics (46).

6.3.1 Primary level

The Lagrangian (46) does not contain higher than linear terms in time derivatives, resulting in a vanishing primary Hessian

$$W_{AB} = 0.$$

From this we can immediately conclude that there are $\dim(W) = \frac{d(d+1)}{2}$ null vectors of W. Furthermore, we can easily convince ourselves that the first two contributions from (27) vanish and we are left with

$$\alpha_B = (\partial^i_B \partial_A \mathcal{L})\partial_i Q^A + (\partial_{\dot{B}} \partial_A \mathcal{L})\dot{Q}^A - \partial_B \mathcal{L}.$$

Straightforward computation yields

$$\begin{aligned}
\alpha_1 &= -\partial_i P^i, \\
\alpha_2 &= \partial_j P^{ji} + \dot{P}^i, \\
\alpha_3 &= -\dot{A}_i + \partial_i A + P_i \mathcal{H}' \\
\alpha_4 &= -\partial_{[i} A_{j]} + \frac{1}{2} P_{ij} \mathcal{H}'.
\end{aligned}$$

We choose the trivial null vectors of W as

$$(\gamma_I)^A = (0, \dots, 0, 1, 0, \dots, 0),$$

where $I = 1, \dots, 4$ and the 1 appears at the A^{th} position. Explicitly,

$$\begin{aligned}
(\gamma_1)^A &= (1, 0, 0, 0), \\
(\gamma_2)^A &= (0, \delta^j_i, 0, 0), \\
(\gamma_3)^A &= (0, 0, \delta^i_j, 0), \\
(\gamma_4)^A &= (0, 0, 0, \delta^{[i}_k \delta^{j]}_l).
\end{aligned}$$

Thus, we find for the primary constraints $\varphi_I = (\gamma_I)^A \alpha_A$

$$\begin{aligned}
\varphi_1 = (\gamma_1)^A \alpha_A &= \alpha_1 = -\partial_i P^i, \\
(\varphi_2)^i = (\gamma_2)^A \alpha_A &= \partial_j P^{ji} + \dot{P}^i, \\
(\varphi_3)_i = (\gamma_3)^A \alpha_A &= \alpha_3 = -\dot{A}_i + \partial_i A + P_i \mathcal{H}', \\
(\varphi_4)_{ij} = (\gamma_4)^A \alpha_A &= \alpha_4 = -\partial_{[i} A_{j]} + \frac{1}{2} P_{ij} \mathcal{H}'.
\end{aligned}$$

Functional independence By explicit evaluation of the Jacobian

$$J_{I\Lambda} := \frac{\partial \varphi_I}{\partial X^\Lambda}, \qquad X^\Lambda = \{Q^A, \dot{Q}^A\},$$

we verify that NLED is irreducible at the primary level, i.e. $M_1' = M_1$ and thus all primary constraints are functionally independent.

Primary constraint surface Setting $\varphi_I \overset{1}{:\approx} 0$ defines the primary constraint submanifold. In the following we have to ensure their stability.

6.3.2 Secondary level

Stability Consistency requires us to not only enforce the primary constraints, but also to ensure that these are preserved under time evolution. Explicitly, we use the secondary Euler–Lagrange equations of motion

$$\widetilde{E}_J = \ddot{Q}^A (\gamma^I)_A \widetilde{W}_{IJ} + \widetilde{\alpha}_J \overset{!}{\underset{1}{\approx}} 0,$$

in order to calculate $\widetilde{E}_J := \dot{\varphi}_J \overset{!}{\underset{1}{\approx}} 0$.

Secondary α's Since the primary Hessian $W_{AB} = 0$, we find that its pseudoinverse M^{AB} vanishes as well and Eq. (34) reduces to

$$\widetilde{\alpha}_J = \dot{Q}^A \partial_A \varphi_J + \partial_i \dot{Q}^A \partial_A^i \varphi_J.$$

Straightforward computation yield

$$\begin{aligned}
\widetilde{\alpha}_1 &= -\partial_i \dot{P}^i, \\
\widetilde{\alpha}_2 &= \partial_i \dot{P}^{ij}, \\
\widetilde{\alpha}_3 &= \dot{P}_i \mathcal{H}' + \dot{P}^k P_i P_k \mathcal{H}'' + \frac{1}{2} \dot{P}^{kl} P_i P_{kl} \mathcal{H}'' + \partial_i \dot{A}, \\
\widetilde{\alpha}_4 &= \frac{1}{2} \dot{P}^k P_{ij} P_k \mathcal{H}'' + \frac{1}{2} \dot{P}_{ij} \mathcal{H}' + \frac{1}{4} \dot{P}^{kl} P_{ij} P_{kl} \mathcal{H}'' - \partial_{[i} \dot{A}_{j]}.
\end{aligned}$$

Secondary Hessian From its defining properties, we can deduce that

$$\widetilde{W}_{1J} = \frac{\partial \varphi_J}{\partial \dot{A}} = 0, \qquad \widetilde{W}_{2J} = \frac{\partial \varphi_J}{\partial \dot{A}_i}, \qquad \widetilde{W}_{3J} = \frac{\partial \varphi_J}{\partial \dot{P}^i}, \qquad \widetilde{W}_{4J} = \frac{\partial \varphi_J}{\partial \dot{P}^{ij}},$$

and the only non–vanishing components are

$$\widetilde{W}_{23} = -\delta_i^j, \qquad \widetilde{W}_{32} = \delta_j^i,$$

or in matrix form

$$\widetilde{W}_{IJ} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -\delta_i^j & 0 \\ 0 & \delta_j^i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We choose as the null vectors of the secondary Hessian, $(\tilde{\gamma}_R)^I \widetilde{W}_{IJ} = 0$, the following objects

$$(\tilde{\gamma}_1)^I = (1, 0, 0, 0), \qquad (\tilde{\gamma}_2)^I = (0, 0, 0, \delta_m^{[i} \delta_n^{j]})$$

Secondary constraints Stability demands that

$$(\tilde{\gamma}_R)^I \widetilde{E}_I = (\tilde{\gamma}_R)^I \widetilde{\alpha}_I \equiv \tilde{\varphi}_R \overset{1}{\approx} 0,$$

and if this is not identically fulfilled we enter the secondary stage of the constraint algorithm. We find

$$\begin{aligned}
\tilde{\varphi}_1 = (\tilde{\gamma}_1)^I \widetilde{\alpha}_I = \widetilde{\alpha}_1 &= -\partial_i \dot{P}^i = -\partial_i (\varphi_2)^i \overset{1}{\approx} 0, \\
\tilde{\varphi}_2 = (\tilde{\gamma}_2)^I \widetilde{\alpha}_I = \widetilde{\alpha}_4 &= \frac{1}{2}\dot{P}^k P_{ij} P_k \mathcal{H}'' + \frac{1}{2}\dot{P}_{ij}\mathcal{H}' + \frac{1}{4}\dot{P}^{kl} P_{ij} P_{kl}\mathcal{H}'' - \partial_{[i}\dot{A}_{j]} = \\
&= \frac{1}{2} P_{ij}\frac{d\mathcal{H}'}{dt} + \frac{1}{2}\dot{P}_{ij}\mathcal{H}' - \partial_{[i}\dot{A}_{j]} \overset{1}{\not\approx} 0.
\end{aligned}$$

It turns out that $\tilde{\varphi}_1$ is equal to the divergence of a primary constraint and when evaluated on the primary constraint submanifold it vanishes weakly. $\tilde{\varphi}_2$ on the other hand does not vanish weakly. It is possible to rewrite it using the primary constraints, but it will never identically vanish. Thus, we only found one true secondary constraint.

Closing of the algorithm The tertiary Hessian is defined as

$$\widehat{W}_{RS} \coloneqq (\tilde{\gamma}_R)^I (\gamma_I)^A \partial_{\dot{A}} \tilde{\varphi}_S,$$

where we only take into account the null vectors on the secondary stage that generate a true constraint, i.e. we exclusively consider $(\tilde{\gamma_2}^I)$. The combination

$$(\tilde{\gamma}_2)^I (\gamma_I)^4,$$

gives a non–zero contribution. Hence, the tertiary Hessian is the 1×1–matrix

$$\widehat{W} = (\tilde{\gamma}_2)^I (\gamma_I)^4 \frac{\partial}{\partial \dot{P}^{mn}} \tilde{\varphi}_2 = \frac{1}{2}\frac{\partial}{\partial \dot{P}^{mn}}\left[P_{ij}\mathcal{H}''\left(\dot{P}^k P_k + \frac{1}{2}\dot{P}^{kl}P_{kl}\right)\right] = \frac{1}{4} P_{ij} P_{mn}\mathcal{H}'',$$

which obviously is of full rank. The algorithm closes and we do not enter a tertiary constraint surface.

6.3.3 Degrees of freedom

We found a total of

$$l = d^2 - d + 1$$

Lagrangian constraints. From the Hamiltonian side we can infer [84, 85] that $g = 1$ and $e = 2$, such that we count

$$n_{dof} = N - \frac{1}{2}(l + g + e) = \frac{1}{2}d(d+1) - \frac{1}{2}(4 - d + d^2) = d - 2.$$

Hence, we corroborate the number of physical degrees of freedom obtained with the Hamiltonian approach.

Off–shell identities The last thing to be done is to derive $e+g$ independently and not borrow it from the Hamiltonian side. We know that the gauge identities will be related to functionally dependent Lagrangian constraints. We therefore inspect

$$\tilde{\varphi}_1 = \frac{\partial}{\partial t}\varphi_1 = -\partial_i \dot{P}^i \equiv \frac{\partial}{\partial t}(E_1^0)$$

and use the functional dependence discovered in the secondary stage

$$-\partial_i \dot{P}^i = -\partial_i (\varphi_2)^i, \equiv -\partial_i (E_2^0)$$

to write

$$G = \frac{\partial}{\partial t}(E_1^0) + \partial_i (E_2^0) = 0. \tag{49}$$

Notice that this is an off–shell relation. Following [84] we multiply Eq. (49) with an arbitrary spacetime function $\epsilon(x,t)$, such that

$$\cancel{\frac{\partial}{\partial t}(\epsilon E_1^0)} + \cancel{\partial_i (\epsilon E_2^0)} - (\partial_i \epsilon) E_2^0 - (\frac{\partial}{\partial t}\epsilon) E_1^0 = 0,$$

where the surface terms vanish because technically we operate under an integral sign [95]. From this we read off the gauge transformation as

$$\delta_\epsilon A = \dot{\epsilon}, \quad \text{and} \quad \delta_\epsilon A_i = \partial_i \epsilon,$$

which we combine to find

$$\delta_\epsilon A_\mu = \partial_\mu \epsilon.$$

Hence, we recover the well–known gauge transformation from Maxwell electrodynamics. Additionally, we find that

$$\delta_\epsilon P_{\mu\nu} = 0,$$

i.e. there exists no gauge symmetry regarding these fields and we confirm $g=1$ and $e=2$.

7 General Relativity

We introduce in detail the general theory of relativity. We comment on curvature, its derivation from an action principle and associated symmetries. Subsequently, we discuss its manifestly first–order formulation — Palatini. In contrast to the $d>2$ case, two–dimensional Palatini is not dynamically equivalent to Einstein–Hilbert. Despite this fact, an investigation of Palatini's constraint structure in two spacetime dimensions is of immediate interest: it closely resembles the structure of constraints in $d>2$ Einstein–Hilbert gravity.

We analyze $d=2$ Palatini in the Hamiltonian, as well as in the Lagrangian, framework. As before, we use Mathematica to complement and double-–check lengthy calculations. The Hamiltonian analysis terminates after the tertiary stage and we find 3 first–class and 12 second—class constraints, resulting in 0 physical degrees of freedom. The Lagrangian analysis terminates after the primary level and we find a total of $l=9$ constraints. We determine the number of gauge identities g and effective gauge parameters e from the off–shell identities. We find $g=3$ and $e=6$, which results in 0 propagating degrees of freedom.

As expected, both approaches yield the same number of physical degrees of freedom. In the end, we comment on a potential tension between our Hamiltonian and Lagrangian methods. This might stem from a conceptual difference between reductional and non–reductional techniques.

General Relativity is the geometric theory of gravitation published by Albert Einstein in 1915 and constitutes one of the pillars of modern physics. It describes gravity as a geometric property of spacetime. The curvature of spacetime and the energy–momentum tensor are directly related through the famous Einstein field equations. The implications of General Relativity are pervading modern astrophysics. To name only a few, gravitational lensing, gravitational waves, black holes and current models of cosmology rely on Einstein's theory. On all other scales than the microscopic world, General Relativity's predictions are strikingly accurate [104].

7.1 Einstein gravity

We model spacetime as the triple (M, g, Γ). M is a d–dimensional, orientable[37], pseudo–Riemannian manifold endowed with a Lorentz metric g, i.e. a metric with signature $(1, d-1)$ [105]. This metric allows tangent vectors to be classified into *timelike*, *null* or *spacelike*. Γ is the *Levi–Civita connection* of (M, g), which is the unique affine connection, with associated covariant derivative ∇, such that it

1. preserves the metric, i.e. $\nabla g = 0$, and
2. is *torsion–free*, i.e. $\nabla_X Y - \nabla_Y X = [X, Y],\ \forall X, Y \in \Gamma(TM)$.

Physically speaking, both of these properties follow from Einstein's principle of equivalence [106]. In the case of the Levi–Civita connection the connection coefficients are called *Christoffel symbols*. The torsion freeness condition implies the symmetry in the lower indices

$$\Gamma^{\lambda}_{\mu\nu} = \Gamma^{\lambda}_{\nu\mu},$$

where the Christoffel symbols are uniquely determined by

$$\Gamma^{\lambda}_{\mu\nu} = \frac{1}{2} g^{\lambda\sigma}(\partial_\nu g_{\sigma\mu} + \partial_\mu g_{\sigma\nu} - \partial_\sigma g_{\mu\nu}).$$

7.1.1 Curvature

Parallel transport The notion of *parallel transport* is a very intuitive one: what does it mean to keep a vector "pointing in the same direction" as it moves along a curve? In flat space, the value of a vector that is parallel transported along a closed curve will always coincide with its initial value, as shown in Fig. 14.

The situation is drastically different once we consider curved spaces. The most obvious example of a curved space might be $\mathcal{S}^2$ — the surface of a sphere. If we picture a closed curve on the sphere and a vector that is parallel transported along, we notice that its initial and final value won't coincide in general. As one can see in Fig. 15, the vector now points in a different direction. This is obviously related to the fact that the surface of the sphere is curved. In other words, in a curved space the result of parallel transporting a vector from one point the another depends on the path taken between the points. This is the crucial difference between flat and curved spaces [45].

In section 3.1.3 we introduced the connection coefficients as functions that specify how the basis vectors of the tangent spaces change from point to point. Thus, it is not surprising that the connection, or rather the covariant derivative, contains all the information we need in order to specify parallel transport [107]. Given a curve $\gamma^\mu(\lambda)$, a (p, q)–tensor is said to be parallel transported if its covariant derivative along that curve vanishes, i.e. [45]

$$\frac{d\gamma^\sigma}{d\lambda} \nabla_\sigma T^{\mu_1,\dots\mu_p}{}_{\nu_1,\dots,\nu_q} = 0.$$

[37] Which is a technical requirement for well–defined integration on the manifold.

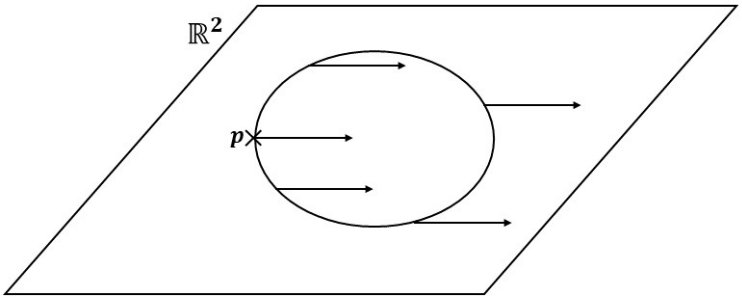

Figure 14: After being parallel transported along the closed curve, the vector will always point in the same direction as it did before.

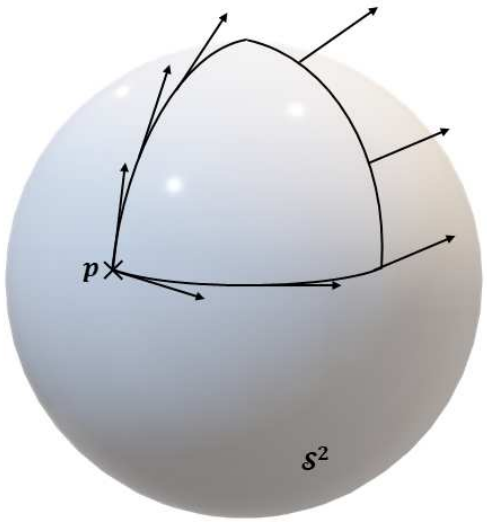

Figure 15: After being parallel transported along the closed curve, the vector will generically point in a different direction.

Riemann curvature tensor Now that we know how to transport vectors (and tensors in general) along a curve, we can use this knowledge to characterize the curvature of our manifold intrinsically. The commutator of two covariant derivatives measures the difference between parallel transporting a tensor first one way and then the other, versus the opposite ordering [45]. Consider a covector field ω_σ and let f be a smooth function. Calculate the action of two derivative operators applied to $(f\omega_\sigma)$ as

$$\begin{aligned}\nabla_\mu\nabla_\nu(f\omega_\sigma) &= \nabla_\mu(\omega_\sigma f + f\nabla_\nu\omega_\sigma)\\ &= (\nabla_\mu\nabla_\nu f)\omega_\sigma + \nabla_\nu f\nabla_\mu\omega_\sigma + \nabla_\mu f\nabla_\nu\omega_\sigma + f\nabla_\mu\nabla_\nu\omega_\sigma,\end{aligned}$$

and subtract from this the tensor $\nabla_\nu\nabla_\mu(f\omega_\sigma)$. We are left with

$$(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)(f\omega_\sigma) = f(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)\omega_\sigma, \tag{50}$$

i.e. we specified the failure of successive differentiation to commute when applied to a covector field [107]. Upon further inspection of Eq. (50), one realizes that $(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)$ defines a linear map

from covectors at the point p to $(0,3)$–tensors at p. That means, its action is that of a $(1,3)$–tensor and there exists a tensor field $R^{\rho}{}_{\mu\nu\sigma}$, such that for all covector fields ω_σ

$$\nabla_\mu \nabla_\nu \omega_\sigma - \nabla_\nu \nabla_\mu \omega_\sigma = R^{\rho}{}_{\mu\nu\sigma}\omega_\rho.$$

One can repeat the above construction for a vector field X^σ to find

$$\nabla_\mu \nabla_\nu X^\sigma - \nabla_\nu \nabla_\mu X^\sigma = -R^{\sigma}{}_{\mu\nu\rho}X^\rho,$$

which can be generalized to arbitrary (p,q)–tensor fields [107]. $R^{\rho}{}_{\mu\nu\sigma}$ is called the *Riemann curvature tensor* and in terms of the Christoffel symbols it reads

$$R^{\mu}{}_{\nu\alpha\beta} = \partial_\alpha \Gamma^{\mu}{}_{\nu\beta} - \partial_\beta \Gamma^{\mu}{}_{\nu\alpha} + \Gamma^{\mu}{}_{\sigma\alpha}\Gamma^{\sigma}{}_{\nu\beta} - \Gamma^{\mu}{}_{\sigma\beta}\Gamma^{\sigma}{}_{\nu\alpha}.$$

One way to think about the Riemann tensor is as a measure of how much the components of a tensor change when it is parallel transported along an infinitesimal closed curve on the manifold [45]. There are several useful objects, which appear often in the context of General Relativity, that one can construct from the Riemann tensor. Its trace over the first and third indices defines the *Ricci curvature tensor*

$$R_{\mu\nu} \coloneqq R^{\alpha}{}_{\mu\alpha\nu}.$$

If we furthermore trace the remaining indices of the Ricci tensor, we obtain the *(Ricci) scalar curvature*

$$R \coloneqq R^{\mu}{}_{\mu}.$$

A particular combination of these curvatures plays an important role in General Relativity, since it will appear directly in the equations of motion. We call this combination,

$$G_{\mu\nu} \coloneqq R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R,$$

the *Einstein tensor*.

7.1.2 Einstein–Hilbert action

Historically, Einstein proposed his theory of General Relativity as a consequence of the *principle of equivalence*, which is basically a statement on local Lorentz and position invariance [108]. A more modern approach is the derivation via a variational principle. The corresponding action functional, the *Einstein–Hilbert action*, was first proposed by Hilbert in 1915 [109]. It can be extended in order to include a cosmological constant Λ and coupling to matter fields, which are collectively denoted by ψ. We define the action of General Relativity as

$$S_{GR}[g_{\mu\nu}, \psi] = \underbrace{\frac{1}{16\pi G}\int (R^{\mu\nu}g_{\mu\nu} - 2\Lambda)\sqrt{-g}d^4x}_{S_{EH}} + \underbrace{\int \mathcal{L}_m(g_{\mu\nu}, \psi)\sqrt{-g}d^4x}_{S_M}, \tag{51}$$

where G is the gravitational constant, $\mathcal{L}_m$ is the minimally coupled matter Lagrangian. The scalar density $g = \det(g_{\mu\nu})$ has to be included to make the volume element invariant under general coordinate transformations [104].

Variation of the action The metric is the only independent structure on the manifold. Variation of Eq. (51) with respect to $g_{\mu\nu}$ can be split into the variation of the first part δS_{EH} and of the second part δS_M. One finds

$$(16\pi)\delta S_{EH} = \int (G_{\mu\nu} + \Lambda g_{\mu\nu})\delta g^{\mu\nu}\sqrt{-g}d^4x - surface\ term^{38}, \tag{52}$$

for the variation of the Einstein–Hilbert term including a cosmological constant [104, 106]. The variation of the second term yields

$$\delta S_M = -\frac{1}{2}\int T_{\mu\nu}\delta g^{\mu\nu}\sqrt{-g}d^4x, \tag{53}$$

where we have introduced the *energy–momentum tensor* as

$$T_{\mu\nu} \coloneqq -2\frac{\partial \mathcal{L}_m}{\partial g^{\mu\nu}} + \mathcal{L}_m g_{\mu\nu}.$$

Diffeomorphism invariance For a diffeomorphism invariant matter action S_M, one can show that $T_{\mu\nu}$ is always conserved by virtue of the matter field equations, i.e.

$$\nabla^\mu T_{\mu\nu} = 0.$$

This reinforces the interpretation of $T_{\mu\nu}$ as the energy–momentum tensor. By the same argument, we can deduce that the diffeomorphism invariance of S_{EH} implies

$$\nabla^\mu G_{\mu\nu} = 0,$$

which are known as *contracted Bianchi identities* [107].

7.1.3 Field equations

By the principle of stationary action we demand

$$\delta S_{GR} = \delta(S_{EH} + S_M) \overset{!}{=} 0,$$

in order to obtain the equations of motion. Combining Eq. (52) and Eq. (53) we thus find the *Einstein field equations*

$$R_{\mu\nu} + \frac{1}{2}g_{\mu\nu}(2\Lambda - R) = 8\pi G T_{\mu\nu}. \tag{54}$$

These equations tell us how the curvature of spacetime reacts to the presence of energy–momentum and can be thought of as a set of second–order differential equations for the metric tensor field $g_{\mu\nu}$ [45]. Since energy–momentum is a conserved quantity, special relativity can be recovered in the neighborhood of every point in spacetime [104]. Even though the metric tensor has 10 independent components in $d = 4$, the contracted Bianchi identities impose four constraints on the Ricci tensor $R_{\mu\nu}(x)$, such that there are only six truly independent field equations. If a metric is a solution to Eq. (54) in a particular coordinate system x^μ, it should also be a solution in any other coordinate system $\tilde{x}^\mu$, i.e. there is some arbitrariness in the choice of coordinates which corresponds to four unphysical degrees of freedom in $g_{\mu\nu}$ [45].

[38] It should be noted that this term only vanishes if M has no boundary. In the case of a boundary ∂M, one has to include an extra boundary term S_B in the action, whose variation δS_B will exactly cancel the surface term [106].

Solutions The Einstein field equations are extremely complicated as differential equations, not to mention their non–linear character. It is very difficult to find a general solution to Eq. (54) without simplifying assumptions. One possibility is to assume symmetries of the metric tensor, which is directly related to our discussion on isometries [45]. But even with the help of symmetries, only a few physically relevant solutions are known at this point.

Most famously, the assumption of a *homogeneous* and *isotropic* universe leads to the *Friedmann–Robertson–Walker* metric as an exact solution. In our language, a spacetime is said to be (spatially) homogeneous if there exists a one–parameter family of spacelike hypersurfaces Σ_t foliating the spacetime, such that for each t and for any points $p, q \in \Sigma_t$ there exists an isometry of the metric tensor $g_{\mu\nu}$, which takes p into q [107].

Another famous exact solution can be found under the assumption of spherical symmetry, i.e. a spacetime metric which is invariant under rotations. This yields the *Schwarzschild solution*, an excellent approximation to the exterior gravitational field of stellar objects, such as our Sun. It even predicts the complete gravitational collapse of sufficiently massive bodies and the arising spacetime singularity, hidden by a *black hole* [107].

7.2 Palatini formulation

In the previous section we outlined Einstein's original formulation of General Relativity, which was solely in terms of the metric tensor $g_{\mu\nu}$. It is only in the case of the Levi–Civita connection that the metric is the only independent structure on the manifold. A connection should be viewed as additional structure on the manifold and can, in principle, be chosen to be independent from the metric. In the context of General Relativity, this idea to treat the connection as a priori independent from $g_{\mu\nu}$ is usually called *Palatini formulation*. In other words, the connection $\tilde{\Gamma}^{\mu}{}_{\nu\lambda}$ is no longer assumed to be metric–compatible, but as it turns out there will be constraints on its torsion [104].

7.2.1 Torsion and non–metricity

It is of course possible to relax either of the defining condition for a Levi–Civita connection. Consider a general connection

$$\tilde{\Gamma}^{\mu}{}_{\nu\lambda} = \Gamma^{\mu}{}_{\nu\lambda} + C^{\mu}{}_{\nu\lambda}, \tag{55}$$

where $\Gamma^{\mu}{}_{\nu\lambda}$ is the Levi–Civita connection and $C^{\mu}{}_{\nu\lambda}$ a not further specified $(1,2)$–tensor [110]. Denote by $\tilde{\nabla}_\mu$ the associated covariant derivative. We define the *torsion tensor* $T^{\mu}{}_{\nu\lambda}$ by

$$[\tilde{\nabla}_\mu, \tilde{\nabla}_\nu]\phi \coloneqq T^{\mu}{}_{\nu\lambda}\partial_\lambda\phi,$$

and the *non–metricity tensor* $Q_{\nu\lambda\mu}$ by

$$\tilde{\nabla}_\mu g_{\nu\lambda} \coloneqq -Q_{\nu\lambda\mu}.$$

In some sense, T and Q quantify by how much a connection fails to be torsion–free and metric–compatible, respectively. In terms of $C^{\mu}{}_{\nu\lambda}$ these tensors can be written as

$$\begin{aligned} T_{\lambda\mu\nu} &= C_{\lambda\mu\nu} - C_{\lambda\nu\mu} = 2C_{\lambda[\mu\nu]}, \\ Q_{\nu\lambda\mu} &= C_{\nu\lambda\mu} + C_{\lambda\nu\mu} = 2C_{(\nu\lambda)\mu}. \end{aligned}$$

If the torsion is zero and the connection is metric–compatible we have $C_{\lambda\mu\nu} = 0$, an it follows that $\tilde{\Gamma}^{\mu}{}_{\nu\lambda} = \Gamma^{\mu}{}_{\nu\lambda}$ [104, 110].

7.2.2 Palatini action

As a starting point, we can use the previous action (51) and add an explicit dependence on $\tilde{\Gamma}^{\mu}{}_{\nu\lambda}$. For simplicity we omit the cosmological constant Λ and the matter fields, which are only allowed to couple to the metric tensor[39]. We write

$$S_P[g_{\mu\nu}, \tilde{\Gamma}^{\mu}{}_{\nu\lambda}] = \frac{1}{16\pi G} \int R^{\mu\nu}(\tilde{\Gamma}) g_{\mu\nu} \sqrt{-g} d^4x, \tag{56}$$

for the *Palatini action* [107]. Most importantly, this formulation is manifestly first order in derivatives, whereas (51) also contains second–order contributions [111]. This makes its use for our purposes clear: since we use a procedure that is only applicable to first–order field theories, we will work in the Palatini formulation of General Relativity.

Variation of the action Variation of the action (56) with respect to the metric is identical to our prior result in Einstein gravity. We find

$$\delta_g S_P = \int \big(R_{\mu\nu}(\tilde{\Gamma}) - \frac{1}{2} g_{\mu\nu} R(\tilde{\Gamma})\big) \delta g^{\mu\nu} \sqrt{-g} d^4x,$$

hence the equations of motion

$$G_{\mu\nu}(\tilde{\Gamma}) \equiv R_{\mu\nu}(\tilde{\Gamma}) - \frac{1}{2} g_{\mu\nu} R(\tilde{\Gamma}) = 0, \tag{57}$$

take the same form as Eq. (54). But, these are not yet the vacuum Einstein field equations because the independent connection $\tilde{\Gamma}$ is not the Levi–Civita connection. In contrast to before, we now have another independent variable $\tilde{\Gamma}$ with respect to which we can vary the action S_P. One finds

$$\delta_{\tilde{\Gamma}} S_P = \int \big(g^{\alpha\beta} C^{\lambda\gamma}{}_{\lambda} + g^{\gamma\beta} C^{\alpha\lambda}{}_{\lambda} - C^{\beta\gamma\alpha} - C^{\alpha\beta\gamma}\big) \delta C_{\gamma\alpha\beta} \sqrt{-g} d^4x,$$

where we used the fact that

$$\begin{aligned} g^{\mu\nu} \delta R_{\mu\nu}(\tilde{\Gamma}) = g^{\mu\nu} \big(& C^{\lambda}{}_{\rho\lambda} \delta C^{\rho}{}_{\mu\nu} - C^{\rho}{}_{\mu\lambda} \delta C^{\lambda}{}_{\rho\nu} - C^{\rho}{}_{\lambda\nu} \delta C^{\lambda}{}_{\mu\rho} + C^{\rho}{}_{\mu\nu} \delta C^{\lambda}{}_{\rho\lambda} \big) \\ & + surface\ terms. \end{aligned}$$

Consequently, the conditions for the action to be stationary with respect to variations of an arbitrary $\tilde{\Gamma}$ read

$$g^{\alpha\beta} C^{\lambda\gamma}{}_{\lambda} + g^{\gamma\beta} C^{\alpha\lambda}{}_{\lambda} - C^{\beta\gamma\alpha} - C^{\alpha\beta\gamma} = 0, \tag{58}$$

which do not determine $C^{\alpha}{}_{\beta\gamma}$ uniquely. Hence, the Palatini action (56) alone does not give rise to well–defined equations of motion. The algebraic equations (58) can be shown to be equivalent to the condition

$$C^{\alpha}{}_{\beta\gamma} = \delta^{\alpha}_{\gamma} U_{\beta},$$

for an arbitrary vector U_β. Apparently, the requirement for the connection to dynamically become the Levi–Civita connection is $U_\beta = 0$. We can rewrite the torsion tensor as

$$T^{\sigma}{}_{\mu\nu} = C^{\sigma}{}_{\mu\nu} - C^{\sigma}{}_{\nu\mu} = \delta^{\sigma}_{\nu} U_{\mu} - \delta^{\sigma}_{\mu} U_{\nu},$$

[39] This is by definition of the Palatini formulation. Its generalization, where matter is allowed to couple to the connection, is known as *metric–affine gravity*.

and by taking its trace we find

$$U_\mu = \frac{1}{d-1} T^\nu{}_{\mu\nu},$$

which should vanish in order for the field equations of the Palatini formulation (57) to be dynamically equivalent to the Einstein field equations (54). Thus, we found the necessary and sufficient assumption to make, such that we recover the correct equations of motion: the trace of the torsion tensor is required to vanish [110]. Consider the family of general affine connections allowed by the Palatini formalism

$$\tilde{\Gamma}^\rho{}_{\mu\nu} = \Gamma^\rho{}_{\mu\nu} + U_\mu \delta^\rho_\nu. \tag{59}$$

It is argued in [112] that there is no reason to assign a preferred status to the Levi–Civita connection, since all Palatini connections lead to the same physics, in the sense that choosing any other connection of the family (59) does not lead to observable effects at the level of the Einstein equations or the trajectories of test particles. Accordingly, the Levi–Civita connection is then the simplest representative of a class of physically indistinguishable connections.

7.2.3 Subtleties

There are some subtleties we have to clarify before we start the canonical analysis of Palatini.

Canonical structure Despite our general discussion of the Palatini formalism above, one should not forget that we want to investigate General Relativity's canonical structure. We only make use of the Palatini fomulation as a means to an end: our approach demands a first–order field theory and Palatini constitutes a manifestly first–order formulation of General Relativity. As such, we of course assume the trace of the torsion tensor to vanish. As discussed in [113], in order to preserve the canonical structure of General Relativity one should add the metric–compatibility condition to the action using Lagrange multipliers. The "naive" Palatini approach, i.e. using on–shell conditions to arrive at the standard metric formulation, really might replace the theory for another, changing the canonical momenta and therefore the dynamics. To avoid this when we consider gravity coupled to matter, we will add the piece

$$S_{LM} = \int \lambda^\mu_{\nu\rho} \nabla_\mu g^{\nu\rho} \sqrt{-g} d^4x,$$

to the Palatini action in order to safeguard the canonical structure of General Relativity. The Lagrange multiplier tensor fields are denoted by $\lambda^\mu_{\nu\rho}$.

The case of $d = 2$ In $d \geq 3$ the covariant constancy of the metric uniquely determines the connection to be the Levi–Civita connection. This is not the case in $d = 2$, since spacetime is not entirely fixed to be a purely metric manifold. In particular, the $\Gamma_\mu \equiv \Gamma^\alpha_{\mu\alpha}$ component of the connection remains undetermined. For $d = 2$, the usual equivalence between the Palatini and the second–order formulation of gravity fails [114]. In particular, this is reflected in their respective gauge transformations: for $d > 2$ one recovers the well–known diffeomorphism invariance [115]

$$\delta g_{\mu\nu} = -\nabla_\mu \xi_\nu - \nabla_\nu \xi_\mu, \qquad \delta\Gamma^\alpha_{\mu\nu} = -\xi^\beta \partial_\beta \Gamma^\alpha_{\mu\nu} + \partial_\beta \xi^\alpha \Gamma^\beta_{\mu\nu} - \partial_\nu \xi^\beta \Gamma^\alpha_{\mu\beta} - \partial_\mu \xi^\beta \Gamma^\alpha_{\nu\beta} - \partial_\mu \partial_\nu \xi^\alpha,$$

which is characterized by two gauge parameters. In the two–dimensional case on the other hand, one does not find a diffeomorphism transformation, since the gauge parameter has three independent components [116]. This does not mean, however, that we cannot investigate its canonical structure in the following.

7.3 Hamiltonian analysis for $d = 2$

General Relativity in its second–order formulation is trivial, since the action only contains a total divergence when the connection is metric compatible [117]. As such, there is no dynamical content and the associated two–dimensional spacetimes are conformally flat. If the Lagrangian can be disregarded as a surface term, there are no derivatives left[40] and the canonical momenta cannot be defined [118]. The situation is less trivial when considering the first–order formulation, since it is not possible to conclude that the connection is compatible with the two–dimensional metric. It is argued in Ref. [119] that, if the general covariance is to be preserved, the triviality of the equations of motion in two dimensions is not a sufficient condition for the Einstein–Hilbert action to be a total divergence. The authors conclude, however, that a first–order formulation different from the affine–metric one has to be used in a canonical approach to the two–dimensional action. This further supports our approach. There are several reasons to investigate gravity in $(1 + 1)$ dimensions. Because of the complexity of General Relativity, it is of pedagogical interest to investigate its properties in lower dimensions. Although the models are unphysical, the mathematical technique and reasoning are similar to the higher dimensional case [120]. In fact, its constraint structure is much closer to the one of higher dimensional first–order gravity than the constraint structure of the second–order $d = 2$ gravity is [121]. It thus allows for deep insights into first–order General Relativity in $d > 2$. But the importance of two–dimensional gravity is not restricted to this fact. As it turns out, it is of academic importance for different fields, including string theory/quantum gravity [122] and black hole physics [123].

7.3.1 Setup

We can rewrite the Palatini action introduced in Eq. (56), following [124, 125], as

$$S_{\tilde{P}} = \int d^d x \underbrace{\; h^{\mu\nu}\big(\partial_\lambda G^\lambda_{\mu\nu} + \frac{1}{d-1} G^\lambda_{\lambda\mu} G^\rho_{\rho\nu} - G^\lambda_{\rho\mu} G^\rho_{\lambda\nu}\big)}_{\mathcal{L}_{\tilde{P}}}, \tag{60}$$

by introducing [126]

$$\begin{aligned} h^{\mu\nu} &= \sqrt{-g} g^{\mu\nu}, \\ G^\lambda_{\mu\nu} &= \Gamma^\lambda_{\mu\nu} - \frac{1}{2}(\delta^\lambda_\nu \Gamma^\rho_{\mu\rho} + \delta^\lambda_\mu \Gamma^\rho_{\nu\rho}). \end{aligned}$$

In the two–dimensional case $\mathcal{L}_{\tilde{P}}$ reduces to a kinematic part and a (non–local) potential,

$$\mathcal{L}_2 = -G\dot{h} - 2G_1\dot{h}^1 - G_{11}\dot{h}^{11} + V(Q^A), \tag{61}$$

where we used the shorthand notation

$$\begin{aligned} h &:= h^{00}, \quad & h^1 &:= h^{01}, \quad & G &:= G^0_{00}, \quad & G_1 &:= G^0_{01}, \\ G_{11} &:= G^0_{11}, \quad & \mathcal{G}^1 &:= G^1_{00}, \quad & \mathcal{G}^1_1 &:= G^1_{01}, \quad & \mathcal{G}^1_{11} &:= G^1_{11}. \end{aligned} \tag{62}$$

The potential,

$$\begin{aligned} V(Q^A) = &- \partial_1 h \mathcal{G}^1 + 2h(G\mathcal{G}^1_1 - G_1\mathcal{G}^1) \\ &- 2\partial_1 h^1 \mathcal{G}^1_1 + 2h^1(G\mathcal{G}^1_{11} - \mathcal{G}^1 G_{11}) \\ &- \partial_1 h^{11}\mathcal{G}^1_{11} + 2h^{11}(G_1\mathcal{G}^1_{11} - G_{11}\mathcal{G}^1_1), \end{aligned}$$

[40] This is still true if we include a cosmological constant

depends only on the generalized coordinates and partial derivatives thereof. The set of generalized coordinates reads

$$Q^A = \{h, h^1, h^{11}, G, G_1, G_{11}, \mathcal{G}^1, \mathcal{G}^1_1, \mathcal{G}^1_{11}\},$$

and the corresponding canonical momenta are

$$P_A = \{P, P_1, P_{11}, \Pi, \Pi^1, \Pi^{11}, \mathbb{P}_1, \mathbb{P}^1_1, \mathbb{P}^1_{11}\}.$$

Canonical Hamiltonian It is straightforward to see that only the coordinates $\{h, h^1, h^{11}\}$ give non–vanishing contributions of the form $\frac{\partial \mathcal{L}_2}{\partial \dot{Q}^A}$, such that the canonical Hamiltonian reduces to

$$H_{can} = \frac{\partial \mathcal{L}_2}{\partial \dot{Q}^A} - \mathcal{L}_2 = -V(Q^A).$$

7.3.2 Primary stage

From our considerations above, we can already deduce the set of primary constraints,

$$\Psi_m = \{P + G, P_1 + 2G_1, P_{11} + G_{11}, \Pi, \Pi^1, \Pi^{11}, \mathbb{P}_1, \mathbb{P}^1_1, \mathbb{P}^{11}_1\}.$$

The demand $\Psi_m \overset{1}{:\approx} 0$ defines the primary constraint submanifold and hence, the primary Hamiltonian reads

$$H_p = H_{can} + \lambda^m \Psi_m.$$

Stability conditions In order to ensure stability of the primary constraints we demand

$$\dot{\Psi}_m = \{\Psi_m, H_p\} \overset{1}{\approx} \{\Psi_m, H_{can}\} + \lambda^r \{\Psi_m, \Psi_r\} \overset{!}{=} 0.$$

We begin by calculating the Poisson brackets of the primaries with themselves and with the canonical Hamiltonian. The only contributions to $\{\Psi_i, \Psi_j\}$ come from

$$\{\Psi_1(x), \Psi_4(y)\} = \delta(x-y), \qquad \{\Psi_2(x), \Psi_5(y)\} = 2\delta(x-y), \qquad \{\Psi_3(x), \Psi_6(y)\} = \delta(x-y),$$

whereas all other combinations vanish. For $\{\Phi_m, H_{can}\}$ we find[41]

$$\begin{aligned}
\{\Psi_1, H_{can}\} &= 2G\mathcal{G}^1_1 - 2G_1\mathcal{G}^1 - \textstyle\int dx[\mathcal{G}^1 f(\partial_1 g)],\\
\{\Psi_2, H_{can}\} &= 2G\mathcal{G}^1_{11} - 2G_{11}\mathcal{G}^1 - \textstyle\int dx 2[\mathcal{G}^1_1 f(\partial_1 g)],\\
\{\Psi_3, H_{can}\} &= 2G_1\mathcal{G}^1_{11} - 2G_{11}\mathcal{G}^1_1 - \textstyle\int dx[\mathcal{G}^1_{11} f(\partial_1 g)],\\
\{\Psi_4, H_{can}\} &= 2h\mathcal{G}^1_1 + 2h^1\mathcal{G}^1_{11},\\
\{\Psi_5, H_{can}\} &= -2h\mathcal{G}^1 + 2h^{11}\mathcal{G}^1_{11},\\
\{\Psi_6, H_{can}\} &= -2h^1\mathcal{G}^1 - 2h^{11}\mathcal{G}^1_1,\\
\{\Psi_7, H_{can}\} &= -\partial_1 h - 2hG_1 - 2h^1 G_{11},\\
\{\Psi_8, H_{can}\} &= -2\partial_1 h^1 + 2hG - 2h^1 1 G_{11},\\
\{\Psi_9, H_{can}\} &= -\partial_1 h^{11} + 2h^1 G + 2h^{11} G_1.
\end{aligned}$$

[41] We encounter non–local Poisson brackets, as introduced in section 4.4.

Now we can put together our stability conditions. We start with the ones that contain a contribution from $\{\Psi_i, \Psi_j\}$, which means that they will determine Lagrange multipliers and can be satisfied inside the primary constraint submanifold.

$$\begin{aligned}
\dot{\Psi}_1 &\overset{!}{=} 0 \quad \rightarrow \lambda^4 = -2G\mathcal{G}_1^1 + 2G_1\mathcal{G}^1 + \int dx[\mathcal{G}^1 f(\partial_1 g)],\\
\dot{\Psi}_2 &\overset{!}{=} 0 \quad \rightarrow \lambda^5 = -G\mathcal{G}_{11}^1 + G_{11}\mathcal{G}^1 + \int dx[\mathcal{G}_1^1 f(\partial_1 g)],\\
\dot{\Psi}_3 &\overset{!}{=} 0 \quad \rightarrow \lambda^6 = -2G_1\mathcal{G}_{11}^1 + 2G_{11}\mathcal{G}_1^1 + \int dx[\mathcal{G}_{11}^1 f(\partial_1 g)],\\
\dot{\Psi}_4 &\overset{!}{=} 0 \quad \rightarrow \lambda^1 = 2h\mathcal{G}_1^1 + 2h^1\mathcal{G}_{11}^1,\\
\dot{\Psi}_5 &\overset{!}{=} 0 \quad \rightarrow \lambda^2 = -h\mathcal{G}^1 + h^{11}\mathcal{G}_{11}^1,\\
\dot{\Psi}_6 &\overset{!}{=} 0 \quad \rightarrow \lambda^3 = -2h^1\mathcal{G}^1 - 2h^{11}\mathcal{G}_1^1.
\end{aligned}$$

The remaining three conditions read

$$\begin{aligned}
\dot{\Psi}_7 &= -\partial_1 h - 2hG_1 - 2h^1 G_{11} \quad &\equiv \chi_1,\\
\dot{\Psi}_8 &= -2\partial_1 h^1 + 2hG - 2h^{11} G_{11} \quad &\equiv \chi_2,\\
\dot{\Psi}_9 &= -\partial_1 h^{11} + 2h^1 G + 2h^{11} G_1 \quad &\equiv \chi_3,
\end{aligned}$$

which are new — secondary — constraints and need to be enforced in the following.

7.3.3 Secondary stage

Setting $\chi_r \overset{2}{\approx} 0$ defines the secondary constraint submanifold. We rewrite the secondary constraints as

$$\begin{aligned}
\chi_1 &= -\partial_1 h - h(\Psi_2 - P_1) - 2h^1(\Psi_3 - P_{11}),\\
\chi_2 &= -2\partial_1 h^1 + 2h(\Psi_1 - P) - 2h^{11}(\Psi_3 - P_{11}),\\
\chi_3 &= -\partial_1 h^{11} + 2h^1(\Psi_1 - P) + h^{11}(\Psi_2 - P_1),
\end{aligned}$$

to keep track of the primary constraints during our calculations. This can be very helpful to make sure that we do not forget any contributions, we just need to be careful not to account for the same contributions twice. The secondary Hamiltonian reads

$$H_s = H_{can} + \lambda^m \Psi_m + \Lambda^r \chi_r,$$

where six of the nine λ^m are already fixed by the stability conditions of the primary stage.

Consistency conditions For consistency we require $\{\Psi_m, H_s\} \overset{2}{\approx} 0$ to hold. We start by calculating the Poisson brackets $\{\Psi_m, \chi_r\}$, which will be needed several times in the following analysis. We find

$$\begin{aligned}
&\{\Psi_1, \chi_1\} = (\Psi_2 - P_1) - \Xi, \quad \{\Psi_1, \chi_2\} = -2(\Psi_1 - P), \quad \{\Psi_1, \chi_3\} = 0,\\
&\{\Psi_2, \chi_1\} = 2(\Psi_3 - P_{11}), \quad \{\Psi_2, \chi_2\} = -2\Xi, \quad \{\Psi_2, \chi_3\} = -2(\Psi_1 - P),\\
&\{\Psi_3, \chi_1\} = 0, \quad \{\Psi_3, \chi_2\} = 2(\Psi_3 - P_{11}), \quad \{\Psi_3, \chi_3\} = -(\Psi_2 - P_1) - \Xi,\\
&\{\Psi_4, \chi_1\} = 0, \quad \{\Psi_4, \chi_2\} = -2h, \quad \{\Psi_4, \chi_3\} = -2h^1,\\
&\{\Psi_5, \chi_1\} = 2h, \quad \{\Psi_5, \chi_2\} = 0, \quad \{\Psi_5, \chi_3\} = -2h^{11},\\
&\{\Psi_6, \chi_1\} = 2h^1, \quad \{\Psi_6, \chi_2\} = 2h^{11}, \quad \{\Psi_6, \chi_3\} = 0,\\
&\{\Psi_7, \chi_r\} = 0, \quad \{\Psi_8, \chi_r\} = 0, \quad \{\Psi_9, \chi_r\} = 0,
\end{aligned}$$

where the non–local contributions are abbreviated as $\Xi \equiv \int dx[(\partial_1 f)g]$. The consistency relations read

$$\begin{aligned}
\{\Psi_1, H_s\} &\stackrel{!}{=} 0 \quad \to \Lambda^1[(\Psi_2 - P_1) - \Xi = 2\Lambda^2(\Psi_1 - P),\\
\{\Psi_2, H_s\} &\stackrel{!}{=} 0 \quad \to \Lambda^1(\Psi_3 - P_{11}) + \Lambda^2\Xi = -\Lambda^3(\Psi_1 - P),\\
\{\Psi_3, H_s\} &\stackrel{!}{=} 0 \quad \to 2\Lambda^2(\Psi_3 - P_{11}) = \Lambda^3[(\Psi_2 - P_1) + \Xi,\\
\{\Psi_4, H_s\} &\stackrel{!}{=} 0 \quad \to \Lambda^2 h = -\Lambda^3 h^1,\\
\{\Psi_5, H_s\} &\stackrel{!}{=} 0 \quad \to \Lambda^1 h = -\Lambda^3 h^{11},\\
\{\Psi_6, H_s\} &\stackrel{!}{=} 0 \quad \to \Lambda^1 h^1 = -\Lambda^2 h^{11}.
\end{aligned}$$

As expected, we found conditions on Lagrange multipliers.

Stability conditions In order to calculate $\dot{\chi}_r = \{\chi_r, H_s\}$ we need several Poisson brackets involving the secondary constraints. We find

$$\begin{aligned}
\{\chi_1, H_{can}\} &= -2h\lambda^5 - 2h^1\lambda^6, \qquad \{\chi_1, \chi_2\} = -2\chi_1,\\
\{\chi_2, H_{can}\} &= 2h\lambda^4 - 2h^{11}\lambda^6, \qquad \{\chi_1, \chi_3\} = -\chi_2,\\
\{\chi_3, H_{can}\} &= 2h^1\lambda^4 + 2h^{11}\lambda^5, \qquad \{\chi_2, \chi_3\} = -2\chi_3.
\end{aligned}$$

Interestingly, observe that after a simple redefinition of the form

$$\sigma_a = \frac{1}{2}(\chi_3 - \chi_1), \qquad \sigma_b = \frac{1}{2}(\chi_1 + \chi_3), \qquad \sigma_c = \frac{1}{2}\chi_2,$$

the algebra of secondaries satisfies

$$\{\sigma_i, \sigma_j\} = \epsilon_{ijk}\sigma_k.$$

By replacing the classical Poisson bracket with the quantum commutator, i.e. $\{\cdot,\cdot\} \to -i \times [\cdot,\cdot]$, this becomes the Lie algebra of $SO(2,1)$ as already noted by the authors of [124]. Combining the above results we can put together the stability conditions as

$$\{\chi_r, H_s\} = \{\chi_r, H_{can}\} + \lambda^m\{\chi_r, \Psi_m\} + \Lambda^s \underbrace{\{\chi_r, \chi_s\}}_{\stackrel{2}{\approx}0},$$

where the last contribution $\Lambda^s\{\chi_r, \chi_s\}$ vanishes weakly on the second constraint submanifold. Thus, we find three new constraints

$$\begin{aligned}
\dot{\chi}_1 &= -2h\lambda^5 - 2h^1\lambda^6 - \lambda^1[(\Psi_2 - P_1) - \Xi] - 2\lambda^2(\Psi_3 - P_{11}) \qquad \equiv \tau_1,\\
\dot{\chi}_2 &= 2h\lambda^4 - 2h^{11}\lambda^6 + 2\lambda^1(\Psi_1 - P) + 2\lambda^2\Xi - 2\lambda^3(\Psi_3 - P_{11}) \quad \equiv \tau_2,\\
\dot{\chi}_3 &= 2h^1\lambda^4 + 4h^{11}\lambda^5 + 2\lambda^2(\Psi_1 - P) + \lambda^3[(\Psi_2 - P_1) - \Xi] \qquad \equiv \tau_3.
\end{aligned} \tag{63}$$

Keep in mind that all Lagrange multipliers present in Eqs. (63) are already fixed. One has to be careful when calculating the stability conditions above: since we included the primaries in our expressions one might be over-counting the contributions to $\dot{\chi}_r$. In fact, we either have to use $\chi_r(P, P_1, P_{11})$ or $\chi_r(G, G_1, G_{11})$ and stay consistent, but not use both for the stability of the secondaries.

7.3.4 Tertiary stage

We define the tertiary constraint submanifold by demanding $\tau_s \overset{3}{\approx} 0$. The tertiary Hamiltonian is defined as

$$H_t = H_{can} + \lambda^m \Psi_m + \Lambda^r \chi_r + \eta^s \tau_s,$$

where η^s are the new Lagrange multipliers appearing in the tertiary stage. Notice, that the tertiary constraints (63) involve the fixed Lagrange multipliers λ^1 to λ^6, which have explicit dependencies on several generalized coordinates and spatial derivatives thereof. This needs to be considered when calculating derivatives of the τ's.

Consistency conditions Before we calculate the stability of the tertiary constraints we check the consistency conditions

$$\{\chi_r, H_t\} \overset{2}{\approx} \eta^s \{\chi_r, \tau_s\},$$

which either vanish identically or further fix some Lagrange multipliers. One finds that, generically, $\{\chi_r, \tau_s\} \not\approx 0$, such that the consistency conditions (partially) fix the new Lagrange multipliers η^s.

Stability Tedious but straightforward algebra unveils that the tertiary constraints are stable inside the tertiary constraint submanifold, i.e. no quaternary constraints arise. This does not happen trivially but requires a lot of rewriting and factoring and there is no value for the reader to show the calculation explicitly. This means, in particular, that the algorithm terminates at the tertiary stage, such that H_t (with fixed η's) is the proper evolution Hamiltonian. As for the other examples in this thesis, the form of the final Hamiltonian is not worked out explicitly, since it is not needed for the degree of freedom count.

7.3.5 Degrees of freedom

The last thing to be done is characterizing all the constraints as first– or second–class in order to count the physical degrees of freedom. From our prior calculations we already know that Ψ_1 to Ψ_6 are second–class constraints. The same is true for the secondary and tertiary constraints. All of them have non–vanishing Poisson brackets with some of the other constraints. Only the set $\{\Psi_7, \Psi_8, \Psi_9\}$ was still first–class in the secondary stage. Indeed, we find that

$$\begin{aligned}
&\{\Psi_7, \tau_1\} \overset{1}{\approx} 0, && \{\Psi_7, \tau_2\} \overset{1}{\approx} 2\chi_1 \overset{2}{\approx} 0, && \{\Psi_7, \tau_3\} \overset{1}{\approx} \chi_2 \overset{2}{\approx} 0,\\
&\{\Psi_8, \tau_1\} \overset{1}{\approx} -2\chi_1 \overset{2}{\approx} 0, && \{\Psi_8, \tau_2\} \overset{1}{\approx} 0, && \{\Psi_8, \tau_3\} \overset{1}{\approx} 2\chi_3 \overset{2}{\approx} 0,\\
&\{\Psi_9, \tau_1\} \overset{1}{\approx} -\chi_2 \overset{2}{\approx} 0, && \{\Psi_9, \tau_2\} \overset{1}{\approx} -2\chi_3 \overset{2}{\approx} 0, && \{\Psi_9, \tau_3\} \overset{1}{\approx} 0,
\end{aligned}$$

i.e. they are still first–class on the final constraint submanifold.

Counting In total we found 15 constraints, 12 of which are second–class on the final constraint submanifold and 3 are first–class. Since we are in $d = 2$, we have $N = 9$ field variables and by means of Eq. (21) we find for the number of physical degrees of freedom

$$n_{dof} = 9 - (3 + \frac{12}{2}) = 0.$$

As was to be expected from $d = 2$ General Relativity (not only Palatini) the theory is non–dynamical, i.e. it is totally constrained and thus does not exhibit any physical degrees of freedom

[127]. It should be noted that our constraint structure is different from the one found in similar treatments (see e.g. [69, 128]). This does not yet imply that the results are contradictory, though. The constraint structure can be altered by starting from different Lagrangians or eliminating fields by virtue of the equations of motion[42] [69]. In the two–dimensional case the number of constraints is even affected by using the metric density $h^{\mu\nu}$ as dynamical variable, rather than the metric tensor $g^{\mu\nu}$ [129]. We will further comment on this situation after the Lagrangian analysis. Despite some differences regarding the constraints, the counting of the physical degrees of freedom yields the same result in all of these cases.

7.4 Lagrangian analysis for $d = 2$

Our setup does not change compared to the Hamiltonian analysis. The main difference is that there are no generalized momenta on the Lagrangian side. Our full set of generalized coordinates reads

$$Q^A = \{h, h^1, h^{11}, G, G_1, G_{11}, \mathcal{G}^1, \mathcal{G}^1_1, \mathcal{G}^1_{11}\}.$$

The Lagrangian density

$$\mathcal{L}_2 = -G\dot{h} - 2G_1\dot{h}^1 - G_{11}\dot{h}^{11} + V(Q^A),$$

is identical to Eq. (61) and the potential is still given by

$$\begin{aligned} V(Q^A) = &- \partial_1 h\mathcal{G}^1 + 2h\big[G\mathcal{G}^1_1 - G_1\mathcal{G}^1\big] \\ &- 2\partial_1 h^1\mathcal{G}^1_1 + 2h^1\big[G\mathcal{G}^1_{11} - \mathcal{G}^1 G_{11}\big] \\ &- \partial_1 h^{11}\mathcal{G}^1_{11} + 2h^{11}\big[G_1\mathcal{G}^1_{11} - G_{11}\mathcal{G}^1_1\big]. \end{aligned}$$

7.4.1 Primary level

Since no products of velocities appear in $\mathcal{L}_2$, the primary Hessian $W_{AB} \coloneqq \partial_{\dot{A}}\partial_{\dot{B}}\mathcal{L}$ vanishes. Accordingly, its Moore–Penrose pseudoinverse M^{AB} is also identically zero. We choose the null vectors of the Hessian $(\gamma_I)^A W_{AB} = 0$ in a way that

$$(\gamma_I)^A = (0, \dots, 0, 1, 0, \dots, 0), \quad I = 1, \dots, 9\ ,$$

where the 1 is at the A^{th} position. The expression for α reduces to

$$\alpha_B = (\partial^1_{\dot{B}}\partial_A\mathcal{L})\partial_1 Q^A + (\partial_{\dot{B}}\partial_A\mathcal{L})\dot{Q}^A - \partial_B\mathcal{L},$$

since no mixed derivative terms appear in $\mathcal{L}$. We find

$$\begin{aligned} \alpha_1 &= -\partial_1\mathcal{G}^1 - \dot{G} - 2\big[G\mathcal{G}^1_1 - G_1\mathcal{G}^1\big], \\ \alpha_2 &= -2\partial_1\mathcal{G}^1_1 - 2\dot{G}_1 - 2\big[G\mathcal{G}^1_{11} - \mathcal{G}^1 G_{11}\big], \\ \alpha_3 &= -\partial_1\mathcal{G}^1_{11} - \dot{G}_{11} - 2\big[G_1\mathcal{G}^1_{11} - G_{11}\mathcal{G}^1_1\big], \\ \alpha_4 &= \dot{h} - 2\big[h\mathcal{G}^1_1 + h^1\mathcal{G}^1_{11}\big], \\ \alpha_5 &= 2\dot{h}^1 + 2\big[h\mathcal{G}^1 - h^{11}\mathcal{G}^1_{11}\big], \\ \alpha_6 &= \dot{h}^{11} + 2\big[h^1\mathcal{G}^1 + h^{11}\mathcal{G}^1_1\big], \\ \alpha_7 &= \partial_1 h + 2\big[hG_1 + h^1 G_{11}\big], \\ \alpha_8 &= 2\partial_1 h^1 + 2\big[h^{11}G_{11} - hG\big], \\ \alpha_9 &= \partial_1 h^{11} - 2\big[h^1 G + h^{11}G_1\big], \end{aligned}$$

[42] This is usually called "weak Hamiltonian" method.

and identify the primary constraints as

$$\varphi_I \equiv (\gamma_I)^A \alpha_A = \alpha_I.$$

In other words, the primary constraints coincide with the α's, because the primary Hessian is zero. Notice the analogy to non–linear electrodynamics.

Functional independence It is important to make sure that our set of primary constraints contains redundant information. We thus check, if the φ_I's are functionally independent among each other. In order to do so we calculate the the Jacobian

$$J_{I\Lambda} \coloneqq \frac{\partial \varphi_I}{\partial X^\Lambda}, \qquad X^\Lambda = \{Q^A, \dot{Q}^A\},$$

and the rank of $J_{I\Lambda}$ equals the number of functionally independent primary constraints. In our case we also allow two rows (or columns) to be dependent via spatial derivatives. By explicit computation we find

$$\mathrm{rank}(J_{I\Lambda}) = 9,$$

which means that our set of primary constraints is functionally independent and does not contain redundant information. We can't and mustn't neglect any of the primaries.

Primary constraint surface By setting

$$\varphi_I \overset{1}{:\approx} 0,$$

we define the primary constraint surface. The dynamics of the system are restricted to this subspace.

7.4.2 Stability

For good health it is not enough to enforce the primaries, we also need to ensure that they do not evolve in time. If their stability is not trivially satisfied on the primary constraint surface we have to set $\dot{\varphi}_I$ to zero explicitly. In order to calculate said time evolution we need

$$\tilde{\alpha}_J = \dot{Q}^A \partial_A \varphi_J + (\partial_1 \dot{Q}^A) \partial^1_A \varphi_J,$$

and the secondary Hessian

$$\tilde{W}_{IJ} = (\gamma_I)^A \partial_{\dot{A}} \varphi_J.$$

Straightforward computation yields[43]

$$\tilde{W} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

[43] Interestingly, this coincides with the $\{\Psi_i, \Psi_j\}$ matrix from the Hamiltonian side.

where it is easy to see that $\text{rank}(\tilde{W}) = 6$ and we choose the null vectors $(\tilde{\gamma}_R)^I \tilde{W}_{IJ} = 0$ as

$$
\begin{aligned}
(\tilde{\gamma}_1)^I &= (0, \dots, 0, 1, 0, 0), \\
(\tilde{\gamma}_2)^I &= (0, \dots, 0, 0, 1, 0), \\
(\tilde{\gamma}_3)^I &= (0, \dots, 0, 0, 0, 1).
\end{aligned}
$$

Obviously $\dim(\tilde{W}) - \text{rank}(\tilde{W}) = 3$, thus the stability of six of the primary constraints is ensured by a set of secondary Euler–Lagrange equations. Only for the three α's that lie inside the kernel of the secondary Hessian we need to demand stability. One finds

$$
\begin{aligned}
\tilde{\alpha}_7 &= \partial_1 \dot{h} + 2\big[\dot{h} G_1 + h \dot{G}_1 + \dot{h}^1 G_{11} + h^1 \dot{G}_{11}\big], \\
\tilde{\alpha}_8 &= 2\partial_1 \dot{h}^1 + 2\big[\dot{h}^{11} G_{11} + h^{11} \dot{G}_{11} - \dot{h} G - h \dot{G}\big], \\
\tilde{\alpha}_9 &= \partial_1 \dot{h}^{11} - 2\big[\dot{h}^1 G + h^1 \dot{G} + \dot{h}^{11} G_1 + h^{11} \dot{G}_1\big],
\end{aligned}
$$

and the secondary constraints *would be* defined as $(\tilde{\gamma}_R)^I \tilde{\alpha}_I \equiv \tilde{\varphi}_R$, such that

$$
\tilde{\varphi}_1 = \tilde{\alpha}_7, \qquad \tilde{\varphi}_2 = \tilde{\alpha}_8, \qquad \tilde{\varphi}_3 = \tilde{\alpha}_9.
$$

However, before we set $\tilde{\varphi}_R \overset{2}{:\approx} 0$ and further restrict ourselves to the secondary constraint surface, we should check if the $\tilde{\varphi}_I$'s do not reduce to identities in the primary constraint surface. We start with $\tilde{\varphi}_1$ and notice that we have a bunch of primary constraints we can insert. In particular, we will use

$$
\begin{aligned}
\alpha_4 \quad &\to \quad \dot{h} \overset{1}{\approx} 2h\mathcal{G}^1_1 + 2h^1 \mathcal{G}^1_{11}, \\
\alpha_5 \quad &\to \quad \dot{h}^1 \overset{1}{\approx} -h\mathcal{G}^1 + h^{11} \mathcal{G}^1_{11}, \\
\alpha_2 \quad &\to \quad \dot{G}_1 \overset{1}{\approx} -\partial_1 \mathcal{G}^1_1 - G\mathcal{G}^1_{11} + \mathcal{G}^1 G_{11}, \\
\alpha_3 \quad &\to \quad \dot{G}_{11} \overset{1}{\approx} -\partial_1 \mathcal{G}^1_{11} - 2G_1 \mathcal{G}^1_{11} + 2G_{11} \mathcal{G}^1_1, \\
\partial_1 \alpha_4 \quad &\to \quad \partial_1 \dot{h} \overset{1}{\approx} 2\mathcal{G}^1_1 \partial_1 h + 2h\partial_1 \mathcal{G}^1_1 + 2\mathcal{G}^1_{11} \partial_1 h^1 + 2h^1 \partial_1 \mathcal{G}^1_{11}.
\end{aligned}
\tag{64}
$$

When inserting the above expressions into $\tilde{\varphi}_1$, many contributions cancel out and we are left with

$$
\tilde{\varphi}_1 \overset{1}{\approx} 2\mathcal{G}^1_1 \underbrace{\big[\partial_1 h + 2hG_1 + 2h^1 G_{11}\big]}_{\alpha_7} + \mathcal{G}^1_{11} \underbrace{2\big[\partial_1 h^1 - hG + h^{11} G_{11}\big]}_{\alpha_8} \overset{1}{\approx} 0.
$$

Hence, $\tilde{\varphi}_1$ is already identically zero inside the first constraint submanifold and it is not a new constraint. The exact same thing happens with $\tilde{\varphi}_2$ and $\tilde{\varphi}_3$, such that we do not find any secondary constraints, but only three Lagrangian identities. Hence, the algorithm terminates here and we found a total of $l = 9$ Lagrangian constraints.

7.4.3 Off–shell identities

As discussed in Sec. 5.2.3, the degree of freedom count on the Lagrangian side is drastically simplified if one already knows the gauge transformations explicitly. In the case of $d = 2$ Palatini they take the form[44] [125]

$$
\delta h^{\mu\nu} = (\epsilon^{\mu\lambda} h^{\sigma\nu} + \epsilon^{\nu\lambda} h^{\sigma\mu}) \xi_{\lambda\sigma}, \qquad \delta G^\rho_{\mu\nu} = -\epsilon^{\rho\lambda} \partial_\lambda \xi_{\mu\nu} - \epsilon^{\lambda\sigma} (G^\rho_{\mu\lambda} \xi_{\sigma\nu} + G^\rho_{\nu\lambda} \xi_{\sigma\mu}),
$$

[44] See e.g. [129] for a component–wise presentation.

where $\epsilon^{\mu\nu}$ and $\xi_{\mu\nu}$ are antisymmetric and symmetric, respectively. Additionally, ϵ satisfies $\epsilon^{01} = -\epsilon^{10} = 1$.

By direct comparison with Eq. (42) we can determine the number of gauge identities g and the number of independent gauge parameters e. We find $g = 3$ and $e = 6$, which is in agreement with [127]. With this information we can readily count the physical degrees of freedom in purely Lagrangian terms.

7.4.4 Degrees of freedom

Since we investigate Palatini in two–dimensional spacetime, the number of field variables is equal to the cardinality of the set Q^A, such that we count $|Q^A| = N = 9$. From our analysis of the off–shell identities above we know that $g = 3$ and $e = 6$. Putting everything together, we can determine the number of propagating modes present in the theory by means of Eq. (22):

$$n_{dof} = N - \frac{1}{2}(l + g + e) = 9 - \frac{1}{2}(9 + 6 + 3) = 0. \tag{65}$$

Our result is in perfect agreement with the counting performed in [124] and [129], where a purely Hamiltonian analysis was done. We thus confirm from a Lagrangian standpoint that this theory is non–dynamical.

Equivalence between Hamiltonian and Lagrangian approaches In our analysis of non–linear electrodynamics we determined g and e from the gauge transformations and complementary from the Hamiltonian method. This direct correspondence between Hamiltonian constraints and gauge parameters and identities was worked out explicitly in [84].

By means of our Hamiltonian analysis we found a total of $N_1 + N_2 = 15$ constraints. More precisely, we obtained $N_1 = 3$ first–class constraints and $N_2 = 12$ second–class constraints. Together with the nine Lagrangian constraints this implies [84] that the number of gauge identities is given by $g = N_1 + N_2 - l = 6$. The total number of gauge parameters e is accordingly given by $N_1 = e = 3$.

This puts our Hamiltonian analysis under some pressure. It seems to contradict our result from the off–shell identities. We elaborate on this in the following.

Possible explanations There is a conceptual difference between our Lagrangian and Hamiltonian constraint algorithms: On the Hamiltonian side *we* do, in general, not use strong zeros in order to evaluate the constraints on a particular submanifold but carry them through the whole algorithm. Our Lagrangian approach, on the other hand, does make use of strong zeros at given submanifolds, as can be seen clearly from Eqs. (64). This is the most likely source for the tension we observe: While technically speaking we do not use a *reductional approach* [130] on the Lagrangian side, we do indeed use strong zeros during the algorithm[45].

It is well–known that eliminating certain fields in the Hamiltonian approach, by setting second–class constraints strongly to zero [86, 130], promotes some of the remaining second–class constraints to being first–class. While this does not change the degree of freedom count, it seems to affect the identification between Lagrangian and Hamiltonian side. Since nearly every canonical analysis of $d = 2$ Palatini was carried out using a reductional method (see e.g. [119, 124, 129]), we can compare these results with our Lagrangian analysis. Using the weak Hamiltonian approach one finds a total of $N_1 = 6$ first–class constraints, $N_2 = 6$ second–class constraints and accordingly $e = 6$ gauge

[45] For a true *Lagrangian reduction* one uses strong zeros in order to rewrite the Lagrangian.

parameters and $g = 3$ gauge identities [127].

With this result the tension is resolved. Notably, this problem did not arise in the case of non–linear electrodynamics. It remains to be investigated if this is a peculiar "feature" of $d = 2$ Palatini or a general characteristic of comparing reductional and non–reductional constraint algorithms.

Part IV
Final considerations

8 Conclusions

We conclude this thesis with a summary of the main results we obtained. We then explain their relevance and usefulness. At last, we discuss the natural follow–up to the thesis. This is work in progress with Verónica Errasti Díez, Julio A. Méndez–Zavaleta and Mojtaba Taslimi Tehrani.

Summary

With the aim of making the thesis self–contained and accessible to every reader, not only the mathematically trained one, we have first established the pertinent mathematical framework.

We have then introduced the Dirac–Bergmann theory of constraints. Additionally, we have put forward a new Lagrangian algorithm. Our Lagrangian approach consistently combines concepts developed throughout several decades and is particularly suited to count physical degrees of freedom in (almost) any first–order field theory. It is algebraically and conceptually much simpler than the usually employed Hamiltonian methods, especially when gravity is involved. This constitutes the main result of the thesis.

At last, we have applied our method to the $\mathcal{H}(\mathcal{P})$–formulation of non–linear electrodynamics and two–dimensional Palatini. Complementary, we have analyzed both theories in terms of the Dirac–Bergmann approach. The comparison between said methods unequivocally shows the strengths of our approach, which we discuss in the following.

Discussion of the results

There are two equivalent formulations of classical mechanics: the Hamiltonian and Lagrangian approaches. Accordingly, one can investigate the constraint structure of a theory in either of those formulations. Which one to use largely depends on the specific purpose of the analysis.

The Hamiltonian approach is better suited to any goal that requires the explicit Poisson algebra of a theory. A remarkable example would be quantization. Indeed, a popular and rigorous quantization technique consists in deforming the Poisson algebra into a Moyal one. The interested reader is referred to [131] for an overview of deformation quantization.

The advantages of our novel Lagrangian method are twofold. First, it is more useful if one wants to determine the physical degrees of freedom in a theory. This can be readily seen in both of our examples. The Lagrangian analysis is conceptually cleaner and lighter. Already in these simple theories, one is confronted with several non–trivialities on the Hamiltonian side. For instance, one encounters non–localities in the constraint algebra and apparent branchings of solutions.

Second, our method yields the explicit form of all constraints. This is in contrast to the ADM analysis[46] [132], where it is shown that requiring linearity in the so–called lapse and shift variables is enough to ensure the stability of the constraints. Notice that this is only the case for General Relativity and thus does not guarantee ghost–freedom in a more general setting, unless the ADM prescription is extended appropriately. In our approach, one can exactly see which term is ghost–full and at which stage the pathology manifests itself. Our method applies in the presented form to (almost) all first–order theories, without modifications.

[46] Which is a Hamiltonian method.

These advantages make our procedure a crucial first step towards establishing a Lagrangian building principle for genuinely ghost–free theories. We discuss how in the following.

Outlook

The most appealing feature of our method is that it admits a reversal of its very logic. Namely, it can be used to determine the necessary and sufficient conditions a Lagrangian must satisfy in order to possess a certain constraint structure. In particular, our ultimate goal is to construct the genuinely ghost–free massive vector Horndeski theory. To this aim, we require:

1. An efficient method to count propagating degrees of freedom.
2. A complete characterization of the constraint structure of generalized Proca fields in flat spacetime.
3. A complete characterization of the constraint structure of Einstein gravity.

Point 1 is met by the main result of this work. Remarkably and as already stated, our approach is so general that it can be applied to most physically sensible theories. Point 2 was established in previous works [20, 21]. Point 3 is subject to work in progress, due to the exponentially bigger algebraic effort compared to the two–dimensional case.

Once all three ingredients are obtained, one can employ the defining features of (constraint) algebras: linearity and additivity. We shall demand that the constraint structure of the massive vector Horndeski theory must be exactly that of generalized Proca in flat spacetime, plus that of Palatini. In other words, the interactions among the fields cannot alter the individual constraint structures of the fields themselves. This requirement is expected to yield non–trivial relations among the various interaction terms.

Importantly, a rigorous construction like this has not been done in previous proposals of couplings between spin–1 fields and gravity. Such existing models are therefore prone to suffer from ghost–like instabilities. This thesis is the first essential step in the construction of a *genuinely ghost–free* massive vector Horndeski theory. Said scenario certainly provides models of substantial interest, not only for the cosmology and black hole communities [133], but to a variety of interdisciplinary fields as well [20].

9 Afterword

The original motivation of this thesis is to investigate the consistent coupling of massive vector fields to gravity. However, the constraint structure analysis methods that we have presented are by no means limited to this scenario. Indeed, they apply to (almost) any first–order field theory. As such, our proposed Lagrangian algorithm can be readily used to count the physical degrees of freedom in a multitude of theories.

Consider bimetric gravity[47] — the consistent theory of two interacting spin–2 fields, one massless and the other massive [135]. The interested reader is referred to [136] for an enlightening overview. The proof of ghost–freedom in bimetric gravity is a celebrated result that follows from a contrived calculation based on the aforementioned ADM analysis. It would be interesting to replicate this proof employing our novel Lagrangian procedure. We expect the calculations to be considerably simplified in our approach.

Another appealing possibility is to look at higher–dimensional descriptions of gravity. For example, consider the celebrated DGP model [137], which exhibits the phenomenologically relevant screening mechanism described in [138]. In this theory, gravity behaves four–dimensional at distances below the so–called cross–over scale r_c. At larger distances $r > r_c$, the effects of the extra dimension become relevant, leading to an effectively massive theory of gravity. The model admits two distinct sectors: the normal and self–accelerating branches. The latter immediately attracted a lot of attention, since it allows for a description of the late–time acceleration of the universe without the need for dark energy, for instance [139, 140]. However, fluctuations around the self–accelerating vacuum were later on found to be plagued by unphysical modes [104]. Our Lagrangian algorithm is by construction an ideal tool to spot possible ghosts in DGP–like models.

As can be seen, the main result of this thesis opens many possibilities worthy of further investigation. Together with the self–contained mathematical exposition in the beginning and the two explicit examples given (those of non–linear electrodynamics and $d = 2$ Palatini), we hope to have made this valuable method accessible to a large community of physicists. May it serve as a catalyst for unambiguously healthy model building.

[47] Not to mention multi–metric theories [134].

References

[1] S. Weinberg. Rev. Mod. Phys. **61**, 1 (1989).

[2] G.W. Horndeski. Int. J. Theor. Phys. **10**, 6 (1974).

[3] A. Nicolis, R. Rattazzi, and E. Trincherini. Phys. Rev. D **79**, 064036 (2009).

[4] L. Heisenberg. JCAP **05**, 015 (2014).

[5] H. Motohashi and T. Suyama. Phys. Rev. D **91**, 085009 (2015).

[6] L. Arturo Ureña–López. J. Phys.: Conf. Ser. **761**, 012076 (2016).

[7] A. Burd and A. Coley. Phys. Lett. B **267**, 330–336 (1991).

[8] R. Kase and S. Tsujikawa. Int. J. Mod. Phys. D **28**, 1942005 (2019).

[9] C.P. Burgess. arXiv:1309.4133 [hep-th] (2013).

[10] G.W. Horndeski. J. Math. Phys. **17**, 1980-–1987 (1976).

[11] D. Momeni et al. Eur. Phys. J. C **77**, No.1, 37 (2017).

[12] J.D. Barrow, M. Thorsrud, and K. Yamamoto. J. High Energ. Phys. **2013**, 146 (2013).

[13] G. Tasinato. Class. Quant. Grav. **31**, 225004 (2014).

[14] R.W. Tucker and C. Wang. Class. Quant. Grav. **15**, 933–954 (1998).

[15] Y.N. Obukhov and E.J. Vlachynsky. Ann. Phys. **8**, 497–510 (1999).

[16] C. Vuille, J. Ipser, and J. Gallagher. Gen. Relativ. Gravit. **34**, 689–696 (2002).

[17] S. Nakamura et al. Phys. Rev. D **99**, 063533 (2019).

[18] A. De Felice et al. JCAP **06**, 048 (2016).

[19] A. Oliveros and M.A. Jaraba. Int. J. Mod. Phys. D **28**, 1950064 (2019).

[20] V. Errasti Díez et al. Phys. Rev. D **101**, 045008 (2020).

[21] V. Errasti Díez et al. Phys. Rev. D **101**, 045009 (2020).

[22] D.W. Cunningham. *Set Theory: A First Course*. 2016.

[23] H.B. Enderton. *Elements of Set Theory*. 1977.

[24] M.H. Mortad. *Introductory Topology*. 2014.

[25] G.E. Bredon. *Topology and Geometry*. 1993.

[26] J.M. Lee. *Introduction to Topological Manifolds*. 2000.

[27] C.J. Isham. *Modern Differential Geometry for Physicists*. 1999.

[28] T. Aubin. *A Course in Differential Geometry*. 2001.

[29] G.F.T. del Castillo. *Differentiable Manifolds — A Theoretical Physics Approach*. 2012.

[30] M. de León and P.R. Rodrigues. *Methods of Differential Geometry in Analytical Mechanics*. North–Holland Mathematics Studies Vol. 158, 1989.

[31] K. Jänich. *Lineare Algebra*. 2008.

[32] S. Roman. *Advanced Linear Algebra*. 2008.

[33] J.M. Lee. *Manifolds and Differential Geometry*. 2009.

[34] G. Rudolph and M. Schmidt. *Differential Geometry and Mathematical Physics — Part I. Manifolds, Lie Groups and Hamiltonian Systems*. 2013.

[35] R. Abraham and J.E. Marsden. *Foundations of Mechanics*. 1987.

[36] P.J. Olver. *Application of Lie Groups to Differential Equations*. Springer Verlag, New York, 1993.

[37] R.W. Sharpe. *Differential Geometry: Cartan's Generalization of Klein's Erlangen Program*. Springer Verlag, New York, 1997.

[38] J.F. Cariñena. Fortschritte der Physik **38**, Nr. 9, 641–679 (1990).

[39] K. Erdmann and M.J. Wildon. *Introduction to Lie Algebras*. Springer Verlag, London, 2006.

[40] J.M. Lee. *Introduction to Smooth Manifolds*. 2003.

[41] P. Libermann and C.–M. Marle. *Symplectic Geometry and Analytical Mechanics*. 1987.

[42] M. Nakahara. *Geometry, Topology and Physics*. 2003.

[43] T. Frankel. *The Geometry of Physics*. 2012.

[44] F. Schuller. *Lectures on the geometric anatomy of theoretical physics*. 2013.

[45] S. Carroll. *Spacetime and Geometry: An Introduction to General Relativity*. Addison Wesley, San Francisco, 2004.

[46] J. Jost. *Riemannian Geometry and Geometric Analysis*. Springer Verlag, 2011.

[47] M.J. Gotay and J.A. Isenberg. Gazette des Mathematiciens **54**, 59–79 (1992).

[48] A.C. da Silva. *Lectures on Symplectic Geometry*. 2008.

[49] M.J. Gotay, J.M. Nester, and G. Hinds. J. Math. Phys. **19**, 2388 (1978).

[50] J.E. Marsden and T.S. Ratiu. *Introduction to Mechanics and Symmetry*. 1999.

[51] K. Sundermeyer. *Symmetries in Fundamental Physics*. 2014.

[52] V.I. Arnold. *Mathematical Methods of Classical Mechanics*. 1989.

[53] A. Deriglazov. *Classical Mechanics — Hamiltonian and Lagrangian Formalism*. 2010.

[54] K. Yano and S. Ishihara. *Tangent and Cotangent Bundles*. 1973.

[55] M. Crampin. J. Phys. A: Math. Gen. **14**, 2567 (1981).

[56] M.C. Muñoz Lecanda and N. Román–Roy. Ann. Inst. Henri Poincaré **57**, 27–45 (1992).

[57] M. Crampin. J. Phys. A: Math. Gen. **16**, 3755 (1983).

[58] J.F. Cariñena, C. López, and N. Román–Roy. J. Math. Phys. **29**, 1143 (1988).

[59] M.J. Gotay and J.M. Nester. Ann. Inst. Henri Poincaré **30**, 129–142 (1979).

[60] C. Batlle et al. J. Phys. A: Math. Gen. **20**, 5113–5123 (1987).

[61] J.F. Cariñena, C. López, and N. Román–Roy. JGP **4**, n.3 (1987).

[62] K. Kamimura. Il Nuevo Cimento B **68**, N.1 (1982).

[63] L.D. Landau and E.M. Lifshitz. *Mechanics*. 1969.

[64] E.C.G. Sudarshan and N. Mukunda. *Classical Dynamics: A Modern Perspective*. World Scientific Publishing, Singapore, 1974.

[65] F. Scheck. *Mechanics — From Newtons Law to Deterministic Chaos*. 2018.

[66] H. Nikolic. arXiv:quant-ph/9812015 (1998).

[67] S.V. Kuzmin and D.G.C. McKeon. Ann. Phys. **318**, 495—496 (2005).

[68] N. Kiriushcheva and S.V. Kuzmin. Eur. Phys. J. C **70**, 389–422 (2010).

[69] D.G.C. McKeon. Int. J. Mod. Phys. A **25**, 3453–3480 (2010).

[70] R. N. Ghalati. arXiv:0901.3344 [gr-qc] (2009).

[71] M. Henneaux and C. Teitelboim. *Quantization of Gauge Systems*. Princeton University Press, Princeton, New Jersey, 1992.

[72] P.A.M. Dirac. Can. J. Math. **2**, 129 (1950).

[73] P.G. Bergmann. Phys. Rev. **75**, 680 (1949).

[74] P.G. Bergmann and J.H.M. Brunings. Rev. Mod. Phys. **21**, 480 (1949).

[75] J.M. Pons. Stud. Hist. Philos. Mod. Phys. **36**, 491–518 (2005).

[76] D.C. Salisbury. Journal of Physics: Conference Series **222** (2010).

[77] D.C. Salisbury and K. Sundermeyer. The European Physical Journal H **42** (2017).

[78] J.L. Anderson and P.G. Bergmann. Phys. Rev. **83**, 1018 (1951).

[79] M.J. Gotay and J.M. Nester. Ann. Inst. Henri Poincaré **32**, 1–13 (1980).

[80] N. Mukunda. Ann. Phys. **9**, 408–433 (1976).

[81] C. Batlle et al. J. Math. Phys. **27**, 2953 (1986).

[82] M. de Leon et al. Int. J. Geom. Meth. Mod. Phys. **2**, 839 (2005).

[83] X. Gràcia, R. Martín, and N. Román–Roy. Int. J. Geom. Meth. Mod. Phys. **6**, 851–872 (2009).

[84] B. Díaz, D. Higuita, and M. Montesinos. J. Math. Phys. **55**, 122901 (2014).

[85] B. Díaz and M. Montesinos. J. Math. Phys. **59**, 052901 (2018).

[86] P.A.M. Dirac. *Lectures on quantum mechanics*. Belfer Graduate School of Science, Yeshiva University, New York, 1964.

[87] G. Date. arXiv:1010.2062 [gr-qc] (2010).

[88] J.B. Pitts. Annals of Physics **351**, pp. 382-406 (2014).

[89] P.A.M. Dirac. Proc. R. Soc. Lond. A **246**, 326-332 (1958).

[90] A. Shirzad. J. Phys. A: Math. Gen. **31**, 2747 (1998).

[91] H. J. Rothe and K. D. Rothe. *Classical and quantum dynamics of constrained Hamiltonian systems*. World Scientific Publishing, Singapore, 2010.

[92] O. Miskovic and J. Zanelli. arXiv:hep-th/0301256 (2003).

[93] O. Miskovic. arXiv:hep-th/0401185 (2004).

[94] T. L. Boullion and P. L. Odell. *Generalized Inverse Matrices*. John Wiley & Sons, New York and Toronto, 1971.

[95] S. Samanta. Int. J. Theor. Phys. **48**, 1436–1448 (2009).

[96] H. J. Rothe R. Banerjee and K. D. Rothe. J. Phys. A **33**, 2059 (2000).

[97] J.M. Martin-Garcia. *xAct: Efficient tensor computer algebra for the Wolfram Language*. URL: `http://www.xact.es/`.

[98] M. Born. Proc. R. Soc. Lond. A **143**, 410 (1934).

[99] N. Bretón. J. Phys.: Conf. Ser. **229**, 012006 (2010).

[100] H. Salazar I., A. García D., and J.F. Plebański. J. Math. Phys. **28**, 2171 (1987).

[101] E. Ayón–Beato and A. García. Phys. Rev. Lett. **80**, 5056–5059 (1998).

[102] E. Ayón–Beato and A. García. Phys. Lett. B **464**, 25 (1999).

[103] A. Alvarez et al. J. High Energ. Phys. **2014**, 041 (2014).

[104] T. Clifton et al. Physics Reports **513**, 1–189 (2012).

[105] R.K. Sachs and H. Wu. *General Relativity for Mathematicians*. Springer Verlag, New York, 1977.

[106] E. Poisson. *A Relativist's Toolkit: The Mathematics of Black-Hole Mechanics*. Cambridge University Press, 2004.

[107] R.M. Wald. *General Relativity*. The University of Chicago Press, Chicago and London, 1984.

[108] M.P. Haugan and C. Lämmerzahl. Lect. Notes Phys. **562**, 195-212 (2001).

[109] C.W. Misner, K.S. Thorne, and J.A. Wheeler. *Gravitation*. W. H. Freeman and Company, 1973.

[110] N. Dadhich and J.M. Pons. Gen. Rel. Grav. **44**, 2337-2352 (2012).

[111] R.N. Ghalati and D.G.C. McKeon. arXiv:0712.2861 [gr-qc] (2008).

[112] A.N. Bernal et al. Phys. Lett. B **768**, 280–287 (2017).

[113] A. Iglesias et al. Phys. Rev. D **76**, 104001 (2007).

[114] S. Deser. Found. Phys. **26**, 617 (1996).

[115] N. Kiriushcheva and S.V. Kuzmin. Central Eur. J. Phys. **9**, 576–615 (2011).

[116] D.G.C. McKeon. Class. Quant. Grav. **23**, 3037–3042 (2006).

[117] J. Gegenberg et al. Phys. Rev. D **37**, 3463-–3471 (1988).

[118] T. Banks and L. Susskind. Int. J. Theor. Phys. **23**, 475–496 (1984).

[119] N. Kiriushcheva and S.V. Kuzmin. Mod. Phys. Lett. A **21**, 899–906 (2006).

[120] P. Collas. Am. J. Phys. **45**, 833–837 (1977).

[121] A.M. Frolov, N. Kiriushcheva, and S.V. Kuzmin. Gen. Rel. Grav. **42**, 1649–1666 (2010).

[122] P. Ginsparg and G.W. Moore. "Lectures on 2D gravity and 2D string theory". 1992. arXiv: `arXiv:hep-th/9304011`.

[123] R.B. Mann. "Lower dimensional black holes: Inside and out". 1995. arXiv: `gr-qc/9501038`.

[124] N. Kiriushcheva, S.V. Kuzmin, and D.G.C. McKeon. Mod. Phys. Lett. A **20**, 1895–1902 (2005).

[125] D.G.C. McKeon. Class. Quant. Grav. **23**, 3037 (2006).

[126] P. Horava. Class. Quant. Grav. **8**, 2069 (1991).

[127] D.G.C. McKeon. arXiv:1607.04132 [hep-th] (2017).

[128] R.N. Ghalati, N. Kiriushcheva, and S.V. Kuzmin. Mod. Phys. Lett. A **22**, 17–28 (2007).

[129] N. Kiriushcheva, S.V. Kuzmin, and D.G.C. McKeon. Mod. Phys. Lett. A **20**, 1961–1972 (2005).

[130] N. Kiriushcheva and S.V. Kuzmin. Ann. Phys. **321**, 958-–986 (2006).

[131] D. Sternheimer. AIP Conf. Proc. **453**: 107–145 (1998).

[132] R. Arnowitt, S. Deser, and C.W. Misner. Gen. Rel. Grav. **40**, 1997 (2008).

[133] L. Heisenberg et al. Phys. Rev. D **96**, 084049 (2017).

[134] N.Khosravi et al. Phys. Rev. D **85**, 024049 (2012).

[135] S.F. Hassan and R.A. Rosen. J. High Energ. Phys. **2012**, 126 (2012).

[136] A. Schmidt-May and M. v. Strauss. J. Phys. A **49**, 183001 (2016).

[137] G. Dvali, G. Gabadadze, and M. Porrati. Phys. Lett. B **485**, 208 (2000).

[138] C. de Rham. Comptes Rendus Physique **13**, 666–681 (2012).

[139] C. Deffayet. Phys. Lett. B **502**, 199–208 (2001).

[140] C. Deffayet, G. Dvali, and G. Gabadadze. Phys. Rev. D **65**, 044023 (2002).

I hereby declare that this thesis is my own work,
and that I have not used any sources and aids other than those stated in the thesis.

Munich, 30.03.2020

Markus Maier