Dynamics of Finite-dimensional Mechanical Systems

Von der Fakultät Konstruktions-, Produktions- und Fahrzeugtechnik
der Universität Stuttgart zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von

Tom Winandy

aus Luxemburg

Hauptberichter:	Prof. Dr. ir. habil. Remco I. Leine
Mitberichter:	Prof. Dr. rer. nat. Uwe Semmelmann

Tag der mündlichen Prüfung: 17.07.2019

Institut für Nichtlineare Mechanik der Universität Stuttgart

2019

To my family.

Preface

This work results from my time at the Institute for Nonlinear Mechanics at the University of Stuttgart. I want to thank Prof. Dr. Remco I. Leine that I could pursue my research interests during the last five years.

I am sincerely grateful to Dr. Simon R. Eugster for guiding[1] my research activities. Your enthusiasm for geometric mechanics provided the initial spark that let me start a journey through the realms of differential geometry. I am very grateful to my travel companion Giuseppe Capobianco. You are not afraid of approaching escarpments in the mathematical landscape. Simon and Giuseppe, the innumerable discussions we had are undoubtedly one of the key elements that made my research endeavours a trip rather than an odyssey.

I want to thank Prof. Dr. Uwe Semmelmann for patiently enduring discussions with me — an autodidact in mathematics. It was with your lecture notes that I started with intrinsic differential geometry. You agreeing to be the co-referee for my thesis means a lot to me.

I am also very grateful to Prof. Dr. Christoph Glocker for his interest, his suggestions, and his enthusiasm for mechanics. Your lectures conveyed the idea of an axiomatic and coordinate-free formulation of mechanics by attributing the central role to the principle of virtual work. This mindset revealed to be a reliable guide in my work. I really appreciated the discussions we had during your visits in Stuttgart.

I thank my colleagues for the pleasant working climate and for sharing good times. I am deeply indebted to Giuseppe Capobianco, Simon R. Eugster, and Matthias Hinze for the painstaking care with which they reviewed my manuscript. Further thanks go to all the students who wrote a thesis under my supervision, in particular to Fabia Bayer, Markus Bergold, and Bassel Katamish.

Finally, I want to thank my family and friends for all their love, support and encouragement. I gratefully acknowledge that my research was funded by the Fonds National de la Recherche, Luxembourg (Proj. Ref. 8864427).

Stuttgart, June 2019

1. Irrespective of his expertise and his efforts, the prevailing doctoral regulations (Promotionsordnung 2011) of the University of Stuttgart do not allow him to serve as a co-referee (Mitberichter) of the present work. Therefore, this footnote is the only official trace of his invaluable contribution.

Contents

Abstract

This monograph deals with the description of mechanical systems having finitely many degrees of freedom using the language of global differential geometry. The mechanical systems may be explicitly time-dependent and involve nonpotential forces. The focus is on the mathematically rigorous formulation of the physical theory dealing with the aforementioned mechanical systems with the objective to introduce the involved physical quantities as well-defined mathematical objects.

The geometric presentation of the physical theory is erected upon a generalized space-time known as Galilean manifold. The state space of a mechanical system is defined as an affine subbundle of the tangent bundle of its associated Galilean manifold. The system's motion is considered to be an integral curve of a second-order vector field on the state space. With the coordinate-free characterization of the motion in terms of second-order vector fields, differential forms appear on stage. A one-to-one correspondence between second-order vector fields and action forms is established. Action forms are differential two-forms with additional properties. The definition of action forms and the derivation of this bijective relation relies on the geometry of double tangent bundles, in which vector bundle homomorphisms and their differential concomitants play an important role.

A coordinate-free definition of forces is given and different geometric interpretations are discussed. With the definition of kinetic energy and of potential forces, the equations of motion are postulated in a coordinate-free way using the action form of the mechanical system. Lagrange's, Hamel's, and Hamilton's equations become local representations of this postulate in terms of a respective chart of the state space. Moreover, the connection between action forms and the concept of virtual work is established. This allows to obtain Lagrange's and Hamel's central equation. This variational perspective is pursued by showing that motions characterized by an exact action form satisfy Hamilton's principle. For this purpose, a coordinate-free definition of the action integral is given.

Finally, constraints are defined as distributions compatible with the time structure of the Galilean manifold on which they are defined. Consequently, the distinction between holonomic and nonholonomic constraints is made using the Frobenius theorem.

Zusammenfassung

Diese Monographie befasst sich mit der Beschreibung von mechanischen Systemen mit endlich vielen Freiheitsgraden mittels globaler Differentialgeometrie. Die mechanischen Systeme dürfen explizit zeitabhängig sein und können Nichtpotentialkräfte beinhalten. Das Hauptaugenmerk liegt auf der mathematischen Durcharbeitung der zugrunde liegenden physikalischen Theorie. Die benötigten physikalischen Größen werden als wohldefinierte mathematische Objekte eingeführt.

Die geometrische Formulierung der physikalischen Theorie baut auf einer verallgemeinerten Raumzeit auf, die als Galilei-Mannigfaltigkeit bekannt ist. Der Zustandsraum eines mechanischen Systems wird als affines Unterbündel des Tangentialbündels der zugehörigen Galilei-Mannigfaltigkeit eingeführt. Die Bewegung eines mechanischen Systems wird als Integralkurve eines Zweitordnungsvektorfeldes auf dem Zustandsraum aufgefasst. Die koordinatenfreie Charakterisierung von Bewegungen durch Zweitordnungsvektorfelder ermöglicht eine Beschreibung mit Hilfe von Differentialformen. Eine eineindeutige Beziehung zwischen Zweitordnungsvektorfeldern und sogenannten Wirkungsformen wird bewiesen. Wirkungsformen sind Zweiformen mit zusätzlichen Eigenschaften. Die Definition von Wirkungsformen und das Aufstellen dieser bijektiven Beziehung basiert auf der Geometrie von Doppeltangentialbündeln, in der Vektorbündelhomomorphismen und deren differentielle Begleiterscheinungen eine zentrale Rolle spielen.

Kräfte werden koordinatenfrei definiert und verschiedene geometrische Interpretationen werden diskutiert. Nach der Einführung der kinetischen Energie und von Potentialkräften werden die Bewegungsgleichungen mit Hilfe der Wirkungsform des mechanischen Systems auf eine koordinatenfreie Weise postuliert. Die Lagrange'schen, die Hamel'schen und die Hamilton'schen Gleichungen sind als lokale Darstellungen dieses Postulates bezüglich einer entsprechenden Karte des Zustandsraumes aufzufassen. Ferner wird die Beziehung zwischen Wirkungsformen und dem Konzept der virtuellen Arbeit untersucht. Dies führt sowohl zur Lagrange'schen wie auch zur Hamel'schen Zentralgleichung. Diese variationelle Sichtweise wird fortgesetzt indem gezeigt wird, dass eine Bewegung, welche durch eine exakte Wirkungsform charakterisiert ist, das Prinzip von Hamilton erfüllt. Hierfür wird das Wirkungsintegral koordinatenfrei definiert.

Schließlich werden Bindungen als mit der Zeitstruktur verträgliche Distributionen auf der Galilei-Mannigfaltigkeit des mechanischen Systems ein-

geführt. Folglich wird das Unterscheiden zwischen holonomen und nichtholonomen Bindungen zu einer Anwendungen des Satzes von Frobenius.

Introduction 1

Die Technik kann ihre theoretischen
Bewegungsprobleme nicht nach
Belieben stellen, wie die rationelle
Mechanik ihre Übungsbeispiele.
— Karl Heun

This monograph deals with the mathematical description of mechanical systems having finitely many degrees of freedom. The considered mechanical systems may be explicitly time-dependent and involve nonpotential forces.[1] The focus lies on the presentation of a theory dealing with the aforementioned mechanical systems in the language of contemporary differential geometry. This approach allows the formulation of a theory for finite-dimensional mechanical systems in which the involved physical quantities are well-defined mathematical objects.

1.1. Motivation

Traditionally, the dynamics of finite-dimensional mechanical systems can be attributed to the area of analytical mechanics which goes back to Lagrange 1788. For many years, analytical mechanics was a field of research for mathematicians. Until the comprehensive treatment of the subject in the book Hamel 1949, the research in mechanics appears to have been rather uniform in the sense that the results have been published in one scientific community.

At some point, engineers started with the aim of making the results from analytical mechanics usable for technical applications. The resulting field is referred to as technical mechanics (or engineering mechanics). The advent of technical mechanics created a new branch of research activities carried out by engineers. Roughly speaking, mathematicians worked on the mathematical foundations of the mechanical theory. Physicists were driven by their interest to understand and describe physical effects that had not been understood before. Engineers worked on finding an efficient way to systematically derive mechanical models for a broad range of technical applications.

1. A nonpotential force is a force that cannot be written as the derivative of a potential.

With differential geometry, mathematicians developed a language that lends itself perfectly to the description of physics. Einstein's general theory of relativity is a prominent example of a theory written in this language. The efforts made by mathematicians to reformulate classical mechanics in the language of differential geometry led to the field of geometric mechanics. While geometric mechanics was driven by the motivation to make a progress in mathematics, technical mechanics had the main objective to develop mechanics in order to solve engineering problems thereby using results from analytical mechanics. In this respect, the study of mechanical systems composed of multiple rigid bodies revealed to be particularly fruitful.

The research activities in geometric and technical mechanics were not only driven by diverging objectives, but also by scientists with a different educational background. The consequence is that nowadays geometric and technical mechanics have drifted widely apart by their underlying mathematical language. Geometric mechanics is based upon contemporary differential geometry, while the mathematics behind technical mechanics dates back to the first half of the twentieth century. Both engineers and mathematicians may encounter severe difficulties in assimilating the results from geometric and technical mechanics, respectively. Engineers usually receive only a shallow formation in mathematics such that it is a huge effort[2] for them to get acquainted with differential geometry. Mathematicians may master their own discipline, but they often ignore the needs arising from practical applications. In the context of finite-dimensional mechanical systems, these are the ability to treat problems involving nonpotential forces and to describe systems that depend explicitly on time. Indeed, in the field of geometric mechanics it is quite common to exclude time-dependence and/or nonpotential forces from the beginning. This monograph tries to bridge the divide between geometric and technical mechanics concerning the description of finite-dimensional mechanical systems.

1.2. An overview of geometric mechanics

In order to give an overview of the different existing approaches in geometric mechanics, we boil the mathematical description of finite-dimensional mechanical systems down to the study of second-order differential equations of the form

$$\mathbf{M}(t,\mathbf{x})\ddot{\mathbf{x}} - \mathbf{h}(t,\mathbf{x},\dot{\mathbf{x}}) = \mathbf{0}. \tag{1.1}$$

2. This hurdle is also identified in Chapter 15 of Glocker 2001.

If the mechanical system has n degrees of freedom, then $\mathbf{x} = (x^1, \ldots, x^n)$ is an $\mathbb{R}^n$-tuple[3] characterizing its position. The components $x^i = x^i(t)$ are known as generalized coordinates. They are real-valued functions that depend on time t. The time-derivative $\dot{\mathbf{x}}$ of the position coordinates $\mathbf{x}$ describes the velocity of the mechanical system. Its acceleration is characterized by the $\mathbb{R}^n$-tuple $\ddot{\mathbf{x}}$ and $\mathbf{M}(t, \mathbf{x})$ denotes the symmetric, positive definite mass matrix, which may depend on the time t and the position coordinates $\mathbf{x}$. The equation of motion (1.1) can be rewritten in first-order form

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{u}, \\ \mathbf{M}(t, \mathbf{x})\dot{\mathbf{u}} &= \mathbf{h}(t, \mathbf{x}, \mathbf{u}). \end{aligned} \tag{1.2}$$

Inspired by Newton's second law, we call the function $\mathbf{h}(t, \mathbf{x}, \mathbf{u})$ a force until we give a precise definition in Section 4.7. The forces may be split into three terms as

$$\mathbf{h}(t, \mathbf{x}, \mathbf{u}) = \mathbf{g}(t, \mathbf{x}, \mathbf{u}) + \mathbf{f}^{\mathrm{p}}(t, \mathbf{x}, \mathbf{u}) + \mathbf{f}^{\mathrm{np}}(t, \mathbf{x}, \mathbf{u}),$$

where $\mathbf{g}$ gathers the gyroscopic forces, $\mathbf{f}^{\mathrm{p}}$ contains the potential forces and $\mathbf{f}^{\mathrm{np}}$ stands for the remaining part, which is referred to as $\mathbb{R}^n$-tuple of non-potential forces. At this stage the reader should not stumble across the velocity-dependence of $\mathbf{f}^{\mathrm{p}}$. We will see in Section 4.9, that a certain type of velocity-dependent forces may be included in the description using a vector potential.

In geometric mechanics, authors often limit their study to mechanical systems that do *not* depend explicitly on time and/or to systems that comprise only a certain type of potential forces. The mathematical consequence is that the resulting theory is formulated on different mathematical spaces. In the following, we try to give an overview on different existing approaches by squeezing the respective equations of motion into matrix notation. We start with time-independent systems.

Time-independent mechanical systems

For mechanical systems that do not depend on time explicitly, the equations of motion (1.2) reduce to

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{u}, \\ \mathbf{M}(\mathbf{x})\dot{\mathbf{u}} &= \mathbf{h}(\mathbf{x}, \mathbf{u}), \end{aligned} \tag{1.3}$$

with

$$\mathbf{h}(\mathbf{x}, \mathbf{u}) = \mathbf{g}(\mathbf{x}, \mathbf{u}) + \mathbf{f}^{\mathrm{p}}(\mathbf{x}) + \mathbf{f}^{\mathrm{np}}(\mathbf{x}, \mathbf{u}).$$

3. The $\mathbb{R}^n$-tuples in (1.1) have to be read as column vectors according to Appendix A.

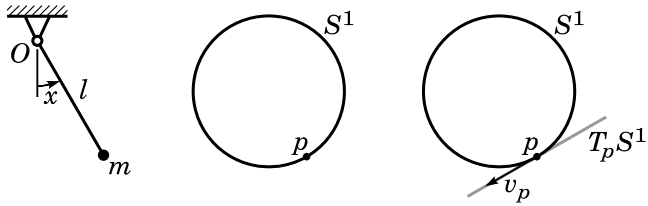


Figure 1.1.: Example of a pendulum consisting of a point mass m which moves at distance l around the point O. The unit circle S^1 can be considered as its configuration manifold. The velocities are represented by tangent vectors to S^1. The angle x between the massless rod of the pendulum and the vertical is a generalized coordinate, which locally describes the circle.

It is well-known[4] that the kinetic energy of such a time-independent system has the form

$$T = \tfrac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{M}(\mathbf{x})\mathbf{u}, \tag{1.4}$$

with a symmetric, positive definite mass matrix $\mathbf{M} = \mathbf{M}(\mathbf{x})$ that may depend on the position $\mathbf{x}$. It is common to denote the generalized coordinates by $\mathbf{q}$ instead of $\mathbf{x}$. They can be interpreted as being the local description of an abstract space of positions. This space is assumed to be a differentiable manifold Q and it is referred to as the configuration manifold of the mechanical system because each point in Q corresponds to a different configuration of the mechanical system. A standard example is given by the pendulum from Figure 1.1, which consists of a point mass m that moves in the plane keeping a constant distance l to the point O. We assume that this is realized by hinging the mass to the point O by a massless rod. The unit circle

$$S^1 := \left\{(a,b) \in \mathbb{R}^2 \,\middle|\, a^2 + b^2 = 1\right\}$$

can be considered as configuration manifold of the pendulum, i.e., $Q = S^1$. Indeed, each point of a circle can be uniquely related to one possible configuration of a pendulum. The velocity of the point mass can be seen as a tangent vector to the configuration manifold as suggested by the right-hand side of Figure 1.1. We will see in Section 3.2, that at each point p of a differentiable manifold Q a vector space T_pQ can be defined, called the tangent space. It represents the abstract space of velocities at a given configuration. We will see with Definition 3.19 that the idea of considering the configuration manifold Q together with all these vector spaces can be made

4. See p. 266 in Hamel 1949.

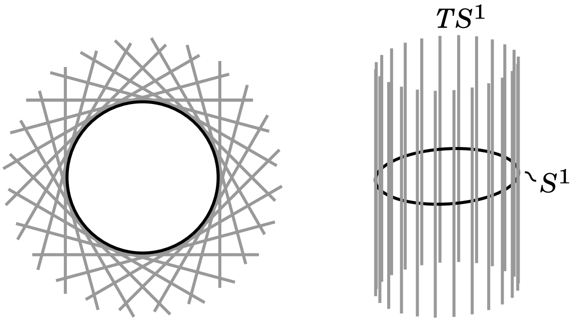


Figure 1.2.: The state space of a time-independent mechanical system is given by the tangent bundle TQ of its configuration manifold Q. For the pendulum from Figure 1.1, this space is denoted by TS^1 and it consists of the circle S^1 as space of positions together with all lines tangent to that circle as one-dimensional vector spaces containing the velocities.

mathematically precise by defining the tangent bundle TQ of the manifold Q. The tangent bundle is said to be the state space of the mechanical system. Figure 1.2 visualizes the tangent bundle of the configuration manifold S^1 of a pendulum.

In general, the mass matrix $\mathbf{M}$ of a time-independent mechanical system endows the configuration manifold Q with a Riemannian metric[5] and the kinetic energy (1.4) is a real-valued function $T\colon TQ \to \mathbb{R}$ on the tangent bundle of the configuration manifold. The definition of a Lagrangian $L\colon TQ \to \mathbb{R}$ of the local form

$$L(\mathbf{x},\mathbf{u}) = T(\mathbf{x},\mathbf{u}) - V(\mathbf{x}) = \tfrac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{M}(\mathbf{x})\mathbf{u} - V(\mathbf{x}) \tag{1.5}$$

and the matrix notations from Appendix A allow us to write the equations of motion (1.3) as

$$\begin{bmatrix} \frac{\partial^2 L}{\partial\mathbf{x}\partial\mathbf{u}} - \frac{\partial^2 L}{\partial\mathbf{x}\partial\mathbf{u}}^{\mathrm{T}} & \frac{\partial^2 L}{\partial\mathbf{u}\partial\mathbf{u}} \\ -\frac{\partial^2 L}{\partial\mathbf{u}\partial\mathbf{u}}^{\mathrm{T}} & \mathbf{0}_{n\times n} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \frac{\partial L}{\partial\mathbf{x}}^{\mathrm{T}} - \frac{\partial^2 L}{\partial\mathbf{x}\partial\mathbf{u}}^{\mathrm{T}}\mathbf{u} \\ -\frac{\partial^2 L}{\partial\mathbf{u}\partial\mathbf{u}}^{\mathrm{T}}\mathbf{u} \end{bmatrix} + \begin{bmatrix} \mathbf{f}^{\mathrm{np}}(\mathbf{x},\mathbf{u}) \\ \mathbf{0} \end{bmatrix}. \tag{1.6}$$

The real-valued function $V(\mathbf{x})$ in the Lagrangian (1.5) is a scalar potential of the potential forces in (1.3) such that

$$\mathbf{f}^{\mathrm{p}}(\mathbf{x}) = -\frac{\partial V}{\partial\mathbf{x}}^{\mathrm{T}}. \tag{1.7}$$

5. See Section 3.10 for the definition.

Because of the local form (1.5), the matrix

$$\frac{\partial^2 L}{\partial \mathbf{u} \partial \mathbf{u}} = \mathbf{M}(\mathbf{x})$$

is regular and, therefore, the matrix on the left-hand side of (1.6) is the local representation of a symplectic form.[6]

The dynamics of a time-independent mechanical system can equally well be studied on the cotangent bundle[7] T^*Q of the configuration manifold Q. This leads to a description of the mechanical system in terms of position and momentum coordinates. The cotangent bundle is said to be the system's phase space. With a Hamiltonian $H\colon T^*Q \to \mathbb{R}$ of the local form

$$H(\mathbf{x}, \mathbf{p}) = \tfrac{1}{2}\mathbf{p}^{\mathrm{T}}\mathbf{M}^{-1}(\mathbf{x})\mathbf{p} + V(\mathbf{x}) \tag{1.8}$$

the equations of motion (1.3) can be expressed as

$$\begin{bmatrix} \mathbf{0}_{n\times n} & \mathbf{I}_{n\times n} \\ -\mathbf{I}_{n\times n} & \mathbf{0}_{n\times n} \end{bmatrix} \begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} -\frac{\partial H}{\partial \mathbf{x}}^{\mathrm{T}} \\ -\frac{\partial H}{\partial \mathbf{p}}^{\mathrm{T}} \end{bmatrix} + \begin{bmatrix} \mathbf{f}^{\mathrm{np}}(\mathbf{x}, \mathbf{p}) \\ \mathbf{0} \end{bmatrix}. \tag{1.9}$$

Again, we observe that the matrix on the left-hand side of (1.9) respresents a symplectic form. This is not astonishing because we will see that the tangent and cotangent bundles of a Riemannian manifold Q are isomorphic[8] bundles. As before, the real-valued function $V(\mathbf{x})$ in the Hamiltonian (1.8) is a scalar potential such that the potential forces are given by (1.7). The restriction to time-independent mechanical systems allows a geometric treatment of the subject within the field of symplectic geometry.

For completeness, we note that there exist alternative approaches to symplectic geometry that can be used for the geometric description of time-independent mechanical systems. The Levi-Civita connection associated with the Riemannian metric defined by the mass matrix $\mathbf{M}(\mathbf{x})$ can be used to define the equations of motion in a coordinate-free way. For this approach we refer the reader to Bullo et al. 2004 and references therein. An alternative view on the Lagrangian picture results from choosing Hamilton's principle[9] as starting point for the description of time-independent mechanical systems involving only potential forces. See for example Section 19 in Arnold 1989 or Theorem 3.8.3 in Abraham and Marsden 1987.

6. Symplectic forms on a differentiable manifold are introduced in Section 3.10.
7. The cotangent bundle of a differentiable manifold is defined by Definition 3.20.
8. See Sections 3.4 and 3.10 as well as equation (3.71).
9. Hamilton's principle is also known as principle of stationary action.

A comparison of equations (1.6) and (1.9) shows that the Hamiltonian formulation appears easier. Indeed, the symplectic form that is represented by the constant matrix on the left-hand side of equation (1.9) is the canonical[10] two-form on the cotangent bundle T^*Q, while the symplectic form on the tangent bundle TQ depends on the choice of a Lagrangian. Therefore, the Hamiltonian side is often preferred to the Lagrangian picture because of its brevity and "mathematical elegance".[11]

In the local expressions (1.6) and (1.9), the nonpotential forces stand on the right-hand side. These forces can be defined in a coordinate-free way[12] as so-called semi-basic differential forms on the respective bundle. However, there are many authors that refrain from dealing with nonpotential forces. The Hamiltonian and Lagrangian picture without nonpotential forces can be found in Chapters 5 and 7 of Marsden et al. 1999, respectively. Alternatively, the reader is referred to Chapter 3 of Abraham and Marsden 1987.

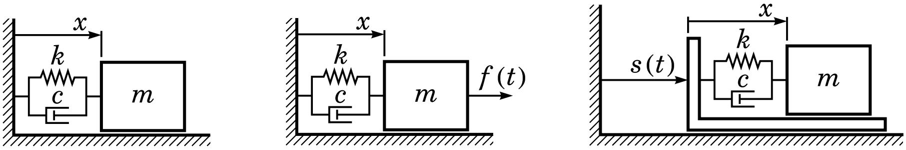
In the recent book Bloch 2015, the author claims in the preface that: *Mechanics has traditionally described the behaviour of free and interacting particles and bodies, the interaction being described by potential forces.* This statement is not only highly questionable, but it implies severe restrictions for the mechanical theory. As an example, we consider the damped harmonic oscillator from Figure 1.3a. It consists of a block with mass m that is attached to a vertical wall by a spring with stiffness k and by a damper with damping coefficient c. The linear displacement of the block with respect to the wall is described by the coordinate x. The spring is undeformed for $x = 0$. Its motion, which is governed by the second-order differential equation

$$m\ddot{x} + c\dot{x} + kx = 0, \tag{1.10}$$

cannot be described within a time-independent geometric theory that excludes nonpotential forces. The examples from Figures 1.3b and 1.3c go beyond the scope of the theory even with $c = 0$ because of their explicit time-dependence.

In classical books such as Hamel 1949 or in the famous work by Landau and E. M. Lifshitz 1969, finite-dimensional mechanical systems may depend on time explicitly. The author thinks that geometric descriptions of finite-dimensional mechanical systems that are limited to the time-independent case have to be seen critically because a modern presentation of a subject should not have a reduced scope compared to more classical treatments.

10. See p. 84 for the definition of the canonical two-form on the cotangent bundle T^*Q.
11. See the introduction of Crampin 1983.
12. We refer to Chapters X and XI in Godbillon 1969 for a treatment of the subject on the tangent bundle.



(a) No excitation (b) Kinetic excitation $f(t)$ (c) Kinematic excitation $s(t)$

Figure 1.3.: Three examples of a damped harmonic oscillator (mass m, spring stiffness k, and damping coefficient c).

Time-dependent mechanical systems

It is well-known that the questions of explicit time-dependence and the appearance of nonpotential forces are related to the definition of the system boundaries.[13] Indeed, if all the bodies of the universe would be included to the model, there would be no need to study (time-dependent) external forces at all. The description of such a time-independent isolated system would not require nonpotential forces in its description.

Souriau gives the following example. Studying the motion of a projectile, air resistance can be taken into account as a velocity-dependent force. But Souriau argues[14] that the consideration of air resistance as a velocity-dependent force acting on the projectile has to be seen as an empirical approximation which is *meant to replace a detailed study of the mechanics of the atmosphere itself.* This observation may be correct, but the exclusion of these “empirical” types of forces, which are common in technical mechanics, artificially reduces the scope of the theory. To underpin this viewpoint, let us consider an example from automotive industry: the dimensioning of a wheel suspension which cushions the shocks caused by potholes. This task can be treated within classical mechanics if the hydraulic shock absorbers that are typically used in such a construction can be approximated as being linear dampers as in the example from Figure 1.3. A complete physical modelling as suggested by Souriau would lead to a model involving fluid mechanics. However, provided that velocity-dependent forces may be dealt with, this engineering problem can be attacked by studying the solutions of a linear ordinary differential equation such as (1.10).

A straightforward approach to incorporate explicit time-dependence is to consider the extended state space given by the Cartesian product $\mathbb{R} \times TQ$ with the Lagrangian $L(t, \mathbf{x}, \mathbf{u})$ depending explicitly on time or to extend the

13. See p. 9 in Arnold 1989.
14. See page 139 in Souriau 1997.

phase space as $\mathbb{R} \times T^*Q$ with the Hamiltonian $H(t, \mathbf{x}, \mathbf{p})$. With $\dot{t} = 1$, the equations of motion (1.2) can be written as

$$\begin{bmatrix} 0 & \frac{\partial L}{\partial \mathbf{x}} - \frac{\partial^2 L}{\partial t \partial \mathbf{u}}^{\mathrm{T}} - \mathbf{u}^{\mathrm{T}} \frac{\partial^2 L}{\partial \mathbf{x} \partial \mathbf{u}} & -\mathbf{u}^{\mathrm{T}} \frac{\partial^2 L}{\partial \mathbf{u} \partial \mathbf{u}} \\ -\frac{\partial L}{\partial \mathbf{x}}^{\mathrm{T}} + \frac{\partial^2 L}{\partial t \partial \mathbf{u}} + \frac{\partial^2 L}{\partial \mathbf{x} \partial \mathbf{u}}^{\mathrm{T}} \mathbf{u} & \frac{\partial^2 L}{\partial \mathbf{x} \partial \mathbf{u}} - \frac{\partial^2 L}{\partial \mathbf{x} \partial \mathbf{u}}^{\mathrm{T}} & \frac{\partial^2 L}{\partial \mathbf{u} \partial \mathbf{u}} \\ \frac{\partial^2 L}{\partial \mathbf{u} \partial \mathbf{u}}^{\mathrm{T}} \mathbf{u} & -\frac{\partial^2 L}{\partial \mathbf{u} \partial \mathbf{u}}^{\mathrm{T}} & \mathbf{0}_{n \times n} \end{bmatrix} \begin{bmatrix} 1 \\ \dot{\mathbf{x}} \\ \dot{\mathbf{u}} \end{bmatrix} = \mathbf{0}$$

and

$$\begin{bmatrix} 0 & -\frac{\partial H}{\partial \mathbf{x}} & -\frac{\partial H}{\partial \mathbf{p}} \\ \frac{\partial H}{\partial \mathbf{x}}^{\mathrm{T}} & \mathbf{0}_{n \times n} & \mathbf{I}_{n \times n} \\ \frac{\partial H}{\partial \mathbf{p}}^{\mathrm{T}} & -\mathbf{I}_{n \times n} & \mathbf{0}_{n \times n} \end{bmatrix} \begin{bmatrix} 1 \\ \dot{\mathbf{x}} \\ \dot{\mathbf{p}} \end{bmatrix} = \mathbf{0}, \tag{1.11}$$

respectively. Equation (1.11) corresponds to the local expression of Theorem 5.1.13 in Abraham and Marsden 1987. Alternatively, it can be also found on pp. 236–237 in Arnold 1989. Note that these descriptions do not include nonpotential forces. Moreover, the physical interpretation of the spaces $\mathbb{R} \times TQ$ and $\mathbb{R} \times T^*Q$ is problematic because their structure as Cartesian product assumes the existence of an absolute space (independent of time). The same holds for a formulation of the dynamics on $T(\mathbb{R} \times Q)$ or $T^*(\mathbb{R} \times Q)$. The core assumption behind (generalized) space-time in classical mechanics (in contrast to the relativistic case) is that we are able to distinguish whether two events happen at the same time or not. As we will see, in time-dependent mechanics a generalized space-time manifold M can be used as underlying space instead of a time-independent configuration manifold Q.

The reader should be aware that there is no general agreement about the terminology: *state space*, *phase space*, *extended state space*, and *extended phase space*. While we refer to the space $\mathbb{R} \times T^*Q$ as extended phase space, the equivalent space $T^*Q \times \mathbb{R}$ is called state space in Section 6.6 of Bishop et al. 1980. If we page forward in the mentioned reference, a similar remark concerning the existing geometric approaches to the dynamics of finite-dimensional mechanical systems can be made. Indeed, Proposition 6.7.1 formulates the mechanics of *time-independent* mechanical systems on the space $T^*Q \times \mathbb{R}$. With $T^*Q \times \mathbb{R}$, Bishop et al. 1980 work on a space that includes time. However, the authors present a theory limited to time-independent mechanics because in their work neither the kinetic energy, nor the forces may depend on time such that the resulting theory does not include the description of time-dependent mechanical systems.

As suggested in Loos 1982, we formulate a mechanical theory for finite-dimensional mechanical systems on a so-called Galilean manifold that was introduced by Dombrowski et al. 1964a. This will allow us to deal with potential and nonpotential forces. The resulting theory allows the description of time-dependent mechanical systems whose motion is governed by an equation of the form (1.1). In particular, this means that the three examples from Figure 1.3 can be described. The motion of the kinetically and the kinematically excited oscillators from Figures 1.3b and 1.3c is respectively governed by

$$m\ddot{x} + c\dot{x} + kx = f(t).$$

and

$$m\ddot{x} + c\dot{x} + kx = -m\ddot{s}(t).$$

Some authors present a geometric blend of the time-independent with the time-dependent case without giving much comments. In Talman 2007, the kinetic energy is defined for time-independent systems on p. 106. In the course of page 176, the time-dependence then silently creeps into the kinetic energy and the Lagrangian. A similar approach can be found in the fifth chapter of Scheck 2007, where the Lagrangian as well as the energy are introduced in Sections 5.5.1 and 5.6.3, respectively, as real-valued functions on the tangent bundle TQ of a time-independent configuration manifold Q. However, the chart representations of these functions erroneously present an explicit time-dependence (see p. 299 and p. 323 in Scheck 2007).

1.3. Aim and scope

The aim of this research monograph is threefold. First, it provides a geometric description of finite-dimensional mechanical systems that does not involve restrictive assumptions such as the limitation to time-independent systems or the exclusion of nonpotential forces. The presented coordinate-free treatment of finite-dimensional mechanical systems, which does not involve restrictions that are considered too restrictive in the field of technical mechanics, should lead to a rapprochement of geometric and technical mechanics as described in Section 1.1. Chapters 2 and 3, which introduce the required mathematical objects, are written in the spirit of this objective. They should allow the reader to get acquainted with the mathematical foundations of Chapter 4, which constitutes the core of this work.

Second, this work aims to introduce the involved physical quantities as mathematically well-defined objects, starting with the notion of space and time. The resulting "heavy" mathematical machinery pays off because we are able to give a precise definition of forces in Section 4.7, which was identified by Hamel as the *chief difficulty*[15] in mechanics.

Third, this work has the intention to unify some classical results about finite-dimensional mechanical systems by deriving them from one common starting point (Postulate 4.8). These are in the order of appearance: Lagrange's equations (4.111), Hamel's equations (4.150), Hamilton's equations (4.162), the virtual work (4.183), Lagrange's central equation (4.190), Hamel's central equation (4.192), and Hamilton's principle (Section 4.11.3). In the proposed mechanical theory Lagrange's, Hamel's, and Hamilton's equations are just different coordinate representations of the differential equations that determine the motion of a given mechanical system. In the usual presentations of time-independent mechanics, the Lagrangian and the Hamiltonian side often appear as "separate" worlds.

Finally, as a byproduct, the geometric treatment of the subject allows us to define constraints in a coordinate-free manner in Section 4.12 such that the distinction between holonomic and nonholonomic constraints becomes an application of the Frobenius theorem. This topic often led to confusion[16] in the literature. Neĭmark et al. 1972 give an overview on the different classical approaches to describe finite-dimensional mechanical system involving nonholonomic constraints.

1.4. Literature and its state

In Section 1.2, we tried to give the reader an overview of existing approaches in geometric mechanics. We have already observed that different authors differ by fundamental assumptions concerning the types of forces/systems they study. Moreover, we observed strong differences in terminology. This makes it a hopeless task to provide a meaningful overview on the available literature on geometric mechanics by using the organization from Section 1.2. Instead, we will present the literature that deals with the description of time-dependent mechanical systems and comment about the underlying results from differential geometry that cannot be considered elementary.

The main reference of this work, Loos 1982, is a typescript related to a seminar that was held in the winter semester 1981/1982 by Ottmar Loos and Josef Rothleitner at the university of Innsbruck in Austria. Since the script was never officially published, it can only be found at an antiquarian bookseller. The declared objective of the script, which is written in German,

15. In 1952, Hamel wrote in a letter to Truesdell that *in the concept of force lies the chief difficulty in the whole of mechanics*. See pp. 523–524 in Truesdell 1984.

16. Wrong definitions of nonholonomic constraints can be found on p. 19 in Päsler 1968 and p. 96 in Roberson et al. 1988. Section 2.2 of Papastavridis 2014 confronts the reader with a terminological flood that does not ease comprehension.

is to make the results known in the French school of mechanics available to the German-speaking scientific community. However, the typescript written by Loos is more than a mere translation. Besides other results, part of the script is contained in the paper Loos 1985 that is written in English. Until the publication of this work in 2019, the latter paper has been ignored.[17] Because of the reduced availability of Loos 1982, the author was forced to restate many of the results it contains.

In the retrospective, it is more than amazing in which way the French school of mechanics systematically opened up the mathematical landscape behind finite-dimensional mechanical systems. The general approach can be traced back to Élie Cartan's lectures on integral invariants (Cartan 1922). The work of Gallissot 1952, suggests to characterize the motion of finite-dimensional mechanical systems using differential two-forms. Including the study of bilaterally and unilaterally constrained mechanical systems, Gallissot demonstrates that the use of differential two-forms leads to a far-reaching approach in the description of finite-dimensional mechanical systems. By his "Maxwell's" principle, Souriau 1970 focusses on the study of mechanical systems that are only subjected to potential forces. Souriau's book clearly continues on the way pursued by Élie Cartan and François Gallissot. Indeed, the link can be formally made because the work of Gallissot is one of the few references given by Souriau. Godbillon 1969 develops the description of time-independent mechanical systems including nonpotential forces by studying the geometry of the double tangent bundle[18] of the configuration manifold. Much of the mathematical structures exposed by Godbillon reappear in the description of time-dependent systems. The works of Lichnerowicz 1945 and of his student Klein 1962 deal with the description of mechanical systems involving nonpotential forces within the calculus of variations.

In the typescript Loos 1982, Loos brought together the generalized space-time developed by Dombrowski et al. 1964a for finite-dimensional mechanical systems and the French results we have just discussed. Loos does not only work out a theory for the description of time-dependent mechanical systems including nonpotential forces that provides the equations of motion, but he is able to give a precise definition of forces. To the best of the author's knowledge, the work of Loos has almost fallen into oblivion. This may be due to several reasons. The English translation Souriau 1997 of the book Souriau 1970 being the exception, many of the cited publications are only available in French or German. The book by Souriau is not only unorthodox by its notation but also by its economy of references. Even if

17. According to Google Scholar it has been cited only twice until May 2019.
18. The double tangent bundle of a manifold is the tangent bundle of its tangent bundle.

the notion of force introduced by Loos is officially published in Loos 1985, it appears as a side result that is not recognizable in the title: *Automorphism Groups of Classical Mechanical Systems*. To the author's knowledge, the discussed results have only partly made their way into the English literature centered around the standard textbook Abraham and Marsden 1987 and the English references therein. Most of the mathematical results are available in English in Libermann et al. 1987 and Morandi et al. 1990. More recent publications on the subject such as Marsden et al. 1999, Oliva 2002, Bullo et al. 2004, Scheck 2007, Bloch 2015 or Cortés et al. 2017 ignore the contributions of Loos. It is not astonishing that one can find recent publications such as Bravetti et al. 2017, which claim to extend the theory while ignoring the existing results.

Because one objective of this work is to bridge the gap between technical and geometric mechanics, the mathematical foundations are presented in Chapters 2 and 3 before finite-dimensional mechanical systems are studied in Chapter 4. Chapter 4 includes many references back to the previous chapters such that readers may dare a direct jump to the mechanical part of the story. For the convenience of the reader, we give a brief overview on the mathematical literature underlying this work. For the algebra behind Chapter 2, we refer to Lang 2005, Artin 2011 and Hornfeck 1969. For linear algebra we refer to Fischer 2010, Hoffman et al. 1971, Lang 2004, and Roman 2008. Multilinear algebra can be found in Bishop et al. 1980, Jeffrey M. Lee 2009, John M. Lee 2013, and Spivak 1999a. For the introduction to affine spaces we refer the reader to Crampin and Pirani 1987. The results from topology are taken from Munkres 2000. For a general introduction to differential geometry, the reader is referred to John M. Lee 2013, Jeffrey M. Lee 2009, and Lang 2001. A compact presentation of the subject is found in Aubin 2001. The five volumes Spivak 1999a, Spivak 1999b, Spivak 1999c, Spivak 1999d, and Spivak 1999e provide a comprehensive treatment of differential geometry including comments about the historical development of differential geometry and the different notations used by physicists and mathematicians. At some points we will need the more specialised books: Abraham, Marsden, and Ratiu 1988, Gallot et al. 1990, Golubitsky et al. 1973, Hermann 1988, and Yano et al. 1973. Results are referenced at their first appearance.

All kinds of algebra 2

By relieving the brain from all unnecessary work, a good notation sets it free to concentrate on more advanced problems.
— Alfred North Whitehead

The present chapter deals with some algebraic concepts that are essential in mechanics. Readers which are familiar with groups, vector spaces, affine spaces, and tensors may skip this chapter. The algebraic part of the presentation is based on Lang 2005 and Artin 2011. References are indicated at the place where they are used.

2.1. Groups

A **law of composition** on a set G is any rule for combining pairs a, b of elements of G to get another element, denoted $a * b$, of G. Formally, a law of composition is a map

$$*\colon G \times G \to G, \quad (a, b) \mapsto a * b. \tag{2.1}$$

We call $(G, *)$ a **group** if the law of composition (2.1) satisfies the following axioms:

G 1. For all $a, b, c \in G$, we have associativity, namely

$$(a * b) * c = a * (b * c).$$

G 2. There exists an element e of G such that $e * a = a * e = a$ for all $a \in G$.

G 3. If a is an element of G, then there exists an element b of G such that $a * b = b * a = e$.

If it holds for all pairs a, b in G that $a * b = b * a$, we call G a **commutative** or **abelian** group.

Depending on the group that is being considered, we may also use, instead of $a * b$, the multiplicative notations

$$ab, \ a \cdot b,$$

or the additive notation

$$a + b$$

for the law of composition. We speak of a **multiplicative** or an **additive** group, respectively.

The element e of G whose existence is asserted by G 2 is uniquely determined. Indeed, if e and e' both satisfy this condition, then $e' = e * e' = e$. For a multiplicative group, e is called the **unit element**, while it is referred to as the **zero element** for an additive group.

The element b from G 3 is also uniquely determined as can be seen from the following consideration. If $c * a = a * c = e$, then

$$c = e * c = (b * a) * c = b * (a * c) = b * e = b.$$

We call b the **inverse** of a. We denote it by a^{-1} in the multiplicative case and by $-a$ in the additive case.

Example 2.1. The set of real numbers with the addition as law of composition is a group $(\mathbb{R}, +)$.

Example 2.2. The set of complex numbers

$$\left\{z \in \mathbb{C} \mid z^n = 1\right\}$$

obtained by taking the n-th roots of unity forms a group with the multiplication of complex numbers as law of composition.

Example 2.3. The **general linear group**[1] over the real numbers is the group of all n-by-n invertible matrices with real entries. It is denoted by $GL(n, \mathbb{R})$. The law of composition is the matrix multiplication.

A subset H of a group G is called a **subgroup** if it contains the unit element, and if, whenever $a, b \in H$, then ab and a^{-1} are also elements of H (resp. $a + b \in H$ and $-a \in H$ in the additive case).

Example 2.4. The **orthogonal group** is defined as

$$O(n, \mathbb{R}) := \left\{\mathbf{A} \in GL(n, \mathbb{R}) \mid \mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{I}\right\},$$

where $\mathbf{I}$ denotes the n-by-n identity matrix and $\mathbf{A}^{\mathrm{T}}$ is the transposed matrix of $\mathbf{A}$. It can be easily checked that the orthogonal group is a subgroup of $GL(n, \mathbb{R})$.

1. Matrix groups can be endowed with the additional structure of a differentiable manifold, which makes them Lie groups. For details, we refer to Hall 2015 or Kühnel 2011.

Example 2.5. The **special orthogonal group**, denoted by $SO(n,\mathbb{R})$, is given by the orthogonal matrices $\mathbf{A}$ in $O(n,\mathbb{R})$ with $\det \mathbf{A} = 1$, i.e.,

$$SO(n,\mathbb{R}) := \left\{\mathbf{A} \in O(n,\mathbb{R}) \mid \det \mathbf{A} = 1\right\}.$$

For $n = 3$, this group is referred to as the **rotation group** because its elements describe all possible rotations around the origin of the three-dimensional space $\mathbb{R}^3$. It is often simply denoted by $SO(3)$ instead of $SO(3,\mathbb{R})$.

Among all the maps between two groups $(G,*)$ and $(G',\star)$, there are maps which preserve the group structure. We call these maps homomorphisms. A **(group) homomorphism** $f\colon G \to G'$ is a map that satisfies

$$f(a*b) = f(a) \star f(b), \tag{2.2}$$

for all a, b in G. If the map f is bijective, i.e., if it is a bijective homomorphism, then it is called a **(group) isomorphism**.

Let e and e' denote the respective unit element of $(G,*)$ and $(G',\star)$, then it holds for any group homomorphism $f\colon G \to G'$ that $e' = f(e)$ and $f(a^{-1}) = f(a)^{-1}$ for all $a \in G$. To prove the first statement, we consider that

$$f(e) = f(e*e) = f(e) \star f(e),$$

from which the desired result is obtained by multiplying both sides with $f(e)^{-1}$. The second assertion follows from the first since

$$f(a) \star f(a)^{-1} = e' = f(e) = f(a*a^{-1}) = f(a) \star f(a^{-1}).$$

The **kernel** of a group homomorphism $f\colon G \to G'$ is the set

$$\ker f := \left\{a \in G \mid f(a) = e'\right\}, \tag{2.3}$$

where $e' \in G'$ denotes the unit (respectively the zero) element of G'.

2.2. Action of a group on a set

We consider a set M. A bijective mapping $f\colon M \to M$ is called a **permutation** of M. The set of all permutations of M, denoted $\mathrm{Perm}(M)$, is a group,[2] the law of composition being the composition of mappings. The **action** of a group G on a set M is a homomorphism

$$\varphi\colon G \to \mathrm{Perm}(M), \quad g \mapsto \varphi_g\,,$$

2. See Lang 2005, Proposition 2.1 on p. 30.

with $\varphi_g \in \mathrm{Perm}(M)$, i.e., $\varphi_g : M \to M$. Let G act on a non-empty set M. For $p \in M$, we consider the subset of M consisting of all elements $\varphi_g(p)$ with $g \in G$, i.e.,

$$\mathrm{Orb}(p) := \left\{ m \in M \mid m = \varphi_g(p), \text{ with } g \in G \right\}$$

and call it the **orbit** of p under G. An action of G on M is said to be **transitive** if there is only one orbit. The statement that M consists of a single orbit is equivalent to the statement that for each pair p, p' in M there exists a g in G such that $\varphi_g(p) = p'$. The action of G on M is called **simply transitive** if for every two p, p' in M there exists *precisely one* g in G such that $\varphi_g(p) = p'$.

Example 2.6. An important[3] example is given by the permutations of a set of k elements $I_k = \{1, \ldots, k\}$. A permutation $s \in S_k := \mathrm{Perm}(I_k)$ can be represented in the form

$$\begin{bmatrix} 1 & 2 & \cdots & k \\ s(1) & s(2) & \cdots & s(k) \end{bmatrix}.$$

The group S_k is called the **symmetric group**. A **transposition** is a permutation which interchanges two numbers and leaves the others fixed. An arbitrary permutation can be represented as a sequence of transpositions.[4] A given permutation allows for infinitely many representations by transpositions. But all these representations either consist of an even or an odd number n of transpositions such that a given permutation is either **even** or **odd**. This lets us introduce the **sign of the permutation** s as $\mathrm{sgn}(s) = (-1)^n$.

We consider the illustrative example of $I_3 = \{1, 2, 3\}$. For three numbers there are six possible permutations. The three even permutations (or permutations with positive sign) can be written as

$$\overset{n=0}{\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}}, \quad \overset{n=2}{\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}}, \quad \overset{n=2}{\begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}}.$$

The three odd ones read

$$\overset{n=1}{\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}}, \quad \overset{n=1}{\begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}}, \quad \overset{n=1}{\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}}.$$

3. This group will be used in Section 2.11 to define alternating tensors by its action on covariant tensors.
4. For the proof we refer to Lang 2005, Theorem 6.1, p. 59.

2.3. Real vector spaces

A **(real) vector space** V is a set V together with two laws of composition:

(i) **addition** of elements from the set V:

$$\dot{+}\colon V \times V \to V,\ \ (v, w) \mapsto v \dot{+} w,$$

(ii) **scalar multiplication** of elements from V by real numbers:

$$\cdot\colon \mathbb{R} \times V \to V,\ \ (\alpha, v) \mapsto \alpha \cdot v.$$

The addition and the scalar multiplication are required to satisfy a number of axioms. In the list below, let u, v and w be arbitrary elements of V and α, β be real numbers:

V 1. $u \dot{+} (v \dot{+} w) = (u \dot{+} v) \dot{+} w$,

V 2. There exists an element $\mathsf{0}$ of V such that $v \dot{+} \mathsf{0} = v$ for all $v \in V$,

V 3. There exists an element $-v$ of V such that $v \dot{+} (-v) = \mathsf{0}$ for all $v \in V$,

V 4. $v \dot{+} w = w \dot{+} v$,

V 5. $\alpha \cdot (\beta \cdot v) = (\alpha\beta) \cdot v$,

V 6. $1 \cdot v = v$ for all $v \in V$,

V 7. $\alpha \cdot (v \dot{+} w) = \alpha \cdot v \dot{+} \alpha \cdot w$,

V 8. $(\alpha + \beta) \cdot v = \alpha \cdot v \dot{+} \beta \cdot v$.

Since we will only consider real vector spaces in this work, we will simply speak of vector space and omit the word real. Elements of a vector space are called **vectors**. Real numbers are called **scalars**. The vector $\mathsf{0}$ is referred to as **zero vector**. We have used the symbol $\mathsf{0}$ to make a notational distinction from the real number 0. It is instructive to show that $0 \cdot v = \mathsf{0}$ for any v in a vector space V. Indeed, the following holds

$$0 \cdot v = (0 + 0) \cdot v = 0 \cdot v \dot{+} 0 \cdot v$$

and $\mathsf{0}$ is the only vector for which holds that $\mathsf{0} = 2 \cdot \mathsf{0}$.

We have used the symbol $\dot{+}$ to distinguish between the addition defined on the vector space V from the addition of real numbers defined on $\mathbb{R}$ (see axiom V 8). For the multiplication, we have paid attention to denote the multiplication between a scalar and a vector by $\cdot$, while the multiplication between two scalars comes without a symbol (see V 5).

Axioms V 1 – V 3 correspond to axioms G 1 – G 3 from Section 2.1 and, therefore, require the set V to have the structure of an additive group. The axiom V 4 requires the additive group to be commutative. Axioms V 5 – V 8 require the two addition and the two multiplication operations to be mutually compatible, respectively. The requirement V 5 guarantees the compatibility of the multiplication between scalars and vectors with the one between scalars. Axiom V 6 makes the identity element of the multiplication on $\mathbb{R}$ also the identity element of the scalar multiplication (ii). The distributivity of scalar multiplication with respect to vector addition is required by V 7 and the one with respect to the addition on $\mathbb{R}$ by V 8.

A subset W of a vector space V is a **subspace** of V if the following three conditions are satisfied for all $v, w \in W$ and all $\alpha \in \mathbb{R}$:

(i) $v \dot{+} w \in W$,

(ii) the zero vector $\mathsf{0} \in V$ is also element of W,

(iii) $\alpha v \in W$.

Conditions (i) and (ii) are equivalent to the claim that W is a subgroup of the additive group V. Moreover, W is a vector space[5] with the addition and scalar multiplication which are induced on W by the algebraic structure of the surrounding space V.

As we did for groups in Section 2.1, we can now define homomorphisms for the algebraic structure of a vector space. A **(vector space) homomorphism** (or **linear map**) from a vector space V to a vector space V' is a map $f\colon V \to V'$ that is compatible with the two laws of composition (addition and scalar multiplication) of the vector spaces, i.e.,

$$f(v \dot{+} w) = f(v) \dot{+} f(w) \quad \text{and} \quad f(\alpha v) = \alpha f(v), \tag{2.4}$$

for all v and w in V and all $\alpha \in \mathbb{R}$. Again, a bijective vector space homomorphism is referred to as **(vector space) isomorphism**. Vector spaces which are related by an isomorphism are said to be **isomorphic**. The diligent reader may have noticed that in (2.4), we made no notational distinction between the addition on V and the one on V', neither did we for the scalar multiplication. Equation (2.4) is structurally similar to equation (2.2). On the left-hand side of the equations stands the operation which is declared on the domain of the homomorphism, that is on V. On the right-hand side of the equations, it is the one defined on the image of the homomorphism,

5. The proof can be found in any book on linear algebra, for example in Fischer 2010, p. 78.

that is on V'. Therefore, there is no danger of causing confusion by using the same notation for both additions and the scalar multiplications.

Let S and T be subspaces of a vector space V. We define the **sum** of S and T to be the subset of V consisting of all sums $u+v$ with $u \in S$ and $v \in T$. This sum is denoted $S+T$ and it is a subspace[6] of V. If it holds that $S+T=V$ and if $S \cap T = \{0\}$, then we write

$$V = S \oplus T$$

and we say that V is the **direct sum** of S and T. If $V = S \oplus T$, then T is called a **complement** of S in V. Of course, the notion of sum and direct sum can be extended to several terms. For this we refer to Chapter 1 in Roman 2008.

2.4. Equivalence relations and quotient sets

An **equivalence relation** on a set M with elements $a, b, c, \ldots$ is a relation that holds between certain pairs of elements of M. We may write it as $a \sim b$ and speak of it as **equivalence** of a and b. An equivalence relation is required to satisfy the axioms:

ER 1. If $a \sim b$ and $b \sim c$, then $a \sim c$.

ER 2. If $a \sim b$, then $b \sim a$.

ER 3. For all $a \in M$, $a \sim a$.

Suppose an equivalence relation $\sim$ is declared on a set M. Then given an element a of M, we consider the subset of M

$$[a] := \left\{ p \in M \mid a \sim p \right\}$$

that consists of all elements of M which are equivalent to a. Because of ER 3, the set $[a]$ is non-empty. Moreover, it follows from the properties ER 1–3 that all elements of $[a]$ are equivalent to one another. Each set $[a]$ is referred to as **equivalence class**. Each element of a class is called a **representative** of the class.

For any pair of elements $a, b \in M$, it either holds that $[a] = [b]$ or the sets $[a]$ and $[b]$ have no element in common. An equivalence relation $\sim$ on a set M determines a decomposition of M into disjoint equivalence classes. These equivalence classes are considered to be elements of a new set

$$M/\!\sim \; := \left\{ [p] \mid p \in M \right\} \tag{2.5}$$

6. See Lang 2004, p. 19.

that is called the **quotient set of M by $\sim$**. By assigning to any element $p \in M$ its equivalence class $[p] \in M/\sim$, we obtain a canonical map

$$f\colon M \to M/\sim, \; p \mapsto [p]. \tag{2.6}$$

Now, we will study a particular example of a quotient set (2.5) in the context where the set M is a group $(M, *)$. If $(N, *)$ is a subgroup[7] of $(M, *)$, then an equivalence relation can be declared on $(M, *)$ as

$$a \sim b \Leftrightarrow \exists n \in N\colon b = a * n. \tag{2.7}$$

The equivalence classes defined by the equivalence relation (2.7)

$$[a] = \left\{ a * n \mid n \in N \right\} =: a * N$$

are referred to as **left cosets**. By swapping a and n in the definition (2.7), we obtain the equivalence relation

$$a \hat{\sim} b \Leftrightarrow \exists n \in N\colon b = n * a \tag{2.8}$$

that comes with the equivalence classes

$$[\hat{a}] = \left\{ n * a \mid n \in N \right\} =: N * a,$$

which are called **right cosets**. Obviously, the equivalence relations (2.7) and (2.8) agree if the group $(M, *)$ is commutative.

A subgroup N of M for which the left and right cosets are identical, i.e.,

$$a * N = N * a \tag{2.9}$$

for all $a \in M$, is referred to as **normal subgroup** of M. In this case, the quotient set (2.5) can be endowed with a group structure.

Theorem 2.7. Let $(M, *)$ be a group and $(N, *)$ be a normal subgroup of M. The set

$$M/\sim \; = \{ a * N \mid a \in M \} = \{ N * a \mid a \in M \} = M/\hat{\sim} \tag{2.10}$$

of cosets[8] defined by the normal subgroup N is a group with the law of composition $\star$ such that

$$(a * N) \star (b * N) := (a * b) * N. \tag{2.11}$$

7. The definition is given on p. 16.

8. The left and right cosets agree because N is required to be a normal subgroup of M. Therefore, we simply speak of cosets.

For the proof, we refer the reader to Theorem 4.5 in Lang 2005. The set of cosets (2.10) is referred to as **quotient group** of M by N and is often denoted by M/N. This notation is preferable to $M/\sim$ or $M/\hat{\sim}$ because it involves the normal subgroup N instead of suggesting that the quotient group would result from choosing the equivalence relation that appears in the notation rather than the other. The unit element of the group M/N is given by the normal subgroup N. Indeed, we see from the law of composition (2.11) that

$$N \star (b*N) = (e*N) \star (b*N) = (e*b)*N = b*N.$$

For the quotient group, the canonical map (2.6) is the map

$$f\colon (M, *) \to (M/N, \star) \tag{2.12}$$

which to each $a \in M$ associates the coset $f(a) = a*N$. It can be easily verified that (2.12) is a (group) homomorphism when the quotient set M/N is equipped with the group structure from Theorem 2.7. Therefore, the map (2.12) is called the **canonical homomorphism** of $(M, *)$ onto the quotient group $(M/N, \star)$.

It holds that the kernel[9] of $f\colon (M, *) \to (M/N, \star)$

$$\ker f := \left\{ a \in M \mid f(a) = N \right\}$$

is just the normal subgroup N. We start by showing first that

$$N \subseteq \ker f \tag{2.13}$$

and then we convince ourselves that

$$\ker f \subseteq N.$$

Let n be an arbitrary element of the normal subgroup N, then it holds by the definition of a subgroup that $f(n) = n*N = N$ which proves (2.13). Conversely, consider an arbitrary element $a \in M$ for which $f(a) = a*N$ is the unit element of M/N, i.e., $f(a) = N$, so $a*N = N$. The latter condition can be rewritten as $a*n \in N$, for all $n \in N$ and by choosing $n = e$, it implies that $a \in N$, which completes the proof.

We have shown that $N = \ker f$ which says that the normal subgroup N is just the kernel of the homomorphism (2.12). So given a normal subgroup N of a group M, we know that there is a group homomorphism such that N is the kernel of that homomorphism. The converse is also true, i.e., the kernel of any group homomorphism $h\colon (M, *) \to (G, \cdot)$ of M into some group G is a normal subgroup of M.

9. See equation (2.3) for the definition.

Let K denote the kernel of the homomorphism $h\colon M \to G$. By definition, h satisfies

$$h\left(a*K*a^{-1}\right) = h(a)\cdot h(K)\cdot h(a^{-1}) = e,$$
$$h\left(a^{-1}*K*a\right) = h(a^{-1})\cdot h(K)\cdot h(a) = e,$$

where e denotes the unit element of the group $(G,\cdot)$. The first equation implies that $a*K*a^{-1}\subseteq K$, while the second implies that $a^{-1}*K*a\subseteq K \Leftrightarrow K\subseteq a*K*a^{-1}$. Hence $a*K*a^{-1}=K$, which proves the kernel K to be a normal subgroup (see equation (2.9)) of $(M,*)$.

Let $a\in M$, then it holds for all $k\in K$ that

$$h(a*k) = h(a)\cdot h(k) = h(a)\cdot e = h(a)$$

or stated equivalently

$$h(a*K) = h(a).$$

This means that all elements in a (left) coset have the same image under the homomorphism h. These observations lead us to the **first isomorphism theorem** for groups.

Theorem 2.8 (Lang 2005, Corollary 4.7)**.** Let $h\colon (M,*)\to(G,\cdot)$ be a homomorphism, and let K be its kernel. Then the association

$$a*K \mapsto h(a*K)$$

is an isomorphism

$$M/K \cong \operatorname{im} h$$

of M/K with the image of h.

In Section 2.3, we saw that the vector space axioms V 1–V 4 require a vector space to be a commutative additive group. Therefore, similar results can be stated for vector spaces. Let V be a vector space and W be a subspace of V. Then the equivalence relation (2.7) reads in additive notation

$$\begin{aligned} u\sim v &\Leftrightarrow \exists w\in W\colon v = u+w \\ &\Leftrightarrow v-u\in W. \end{aligned} \tag{2.14}$$

Since, a vector space is a *commutative* additive group, the equivalence relations (2.7) and (2.8) agree. The equivalence classes defined by (2.14) are given by

$$[u] := \left\{u+w \mid w\in W\right\} = u+W$$

and they are called **cosets** again. The quotient set of V by $\sim$ can be equipped with the structure of a vector space.

Theorem 2.9 (Roman 2008, Theorem 3.1)**.** Let V be a vector space and W be a subspace V. The set

$$V/\!\sim\, = V/W = \left\{[u] \mid u \in V\right\} = \left\{u + W \mid u \in V\right\}$$

of cosets defined by the subspace W is a vector space with the laws of composition

$$(u+W) + (v+W) = (u+v) + W$$

and

$$\alpha \cdot (u+W) = (\alpha \cdot u) + W.$$

The zero vector in V/W is the coset $0 + W = W$.

The vector space V/W from Theorem 2.9 is called the **quotient space** of V by W. As a quotient set, the quotient space V/W comes with the canonical map (2.6), i.e.,

$$f\colon V \to V/W,\ u \mapsto [u] = u + W \tag{2.15}$$

which sends each vector to the coset containing it. Similar to the situation for groups. The map (2.15) can be shown to be a (vector space) homomorphism, when the quotient set V/W is endowed with the vector space structure from Theorem 2.9. Therefore, the map (2.15) is again called **canonical homomorphism**. Moreover, the kernel of $f\colon V \to V/W$ is nothing but the subspace W. For the proof of the previous two statements, the reader is referred to Theorem 3.2 in Roman 2008. Finally, we can state the **first isomorphism theorem** for vector spaces that is the analogue result for vector spaces as Theorem 2.8 was for groups.

Theorem 2.10 (Roman 2008, Theorem 3.5)**.** Let $h: V \to T$ be a (vector space) homomorphism, and let K be its kernel. Then the association

$$u + K \mapsto h(u + K)$$

is an isomorphism

$$V/K \cong \operatorname{im} h$$

of V/K with the image of h.

Let S be a subspace of a vector space V. The first isomorphism theorem can be used to show[10] that whenever

$$V = W \oplus T,$$

then it holds that

$$T \cong V/W. \tag{2.16}$$

This means that all complements of W in V are isomorphic to V/W and hence to each other.

10. See Theorem 3.6 in Roman 2008.

2.5. Basis of a vector space

So far, we have considered sets endowed with some additional algebraic structure. In a set, the order of elements does not matter. If we are given two vectors v, w from a vector space V, then the sets $\{v, w\}$ and $\{w, v\}$ are equal because they contain the same elements. Now, we consider **tuples** of vectors, i.e., collections of objects where the order matters. Consequently, the tuples (v, w) and (w, v) are *not* equal. In this work, we use the term tuple to designate any ordered collection of objects. A typical example is given by the elements of $\mathbb{R}^n$, which are n-tuples of real numbers.

Let V be a vector space and let $S = (v_1, \dots, v_n)$ be an n-tuple of elements of V. A **linear combination** of S is a vector of the form

$$w = \alpha^1 \cdot v_1 \dot{+} \cdots \dot{+} \alpha^n \cdot v_n = \alpha^i \cdot v_i,$$

with $\alpha^i \in \mathbb{R}$. The last equality follows by adopting Einstein's summation convention, which says that a summation is understood over any index appearing once as a lower and once as an upper index. The latin index i thereby runs from 1 to n. Let W be the set of all such elements $\alpha^i \cdot v_i$ with $\alpha^i \in \mathbb{R}$. It can be easily verified that W is a subspace of V. We say that W is the subspace **spanned** by the vectors $v_1, \dots, v_n$ and write

$$W = \operatorname{span}\{v_1, \dots, v_n\}.$$

A vector space V is **finite-dimensional** if some finite set of vectors spans V. Otherwise, V is **infinite-dimensional**. We say that $v_1, \dots, v_n$ are **linearly dependent** if there exist elements $\alpha^1, \dots, \alpha^n$ not all equal to 0 such that

$$\alpha^i \cdot v_i = \alpha^1 \cdot v_1 \dot{+} \cdots \dot{+} \alpha^n \cdot v_n = \mathsf{0}.$$

If no such elements do exist, then we qualify the vectors $v_1, \dots, v_n$ to be **linearly independent**.

A **basis** of a vector space V is a tuple $B = (e_1, \dots, e_n)$ of linearly independent vectors e_i from V that generate V, i.e., $V = \operatorname{span}\{e_1, \dots, e_n\}$. The **dimension** of a finite-dimensional vector space V is the number n of vectors in a basis. It is denoted by dim V. The choice of a basis in a vector space V provides a one-to-one correspondence between vectors and n-tuples of real numbers. Each element $v \in V$ can be written as linear combination

$$v = x^i \cdot e_i = x^1 \cdot e_1 \dot{+} \cdots \dot{+} x^n \cdot e_n$$

of the basis vectors $e_1, \dots, e_n$. The n-tuple of real numbers $(x^1, \dots, x^n)$ gathers the **coordinates** of v with respect to the basis B. In other words, each element v of the vector space can be uniquely represented by an element

$\mathbf{x} := (x^1, \dots, x^n)$ of $\mathbb{R}^n$. This one-to-one correspondence can be formally written as a *linear* bijective map called **chart** or **coordinate map**

$$\phi\colon V \to \mathbb{R}^n,\ v \mapsto \phi(v) = (x^1, \dots, x^n), \tag{2.17}$$

with $v = x^i \cdot e_i$. If we consider another basis $B' = (e'_1, \dots, e'_n)$, then we get another n-tuple of real numbers $(y^1, \dots, y^n)$ as representation of the vector v

$$v = y^i \cdot e'_i = y^1 \cdot e'_1 \dotplus \cdots \dotplus y^n \cdot e'_n$$

in $\mathbb{R}^n$. The corresponding chart is given by

$$\psi\colon V \to \mathbb{R}^n,\ v \mapsto \psi(v) = (y^1, \dots, y^n). \tag{2.18}$$

Given the charts (2.17) and (2.18), we can perform a **change of coordinates**, that is, given the coordinates of a vector with respect to the basis B, we can compute its coordinates with respect to the basis B'. The change of coordinates from the basis B to the basis B' is given by the map

$$\psi \circ \phi^{-1}\colon \mathbb{R}^n \to \mathbb{R}^n,\ (x^1, \dots, x^n) \mapsto (y^1, \dots, y^n),$$

which is a linear bijective map.

The space $\mathbb{R}^n$ can be endowed with the component-wise addition

$$\mathbf{v} + \mathbf{w} := (v^1 + w^1, \dots, v^n + w^n) \tag{2.19}$$

and the scalar multiplication

$$\alpha \bullet \mathbf{v} = \alpha\mathbf{v} := (\alpha v^1, \dots, \alpha v^n), \tag{2.20}$$

which have to hold for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and all $\alpha \in \mathbb{R}$. The zero vector is defined as $\mathbf{0} := (0, \dots, 0)$ and the inverse element of a vector $\mathbf{v} = (v^1, \dots, v^n)$ is declared as $-\mathbf{v} := (-v^1, \dots, -v^n)$. It is easy to verify that $\mathbb{R}^n$ endowed with the zero vector, the inverse element and with the laws of composition (2.19) and (2.20) is indeed a vector space. With the vector space structure on $\mathbb{R}^n$, the coordinate maps (2.17) and (2.18) become isomorphisms between the vector spaces V and $\mathbb{R}^n$. This means that the two diagrams

$$\begin{tikzcd} V \times V \arrow[r, "{(\phi,\phi)}"] \arrow[d, "\dotplus"'] & \mathbb{R}^n \times \mathbb{R}^n \arrow[d, "+"] \\ V \arrow[r, "\phi"] & \mathbb{R}^n \end{tikzcd} \qquad \begin{tikzcd} \mathbb{R} \times V \arrow[r, "{(\mathrm{id}_{\mathbb{R}},\phi)}"] \arrow[d, "\cdot"'] & \mathbb{R} \times \mathbb{R}^n \arrow[d, "\bullet"] \\ V \arrow[r, "\phi"] & \mathbb{R}^n \end{tikzcd}$$

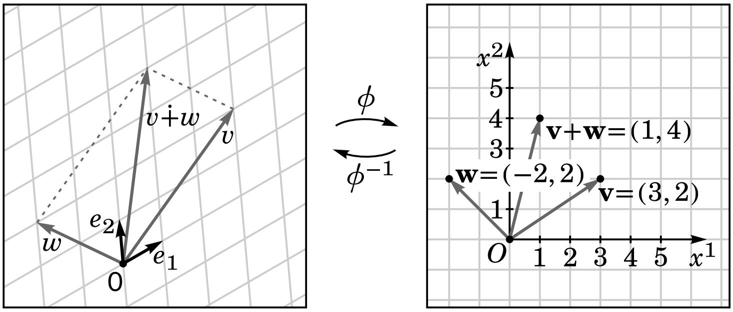


Figure 2.1.: The vector space of arrows in the plane and its coordinate representation on $\mathbb{R}^2$.

commute, that is

$$+\circ(\phi,\phi) = \phi\circ\dot{+}$$

and

$$\bullet\circ(\mathrm{id}_{\mathbb{R}},\phi) = \phi\circ\cdot\,,$$

where $\mathrm{id}_{\mathbb{R}}:\mathbb{R}\to\mathbb{R},\ \alpha\mapsto\alpha$ denotes the identity map on $\mathbb{R}$ and $\circ$ the composition of mappings.

So far, we have used $\dot{+}$ to denote the addition on the vector space V and an ordinary plus symbol to denote the addition (2.19) defined on $\mathbb{R}^n$. The scalar multiplication on V is denoted by $\cdot$, while the one on $\mathbb{R}^n$ is written as $\bullet$. From now on, this didactically motivated distinction will be dropped because it will always be clear from the objects that are to be added/multiplied, which operation is to be used.

Example 2.11. To become familiar with the concepts from Section 2.3, we consider the vector space of arrows that can be drawn in the plane. We endow this set with an addition and a scalar multiplication that make it a real two-dimensional vector space. The addition of two arrows v and w can be defined graphically using the "parallelogram rule" as shown in Figure 2.1. The sum $v+w$ is the arrow that starts from the same point 0 as v and w and that ends in the opposite corner of the parallelogram spanned by v and w. The scalar multiplication of an arrow v by α is given graphically by constructing an arrow αv which is stretched by the factor $\alpha\in\mathbb{R}$ compared to v. The choice of a basis (e_1,e_2) provides a coordinate map from the space of arrows to $\mathbb{R}^2$. The component-wise addition on $\mathbb{R}^2$ is compatible with the addition declared on the space of arrows. In other words, the corresponding diagram commutes. The same holds for the scalar multiplications.

2.6. The dual space of a vector space

Let V be a real vector space. The set V^* of all linear maps from V to $\mathbb{R}$ is the **dual space of** V. An element $\rho \in V^*$ is called a **covector**. The dual space V^* itself becomes a real vector space if it is equipped with the following two laws of composition defining addition and scalar multiplication. The addition $+\colon V^* \times V^* \to V^*$ is declared such that for all ρ, $\sigma \in V^*$ and $v \in V$

$$(\rho + \sigma)(v) := \rho(v) + \sigma(v).$$

The scalar multiplication is defined by

$$(\alpha\rho)(v) := \alpha\rho(v),$$

for all $\rho \in V^*$, $v \in V$, and $\alpha \in \mathbb{R}$. The definitions of these composition laws rely on the addition and multiplication of real numbers and on the definition of covectors as linear maps of vectors to the real numbers.

If the vector space V is finite-dimensional with basis $(e_1, \dots, e_n)$, then V^* has the same dimension[11] as V, i.e., $\dim V = \dim V^* = n$. A basis $(e^1, \dots, e^n)$ of V^* is called the **dual basis** of the basis $(e_1, \dots, e_n)$ of V if

$$e^i(e_j) = \delta^i_j := \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

with $i, j = 1, \dots, n$. The symbol δ^i_j is known as **Kronecker delta**. Using the dual basis, a covector ρ can be expressed as

$$\rho = \rho_i\, e^i, \tag{2.21}$$

with $\rho_i = \rho(e_i)$ as can be seen by applying ρ to the primal basis vectors $e_1, \dots, e_n$.

The **bidual space** $V^{**} := (V^*)^*$ is the space of linear real-valued maps on the dual space V^*. In the case where V is finite-dimensional, the vector spaces V^{**} and V are canonically isomorphic in the sense that the spaces are related by the following vector space isomorphism[12]

$$\iota\colon V \to V^{**},\ v \mapsto \iota_v \quad \text{with } \iota_v(\rho) := \rho(v)\ \forall \rho \in V^*, \tag{2.22}$$

which relates vectors $v \in V$ to elements $\iota_v \in V^{**}$. We identify V^{**} with V and write $v(\rho) = \rho(v)$. This justifies the dot notation

$$v \cdot \rho = \rho \cdot v = v(\rho) = \rho(v), \tag{2.23}$$

to which we refer as **duality pairing**.

11. The proof can be found in Bishop et al. 1980, Proposition 2.7.1 on p. 75.
12. See Theorem 2.9.1 in Bishop et al. 1980 for the proof that (2.22) defines an isomorphism.

A subspace W of the vector space V defines the subspace

$$W^\circ := \{\rho \in V^* \mid \rho(w) = 0, \text{ for all } w \in W\} \tag{2.24}$$

in V^* that is called the **annihilator of** W. Note that the dimension[13] of W° is given by

$$\dim W^\circ = \dim V - \dim W. \tag{2.25}$$

2.7. Bilinear forms

A **bilinear form** on a real vector space V is a real-valued map

$$B\colon V \times V \to \mathbb{R} \tag{2.26}$$

that is linear in each argument separately such that

$$\begin{aligned} B(\alpha u + \beta v, w) &= \alpha B(u, w) + \beta B(v, w), \\ B(u, \alpha v + \beta w) &= \alpha B(u, v) + \beta B(u, w), \end{aligned}$$

for all vectors $u, v, w \in V$ and real numbers $\alpha, \beta \in \mathbb{R}$. The bilinear form $B\colon V \times V \to \mathbb{R}$ is said to be **non-degenerate** if

$$B(u, v) = 0, \ \forall v \in V \ \Rightarrow \ u = 0. \tag{2.27}$$

The bilinear form $B\colon V \times V \to \mathbb{R}$ is called **symmetric** if $B(v, w) = B(w, v)$ for all $v, w \in V$ and it is said to be **alternating** (or **skew-symmetric**) if $B(v, w) = -B(w, v)$ for all $v, w \in V$. Among all possible bilinear forms, the symmetric and alternating forms are the ones that have a predictable behaviour if their arguments are interchanged.

Symplectic vector space

A **symplectic form** on a real vector space V is a real-valued map

$$\omega\colon V \times V \to \mathbb{R} \tag{2.28}$$

that satisfies the following four axioms for all $u, v, w \in V$ and all $\alpha, \beta \in \mathbb{R}$:

SF 1. $\omega(u + v, w) = \omega(u, w) + \omega(v, w)$,

SF 2. $\omega(\alpha v, w) = \alpha\omega(v, w)$,

SF 3. $\omega(v, w) = -\omega(w, v)$,

13. See Fischer 2010, pp. 333–334 for the proof.

SF 4. $\omega(v, w) = 0$ for all $w \in V \;\Rightarrow\; v = 0$.

By axioms SF 1 – SF 3, the map $\omega\colon V \times V \to \mathbb{R}$ is an alternating bilinear form that has to be non-degenerate by axiom SF 4. A real vector space V together with a symplectic form (2.28) is a **symplectic vector space**.

Inner product space

A real vector space V can be equipped with an **inner product**,[14] which is a real-valued map

$$\langle \cdot, \cdot \rangle\colon V \times V \to \mathbb{R} \tag{2.29}$$

that satisfies the following four axioms for all $u, v, w \in V$ and all $\alpha, \beta \in \mathbb{R}$:

IP 1. $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$,

IP 2. $\langle \alpha v, w \rangle = \alpha \langle v, w \rangle$,

IP 3. $\langle v, w \rangle = \langle w, v \rangle$,

IP 4. $\langle v, v \rangle > 0 \quad \text{if} \quad v \neq 0$.

Axioms IP 1 – IP 3 require the inner product to be a symmetric bilinear form that is non-degenerate by axiom IP 4. A real vector space V together with an inner product (2.29) is an **inner product space**.

An inner product allows to define the **length** of a vector $v \in V$ as

$$\|v\| := \sqrt{\langle v, v \rangle}$$

and the **angle** $\measuredangle(v, w)$ that is spanned by a pair of nonzero vectors $v, w \in V$ as

$$\cos \measuredangle(v, w) := \frac{\langle v, w \rangle}{\|v\| \|w\|}. \tag{2.30}$$

Example 2.12. On the vector space $\mathbb{R}^n$, the inner product that is defined for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ as

$$\langle \mathbf{v}, \mathbf{w} \rangle := \sum_{i=1}^{n} v^i w^i$$

is called the **standard inner product of** $\mathbb{R}^n$.

14. See Hoffman et al. 1971, p. 271.

Orthogonal complement

Inspired by the notion of angle (2.30), we define the set of vectors that are orthogonal to a given subspace W of a vector space V. However, we generalize the concept from an inner product on V to an arbitrary (possibly degenerate) bilinear form B on V that is either symmetric or alternating. The **orthogonal complement of** W is the set

$$W^\perp := \Big\{ v \in V \mid B(v,w) = 0, \text{ for all } w \in W \Big\}.$$

If the bilinear form B is given by an inner product, i.e., if B is symmetric and positive definite (and thereby non-degenerate), then $W \cap W^\perp = \{0\}$ and $V = W \oplus W^\perp$ and obviously

$$\dim V = \dim W + \dim W^\perp. \tag{2.31}$$

Let

$$B|_W \colon W \times W \to \mathbb{R}$$

denote the restriction of the bilinear form (2.26) to the subspace W. Indeed, the intersection $W \cap W^\perp = \{0\}$ if and only if $B|_W$ is non-degenerate. In particular, the bilinear form B is non-degenerate if and only if $V^\perp = \{0\}$. The dimension formula (2.31) is generalized to the case of a possibly non-degenerate bilinear form by the following proposition.

Proposition 2.13. Let V be a real vector space and $W \subseteq V$ be a subspace in V. Let V be equipped with a symmetric or alternating bilinear form $B \colon V \times V \to \mathbb{R}$. Then

$$\dim W + \dim W^\perp = \dim V + \dim W \cap V^\perp.$$

In particular, if B is non-degenerate, then

$$\dim W + \dim W^\perp = \dim V.$$

Proof. Consider the linear map

$$f \colon V \to V^*, \ v \mapsto B(v, \cdot)$$

induced by the bilinear form $B \colon V \times V \to \mathbb{R}$ and let

$$g := f|_W \colon W \to V^*, \ w \mapsto B(w, \cdot)$$

denote its restriction to $W \subseteq V$. By the rank-nullity theorem, it holds that

$$\dim W = \dim \ker g + \dim \operatorname{im} g. \tag{2.32}$$

By comparing

$$\ker g = \left\{ w \in W \mid g(w) = 0 \in V^* \right\}$$

and

$$\begin{aligned} V^\perp &= \left\{ v \in V \mid B(v, w) = 0, \text{ for all } w \in V \right\} \\ &= \left\{ v \in V \mid B(v, \cdot) = 0 \in V^* \right\}, \end{aligned}$$

we see that

$$\ker g = W \cap V^\perp.$$

By (2.24), the annihilator of $\operatorname{im} g$ is given by

$$(\operatorname{im} g)^\circ = \left\{ v \in V \mid \rho(v) = 0, \text{ for all } \rho \in \operatorname{im} g \right\}$$

because we identified V^{**} with V. Every $\rho \in \operatorname{im} g$ is of the form $\rho = B(w, \cdot)$ for some $w \in W$ and, therefore,

$$(\operatorname{im} g)^\circ = \left\{ v \in V \mid B(w, v) = 0, \text{ for all } w \in W \right\} = W^\perp,$$

where the last equality follows by the (skew)-symmetry of the bilinear form since it implies that $B(w, v) = B(v, w)$ (respectively $B(w, v) = -B(v, w)$). Hence, it follows by the dimension formula (2.25) that

$$\dim \operatorname{im} g = \dim V^* - \dim\, (\operatorname{im} g)^\circ = \dim V - \dim W^\perp$$

and, by (2.32), we conclude that

$$\dim W = \dim W \cap V^\perp + \left(\dim V - \dim W^\perp \right).$$

□

2.8. Lie algebra

A **Lie bracket** on a real vector space V is a law of composition

$$[\cdot, \cdot] \colon V \times V \to V, \ \ (u, v) \mapsto [u, v]$$

that satisfies the following axioms for all $u, v, w \in V$ and all $\alpha, \beta \in \mathbb{R}$:

LB 1. (Bilinearity)

$$\begin{aligned} [\alpha u + \beta v, w] &= \alpha [u, w] + \beta [v, w], \\ [u, \alpha v + \beta w] &= \alpha [u, v] + \beta [u, w], \end{aligned}$$

LB 2. (Anticommutativity)

$$[u, v] = -[v, u],$$

LB 3. (Jacobi identity)

$$[u, [v, w]] + [v, [w, u]] + [w, [u, v]] = 0.$$

A vector space V together with a Lie bracket $[\cdot, \cdot]$ is called a **Lie algebra**. A subspace $W \subseteq V$ of a Lie algebra $(V, [\cdot, \cdot])$ is a **Lie subalgebra** if W is closed under the law of composition defined by the Lie bracket, i.e., if $[u, v] \in W$ for all $u, v \in W$.

Example 2.14. The vector space $\mathbb{R}^3$ (component-wise addition (2.19) and scalar multiplication (2.20)) becomes a Lie algebra when it is equipped with the cross product as Lie bracket such that for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ with

$$[\mathbf{u}, \mathbf{v}] := \mathbf{u} \times \mathbf{v} = \begin{bmatrix} u^1 \\ u^2 \\ u^3 \end{bmatrix} \times \begin{bmatrix} v^1 \\ v^2 \\ v^3 \end{bmatrix} = \begin{bmatrix} u^2 v^3 - u^3 v^2 \\ u^3 v^1 - u^1 v^3 \\ u^1 v^2 - u^2 v^1 \end{bmatrix}.$$

The reader might easily check that $\mathbf{u} \times \mathbf{v}$ satisfies the axioms LB 1–LB 3 and that every one-dimensional subspace of $\mathbb{R}^3$ (lines through the origin) is a Lie subalgebra of $\mathbb{R}^3$.

2.9. Affine spaces

If we try to use a two-dimensional real vector-space to make a mathematical abstraction of a piece of paper on which we have drawn the arrows from Example 2.11, then we might stumble on finding the origin. Indeed, among all the material points of the sheet, there is no isolated outsider which could be identified as *the* origin. However, as soon as we have designated one material point p as origin we can identify any other point of the sheet with the arrow that points from the origin to this particular point (see Figure 2.2). If we put it differently, an arrow v can be "applied" to each point of the sheet such that this point is translated by the arrow (of course some points may be mapped outside of the sheet). The underlying algebraic structure is the one of an affine space.

A set A is called an **affine space**[15] modelled on the real vector space V if there is a map, called an **affine structure**,

$$A \times V \to A, \quad (p, v) \mapsto p \hat{+} v \tag{2.33}$$

that satisfies the following axioms:

15. See Crampin and Pirani 1987, p. 9.

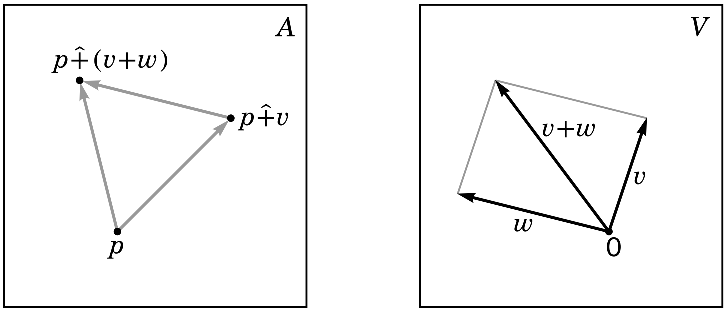


Figure 2.2.: Two-dimensional affine space modelled over the vector space of arrows in the plane from Example 2.11.

A 1. $(p \hat{+} v) \hat{+} w = p \hat{+} (v + w)$ for all $p \in A$ and all $v, w \in V$,

A 2. $p \hat{+} 0 = p$ for all $p \in A$, where $0 \in V$ is the zero vector,

A 3. for any pair of points $p, p' \in A$ there is a unique element of V, denoted $p' \hat{-} p$, such that $p \hat{+} (p' \hat{-} p) = p'$.

If $\dim V = n$, then we say that the affine space A is of dimension n.

The diligent reader may have observed that the map (2.33) together with axioms A 1 and A 2 defines the action[16] of V on A as an additive group. The map $\varphi_v \in \mathrm{Perm}(A)$ is given by $\varphi_v : A \to A,\ p \mapsto \varphi_v(p) = p \hat{+} v$. Axiom A 3 requires the group action of V on A to be simply transitive.

2.10. Tensors

In Section 2.3, we saw that linear maps between vector-spaces are special because they are compatible with the algebraic structure of the spaces they relate. This section studies maps from Cartesian products of vector spaces to the real numbers that are linear in each argument. Our first contact with this species called tensor takes place in the protected area of finite dimensional vector spaces. In Chapter 3, we will see that these animals can be resettled point by point to a differentiable manifold.

Let $V_1, \dots, V_k$ and W be vector spaces. A map $f\colon V_1 \times \dots \times V_k \to W$ is **multilinear** if it is linear in each argument, i.e.,

$$f({}_1v, \dots, \alpha\, {}_iv + \beta\, {}_iu, \dots, {}_kv) = \alpha f({}_1v, \dots, {}_iv, \dots, {}_kv) + \beta f({}_1v, \dots, {}_iu, \dots, {}_kv),$$

16. The action of a group on a set is defined in Section 2.2.

for all $\alpha, \beta \in \mathbb{R}$ and $i = 1, \dots, k$. Let $L(V_1, \dots, V_k; W)$ denote the set of all multilinear maps from $V_1, \dots, V_k$ to W. It becomes a real vector space when endowed with the laws of composition declared such that

$$\begin{aligned}(f+g)({}_1v, \dots, {}_kv) &:= f({}_1v, \dots, {}_kv) + g({}_1v, \dots, {}_kv),\\ (\alpha f)({}_1v, \dots, {}_kv) &:= \alpha f({}_1v, \dots, {}_kv),\end{aligned}$$

for all $f, g \in L(V_1, \dots, V_k; W)$ and all $\alpha \in \mathbb{R}$.

If $W = \mathbb{R}$, then a map $f \in L(V_1, \dots, V_k; \mathbb{R})$ is called a **k-form**, a **tensor of k-th order** or simply a **tensor**. Let $V_1, \dots, V_k$ and $W_1, \dots, W_l$ be vector spaces. If $f \in L(V_1, \dots, V_k; \mathbb{R})$ and $g \in L(W_1, \dots, W_l; \mathbb{R})$ are two tensors, then the **tensor product of f and g** is the real-valued function

$$f \otimes g \colon V_1 \times \dots \times V_k \times W_1 \times \dots \times W_l \to \mathbb{R} \tag{2.34}$$

defined by

$$f \otimes g({}_1v, \dots, {}_kv, {}_1w, \dots, {}_lw) := f({}_1, \dots, {}_kv) g({}_1w, \dots, {}_lw).$$

It holds that $f \otimes g$ is a tensor, i.e., $f \otimes g \in L(V_1, \dots, V_k, W_1, \dots, W_l; \mathbb{R})$.

Theorem 2.15 (John M. Lee 2013, Proposition 12.4)**.** Let $V_1, \dots, V_k$ be vector spaces of the respective dimension $n_1, \dots, n_k$. For $j = 1, \dots, k$, let

$$({}_je_1, \dots, {}_je_{n_j})$$

be a basis of V_j and let

$$({}_je^1, \dots, {}_je^{n_j})$$

designate the corresponding dual basis[17] of V_j^*. Then the set

$$B := \left\{ {}_1e^{i_1} \otimes \dots \otimes {}_ke^{i_k} \;\middle|\; 1 \le i_1 \le n_1, \dots, 1 \le i_k \le n_k \right\}$$

is a basis of $L(V_1, \dots, V_k; \mathbb{R})$ and, therefore, has the dimension $n_1 \cdots n_k$.

Using the **multi-index** $I := (i_1, \dots, i_k)$, a tensor $f \in L(V_1, \dots, V_k; \mathbb{R})$ can be written as

$$\begin{aligned} f &= \sum_{i_1=1}^{n_1} \cdots \sum_{i_k=1}^{n_k} f_{i_1,\dots,i_k}\, {}_1e^{i_1} \otimes \dots \otimes {}_ke^{i_k}\\ &= \sum_I f_I\, {}_1e^{i_1} \otimes \dots \otimes {}_ke^{i_k}, \end{aligned}$$

with $f_I = f_{i_1,\dots,i_k} := f({}_1e_{i_1}, \dots, {}_ke_{i_k})$.

17. The dual basis to a given basis of a vector space is defined in Section 2.6.

Proof. We consider arbitrary vectors

$$_jv = \sum_{i_j=1}^{n_j} {_jv^{i_j}} \, {_je_{i_j}} \in V_j$$

with $j = 1, \ldots, k$ and a tensor $f \in L(V_1, \ldots, V_k; \mathbb{R})$. The following holds

$$f({_1v}, \ldots, {_kv}) = \sum_{i_1=1}^{n_1} \cdots \sum_{i_k=1}^{n_k} {_1v^{i_1}} \cdots {_kv^{i_k}} \, f({_1e_{i_1}}, \ldots, {_ke_{i_k}}) \tag{2.35}$$

because f is multilinear. We assume that

$$f = \sum_I f_I \, {_1e^{i_1}} \otimes \cdots \otimes {_ke^{i_k}}. \tag{2.36}$$

In this case, it holds that

$$\begin{aligned} f({_1v}, \ldots, {_kv}) &= \sum_I \sum_{j_1=1}^{n_1} \cdots \sum_{j_k=1}^{n_k} f_I \, {_1e^{i_1}} \left({_1v^{j_1}} {_1e_{j_1}} \right) \cdots {_ke^{i_k}} \left({_kv^{j_k}} {_ke_{j_k}} \right) \\ &= \sum_I \sum_{j_1=1}^{n_1} \cdots \sum_{j_k=1}^{n_k} f_I \, {_1v^{j_1}} \cdots {_kv^{j_k}} \, \delta^{i_1}_{j_1} \cdots \delta^{i_k}_{j_k} \\ &= \sum_I f_I \, {_1v^{i_1}} \cdots {_kv^{i_k}}. \end{aligned} \tag{2.37}$$

Comparing expressions (2.35) and (2.37), it follows that f has the form (2.36) if and only if $f_I = f({_1e_{i_1}}, \ldots, {_ke_{i_k}})$. Thus, B spans $L(V_1, \ldots, V_k; \mathbb{R})$.

We still need to show the linear independence of the ${_1e^{i_1}} \otimes \cdots \otimes {_ke^{i_k}}$. For this, we set

$$f = \sum_I f_I \, {_1e^{i_1}} \otimes \cdots \otimes {_ke^{i_k}} \stackrel{!}{=} 0,$$

that is

$$f({_1v}, \ldots, {_kv}) = \sum_I f_I \, {_1e^{i_1}} \otimes \cdots \otimes {_ke^{i_k}} ({_1v}, \ldots, {_kv}) \stackrel{!}{=} 0,$$

for all ${_jv} \in V_j$ with $j = 1, \ldots, k$. By equation (2.37), this implies that

$$f_I = 0,$$

for all multi-indices I and this proves that the ${_1e^{i_1}} \otimes \cdots \otimes {_ke^{i_k}}$ are linearly independent. Thus B is a basis. The number of elements in B corresponds to the dimension, therefore,

$$\dim \left(L(V_1, \ldots, V_k; \mathbb{R}) \right) = n_1 \cdots n_k.$$

□

Now, we consider the special case where the vector spaces $V_1, \dots, V_k$ and $W_1, \dots, W_l$ are replaced by k copies of the dual space V^* and by l copies of the vector space V itself, respectively. The **space of mixed tensors on V of type** (k, l) is the vector space

$$\otimes^k_l V := L(\underbrace{V^*, \dots, V^*}_{k}, \underbrace{V, \dots, V}_{l}; \mathbb{R}) =: \underbrace{V \otimes \cdots \otimes V}_{k} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{l}. \tag{2.38}$$

An element $f \in \otimes^k_l V$ is a (k, l)**-tensor**. There exist other definitions of tensors for which we refer to Chapter 7 in Jeffrey M. Lee 2009. In the finite-dimensional case, these definitions are equivalent because the different tensor spaces are isomorphic. Note that the position of the asterisks in equation (2.38) is correct. By definition, elements from V^* are linear real-valued maps on V and because of the isomorphism $V \cong V^{**}$ from (2.22), vectors from V are identified as linear real-valued maps on V^* (i.e., as elements of V^{**}).

The two special cases $k = 0$ and $l = 0$ play an important role and, therefore, have own names. For $l = 0$, the resulting **space of contravariant** k**-tensors on** V is written as

$$\otimes^k V := \otimes^k_0 V = L(\underbrace{V^*, \dots, V^*}_{k}; \mathbb{R}) = \underbrace{V \otimes \cdots \otimes V}_{k}.$$

An element $f \in \otimes^k V$ is called a **contravariant** k**-tensor** or a **contravariant tensor of rank** k. For $k = 0$, the **space of covariant** l**-tensors on** V is denoted by

$$\otimes^l V^* := \otimes^0_l V = L(\underbrace{V, \dots, V}_{l}; \mathbb{R}) = \underbrace{V^* \otimes \cdots \otimes V^*}_{l}.$$

An element $f \in \otimes^l V^*$ is referred to as **covariant** l**-tensor** or as **covariant tensor of rank** l. The bilinear forms from Section 2.7 are covariant tensors of rank two.

A covariant l-tensor f on a vector space V assigns a real number to l elements of V. Among these tensors, two special types of tensors can be singled out, the symmetric and the alternating tensors. The tensor f is **symmetric** if its value remains unchanged by interchanging any pair of its arguments, i.e.,

$$f(v_1, \dots, v_i, \dots, v_j, \dots, v_l) = f(v_1, \dots, v_j, \dots, v_i, \dots, v_l),$$

whenever $1 \le i < j \le l$ and for all $v_1, \dots, v_l \in V$. The tensor f is **alternating** (**antisymmetric** or **skew-symmetric**) if it changes sign whenever two of its arguments are interchanged. This means that

$$f(v_1, \dots, v_i, \dots, v_j, \dots, v_l) = -f(v_1, \dots, v_j, \dots, v_i, \dots, v_l),$$

whenever $1 \leq i < j \leq l$ and for all $v_1, \dots, v_l \in V$. The common property of symmetric and alternating tensors is that their value changes predictably if their arguments are rearranged.

It can be shown that the sets of symmetric and alternating tensors of order l constitute respective subspaces of the tensor space $\otimes^l V^*$. This means in particular that the sum of two symmetric (alternating) tensors is a symmetric (alternating) tensor. There are natural projections on these two subspaces. We will see that tensors and, especially tensor fields, play an eminent role in mechanics. We continue our study by focussing on alternating tensors.

2.11. Alternating forms and their exterior algebra

First, we want to state the projection map which projects a covariant tensor $f \in \otimes^l V^*$ to its alternating part. The symmetric group S_l, which we studied in Example 2.6, reveals to be useful in this context. We define the **action of a permutation on a tensor** $f \in \otimes^l V^*$ as the *new* tensor $\varphi_s f \in \otimes^l V^*$ that is given by

$$\varphi_s f\colon\ (v_1, \dots, v_l) \mapsto f(v_{s(1)}, \dots, v_{s(l)}).$$

We can think of f and $\varphi_s f$ as the same animals with l mouths to eat the vectors $v_1, \dots, v_l$. Only, their eating habits may differ in the way which vector is fed to which mouth depending on the element $s \in S_l$.

An **alternating l-form on V** is an alternating covariant tensor of rank l

$$\omega\colon\ \underbrace{V \times \cdots \times V}_{l} \to \mathbb{R},$$

where l is called the **degree** of the form. The **set of alternating l-forms on V**, denoted by $\bigwedge^l V^*$, is a vector subspace of $\otimes^l V^*$. For any $\eta \in \otimes^l V^*$ we define the **alternation of** η as

$$\operatorname{Alt}(\eta) \coloneqq \frac{1}{l!} \sum_{s \in S_l} \operatorname{sgn}(s)\, \varphi_s \eta. \tag{2.39}$$

By the linearity of the summation and by straightforward computation, it can be seen that the alternation (2.39) is linear. Additionally, it has the following properties.

Theorem 2.16 (Spivak 1999a, Proposition 1, p. 203)**.**

(i) If $\eta \in \otimes^l V^*$, then $\operatorname{Alt}(\eta) \in \bigwedge^l V^*$.

(ii) If $\omega \in \bigwedge^l V^*$, then $\mathrm{Alt}(\omega) = \omega$.

(iii) If $\eta \in \otimes^l V^*$, then $\mathrm{Alt}\big(\mathrm{Alt}(\eta)\big) = \mathrm{Alt}(\eta)$.

To tame these alternating animals and to clarify the use of (2.39), we have a look at the following example.

Example 2.17. Let us consider a covariant 2-tensor $\eta \in \otimes^2 V^*$ on a two-dimensional vector space V, i.e., $l = 2$ and $\dim V = 2$. Let (e^1, e^2) be a basis of V^*, then the tensor η can be written as

$$\eta = \eta_{11}\, e^1 \otimes e^1 + \eta_{12}\, e^1 \otimes e^2 + \eta_{21}\, e^2 \otimes e^1 + \eta_{22}\, e^2 \otimes e^2,$$

according to Theorem 2.15. We can calculate $\mathrm{Alt}(\eta)$ as

$$\begin{aligned}
\mathrm{Alt}(\eta) &= \eta_{11}\, \mathrm{Alt}(e^1 \otimes e^1) + \eta_{12}\, \mathrm{Alt}(e^1 \otimes e^2) \\
&\qquad + \eta_{21}\, \mathrm{Alt}(e^2 \otimes e^1) + \eta_{22}\, \mathrm{Alt}(e^2 \otimes e^2) \\
&= \eta_{11}\, \tfrac{1}{2}(e^1 \otimes e^1 - e^1 \otimes e^1) + \eta_{12}\, \tfrac{1}{2}(e^1 \otimes e^2 - e^2 \otimes e^1) \\
&\qquad + \eta_{21}\, \tfrac{1}{2}(e^2 \otimes e^1 - e^1 \otimes e^2) + \eta_{22}\, \tfrac{1}{2}(e^2 \otimes e^2 - e^2 \otimes e^1) \\
&= \tfrac{1}{2}(\eta_{12} - \eta_{21})(e^1 \otimes e^2 - e^2 \otimes e^1).
\end{aligned}$$

The alternation (2.39) and the tensor product (2.34) allow to define a product that assigns an alternating $(k+l)$-form to a pair consisting of an alternating k-form and an alternating l-form. The **wedge product** (or **exterior product**) of two alternating forms $\eta \in \bigwedge^k V^*$ and $\omega \in \bigwedge^l V^*$ is defined as

$$\eta \wedge \omega := \frac{(k+l)!}{k!\, l!}\, \mathrm{Alt}(\eta \otimes \omega) \overset{(2.39)}{=} \frac{1}{k!\, l!} \sum_{s \in S_{k+l}} \mathrm{sgn}(s)\, \varphi_s(\eta \otimes \omega), \qquad (2.40)$$

with $\eta \wedge \omega \in \bigwedge^{k+l} V^*$. The **set of alternating forms of arbitrary degree**, denoted by $\bigwedge^\star V^*$, becomes an exterior algebra (Graßmann Algebra) when endowed with the wedge product.

Example 2.18. It is easy to see that the alternation (2.39) maps covectors to themselves, such that the spaces $\otimes^1 V^*$ and $\bigwedge^1 V^*$ are identical and the terms covector and 1-form are synonyms. We consider the wedge product of the basis (co-)vectors from Example 2.17, i.e.,

$$\begin{aligned}
\left(e^1 \wedge e^2\right)(v_1, v_2) &\overset{(2.40)}{=} \left(e^1 \otimes e^2\right)(v_1, v_2) - \left(e^1 \otimes e^2\right)(v_2, v_1) \\
&= \left(e^1 \otimes e^2 - e^2 \otimes e^1\right)(v_1, v_2).
\end{aligned}$$

We see that the tensor $\mathrm{Alt}(\eta)$ from Example 2.17 is given by

$$\mathrm{Alt}(\eta) = \tfrac{1}{2}(\eta_{12} - \eta_{21})\, e^1 \wedge e^2.$$

Proposition 2.19 (Properties of the wedge product)**.** For all $\xi, \eta, \omega \in \Lambda^\star V^*$ and all real numbers $a, b \in \mathbb{R}$, the wedge product satisfies

(i) $(\xi \wedge \eta) \wedge \omega = \xi \wedge (\eta \wedge \omega)$,

(ii) $(a\xi + b\eta) \wedge \omega = a\xi \wedge \omega + b\eta \wedge \omega$,

(iii) $(\eta \wedge \omega) = (-1)^{kl} (\omega \wedge \eta)$, when $\eta \in \Lambda^k V^*$ and $\omega \in \Lambda^l V^*$,

(iv) $\omega \wedge \omega = 0$, whenever the degree of ω is odd.

The proof[18] of these properties is left to the reader. We introduce the following shorthand notation for the wedge product between basis vectors. Let $I = (i_1, \dots, i_k)$ be a multi-index and let $(e^1, \dots, e^n)$ be a basis of V^*, then we define

$$e^I = e^{(i_1,\dots,i_k)} := e^{i_1} \wedge \cdots \wedge e^{i_k} \overset{(2.40)}{=} k!\,\mathrm{Alt}\left(e^{i_1} \otimes \cdots \otimes e^{i_k}\right).$$

The following theorem states that the vector space $\Lambda^k V^*$ is spanned by the vectors e^I with increasing multi-indices I.

Theorem 2.20. Let V be a vector space with basis $(e_1, \dots, e_n)$ and V^* be its dual space with the dual basis $(e^1, \dots, e^n)$. Then, the set

$$B = \left\{ e^I = e^{i_1} \wedge \cdots \wedge e^{i_k} \;\middle|\; i_1 < i_2 < \cdots < i_k \right\}$$

is a basis of $\Lambda^k V^*$ for $k \le n$. Thus, the dimension of $\Lambda^k V^*$ is given for $k \le n$ by

$$\dim \Lambda^k V^* = \binom{n}{k} = \frac{n!}{k!(n-k)!}\,.$$

Before we prove Theorem 2.20, we define the Kronecker delta for multi-indices. The **determinant** of a real n-by-n matrix $\mathbf{A}$ is defined as

$$\det \mathbf{A} := \sum_{s \in S_n} \mathrm{sgn}(s)\, A_1^{s(1)} \cdots A_n^{s(n)} = \sum_{s \in S_n} \mathrm{sgn}(s)\, A_{s(1)}^1 \cdots A_{s(n)}^n\,.$$

With this definition, it is easy to see that

$$e^I(v_1, \dots, v_k) = \det \begin{bmatrix} v_1^{i_1} & \cdots & v_k^{i_1} \\ \vdots & \ddots & \vdots \\ v_1^{i_k} & \cdots & v_k^{i_k} \end{bmatrix},$$

18. The proof can be found in John M. Lee 2013, Proposition 14.11.

with $e^I \in \bigwedge^k V^*$ and $v_j = v^i_j\, e_i \in V$ with $j = 1, \dots, k$ and $i = 1, \dots, \dim V$. We define the Kronecker delta for multi-indices as

$$\delta^I_J := e^I(e_{j_1}, \dots, e_{j_k}) = \det \begin{bmatrix} \delta^{i_1}_{j_1} & \cdots & \delta^{i_1}_{j_k} \\ \vdots & \ddots & \vdots \\ \delta^{i_k}_{j_1} & \cdots & \delta^{i_k}_{j_k} \end{bmatrix}$$

$$= \begin{cases} \operatorname{sgn}(s) & \text{if no repeated indices in } I, J \text{ and } \exists\, s\colon s(I) = J, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Since $\bigwedge^k V^* \subset \bigotimes^k V^*$, an alternating k-form $\omega \in \bigwedge^k V^*$ can be expressed with respect to the basis $e^{i_1} \otimes \cdots \otimes e^{i_k}$ of $\bigotimes^k V^*$ as

$$\omega = \sum_I \omega_I\, e^{i_1} \otimes \cdots \otimes e^{i_k}.$$

By the second property from Theorem 2.16, we know that

$$\begin{aligned} \omega = \operatorname{Alt}(\omega) &= \operatorname{Alt}\Big(\sum_I \omega_I\, e^{i_1} \otimes \cdots \otimes e^{i_k} \Big) \\ &= \sum_I \omega_I \operatorname{Alt}\big(e^{i_1} \otimes \cdots \otimes e^{i_k}\big) \\ &= \frac{1}{k!} \sum_I \omega_I\, e^I. \end{aligned}$$

At this point, it is reasonable to introduce a summation over increasing multi-indices. If we write a summation symbol that is decorated with an arrow $\nearrow$, then there is only a summation over multi-indices $I = (i_1, \dots, i_k)$ for which holds $i_1 < \cdots < i_k$. This summation allows us to write

$$\begin{aligned} \omega = \operatorname{Alt}(\omega) = \frac{1}{k!} \sum_I \omega_I\, e^I &= \frac{1}{k!} \sum_I^{\nearrow} \sum_{s \in S_k} \omega_{s(I)}\, e^{s(I)} \\ &= \sum_I^{\nearrow} \underbrace{\frac{1}{k!} \sum_{s \in S_k} \omega_{s(I)} \operatorname{sgn}(s)}_{=:\, \tilde{\omega}_I}\, e^I. \end{aligned}$$

Note that $e^I = 0$ if the multi-index I contains repeated indices (see Example 2.21). Moreover, it holds that $e^{s(I)} = \operatorname{sgn}(s)\, e^I$.

To prove linear independence of the e^I with increasing multi-indices I, we consider that

$$\sum_I^{\nearrow} \omega_I \, e^I \stackrel{!}{=} 0 \quad \Leftrightarrow \quad \sum_I^{\nearrow} \omega_I \, e^I(v_1, \dots, v_k) \stackrel{!}{=} 0, \;\; \forall v_j \in V. \tag{2.41}$$

Equation (2.41) implies that

$$\sum_I^{\nearrow} \omega_I \; \underbrace{e^I(e_{j_1}, \dots, e_{j_k})}_{\delta^I_J} = \alpha_J \stackrel{!}{=} 0,$$

for all $J = (j_1, \dots, j_k)$ with $j_1 < \dots < j_k$. This proves the linear independence of the vectors e^I with increasing multi-indices I.

Since the space $\bigwedge^k V^*$ is spanned by the vectors e^I with increasing multi-indices I, the computation of its dimension boils down to the combinatorial question: How many different ways do exist of picking k different indices from a index set of n elements (without putting back and ignoring the order)? As is well-known from elementary combinatorics, the answer is given by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

such that $\dim \bigwedge^k V^* = \binom{n}{k}$. □

Example 2.21. To complete the section, let us have a look at the space of alternating second-order tensors $\bigwedge^2 V^* \subset \bigotimes^2 V^*$ on a 3-dimensional vector space V (i.e., $k = 2$ and $n = 3$). We start with a covariant 2-tensor $\omega \in \bigotimes^2 V^*$ that is given by

$$\begin{aligned}\omega = \omega_{11}\, e^1 \otimes e^1 + \omega_{12}\, e^1 \otimes e^2 + \omega_{21}\, e^2 \otimes e^1 + \omega_{22}\, e^2 \otimes e^2 + \omega_{13}\, e^1 \otimes e^3 \\ + \omega_{31}\, e^3 \otimes e^1 + \omega_{23}\, e^2 \otimes e^3 + \omega_{32}\, e^3 \otimes e^2 + \omega_{33}\, e^3 \otimes e^3.\end{aligned}$$

Considering that $\mathrm{Alt}(e^i \otimes e^j) = \frac{1}{2!} e^{ij}$, we can write

$$\begin{aligned}\mathrm{Alt}(\omega) = \tfrac{1}{2}\big(\omega_{11}\, e^{11} + \omega_{12}\, e^{12} + \omega_{21}\, e^{21} + \omega_{22}\, e^{22} + \omega_{13}\, e^{13} \\ + \omega_{31}\, e^{31} + \omega_{23}\, e^{23} + \omega_{32}\, e^{32} + \omega_{33}\, e^{33}\big).\end{aligned}$$

Taking into account that $e^{11} = e^{22} = e^{33} = 0$ (see Example 2.17), we write

$$\begin{aligned}
\mathrm{Alt}(\omega) &= \tfrac{1}{2} \sum_I \omega_I \, e^I \quad \text{with } I \in \{12, 21, 13, 31, 23, 32\} \\
&= \underbrace{\tfrac{1}{2}(\omega_{12} - \omega_{21})}_{\tilde{\omega}_{12}} \, e^{12} + \underbrace{\tfrac{1}{2}(\omega_{13} - \omega_{31})}_{\tilde{\omega}_{13}} \, e^{13} + \underbrace{\tfrac{1}{2}(\omega_{23} - \omega_{32})}_{\tilde{\omega}_{23}} \, e^{23} \\
&= \sum_I^{\nearrow} \underbrace{\sum_{s \in S_2} \tfrac{1}{2!} \, \omega_{s(I)} \, \mathrm{sgn}(s)}_{\tilde{\omega}_I} \, e^I \\
&= \sum_J \underbrace{\sum_{s \in S_2} \tfrac{1}{2!} \, \omega_{s(J)} \, \mathrm{sgn}(s)}_{\tilde{\omega}_J} \, e^J \quad \text{with } J \in \{12, 13, 23\}.
\end{aligned}$$

Differential geometry 3

[...] no one denies that modern definitions are clear, elegant, and precise; it's just that it's impossible to comprehend how any one ever thought of them.
— Michael Spivak

This chapter deals with differential geometry. It presents the mathematical foundation of the next chapter that deals with finite-dimensional mechanical systems. For the topological concepts, we refer to Part I in Munkres 2000. As general references for differential geometry, we recommend the books: Jeffrey M. Lee 2009, John M. Lee 2013, and Spivak 1999a. The more elaborate geometric concepts that we introduce are referenced at the place of appearance.

3.1. Differentiable manifolds

Definition 3.1. A **topology** on a set M is a collection[1] $\mathcal{T}$ of subsets of M having the following properties:

T 1. $\emptyset$ and M are in $\mathcal{T}$.

T 2. The union of the elements of any subcollection of $\mathcal{T}$ is in $\mathcal{T}$.

T 3. The intersection of the elements of any finite subcollection of $\mathcal{T}$ is in $\mathcal{T}$.

A set M for which a topology has been specified is called a **topological space**.

If M is a topological space with topology $\mathcal{T}$, then a subset U is said to be **open** whenever $U \in \mathcal{T}$ and it is said to be **closed** if $M \setminus U$ is open.[2] Therefore, the choice of a topology on a set M specifies which subsets of M are open and which are closed sets.

1. A set whose elements are sets is referred to as a **collection**. Here $\mathcal{T}$ is just a subset of $\mathcal{P}(M)$ (the power set of M, i.e., the set of all subsets of M) that satisfies axioms T 1–T 3.
2. $M \setminus U$ denotes the set difference, i.e., the set $\{x \in M \mid x \notin U\}$.

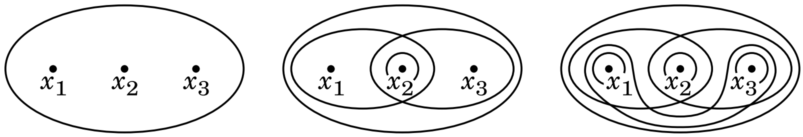


Figure 3.1.: Three different topologies for a set of three elements. The trivial topology is shown on the left. The right diagram depicts the discrete topology, which has the maximal number of elements. The middle diagram visualizes a topology which is finer than the trivial topology but coarser than the discrete topology.

Example 3.2. Depending on their number of elements, topologies can be finer or coarser. The coarsest possible topology on a set M is known as the **trivial topology** of M. It consists of the two elements $\emptyset$ and M, which are required by property T 1. The finest possible topology on M, known as the **discrete topology**, is given by the powerset $\mathcal{P}(M)$, i.e., by the set of all subsets of M. The check that both topologies satisfy axioms T 1–T 3 is left to the reader. These topologies are represented schematically for a set of three elements by the left and right diagram in Figure 3.1.

Let M be a topological space with topology $\mathcal{T}$. If V is a subset of M, the collection

$$\mathcal{T}_V := \left\{ V \cap U \mid U \in \mathcal{T} \right\}$$

is a topology[3] on V, called the **subspace topology**.

An important class of topological spaces is given by the metric spaces. A **metric space** (M, d) is a set M together with a **distance function**

$$d \colon M \times M \to \mathbb{R}$$

having the following properties:

D 1. $d(x, y) \geq 0$ for all $x, y \in M$; equality holds if and only if $x = y$.

D 2. $d(x, y) = d(y, x)$ for all $x, y \in M$.

D 3. $d(x, y) + d(y, z) \geq d(x, z)$ for all $x, y, z \in M$.

Let $B(x, r)$ be the ball around x with radius $r > 0$ defined as subset of M by

$$B(x, r) := \left\{ y \in M \mid d(x, y) < r \right\}.$$

3. The proof of this statement can be found in Munkres 2000, §16, p. 89.

The open sets, i.e., the elements of the topology, are then defined as follows. A subset U of M is called *open* if and only if

$$\forall x \in U,\ \exists\, r > 0,\ \text{such that}\ B(x,r) \subset U.$$

This topology is called the **metric topology** induced by d. Since the norm on a normed vector space induces a metric, the normed vector spaces are topological spaces.

Example 3.3. The **standard topology** on $\mathbb{R}^n$ is given by the metric topology that uses the balls

$$B(\mathbf{x},r) := \left\{\mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{x}-\mathbf{y}\|_2 < r\right\}. \tag{3.1}$$

around $\mathbf{x} \in \mathbb{R}^n$ with radius $r > 0$. The distance function in (3.1) is the **Euclidean distance**

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2, \quad \text{with}\ \|\mathbf{z}\|_2 := \sqrt{\left(z^1\right)^2 + \cdots + \left(z^n\right)^2}\,,$$

for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$. Note that the norm $\|\cdot\|_2$ is just the norm that is induced by the standard inner product from Example 2.12.

Let $f\colon M \to N$ be a function between two sets M and N. If the sets are endowed with respective topologies $\mathcal{T}_M$ and $\mathcal{T}_N$, then we can define the notion of continuity for the function f. A function $f\colon M \to N$ between two topological spaces M and N is said to be **continuous** if for each open subset V of N, the set $f^{-1}(V)$ is an open subset of M. The set

$$f^{-1}(V) := \left\{m \in M \mid f(m) \in V\right\}$$

is called the **preimage** of V under f.

Example 3.4. The trivial and the discrete topology from Example 3.2 serve as illustrative examples. If the set M is equipped with the discrete topology, then all functions $f\colon M \to N$ are continuous irrespective of the topology on N. The discrete topology declares any possible subset of M to be open. Similarly, if N is endowed with the trivial topology, then all functions are continuous irrespective of the topology on M. With the trivial topology on N, only $f^{-1}(\emptyset) = \emptyset$ and $f^{-1}(N) = M$ need to be open sets of M and they are (for any topology on M) by axiom T 1. This example illustrates that the concept of continuity of a function is an interplay of the topologies on the domain and on the codomain of the function.

We have seen that functions between topological spaces are continuous *with respect to* the specific topologies of these spaces. Therefore, continuous functions may serve to relate topological spaces.

Definition 3.5. Let M and N be topological spaces and let $f\colon M \to N$ be a bijective map. If both the function f and the inverse function $f^{-1}\colon N \to M$ are continuous, then f is called a **homeomorphism**.

In Chapter 2, we introduced isomorphisms between groups and isomorphisms between vector spaces (see Sections 2.1 and 2.3, respectively). We have seen that an isomorphism is a bijective correspondence between two algebraic objects and that it is compatible with the algebraic structure of the objects that it relates. In topology, a homeomorphism is the analogue; it is a bijective correspondence which is compatible with the topologies on the spaces it relates.

Let x be a point in a topological space M. A **neighbourhood of** x is an open subset of M containing x. A topological space M is said to be **separated** if for each pair x_1, x_2 of distinct points of M, there exist disjoint neighbourhoods U_1 and U_2 of x_1 and x_2, respectively. Separated topological spaces are also known as **Hausdorff spaces**. Among the topologies depicted in Figure 3.1, only the discrete topology shown on the right yields a separated topological space.

A collection $\mathcal{A}$ of subsets of a space M is said to **cover** M, or to be a **covering** of M, if the union of the elements of $\mathcal{A}$ is equal to M. It is called an **open covering** of M if its elements are open subsets of M. A space M is said to be **compact** if every open covering $\mathcal{A}$ of M contains a finite subcollection that also covers M.

Let M be a topological space. A collection $\mathcal{A}$ of subsets of M is said to be **locally finite** in M if every point of M has a neighbourhood that intersects only finitely many elements of $\mathcal{A}$.

Let $\mathcal{A}$ be a collection of subsets of the space M. A collection $\mathcal{B}$ of subsets of M is said to be a **refinement** of $\mathcal{A}$ (or is said to **refine** $\mathcal{A}$) if for each element B of $\mathcal{B}$, there is an element A of $\mathcal{A}$ containing B. If the elements of $\mathcal{B}$ are open sets, we call $\mathcal{B}$ an **open refinement** of $\mathcal{A}$. A space M is **paracompact** if every open covering $\mathcal{A}$ of M has a locally finite open refinement $\mathcal{B}$ that covers M.

Definition 3.6. An **n-dimensional topological manifold** M is a paracompact separated topological space M such that every point p of M has a neighbourhood that is homeomorphic to an open subset of $\mathbb{R}^n$. The local homeomorphisms $\phi_\alpha\colon M \supseteq U \to V \subseteq \mathbb{R}^n$ are called **charts**. A set $\mathcal{A}$ of charts (U_α, ϕ_α) that cover M is called **atlas**.

If we are given two charts (U_α, ϕ_α) and (U_β, ϕ_β) with $U_\alpha \cap U_\beta \neq \emptyset$, then the **chart transition map** $\phi_{\alpha\beta}$ is defined as

$$\phi_{\alpha\beta} := \phi_\beta \circ \phi_\alpha^{-1}\colon \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta). \tag{3.2}$$

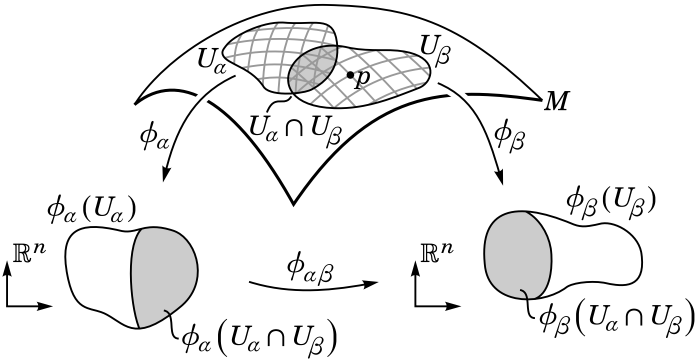


Figure 3.2.: Illustration of the chart transition map between the charts (U_α, ϕ_α) and (U_β, ϕ_β) of a topological manifold M.

Chart transition maps are homeomorphisms between open subsets of $\mathbb{R}^n$. The inverse function of $\phi_{\alpha\beta}$ is

$$\phi_{\beta\alpha} = \phi_{\alpha\beta}^{-1} = \phi_\alpha \circ \phi_\beta^{-1}.$$

For the author it has revealed beneficial to draw sketches such as Figure 3.2 to memorize the concepts.

Let M be an m-dimensional and N be an n-dimensional topological manifold. Because M and N are topological spaces, we know the concept of continuity for a map $f\colon M \to N$. If we are given respective charts (U, ϕ) and (V, ψ) of M and N, then we can consider the **chart representation**

$$\mathbf{f} := \psi \circ f \circ \phi^{-1}\colon \mathbb{R}^m \supseteq \phi\big(f^{-1}(V) \cap U\big) \to \psi(V) \subseteq \mathbb{R}^n \tag{3.3}$$

of the map f (see Figure 3.3). Because the charts ϕ and ψ are homeomorphisms, the topologies on the spaces M and $\mathbb{R}^m$ respectively on N and $\mathbb{R}^n$ are compatible. Therefore, the function $f\colon M \to N$ is continuous if and only if its chart representation (3.3) is.

Two particular types of functions $f\colon M \to N$, which will become important in what follows, are curves ($M = I \subseteq \mathbb{R}$) and real-valued functions ($N = \mathbb{R}$). Figure 3.4 shows a visualization thereof. A **curve** γ on N is a map from a subset I of the real numbers to the manifold N, that is

$$\gamma\colon \mathbb{R} \supseteq I \to N,\ \tau \mapsto \gamma(\tau). \tag{3.4}$$

A **(real-valued) function** f on the manifold M is a map

$$f\colon M \to \mathbb{R}. \tag{3.5}$$

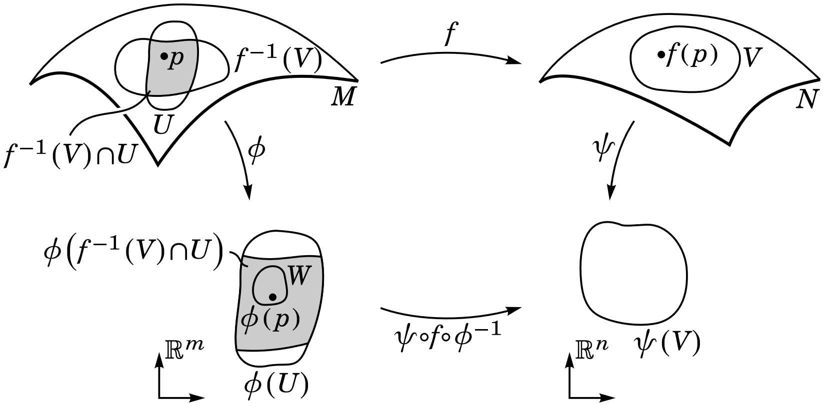


Figure 3.3.: A map f between two topological manifolds M and N and its chart representation $\psi \circ f \circ \phi^{-1}$.

The chart representation (3.3) of the curve (3.4) with respect to the chart (V, ψ) is given by

$$\mathbf{x} := \psi \circ \gamma \colon I \cap \gamma^{-1}(V) \to \psi(V) \subseteq \mathbb{R}^n.$$

The chart representation (3.3) of the function (3.5) with respect to the chart (U, ϕ) reads

$$f \circ \phi^{-1} \colon \mathbb{R}^m \supseteq \phi(U) \to \mathbb{R}. \tag{3.6}$$

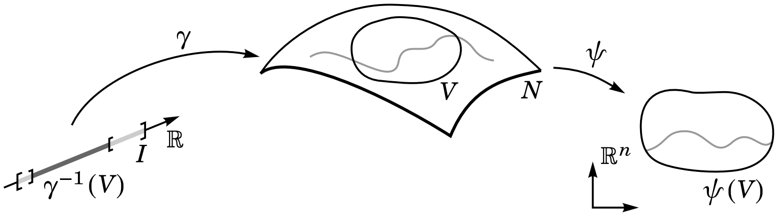
The next step is to endow the topological manifolds M and N with a differentiable structure that allows to study the differentiability of functions $f \colon M \to N$.

Definition 3.7. A bijective map $f \colon \mathbb{R}^n \supseteq U \to V \subseteq \mathbb{R}^n$ between open subsets U and V of $\mathbb{R}^n$ is said to be a **diffeomorphism** if both the function f and its inverse function f^{-1} are infinitely differentiable.

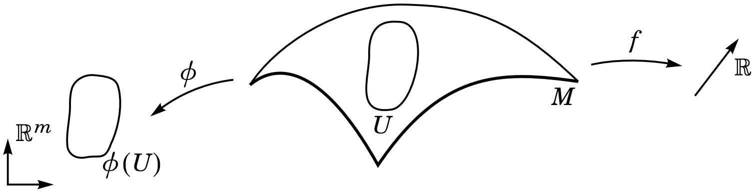
An atlas of a topological manifold M for which all chart transition maps[4] are diffeomorphisms is said to be a **smooth atlas**. Two smooth atlases are said to be **equivalent** if their union is again a smooth atlas.

Definition 3.8. An n-**dimensional differentiable manifold** is a topological manifold of dimension n together with an equivalence class of smooth atlases.

4. See equation (3.2) for the definition.



(a) A curve $\gamma\colon \mathbb{R} \supseteq I \to N$ on a topological manifold N and its chart representation $\psi \circ \gamma\colon I \cap \gamma^{-1}(V) \to \mathbb{R}^n$.



(b) A function $f\colon M \to \mathbb{R}$ on a topological manifold M and its chart representation $f \circ \phi^{-1}\colon \mathbb{R}^m \supseteq \phi(U) \to \mathbb{R}$.

Figure 3.4.: Curve and function on a differentiable manifold.

An equivalence class of smooth atlases is also called a **differentiable structure**. Differentiable manifolds are also referred to as **smooth manifolds**.

Example 3.9. Every open subset $U \subseteq \mathbb{R}^n$ is an n-dimensional differentiable manifold. A smooth atlas is given by

$$\mathcal{A} = \left\{(U, \mathrm{id}_{\mathbb{R}^n})\right\}.$$

Example 3.10. The n-sphere

$$S^n := \left\{\mathbf{x} \in \mathbb{R}^{n+1} \mid \langle \mathbf{x}, \mathbf{x} \rangle = 1\right\}$$

with two different stereographic projections[5] as charts is an n-dimensional differentiable manifold. Note that $\langle \cdot, \cdot \rangle$ denotes the standard inner product of $\mathbb{R}^{n+1}$ from Example 2.12.

Now, we are ready to study differentiable maps. Let M and N be differentiable manifolds of the respective dimensions m and n. A map $f\colon M \to N$ is said to be **differentiable** (or **smooth**) if for each $p \in M$ there exist charts

5. See John M. Lee 2013, Problem 1-7, p. 30.

(U, ϕ) around p and (V, ψ) around $f(p)$ such that the chart representation of f given by

$$\psi \circ f \circ \phi^{-1} \colon W \to \psi(V),$$

which is defined on a neighbourhood $W \subseteq \phi\big(f^{-1}(V) \cap U\big)$ around $\phi(p)$, is infinitely differentiable (see Figure 3.3). We denote the set of differentiable maps $f \colon M \to N$ by $C^\infty(M; N)$. In the special case of smooth functions $f \colon M \to \mathbb{R}$, we will use the abbreviation $C^\infty(M)$ instead of writing $C^\infty(M; \mathbb{R})$.

Definition 3.11. A **diffeomorphism** $f \colon M \to N$ is a homeomorphism for which the bijective function f and its inverse function f^{-1} are both differentiable. In this case, the differentiable manifolds M and N are said to be **diffeomorphic**.

3.2. Tangent and cotangent space

When it comes to the concept of tangent vectors to a differentiable manifold, several definitions exist (see for example Chapter 3 in John M. Lee 2013). The author thinks that the most intuitive approach is to relate tangent vectors with curves. People aware of elementary physics would not feel any discomfort in saying that the velocity of a particle moving in space can be represented by a vector that is tangent to the time-parametrized curve describing the motion of the particle. Therefore, we will start by introducing the tangent vectors as equivalence classes of curves. However, it is the algebraic definition of tangent vectors as derivations on smooth functions that reveals to be the work horse in doing computations such that we cannot pass over this interpretation.

Tangent vectors as equivalence classes of curves

Two curves[6] $\gamma_1, \gamma_2 \colon]-r, r[\to M$ with $\gamma_1(0) = \gamma_2(0) = p \in M$ and $r > 0$ are said to be equivalent,

$$\gamma_1 \sim_p \gamma_2, \tag{3.7}$$

if it holds for some chart (U, ϕ) around p that

$$\left.\frac{\mathrm{d}}{\mathrm{d}\tau}\right|_{\tau=0} (\phi \circ \gamma_1)(\tau) = \left.\frac{\mathrm{d}}{\mathrm{d}\tau}\right|_{\tau=0} (\phi \circ \gamma_2)(\tau). \tag{3.8}$$

We can easily see that the equivalence criterion (3.8) is chart independent, i.e., if it holds for one chart, then it holds for every chart around p. To see

6. See equation (3.4).

this, we consider a second chart (V, ψ) around p. Then, for any $p \in U \cap V$, it holds that $\phi \circ \gamma_i = (\phi \circ \psi^{-1}) \circ \psi \circ \gamma_i$ with $i = 1, 2$ and therefore

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\tau}\bigg|_{\tau=0} (\phi \circ \gamma_i)(\tau) &= \frac{\partial \phi \circ \psi^{-1}}{\partial x^j}\bigg|_{\psi(p)} \frac{\mathrm{d}}{\mathrm{d}\tau}\bigg|_{\tau=0} (\psi^j \circ \gamma_i)(\tau) \\ &= \frac{\partial \phi \circ \psi^{-1}}{\partial \mathbf{x}}\bigg|_{\psi(p)} \frac{\mathrm{d}}{\mathrm{d}\tau}\bigg|_{\tau=0} (\psi \circ \gamma_i)(\tau), \end{aligned} \tag{3.9}$$

where $\psi^j : V \to \mathbb{R}$ denotes the j-th coordinate function of the chart

$$\psi : M \supseteq V \to \mathbb{R}^n.$$

The last equality in (3.9) follows by using matrix notation (see Appendix A). Since the chart transition map $\phi \circ \psi^{-1}$ is a diffeomorphism, the linear map given by the matrix

$$\frac{\partial \phi \circ \psi^{-1}}{\partial \mathbf{x}}\bigg|_{\psi(p)}$$

is a vector space isomorphism of $\mathbb{R}^n$. Therefore, it is clear that if the equality (3.8) holds in one chart of an atlas $\mathcal{A}$, then it holds in every chart of $\mathcal{A}$ that has p in its domain.

Definition 3.12. Let M be a differentiable manifold. **Tangent vectors** to M at a point $p \in M$ are equivalence classes with respect to $\sim_p$ of curves in M through p. The **tangent space of M in p** is the set of all these equivalence classes

$$\mathsf{T}_p M := \left\{\gamma : \,]-r, r[\, \to M \text{ differentiable}, \ \gamma(0) = p\right\} \big/ \sim_p .$$

As introduced in Section 2.4, we use $[\gamma]$ to denote the equivalence class of a curve $\gamma : \,]-r, r[\, \to M$ with respect to (3.7). So far, the tangent space $\mathsf{T}_p M$ is just a set. But it can be endowed with the structure of a real vector space of the same dimension than the manifold M.

Theorem 3.13. Let M be an n-dimensional differentiable manifold. Then the tangent space $\mathsf{T}_p M$ at some point $p \in M$ is a real n-dimensional vector space.

Proof. Let (U, ϕ) be a chart around p, then we define the map

$$\mathrm{d}\phi_p : \mathsf{T}_p M \to \mathbb{R}^n, \ [\gamma] \mapsto \frac{\mathrm{d}}{\mathrm{d}\tau}\bigg|_{\tau=0} (\phi \circ \gamma)(\tau). \tag{3.10}$$

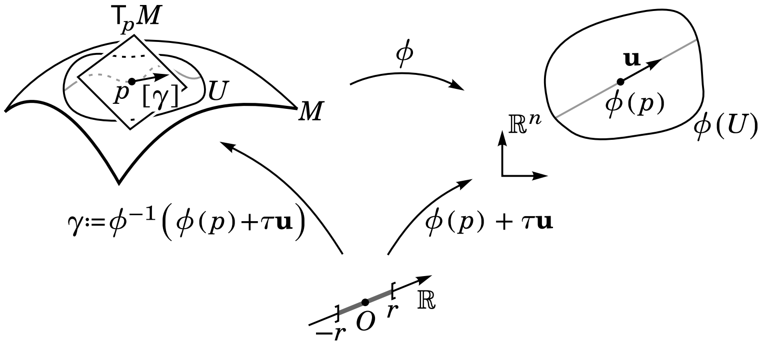


Figure 3.5.: Construction of curves through the point $p \in M$.

The idea is to prove that $\mathrm{d}\phi_p$ is a bijective map, which then can be used to import the vector space structure[7] from $\mathbb{R}^n$ to the tangent space.

The map (3.10) is injective by the definition of the equivalence relation. Indeed, the function $\mathrm{d}\phi_p$ maps elements of T_pM to the real numbers that have been used to divide the curves into equivalence classes. To prove surjectivity, we define the curve

$$\gamma(\tau) := \phi^{-1}\big(\phi(p) + \tau\mathbf{u}\big) \tag{3.11}$$

for arbitrary n-tuples $\mathbf{u} \in \mathbb{R}^n$ (see Figure 3.5). For the curves (3.11), we can calculate

$$\begin{aligned} \mathrm{d}\phi_p([\gamma]) &= \frac{\mathrm{d}}{\mathrm{d}\tau}\bigg|_{\tau=0} \phi \circ \phi^{-1}\big(\phi(p) + \tau\mathbf{u}\big) \\ &= \frac{\mathrm{d}}{\mathrm{d}\tau}\bigg|_{\tau=0} \big(\phi(p) + \tau\mathbf{u}\big) = \mathbf{u}, \end{aligned}$$

which proves the surjectivity of $\mathrm{d}\phi_p$ because we have shown that the whole $\mathbb{R}^n$ can be reached. The addition and scalar multiplication that endow T_pM with the structure of a vector space are defined as

$$\begin{aligned} [\gamma_1] + [\gamma_2] &:= \mathrm{d}\phi_p^{-1}\big(\mathrm{d}\phi_p([\gamma_1]) + \mathrm{d}\phi_p([\gamma_2])\big), \\ \alpha[\gamma_1] &:= \mathrm{d}\phi_p^{-1}\big(\alpha\mathrm{d}\phi_p([\gamma_1])\big), \end{aligned} \tag{3.12}$$

for all $[\gamma_1], [\gamma_2] \in \mathsf{T}_pM$ and all $\alpha \in \mathbb{R}$. Note that on the right-hand side of these definitions stands the addition (2.19) and scalar multiplication (2.20) of $\mathbb{R}^n$. □

7. See equations (2.19) and (2.20).

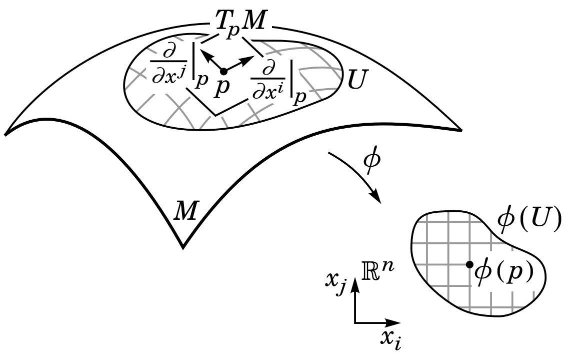


Figure 3.6.: Basis vectors of the tangent space T_pM that are induced by the chart (U, ϕ) of the differentiable manifold M.

Tangent vectors as derivations

The algebraic approach to tangent vectors of a differentiable manifold M is to consider them as operators on the set of smooth functions $C^\infty(M)$ on M. A linear map $u_p : C^\infty(M) \to \mathbb{R}$ is said to be a **derivation in** $p \in M$ if it has the two properties:

D 1. Locality: For all open subsets $U \subseteq V$ of M with $p \in U$ and functions $f \in C^\infty(V)$

$$u_p[f] = u_p[f|_U],$$

where $f|_U$ denotes the restriction of the function $f\colon V \to \mathbb{R}$ to the neighbourhood $U \subseteq V$.

D 2. Product rule: Let $f, g \in C^\infty(M), p \in M$, then

$$u_p[fg] = f(p)u_p[g] + g(p)u_p[f],$$

where $fg\colon M \to \mathbb{R}, p \mapsto f(p)g(p)$.

The concept of a derivation in p is an abstraction of the derivative operator. Property D 1 guarantees that the derivation is a local operator similar to the derivative. From the locality property, it follows that

$$f|_U = g|_U \quad \Rightarrow \quad u_p[f] = u_p[f|_U] = u_p[g|_U] = u_p[g].$$

Note that we used square brackets in D 1 and D 2 to denote the function argument of the map $u_p : C^\infty(M) \to \mathbb{R}$. This notation should not be confounded with the one used for equivalence classes.

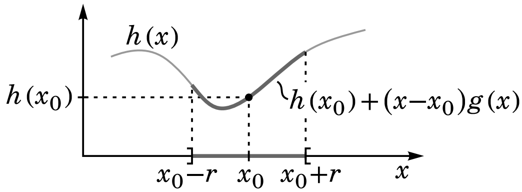


Figure 3.7.: Visualization of Lemma 3.16 for the scalar case.

Definition 3.14. Let M be a differentiable manifold. The set T_pM of derivations in $p \in M$ is said to be the **tangent space at the point** p. A derivation $u_p \in T_pM$ is called **tangent vector at the point** p.

We will show in Theorem 3.17 that the sets T_pM and T_pM from Definitions 3.12 and 3.14 are isomorphic vector spaces. Therefore it make sense to use the same terminology for these spaces.

Theorem 3.15. Let M be an n-dimensional differentiable manifold and let (U, ϕ) be a chart of M such that $\phi\colon p \mapsto \phi(p) = (x^1, \dots, x^n)$. The tangent space T_pM in $p \in U \subseteq M$ is an n-dimensional real vector space. The derivations

$$\left.\frac{\partial}{\partial x^i}\right|_p [f] := \left.\frac{\partial}{\partial x^i}\right|_{\phi(p)} (f \circ \phi^{-1}), \tag{3.13}$$

with $f \in C^\infty(M)$ form a basis of T_pM such that a tangent vector $u_p \in T_pM$ can be written as

$$u_p = u_p[\phi^i] \left.\frac{\partial}{\partial x^i}\right|_p . \tag{3.14}$$

Theorem 3.15 is depicted in Figure 3.6. Its proof relies on the following lemma, which is visualized for the scalar case in Figure 3.7.

Lemma 3.16. Let $h \in C^\infty(\mathbb{R}^n)$, then there exist functions $g_1, \dots, g_n \in C^\infty(V)$ with[8] $V = B(\mathbf{x}_0, r) \subset \mathbb{R}^n$ such that the function $h\colon \mathbf{x} \mapsto h(\mathbf{x})$ can be locally written as

$$h(\mathbf{x}) = h(\mathbf{x}_0) + (x^i - x_0^i)\, g_i(\mathbf{x}) \quad \forall \mathbf{x} \in V,$$

with

$$g_i(\mathbf{x}_0) = \frac{\partial h}{\partial x^i}(\mathbf{x}_0). \tag{3.15}$$

8. $B(\mathbf{x}_0, r)$ denotes the open ball with radius $r > 0$ centered at the point $\mathbf{x}_0$ in $\mathbb{R}^n$ as defined in (3.1).

Proof of Lemma 3.16. The lemma follows from the fundamental theorem of calculus, by which we can write

$$\begin{aligned} h(\mathbf{x}) - h(\mathbf{x}_0) &= h\big(t\mathbf{x} + (1-t)\mathbf{x}_0\big)\Big|_{t=0}^{t=1} \\ &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} h\big(t\mathbf{x} + (1-t)\mathbf{x}_0\big)\,\mathrm{d}t \\ &= \int_0^1 \frac{\partial h}{\partial x^i}\big(t\mathbf{x} + (1-t)\mathbf{x}_0\big)(x^i - x^0)\,\mathrm{d}t \\ &= (x^i - x^0) \underbrace{\int_0^1 \frac{\partial h}{\partial x^i}\big(t\mathbf{x} + (1-t)\mathbf{x}_0\big)\mathrm{d}t}_{=:g_i(\mathbf{x})}. \end{aligned}$$

Finally, we can convince ourselves that with this definition of $g_i(\mathbf{x})$

$$g_i(\mathbf{x}_0) = \int_0^1 \frac{\partial h}{\partial x^i}(\mathbf{x}_0)\,\mathrm{d}t = \frac{\partial h}{\partial x^i}(\mathbf{x}_0).$$

□

Proof of Theorem 3.15. We start with a list of what needs to be shown:

1. The objects $\partial/\partial x^i|_p$ are derivations.
2. The derivations $\partial/\partial x^i|_p$ are linearly independent.
3. A tangent vector $u_p \in T_pM$ can be written as

$$u_p = u_p[\phi^i]\,\frac{\partial}{\partial x^i}\bigg|_p.$$

To prove 1, we show that $\partial/\partial x^i|_p$ are linear real-valued maps on $C^\infty(M)$ and that they have properties D 1 (locality) and D 2 (product rule). Let $\alpha, \beta \in \mathbb{R}$ and $f, g \in C^\infty(M)$. To show linearity, we consider

$$\begin{aligned} \frac{\partial}{\partial x^i}\bigg|_p [\alpha f + \beta g] &= \frac{\partial}{\partial x^i}\bigg|_{\phi(p)} \Big(\big(\alpha f + \beta g\big)\circ\phi^{-1}\Big) \\ &= \alpha\frac{\partial}{\partial x^i}\bigg|_{\phi(p)}\big(f\circ\phi^{-1}\big) + \beta\frac{\partial}{\partial x^i}\bigg|_{\phi(p)}\big(g\circ\phi^{-1}\big) \\ &= \alpha\frac{\partial}{\partial x^i}\bigg|_p [f] + \beta\frac{\partial}{\partial x^i}\bigg|_p [g], \end{aligned}$$

which proves linearity. Property D 2 is induced by the product rule of partial differentiation on $\mathbb{R}^n$ as follows

$$\begin{aligned}
\frac{\partial}{\partial x^i}\bigg|_p [fg] &= \frac{\partial}{\partial x^i}\bigg|_{\phi(p)} \big((fg)\circ\phi^{-1}\big) \\
&= \frac{\partial}{\partial x^i}\bigg|_{\phi(p)} \Big(\big(f\circ\phi^{-1}\big)\big(g\circ\phi^{-1}\big)\Big) \\
&= g(p)\frac{\partial}{\partial x^i}\bigg|_{\phi(p)} \big(f\circ\phi^{-1}\big) + f(p)\frac{\partial}{\partial x^i}\bigg|_{\phi(p)} \big(g\circ\phi^{-1}\big) \\
&= g(p)\frac{\partial}{\partial x^i}\bigg|_p [f] + f(p)\frac{\partial}{\partial x^i}\bigg|_p [g].
\end{aligned}$$

Hence $\partial/\partial x^i|_p$ is a derivation because property D 1 is induced by the locality of the partial derivative on $\mathbb{R}^n$.

To prove 2, we need to check that for all $\alpha^i \in \mathbb{R}$ with $i = 1, \dots, n$ the claim

$$\alpha^i \frac{\partial}{\partial x^i}\bigg|_p \overset{!}{=} 0 \quad \Leftrightarrow \quad \alpha^i \frac{\partial}{\partial x^i}\bigg|_p [f] \overset{!}{=} 0 \quad \forall f \in C^\infty(M) \tag{3.16}$$

implies $\alpha^i = 0$. Equation (3.16) needs to hold in particular for $f = \phi^j$ with $j = 1, \dots, n$, that is

$$0 \overset{!}{=} \alpha^i \frac{\partial}{\partial x^i}\bigg|_p [\phi^j] = \alpha^i \frac{\partial}{\partial x^i}\bigg|_{\phi(p)} \big(\phi^j \circ \phi^{-1}\big) = \alpha^i \delta_i^j = \alpha^j$$

for $j = 1, \dots, n$. Therefore, the $\partial/\partial x^i|_p$ are linearly independent.

To prove 3, we start by showing that the derivation of a constant function vanishes. Let $u_p \in T_pM$ and $c_\alpha : M \to \mathbb{R}, p \mapsto \alpha$, then we can calculate

$$u_p[c_1] = u_p[c_1 c_1] = 1\, u_p[c_1] + u_p[c_1]\, 1 = 2\, u_p[c_1] \quad \Rightarrow \quad u_p[c_1] = 0$$

and

$$u_p[c_\alpha] = u_p[\alpha\, c_1] = \alpha\, u_p[c_1] = 0. \tag{3.17}$$

Let (U, ϕ) be the given chart and let $f \in C^\infty(M)$ be a smooth function on M (see Figure 3.8). The set $\phi(U)$ is open in $\mathbb{R}^n$ endowed with the standard topology because the map $\phi\colon U \to \phi(U)$ is a homeomorphism (see Definition 3.5), i.e., it maps open sets to open sets. By the definition of the metric topology (see p. 47), there exists an open set around $\phi(p)$, a ball $V = B(\phi(p), r) \subseteq \phi(U)$. We consider the chart representation of the function $f\colon M \to \mathbb{R}$ (see equation (3.6)), that is

$$h := f \circ \phi^{-1} \colon \mathbb{R}^n \supseteq \phi(U) \to \mathbb{R},\ \mathbf{x} \mapsto h(\mathbf{x}).$$

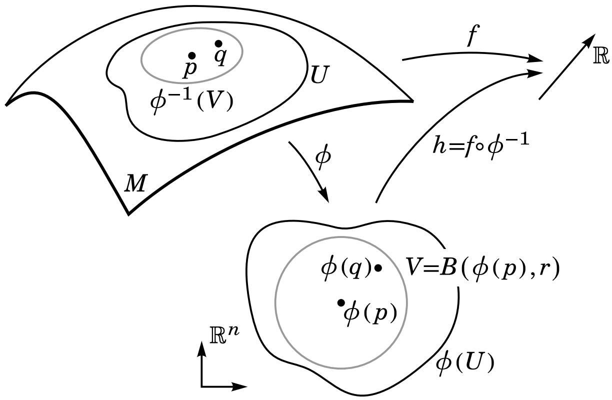


Figure 3.8.: Representation of the sets used in the proof of Theorem 3.15.

According to Lemma 3.16, the function $\mathbf{x} \mapsto h(\mathbf{x})$ can be written on V as

$$h(\mathbf{x}) = h(\phi(p)) + (x^i - \phi^i(p))\, g_i(\mathbf{x}) \tag{3.18}$$

for all $\mathbf{x}_0 \in V$. Thus, f can be written on $\phi^{-1}(V)$ as

$$f(q) = h \circ \phi(q) \overset{(3.18)}{=} h(\phi(p)) + (\phi^i(q) - \phi^i(p))\, g_i(\phi(q)) \tag{3.19}$$

for all $q \in \phi^{-1}(V)$. Next, we apply a tangent vector $u_p \in T_pM$ at the point $p \in M$ to the function f as follows

$$\begin{aligned}
u_p[f] &\overset{\text{D1}}{=} u_p\left[f|_{\phi^{-1}(V)}\right] \\
&\overset{(3.19)}{=} u_p\Big[h\big(\phi(p)\big)\Big] + u_p\Big[(\phi^i(\cdot) - \phi^i(p))\, g_i(\phi(\cdot))\Big] \\
&\overset{(3.17)}{=} u_p\Big[(\phi^i(\cdot) - \phi^i(p))\, g_i(\phi(\cdot))\Big] \\
&\overset{\text{D2}}{=} g_i\big(\phi(p)\big) \cdot u_p\Big[\phi^i(\cdot) - \phi^i(p)\Big] + (\phi^i(p) - \phi^i(p)) \cdot u_p\Big[g_i(\phi(\cdot))\Big] \\
&= g_i\big(\phi(p)\big) \cdot u_p\Big[\phi^i(\cdot)\Big] \\
&\overset{(3.15)}{=} u_p\Big[\phi^i(\cdot)\Big] \frac{\partial h}{\partial x^i}\bigg|_{\phi(p)} \\
&\overset{(3.19)}{=} u_p\Big[\phi^i(\cdot)\Big] \frac{\partial}{\partial x^i}\bigg|_{\phi(p)} \left(f \circ \phi^{-1}\right) \\
&\overset{(3.13)}{=} u_p\Big[\phi^i(\cdot)\Big] \frac{\partial}{\partial x^i}\bigg|_{p} [f].
\end{aligned}$$

Hence the derivations $\partial/\partial x^i|_p$ with $i = 1, \dots, n$ generate the vector space T_pM, which therefore has dimension n. □

Let (U, ϕ) and (V, ψ) be two charts around a point $p \in M$ such that $\phi\colon p \mapsto \phi(p) = \mathbf{x}$ and $\psi\colon p \mapsto \psi(p) = \mathbf{y}$. Both charts provide a respective basis of the tangent space T_pM that is given by the induced derivations $\partial/\partial x^1|_p, \dots, \partial/\partial x^n|_p$ and $\partial/\partial y^1|_p, \dots, \partial/\partial y^n|_p$, respectively. If we are given a tangent vector $u_p \in T_pM$, then by (3.14) it can be represented with respect to both bases as

$$u_p = u_p[\phi^i] \left.\frac{\partial}{\partial x^i}\right|_p = u_p[\psi^j] \left.\frac{\partial}{\partial y^j}\right|_p. \tag{3.20}$$

Considering (3.20) for $u_p = \partial/\partial x^i\big|_p$, we find that the basis vectors are related by

$$\left.\frac{\partial}{\partial x^i}\right|_p = \left.\frac{\partial}{\partial x^i}\right|_p \left[\psi^j\right] \left.\frac{\partial}{\partial y^j}\right|_p. \tag{3.21}$$

Inserting the transformation rule (3.21) into equation (3.20), it follows that

$$u_p[\psi^j] = \left.\frac{\partial\left(\psi^j \circ \phi^{-1}\right)}{\partial x^i}\right|_{\phi(p)} u_p[\phi^i].$$

Relation between the two notions of tangent vectors

For both definitions[9] of tangent vectors, we showed that the respective tangent spaces T_pM and T_pM can be endowed with the structure of a vector space of the same dimension than the manifold M (Theorems 3.13 and 3.15). The tangent spaces T_pM and T_pM are related by a vector space isomorphism. Let $[\gamma] \in \mathsf{T}_pM$ and $f \in C^\infty(M)$. The **Lie derivative of f in the direction** $[\gamma]$ is defined as

$$L_{[\gamma]}f := \left.\frac{\mathrm{d}}{\mathrm{d}\tau}\right|_{\tau=0} f \circ \gamma(\tau). \tag{3.22}$$

The reader may easily convince himself that the Lie derivative

$$L_{[\gamma]}\colon C^\infty(M) \to \mathbb{R}$$

is a derivation and, therefore, it holds that $L_{[\gamma]} \in T_pM$ for all $[\gamma] \in \mathsf{T}_pM$.

Theorem 3.17. Let M be a differentiable manifold and let p be a point of M. The map

$$\mathsf{T}_pM \ni [\gamma] \mapsto L_{[\gamma]} \in T_pM \tag{3.23}$$

defines a vector space isomorphism between the tangent spaces T_pM and T_pM, i.e., $\mathsf{T}_pM \cong T_pM$.

9. See Definitions 3.12 and 3.14.

Proof. Consider a chart (U, ϕ) around $p \in M$. Let $[\gamma] \in \mathsf{T}_pM$ be a tangent vector. For a function $f \in C^\infty(M)$, it follows by the chain rule that

$$\begin{aligned} L_{[\gamma]}f &= \frac{\mathrm{d}}{\mathrm{d}\tau}\bigg|_{\tau=0} \left(f \circ \phi^{-1} \circ \phi \circ \gamma\right)(\tau) \\ &= \frac{\mathrm{d}}{\mathrm{d}\tau}\bigg|_{\tau=0} \left(\phi^i \circ \gamma\right)(\tau) \cdot \frac{\partial f \circ \phi^{-1}}{\partial x^i}\bigg|_{\phi(\gamma(0))}. \end{aligned}$$

The first factor of each summand is just a real number, i.e.,

$$\mathbb{R} \ni u_p^i := \frac{\mathrm{d}}{\mathrm{d}\tau}\bigg|_{\tau=0} \left(\phi^i \circ \gamma\right)(\tau).$$

Because $\gamma(0) = p$, the second factor in each term can be recognized as basis vectors (3.13) being applied to f such that

$$L_{[\gamma]}f = u_p^i \frac{\partial f \circ \phi^{-1}}{\partial x^i}\bigg|_{\phi(p)} = u_p^i \frac{\partial}{\partial x^i}\bigg|_p [f]. \tag{3.24}$$

The injectivity of the map (3.23) can be seen from equation (3.24). If $L_{[\gamma]} = 0$ for all $f \in C^\infty(M)$, then it holds in particular for $f = \phi^j$ with $j = 1, \dots, n$ from which follows that $u_p^j = 0$ for $j = 1, \dots, n$. Because the tangent spaces T_pM and T_pM have the same dimension, the surjectivity of the map (3.23) follows from its injectivity by the rank-nullity theorem of linear algebra. Therefore, the map (3.23) is bijective. The linearity of (3.23) follows from the definition (3.12) and the linearity of differentiation. □

Similar to the map (3.10), we define the map

$$\begin{aligned} &\mathrm{d}\phi_p \colon T_pM \to \mathbb{R}^n, \\ &u_p = u^i \frac{\partial}{\partial x^i}\bigg|_p \mapsto \mathrm{d}\phi_p(u_p) := \mathbf{u} = (u^1, \dots, u^n). \end{aligned} \tag{3.25}$$

The map (3.25) is bijective because its inverse can be written explicitly with $\mathbf{x} := \phi(p)$ as

$$\mathrm{d}\phi_p^{-1}(\mathbf{u}) = u^i \frac{\partial}{\partial x^i}\bigg|_{\phi^{-1}(\mathbf{x})}.$$

Theorem 3.17 tells us that the diagram

$$\begin{tikzcd} \mathsf{T}_pM \arrow[rr, "{[\gamma] \mapsto L_{[\gamma]}}"] \arrow[dr, "\mathrm{d}\varphi_p"'] & & T_pM \arrow[dl, "\mathrm{d}\phi_p"] \\ & \mathbb{R}^n & \end{tikzcd}$$

commutes, i.e., that $\mathrm{d}\varphi_p = \mathrm{d}\phi_p \circ L_{(\cdot)}$. In what follows, we will mainly focus on the algebraic view of tangent vectors as derivations.

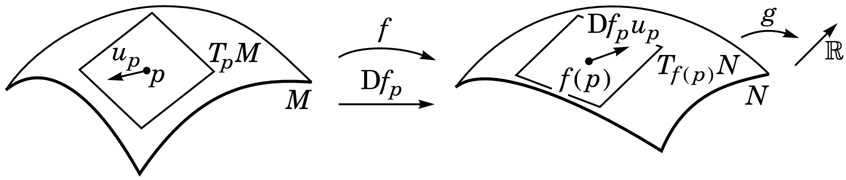


Figure 3.9.: The differential of the function $f: M \to N$ at the point p maps the tangent space T_pM at the point p of M to the tangent space $T_{f(p)}N$ at the point $f(p)$ of N.

The differential of a map between manifolds

The tangent vector to a curve is just a linear approximation of the curve. In this sense, the tangent space to a differentiable manifold at some point $p \in M$ can be seen as a linear approximation of the manifold around the point p. Now, we study the linear approximation of differentiable maps $f: M \to N$ between two differentiable manifolds M and N (see Figure 3.9). The **differential of f in $p \in M$** is the map

$$\mathrm{D}f_p : T_pM \to T_{f(p)}N,\ u_p \mapsto \mathrm{D}f_p u_p \tag{3.26}$$

that is defined by

$$\mathrm{D}f_p u_p[g] := u_p[g \circ f]$$

for all $g \in C^\infty(N)$.

Proposition 3.18 (John M. Lee 2013, Proposition 3.6)**.** Let M, N, and P be differentiable manifolds, let $f: M \to N$ and $g: N \to P$ be differentiable functions, let $p \in M$.

(i) $\mathrm{D}f_p : T_pM \to T_{f(p)}N$ is linear.

(ii) $\mathrm{D}(g \circ f)_p = \mathrm{D}g_{f(p)} \circ \mathrm{D}f_p : T_pM \to T_{g \circ f(p)}P$.

(iii) $\mathrm{D}(\mathrm{id}_M)_p = \mathrm{id}_{T_pM} : T_pM \to T_pM$.

(iv) If f is a diffeomorphism, then the linear map $\mathrm{D}f_p : T_pM \to T_{f(p)}N$ is an isomorphism, and $(\mathrm{D}f_p)^{-1} = \mathrm{D}(f^{-1})_{f(p)}$.

The cotangent space

In Section 2.6, we introduced the dual space V^* of a vector space V. If we take V to be the tangent space T_pM of a differentiable manifold M at some point $p \in M$, then we can define the **cotangent space of** M **at** p, denoted by T_p^*M, to be the dual space of the tangent space, that is

$$T_p^*M := (T_pM)^*. \tag{3.27}$$

The dual basis $(\mathrm{d}x_p^1, \dots, \mathrm{d}x_p^n)$ to the basis $(\partial/\partial x^1|_p, \dots, \partial/\partial x^n|_p)$ of the tangent space is defined by

$$\mathrm{d}x_p^i \left(\frac{\partial}{\partial x^j}\bigg|_p \right) \overset{!}{=} \delta_j^i, \tag{3.28}$$

such that a covector $\omega \in T_p^*M$ can be written as

$$\omega = \omega_i \, \mathrm{d}x_p^i.$$

3.3. Immersions, submersions and embeddings

Let $f\colon M \to N$ be a differentiable map between two differentiable manifolds M and N. The **rank of** f **at** p is defined as the rank of the differential of f in p, i.e., as the rank of the linear map $\mathrm{D}f_p : T_pM \to T_{f(p)}N$. If f has the same rank r at every point $p \in M$, we say that it has **constant rank** and we write $\operatorname{rank} f = r$. The rank of a linear map is never higher than the dimension of either its domain or codomain. The most important maps of constant rank are those of maximal rank, i.e., for which $\operatorname{rank} f = \min\{\dim M, \dim N\}$. We say that a differentiable map $f\colon M \to N$ is a **submersion** if its differential in p is surjective for each $p \in M$ (or equivalently, if $\operatorname{rank} f = \dim N$). We call it an **immersion** if its differential in p is injective for each $p \in M$ (equivalently, $\operatorname{rank} f = \dim M$). If $f\colon M \to N$ is not only an immersion but also a topological embedding, i.e., a homeomorphism onto its image $f(M) \subseteq N$ in the subspace topology (see p. 46), then it is called an **embedding of** M **into** N.

An **immersed submanifold of** N is a subset $M \subseteq N$ endowed with a topology (not necessarily the subspace topology) with respect to which it is a topological manifold, and with a differentiable structure with respect to which the inclusion map[10] $\iota\colon M \hookrightarrow N$ is an immersion. An **embedded submanifold of** N is a subset $M \subseteq N$ that is a manifold in the subspace topology, endowed with a differentiable structure with respect to which the inclusion map $\iota\colon M \hookrightarrow N$ is an embedding. If $M \subseteq N$ is a submanifold of N (immersed or embedded), then we call the difference $\dim N - \dim M$ the **codimension of** M **in** N. For a detailed treatment of submanifolds, we refer to Chapter 5 in John M. Lee 2013.

10. If A is a subset of B, then the identity map $x \mapsto x$ for all $x \in A$ viewed as a mapping $A \to B$ is called **inclusion map**. It is usually denoted by $\iota : A \hookrightarrow B, x \mapsto \iota(x) = x$ (see p. 28 of Lang 2005).

3.4. Vector bundles

In Section 3.2, we saw that the tangent space T_pM at an arbitrary point p of a differentiable manifold M is a vector space of the same dimension as the manifold M. Moreover, we introduced the cotangent space T_p^*M as the dual space of the tangent space T_pM. Now, we consider the differentiable manifold M together with all its tangent, respectively cotangent spaces. This leads us to the following two definitions.

Definition 3.19. The **tangent bundle** TM of a differentiable manifold M is the disjoint union of all tangent spaces T_pM, i.e.,

$$TM := \bigcup_{p\in M} \big(\{p\}\times T_pM\big).$$

Definition 3.20. The **cotangent bundle** T^*M of a differentiable manifold M is the disjoint union of all cotangent spaces T_p^*M, i.e.,

$$T^*M := \bigcup_{p\in M} \big(\{p\}\times T_p^*M\big).$$

A point u in the tangent bundle TM has the form

$$u = (p, u_p) \text{ with } p \in M \text{ and } u_p \in T_pM, \tag{3.29}$$

while a point σ in the cotangent bundle T^*M is given by

$$\sigma = (p, \sigma_p) \text{ with } p \in M \text{ and } \sigma_p \in T_p^*M.$$

By definition, the sets TM and T^*M come with the natural projections

$$\pi_{TM}\colon TM \to M,\ \ (p, u_p) \mapsto p. \tag{3.30}$$

and

$$\pi_{T^*M}\colon T^*M \to M,\ \ (p, \sigma_p) \mapsto p, \tag{3.31}$$

respectively. We will see that TM and T^*M can be endowed with a vector bundle structure that can be constructed from the differentiable structure of the base manifold M using the natural projections (3.30) and (3.31), respectively.

Definition 3.21. Let E, F, and M be smooth manifolds and let $\pi\colon E \to M$ be a differentiable surjective map. The quadruple (E, π, M, F) is called a **fibre bundle**[11] if for each point $p \in M$ there is an open set U containing p and a diffeomorphism $\theta\colon \pi^{-1}(U) \to U\times F$ such that the diagram

$$\begin{tikzcd} \pi^{-1}(U) \arrow[r, "\theta"] \arrow[dr, "\pi"'] & U \times F \arrow[d, "\mathrm{pr}_1"] \\ & U \end{tikzcd}$$

commutes, i.e., if $\mathrm{pr}_1 \circ \theta = \pi$. The diffeomorphism $\theta\colon \pi^{-1}(U) \to U \times F$ is called **local trivialization** and $\pi\colon E \to M$ is said to be the **bundle projection**. The manifolds E, F, and M are called **total space**, **typical fibre**, and **base manifold**, respectively. The projection $\mathrm{pr}_1\colon U \times F \to U$ is referred to as **natural projection on the first factor**. The set $E_p := \pi^{-1}(p)$ is called the **fibre over** p.

A fibre bundle is **trivial** if there exists a global trivialization such that the diagram

$$\begin{tikzcd} E \arrow[r, "\theta"] \arrow[dr, "\pi"'] & M \times F \arrow[d, "\mathrm{pr}_1"] \\ & M \end{tikzcd}$$

commutes.

Definition 3.22. A **(real) vector bundle**[11] of rank k over M is a fibre bundle $(E, \pi, M, \mathbb{R}^k)$ with typical fibre $\mathbb{R}^k$ such that:

(i) for each $p \in M$ the fibre $E_p = \pi^{-1}(p)$ over p is endowed with the structure of a k-dimensional real vector space.

(ii) for each $p \in U$, the restriction of θ to E_p

$$\theta\big|_{E_p} : E_p \to U \times \mathbb{R}^k$$

is a vector space isomorphism from E_p to $\{p\} \times \mathbb{R}^k \cong \mathbb{R}^k$.

A vector bundle $(D, \pi_D, M, \mathbb{R}^l)$ of rank l over M is said to be a **subbundle** of a vector bundle $(E, \pi_E, M, \mathbb{R}^k)$ of rank $k \geq l$ over M if D is an embedded submanifold of E and the bundle projection $\pi_D : D \to M$ is the restriction of $\pi_E : E \to M$ to D, such that for each $p \in M$, the subset $D_p = D \cap E_p$ is a vector subspace of E_p, and the vector space structure on D_p is the one inherited from E_p.

A **vector bundle homomorphism**[12] between two vector bundles $(E_1, \pi_1, M_1, \mathbb{R}^k)$ and $(E_2, \pi_2, M_2, \mathbb{R}^l)$ is a pair $(\hat{f}, f)$ of differentiable maps $\hat{f} : E_1 \to E_2$ and $f\colon M_1 \to M_2$, such that the diagram

11. See Chapter 6 in Jeffrey M. Lee 2009 for a detailed treatment of fibre and vector bundles.
12. See Definition 6.25 in Jeffrey M. Lee 2009.

$$\begin{array}{ccc} E_1 & \xrightarrow{\hat{f}} & E_2 \\ \pi_1 \downarrow & & \downarrow \pi_2 \\ M_1 & \xrightarrow{f} & M_2 \end{array}$$

commutes, i.e., $f \circ \pi_1 = \pi_2 \circ \hat{f}$ and where the restrictions to the fibres

$$\hat{f}|_{\pi_1^{-1}(p)} : \pi_1^{-1}(p) \to \pi_2^{-1}\big(f(p)\big) \tag{3.32}$$

are vector space homomorphisms, i.e., linear maps for all $p \in M_1$. If $\hat{f}$ is a diffeomorphism such that its inverse is also a vector bundle homomorphism, then it is called a **vector bundle isomorphism**.

We define the **rank of $\hat{f}$ at** p for each $p \in M_1$ as the rank of the linear map (3.32). If the rank is the same for all points $p \in M_1$, then $\hat{f}$ is said to have **constant rank**. Vector bundle homomorphisms define subbundles according to the following proposition.

Proposition 3.23 (Golubitsky et al. 1973, Proposition 5.14, p. 26)**.** Let $(\hat{f}, f)$ be a vector bundle homomorphism between the two vector bundles $(E_1, \pi_1, M_1, \mathbb{R}^k)$ and $(E_2, \pi_2, M_2, \mathbb{R}^l)$. Suppose that $\hat{f}$ has constant rank. Then

$$\ker \hat{f} := \bigcup_{p \in M_1} \big(\{p\} \times \ker \hat{f}_p\big)$$

is a subbundle of E_1.

If the vectors bundles E_1 and E_2 have the same base space M, then the above considerations can be specialized to $f = \mathrm{id}_M$ and we call the map $\hat{f}: E_1 \to E_2$ a **vector bundle homomorphism over** M, respectively **vector bundle isomorphism over** M if $\hat{f}$ is a diffeomorphism such that its inverse is also a vector bundle homomorphism over M.

The following theorem shows that the tangent bundle of a differentiable manifold M that we defined in Definition 3.19 is a vector bundle.

Theorem 3.24. The tangent bundle TM of an n-dimensional differentiable manifold M is a vector bundle $(TM, \pi_{TM}, M, \mathbb{R}^n)$ of rank n over M.

For the proof, we refer to John M. Lee 2013, Propositions 3.18 and 10.4. Lee proves that the tangent bundle is a smooth manifold by constructing an atlas of TM that is induced by the atlas (U_α, ϕ_α) of the base manifold M. For this purpose he uses the map (3.25) to define the charts $(\pi^{-1}(U_\alpha), \Phi_\alpha)$ of TM

$$\Phi_\alpha : \pi^{-1}(U_\alpha) \to \phi_\alpha(U_\alpha) \times \mathbb{R}^n \subseteq \mathbb{R}^{2n},$$
$$\left(p, u^i \frac{\partial}{\partial x^i}\bigg|_p\right) \mapsto \big(\phi_\alpha(p), \mathrm{d}\phi_{\alpha p}(u_p)\big) = (\mathbf{x}, \mathbf{u})$$

with $p \in U_\alpha$ and $u_p \in T_pM$ (see equation (3.29)). The projection (3.30) is differentiable because all of its chart representations

$$\phi_\alpha \circ \pi_{TM} \circ \Phi_\alpha^{-1}(\mathbf{x}, \mathbf{u}) = \mathbf{x}$$

are differentiable (see p. 51). Since the tangent spaces T_pM are isomorphic to $\mathbb{R}^n$, the typical fibre is $\mathbb{R}^n$. In the proof of Proposition 10.4, Lee shows that the maps defined as

$$\begin{aligned} \theta_{TM} &\colon TM \to U_\alpha \times \mathbb{R}^n \\ \left(p, u^i \frac{\partial}{\partial x^i}\bigg|_p\right) &\mapsto \big(p, \mathrm{d}\phi_{\alpha p}(u_p)\big) = (p, \mathbf{u}) \end{aligned} \tag{3.33}$$

for all $p \in U_\alpha$ are indeed local trivializations. Note that the fibre over $p \in M$ is given by $E_p = \pi_{TM}^{-1}(p) = \{p\} \times T_pM$. Let $e_p = (p, v_p)$ and $e'_p = (p, v'_p)$ be two elements of E_p. Then v_p and v'_p are tangent vectors from T_pM. If $\dot{+}$ denotes the addition on E_p and $+$ the one on T_pM, then

$$e_p \dot{+} e'_p = (p, v_p) \dot{+} (p, v'_p) = (p, v_p + v'_p).$$

The respective scalar multiplications $\bullet$ and $\cdot$ of E_p and T_pM are related as

$$\alpha \bullet e_p = \alpha \bullet (p, v_p) = (p, \alpha \cdot v_p),$$

for all real numbers α. A similar construction of charts and local trivializations can be used to endow the cotangent bundle T^*M of an n-dimensional differentiable manifold with the structure of a vector bundle of rank n over M.

The results from Sections 2.10 and 2.11 can be used to introduce tensor bundles. We define the **bundle of mixed tensors of type** (k, l) as

$$\otimes_l^k TM := \bigcup_{p \in M} \big(\{p\} \times \otimes_l^k T_pM\big), \tag{3.34}$$

the **bundle of covariant l-tensors** as

$$\otimes^l T^*M := \bigcup_{p \in M} \big(\{p\} \times \otimes^l T_p^*M\big) \tag{3.35}$$

and the **bundle of alternating l-forms on M** as

$$\Lambda^l T^*M := \bigcup_{p \in M} \big(\{p\} \times \Lambda^l T_p^*M\big). \tag{3.36}$$

Such as TM and T^*M, the bundles (3.34), (3.35) and (3.36) are vector bundles. Their respective atlas can be constructed similarly to the one

Designation	Vector bundle	Set of sections
Vector field	TM	$\Gamma(TM)$, $\mathrm{Vect}(M)$
Covector field	T^*M	$\Gamma(T^*M)$, $\Omega^1(M)$
(k,l)-tensor field	$\otimes^k_l TM$	$\Gamma\big(\otimes^k_l TM\big)$
Covariant l-tensor field	$\otimes^l T^*M$, $\otimes^0_l TM$	$\Gamma\big(\otimes^l T^*M\big)$, $\Gamma\big(\otimes^0_l TM\big)$
Differential l-form	$\bigwedge^l T^*M$	$\Gamma\big(\bigwedge^l T^*M\big)$, $\Omega^l(M)$

Table 3.1.: Smooth sections of tensor bundles.

of the tangent bundle by using the chart of the base manifold to represent the points $p \in M$ together with its induced bases $(\partial/\partial x^1|_p, \dots, \partial/\partial x^n|_p)$ and $(\mathrm{d}x^1_p, \dots, \mathrm{d}x^n_p)$ for the representation of the tensor at each point. The local trivializations can be deduced from the ones of the tangent bundle (see equation (3.33)) in an analogue way. Roughly speaking, the charts and the trivialization simply "readout" the coefficients of the tensor from its representation with respect to the coordinates that are induced by the charts of the base manifold M.

A tensor field on a differentiable manifold M is just a smooth assignment of a tensor to each point p of M. This concept can be defined in general for fibre bundles as follows. A **(smooth) section of a fibre bundle** (E, M, ϕ, F) is a smooth map $s\colon M \supseteq U \to E$ such that

$$\pi \circ s = \mathrm{id}_U, \tag{3.37}$$

i.e., $\pi(s(p)) = p$ for all $p \in U$. The **set of all sections of** E is denoted by $\Gamma(E)$. The criterion (3.37) guarantees that a point p is mapped to a pair $(p, f_p) \in \pi^{-1}(p)$, i.e., to an element of the fibre over p. We can define sections through the different bundles that we have seen so far. The result is summarized in Table 3.1.

3.5. Vector fields

A **vector field** $v \in \Gamma(TM)$ on a differentiable manifold M is a section of the tangent bundle TM, i.e., a map

$$v\colon M \supseteq U \to TM,\ p \mapsto (p, v_p),$$

with $v_p \in T_pM$ according to (3.37). The set of vector fields on M, denoted by $\mathrm{Vect}(M) \coloneqq \Gamma(TM)$, is an (infinite-dimensional) vector space in virtue of the addition and scalar multiplication defined point by point as

$$\begin{aligned}(u+v)(p) &:= (p, u_p + v_p),\\ \alpha u(p) &:= (p, \alpha u_p),\end{aligned} \tag{3.38}$$

for all $u, v \in \mathrm{Vect}(M)$, all $\alpha \in \mathbb{R}$ and all $p \in M$. Additionally to the multiplication by real numbers from (3.38), the set of vector fields can be equipped with a multiplication between real-valued functions and vector fields. Let $f \in C^\infty(M)$ and $v \in \mathrm{Vect}(M)$, then this multiplication is defined as

$$(fv)(p) := (p, f(p)v_p) \tag{3.39}$$

for all points $p \in M$.

Analogously, a **covector field** $\rho \in \Gamma(T^*M)$ on a differentiable manifold M is a section of the cotangent bundle T^*M, i.e., a map

$$\rho\colon M \supseteq U \to T^*M,\ p \mapsto (p, \rho_p),$$

with $\rho_p \in T_p^*M$ according to (3.37). The set of covector fields on M, which is also denoted by $\Omega^1(M) := \Gamma(T^*M)$, can be endowed with a vector space structure by defining

$$\begin{aligned}(\rho+\sigma) &:= (p, \rho_p + \sigma_p),\\ \alpha\rho(p) &:= (p, \alpha\rho_p)\end{aligned} \tag{3.40}$$

for all $\alpha \in \mathbb{R}$ and all $\rho, \sigma \in \Omega^1(M)$ with $\rho(p) = (p, \rho_p)$ and $\sigma(p) = (p, \sigma_p)$. For covector fields, the analogue to multiplication (3.39) is defined as follows. Let $f \in C^\infty(M)$ and $\rho \in \Omega^1(M)$, then

$$(f\rho)(p) := (p, f(p)\rho_p) \tag{3.41}$$

for all points $p \in M$. The operation of a covector field $\rho \in \Omega^1(M)$ on a vector field $v \in \mathrm{Vect}(M)$ is denoted by

$$\rho(v) = \rho \cdot v = v \cdot \rho \tag{3.42}$$

and it is defined using the pointwise duality pairing (2.23) as

$$(\rho(v))(p) := \rho_p \cdot v_p.$$

In Section 3.2, we saw that a tangent vector at some point p of a differentiable manifold M can operate as a derivation in p[13] on a function $f \in C^\infty(M)$. This pointwise property can be transferred to vector fields. A linear map

13. See p. 55 for the definition.

$D\colon C^\infty(M) \to C^\infty(M)$ is said to be a **derivation** if it fulfils the product rule, i.e., if

$$D[fg] = fD[g] + gD[f]$$

for all $f, g, \in C^\infty(M)$. Vector fields $v \in \mathrm{Vect}(M)$ operate pointwise on functions $f \in C^\infty(M)$ as

$$(v[f])(p) \coloneqq v_p[f], \tag{3.43}$$

with $v(p) = (p, v_p)$ because of (3.29). In particular, because of the locality property D 1 of v_p as derivation in p, it holds that

$$v[f]|_V = v[f|_V]$$

for all open subsets $V \subset U \subseteq M$.

Theorem 3.25 (John M. Lee 2013, Proposition 8.15)**.** Let M be a differentiable manifold. Then, the derivations on $C^\infty(M)$ can be identified with the vector fields $\mathrm{Vect}(M)$ on M.

If we are given two derivations D_1 and D_2 on $C^\infty(M)$ and two functions $f, g \in C^\infty(M)$, then we know that $D_1[f] \in C^\infty(M)$. The question arises whether the concatenation D_2D_1 of two derivations is a derivation. The answer is no, as the following calculation shows:

$$\begin{aligned} D_2D_1[fg] &= D_2\big[fD_1[g] + gD_1[f]\big] \\ &= fD_2D_1[g] + D_2[f]D_1[g] + gD_2D_1[f] + D_2[g]D_1[f] \\ &\neq fD_2D_1[g] + gD_2D_1[f]. \end{aligned} \tag{3.44}$$

However, we may observe from (3.44) that $D_2D_1 - D_1D_2$ satisfies the product rule and, therefore, is a derivation.

Theorem 3.25 allows us to transfer this observation to the space $\mathrm{Vect}(M)$ of vector fields on M. For any two vector fields $u, v \in \mathrm{Vect}(M)$, we define the vector field $[\![u, v]\!] \in \mathrm{Vect}(M)$ by its operation on functions $f \in C^\infty(M)$ as

$$[\![u, v]\!][f] \coloneqq u\big[v[f]\big] - v\big[u[f]\big]. \tag{3.45}$$

The reader may check that the bracket (3.45) satisfies axioms LB 1–LB 3 from the definition of a Lie bracket on p. 33. The vector space $\mathrm{Vect}(M)$ becomes a Lie algebra when it is considered together with the Lie bracket from (3.45).

In Section 3.2, we saw that if we are given a differentiable map $f\colon M \to N$ between two differentiable manifolds M and N, then the differential of f in a point p of M (3.26) maps the tangent space T_pM to the tangent space $T_{f(p)}N$.

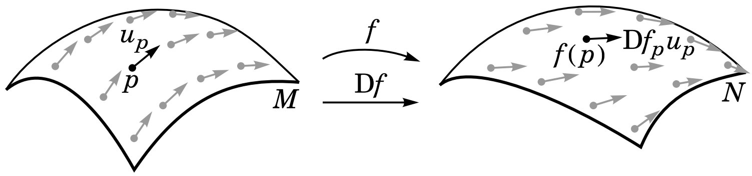


Figure 3.10.: The differential of the function $f\colon M \to N$ maps tangent vectors on M to tangent vectors on N.

This pointwise map can be extended to the tangent bundles TM and TN. The **differential of** f is defined as the map

$$\mathrm{D}f\colon TM \to TN,\ \ (p, u_p) \mapsto \big(f(p), \mathrm{D}f_p u_p\big).$$

The differential of f can be used to relate vector fields on M with vector fields on N (see Figure 3.10). We say that the vector fields $v \in \mathrm{Vect}(M)$ and $w \in \mathrm{Vect}(N)$ are f**-related** if

$$w \circ f(p) = \mathrm{D}f\, v\,(p) \text{ for all } p \in M$$

or, equivalently, if

$$w[g] \circ f = v[g \circ f] \text{ for all } g \in C^\infty(N).$$

In the case where $f\colon M \to N$ is a diffeomorphism, it induces the map

$$f_{\rightarrow}\colon \mathrm{Vect}(M) \to \mathrm{Vect}(N),\ \ v \mapsto f_{\rightarrow}(v)$$

that is defined by

$$(f_{\rightarrow} v)(q) := \mathrm{D}f_{f^{-1}(q)} v\big(f^{-1}(q)\big) = \mathrm{D}f\, v \circ f^{-1}(q),$$

for all $q \in N$. The vector field $f_{\rightarrow}(v) \in \mathrm{Vect}(N)$ is called the **pushforward of the vector field** $v \in \mathrm{Vect}(M)$ **with f**. The inverse of the map $f\colon M \to N$ allows to define the **pullback** $f_{\leftarrow}$ **of a vector field** $w \in \mathrm{Vect}(N)$ **with** f as the vector field

$$f_{\leftarrow} w := \big(f^{-1}\big)_{\rightarrow} w$$

on M. Overloading notation, we consider the map

$$f_{\leftarrow}\colon \Omega^1(N) \to \Omega^1(M),\ \ \rho \mapsto f_{\leftarrow}\rho,$$

defined using the diffeomorphism $f\colon M \to N$ by

$$(f_{\leftarrow}\rho)(v) := \rho(f_{\rightarrow} v)$$

for all $\rho \in \Omega^1(N)$ and all $v \in \mathrm{Vect}(M)$. The covector field $f_{\leftarrow}\rho \in \Omega^1(M)$ is called the **pullback of the covector field** $\rho \in \Omega^1(N)$ **with** f. Finally, the **pushforward of the covector field** $\sigma \in \Omega^1(M)$ **with** f is the covector field on N that is given by

$$f_{\rightarrow}\sigma := \left(f^{-1}\right)_{\leftarrow}\sigma.$$

Note that a vector field $v \in \mathrm{Vect}(M)$ and its pushforward $f_{\rightarrow}v$ are f-related. The f-relation of vector fields is compatible with the Lie algebra structure on $\mathrm{Vect}(M)$ and $\mathrm{Vect}(N)$ according to the following theorem.

Theorem 3.26 (John M. Lee 2013, Proposition 8.30)**.** Let $v_i \in \mathrm{Vect}(M)$ and $w_i \in \mathrm{Vect}(N)$ with $i = 1, 2$ such that v_i and w_i are f-related for all i. Then the Lie brackets $\llbracket v_1, v_2 \rrbracket$ and $\llbracket w_1, w_2 \rrbracket$ are also f-related, i.e.,

$$\llbracket w_1, w_2 \rrbracket \circ f(p) = \mathrm{D}f\, \llbracket v_1, v_2 \rrbracket(p) \text{ for all } p \in M.$$

The **coordinate fields** induced by a chart (U, ϕ) of a differentiable manifold M with $\phi\colon p \mapsto \phi(p) = (x^1, \dots, x^n)$ are defined as the sections

$$\frac{\partial}{\partial x^i}\colon U \to TM,\ p \mapsto \left(p, \left.\frac{\partial}{\partial x^i}\right|_p\right). \tag{3.46}$$

By equations (3.38), (3.39), (3.43), and (3.46), every vector field $v \in \mathrm{Vect}(M)$ can be locally written as

$$v = v[\phi^i]\frac{\partial}{\partial x^i} \tag{3.47}$$

because

$$v(p) = \left(p, v_p[\phi^i]\left.\frac{\partial}{\partial x^i}\right|_p\right)$$

for all $p \in U \subseteq M$.

The **dual coordinate fields** on M induced by the chart (U, ϕ) are defined to be the sections

$$\mathrm{d}x^i\colon U \to T^*M,\ p \mapsto \left(p, \mathrm{d}x^i_p\right), \tag{3.48}$$

where the basis covectors $(\mathrm{d}x^1_p, \dots, \mathrm{d}x^n_p)$ are defined by equation (3.28). With equations (2.21), (3.40), (3.41), (3.42), and (3.48), every covector field $\rho \in \Omega^1(M)$ can be locally expressed as

$$\rho = \rho\left(\frac{\partial}{\partial x^i}\right)\mathrm{d}x^i, \tag{3.49}$$

because

$$\rho(p) = \left(p, \rho_p\left(\left.\frac{\partial}{\partial x^i}\right|_p\right)\mathrm{d}x^i_p\right)$$

for all $p \in U \subseteq M$.

3.6. Flow of a vector field

In Definition 3.12, we used equivalence classes of curves to define the tangent space T_pM to a differentiable manifold M at some point $p \in M$. With Definition 3.14 we saw that tangent vectors in $p \in T_pM$ can operate on functions $f \in C^\infty(M)$ as derivations in p. Theorem 3.17 tells us that both definitions are equivalent because the resulting tangent spaces T_pM and T_pM are isomorphic.

In this section, we change our perspective. Instead of considering a single point $p \in M$, we are interested in all the points which lie along some curve $\gamma\colon \mathbb{R} \supset I \to M,\ \tau \mapsto \gamma(\tau)$ that passes through p such that $\gamma(\tau_0) = p \in M$. At each point of the curve, we consider the tangent vector defined by the curve in that point. If $0 \in I$, then the tangent vector defined by the curve γ is just $[\gamma]$ (the equivalence class of γ), which is an element of $\mathsf{T}_{\gamma(0)}M$. The tangent vector defined by γ for any other value $\tau^* \in I$ with $\tau^* \neq 0$ is defined to be the equivalence class of the curve $\gamma_{\tau^*}(\tau) := \gamma(\tau + \tau^*)$, i.e., the tangent vector $[\gamma_{\tau^*}] \in \mathsf{T}_{\gamma(\tau^*)}M$. Note that the equivalence relation (3.7) is defined for curves with intervals $I =]-r, r[$ centred around 0. Given an arbitrary curve $\gamma\colon \mathbb{R} \supset I \to M$ and some point $q \in \gamma(\tau^*)$ on it, the reparametrized curve γ_{τ^*} satisfies $\gamma_{\tau^*}(0) = q$.

With this construction, we get a vector field along the curve γ, called the **tangent field along** γ, that we denote by $\dot\gamma$. By Theorem 3.25, the tangent field along γ can operate as derivation on functions $f \in C^\infty(M)$. The operation is defined pointwise according to (3.43) as

$$
\begin{aligned}
\big(\dot\gamma[f]\big)(\gamma(\tau^*)) = \dot\gamma_{\gamma(\tau^*)}[f] &\overset{(3.23)}{=} L_{[\gamma_{\tau^*}]}f \\
&\overset{(3.22)}{=} \left.\frac{\mathrm{d}}{\mathrm{d}\tau}\right|_{\tau=0} \big(f \circ \gamma_{\tau^*}\big)(\tau) \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}\tau}\right|_{\tau=\tau^*} (f \circ \gamma)(\tau).
\end{aligned}
\tag{3.50}
$$

Let $\gamma\colon \mathbb{R} \supset I \to M$ be a curve passing through $p \in M$. If we are given a vector field v on M, then we can check whether or not the tangent field $\dot\gamma$ along the curve γ corresponds to the vector field v evaluated along the curve. If this is the case, i.e., if

$$
\dot\gamma(\gamma(\tau)) = v(\gamma(\tau)),
\tag{3.51}
$$

then the curve $\gamma(\tau)$ is said to be an **integral curve** of the vector field v through p (see Figure 3.11). Let (U, ϕ) be a chart that contains part of the curve, i.e., $\gamma(\tau) \in U$ for all $\tau \in I \cap \gamma^{-1}(U) \neq \emptyset$. Let the vector field

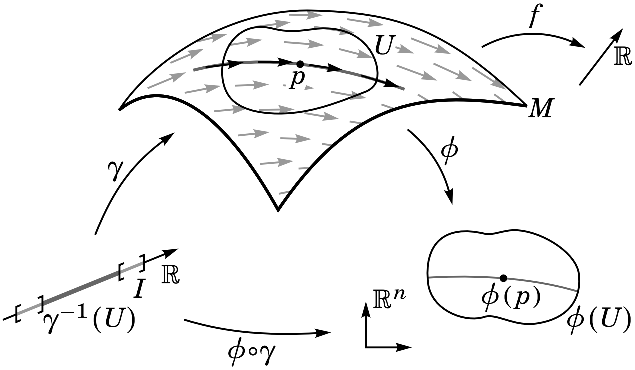


Figure 3.11.: An integral curve of a vector field passing through the point p.

$v \in \mathrm{Vect}(M)$ have the local expression

$$v = v^i \frac{\partial}{\partial x^i},$$

with $v^i \in C^\infty(U)$. Equation (3.51) can be expressed with respect to the chart (U, ϕ) according to (3.47) as

$$\dot{\gamma}[\phi^i](\gamma(\tau)) = v[\phi^i](\gamma(\tau)), \tag{3.52}$$

with $i = 1, \dots, \dim M$. With equation (3.50), we recognize that (3.52) defines the following system of ordinary differential equations in the chart (U, ϕ)

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\phi^i \circ \gamma\right)(\tau) = v^i(\gamma(\tau)). \tag{3.53}$$

The determination of an integral curve γ that passes through a given point $p \in M$ boils down to solving the system of ordinary differential equations (3.53) with the initial condition

$$\gamma(\tau_0) = p. \tag{3.54}$$

Therefore, many results about ordinary differential equations can be applied to integral curves. The theorem of Picard-Lindelöf guarantees that for each initial condition (3.54) there is some value $r > 0$ such that there exists a unique solution $\gamma(\tau)$ to the initial value problem (3.53)–(3.54) on the interval $]\tau_0 - r, \tau_0 + r[$ that passes through p for $\tau = \tau_0$. In particular, this means that integral curves do not intersect. Let $\gamma_p : \,]\tau_0 - r, \tau_0 + r[\, \to M$ be the integral curve of a vector field $v \in \mathrm{Vect}(M)$ that passes through

$p \in U \subseteq M$. The map

$$\begin{aligned}\varphi\colon\ ]\tau_0 - r, \tau_0 + r[\times U &\to M, \\ (\tau, p) &\mapsto \varphi(\tau, p) := \gamma_p(\tau).\end{aligned} \tag{3.55}$$

with $U \subseteq M$ is said to be the **(local) flow of** v. The local flow of v provides the map

$$\begin{aligned}\varphi_\tau\colon U &\to M, \\ p &\mapsto \varphi_\tau(p) := \gamma_p(\tau).\end{aligned} \tag{3.56}$$

defined on a neighbourhood of p.

Theorem 3.27. The maps (3.56) are local diffeomorphisms and it holds that

$$\varphi_{\tau_1+\tau_2} = \varphi_{\tau_1} \circ \varphi_{\tau_2}$$

for all parameters τ_1, τ_2 for which φ_{τ_1}, φ_{τ_2}, and $\varphi_{\tau_1+\tau_2}$ are defined. The set $\{\varphi_\tau\}$ is a one-parameter group of local diffeomorphisms.

Proof. It follows from the definition (3.56) that

$$\begin{aligned}\varphi_{\tau_1+\tau_2}(p) &= \varphi(\tau_1 + \tau_2, p) = \gamma_p(\tau_1 + \tau_2) \\ &= \gamma_{\gamma_p(\tau_2)}(\tau_1) = \varphi_{\tau_1}(\gamma_p(\tau_2)) = \varphi_{\tau_1} \circ \varphi_{\tau_2}(p).\end{aligned} \tag{3.57}$$

In particular, we can see from (3.57) that $\mathrm{id}_U = \varphi_0 = \varphi_{\tau+(-\tau)} = \varphi_\tau \circ \varphi_{-\tau}$ and, therefore, $\varphi_\tau^{-1} = \varphi_{-\tau}$. Hence, the map φ_τ is invertible and its inverse map φ_τ^{-1} is differentiable such that φ_τ is a local diffeomorphism. □

3.7. Tensor fields

A tensor field of type (k, l)

$$G \in \Gamma\big(\otimes^k_l TM\big)$$

is a section such that

$$G(p) = (p, G_p) \text{ with } G_p \in \otimes^k_l T_pM.$$

Its application to k covector fields and l vector fields is defined point by point as

$$G({}_1\sigma, \dots, {}_k\sigma, {}_1u, \dots, {}_lu)(p) := G_p({}_1\sigma_p, \dots, {}_k\sigma_p, {}_1u_p, \dots, {}_lu_p) \tag{3.58}$$

for all ${}_\alpha\sigma \in \Gamma(T^*M) = \Omega^1(M)$ with $\alpha = 1, \dots, k$, all ${}_\beta u \in \mathrm{Vect}(M) = \Gamma(TM)$ with $\beta = 1, \dots, l$, and all $p \in M$. The set of tensor fields of type (k, l) can be endowed with the structure of a vector space by defining the addition and scalar multiplication point by point for all $p \in M$ as

$$\begin{aligned} (G+H)(p) &:= (p, G_p + H_p), \\ \alpha G(p) &:= (p, \alpha G_p), \end{aligned} \tag{3.59}$$

for all

$$G, H \in \Gamma\big(\otimes^k_l TM\big)$$

and all $\alpha \in \mathbb{R}$. Furthermore, we define a multiplication between functions $f \in C^\infty(M)$ and tensor fields G as

$$(fG)(p) := \big(p, f(p)G_p\big) \tag{3.60}$$

for all points $p \in M$. The vector space structures of $\mathrm{Vect}(M)$ and $\Omega^1(M)$ that are declared respectively by equations (3.38) and (3.40) are special cases of (3.59). Similarly, the multiplications (3.39) and (3.41) follow from the general definition (3.60).

In Section 2.10, we defined the tensor product with equation (2.34). Theorem 2.15 tells us that this tensor product can be used to construct a basis of a given tensor space. The tensor product can be extended to tensor fields. Let

$$\begin{aligned} F &\in \Gamma\big(\otimes^k_l TM\big), \\ G &\in \Gamma\big(\otimes^r_s TM\big) \end{aligned}$$

be two tensor fields with

$$\begin{aligned} F(p) &= (p, F_p) \text{ with } F_p \in \otimes^k_l T_pM, \\ G(p) &= (p, G_p) \text{ with } G_p \in \otimes^r_s T_pM. \end{aligned}$$

The **tensor product of the tensor fields F and G** is the map

$$F \otimes G\colon \underbrace{W \times \cdots \times W}_{k} \times \underbrace{V \times \cdots \times V}_{l} \times \underbrace{W \times \cdots \times W}_{r} \times \underbrace{V \times \cdots \times V}_{s} \to C^\infty(M),$$

with $V = \mathrm{Vect}(M)$ and $W = \Omega^1(M)$ that is defined by

$$\begin{aligned} &F \otimes G(\rho_1, \dots, \rho_k, u_1, \dots, u_l, \sigma_1, \dots, \sigma_r, v_1, \dots, v_s)(p) \\ &\quad := \big(p, F_p(\rho_{1p}, \dots, \rho_{kp}, u_{1p}, \dots, u_{lp}) G_p(\sigma_{1p}, \dots, \sigma_{rp}, v_{1p}, \dots, v_{sp})\big) \end{aligned}$$

for all $\rho_\alpha, \sigma_\gamma \in \Omega^1(M)$ with $\alpha = 1, \dots, k$, $\gamma = 1, \dots, r$, all $u_\beta, v_\delta \in \mathrm{Vect}(M)$ with $\beta = 1, \dots, l$, $\delta = 1, \dots, s$, and all $p \in M$.

Because of this pointwise construction, the coordinate fields (3.46) and the dual coordinate fields (3.48) can be used together with the tensor product to construct a basis of the space of tensor fields in a way that is analogue to Theorem 2.15. A similar reasoning can be used to define the wedge product[14] of tensor fields.

By their point-by-point operation on vectors and covectors, smooth tensor fields define maps on the corresponding spaces of vector and covector fields on M that are multilinear over $C^\infty(M)$. Because we are primarily interested in covariant tensor fields, we state the following lemma.

Lemma 3.28 (John M. Lee 2013, Lemma 12.24)**.** A map

$$A\colon \underbrace{\mathrm{Vect}(M)\times\cdots\times\mathrm{Vect}(M)}_{l\text{ copies}} \to C^\infty(M),$$

is induced by a smooth covariant l-tensor field as above if and only if it is multilinear over $C^\infty(M)$.

Pushforward and pullback of tensor fields

On pages 71–72, we defined the pushforward and pullback of vector and covector fields. These concepts can be used to define the pushforward and the pullback for arbitrary tensor fields. Let $f\colon M \to N$ be a diffeomorphism between the manifolds M and N. The **pullback of a (k,l)-tensor field G on N** is the tensor field $f_{\leftarrow}G$ on M defined by

$$\big(f_{\leftarrow}G\big)\left(\rho_1,\dots,\rho_k,v_1,\dots,v_l\right) \coloneqq G\big(f_{\rightarrow}\rho_1,\dots,f_{\rightarrow}\rho_k,f_{\rightarrow}v_1,\dots,f_{\rightarrow}v_l\big),$$

with $\rho_1,\dots,\rho_k \in \Omega^1(M)$ and $v_1,\dots,v_l \in \mathrm{Vect}(M)$ such that

$$f_{\leftarrow}\colon \Gamma\big(\otimes^k_l TN\big) \to \Gamma\big(\otimes^k_l TM\big).$$

In the opposite direction, the **pushforward of a (k,l)-tensor field F on M** is the tensor field $f_{\rightarrow}F$ on N defined by

$$\big(f_{\rightarrow}F\big)\left(\sigma_1,\dots,\sigma_k,w_1,\dots,w_l\right) \coloneqq F\big(f_{\leftarrow}\sigma_1,\dots,f_{\leftarrow}\sigma_k,f_{\leftarrow}w_1,\dots,f_{\leftarrow}w_l\big),$$

with $\sigma_1,\dots,\sigma_k \in \Omega^1(N)$ and $w_1,\dots,w_l \in \mathrm{Vect}(N)$ such that

$$f_{\rightarrow}\colon \Gamma\big(\otimes^k_l TM\big) \to \Gamma\big(\otimes^k_l TN\big).$$

14. See equation (2.40) on p. 40.

3.8. Differential forms

Equation (3.36) defines the bundle of alternating l-forms on a differentiable manifold M. A **differential l-form** on M is a section of the bundle of alternating l-forms, i.e., a map

$$\omega: M \to \Lambda^l T^*M,$$

with $\pi \circ \omega = \mathrm{id}_M$. For $l = 0$, we define $\Omega^0(M) := C^\infty(M)$ and denote the **set of differential forms of arbitrary degree** on M by

$$\Omega^\star(M) := \bigoplus_{l=0}^{\dim M} \Omega^l(M).$$

The differential forms of degree $l > \dim M$ are zero. This can be easily seen from Proposition 2.19(iv) because in the local expression of such a differential l-form with respect to a chart (U, ϕ) of M at least one of the basis covector fields $\mathrm{d}x^1, \dots, \mathrm{d}x^n$ would appear more than once.

A differential l-form $\omega \in \Omega^l(M)$ induces a $C^\infty(M)$-multilinear, alternating map

$$\omega: \mathrm{Vect}(M) \times \cdots \times \mathrm{Vect}(M) \to C^\infty(M),\ (v_1, \dots, v_l) \mapsto \omega(v_1, \dots, v_l)$$

that is defined point by point as

$$\omega(v_1, \dots, v_l)(p) := \omega(p)(v_1(p), \dots, v_l(p)).$$

The converse turns out to be also true[15] such that there is a one-to-one correspondence between differential forms and the $C^\infty(M)$-multilinear, alternating maps on vector fields. In what follows, we will make no notational distinction between differential forms and the map they induce on vector fields.

Let $f: M \to N$ be a smooth map between the differentiable manifolds M and N and let $\omega \in \Omega^l(N)$ be a differential l-form on N. The **pullback** $f_\leftarrow\omega \in \Omega^l(M)$ **of** ω **with** f is the differential l-form on M defined by

$$\left(f_\leftarrow\omega\right)_p (v_1, \dots, v_l) = \omega_{f(p)}\left(\mathrm{D}f_p(v_1), \dots, \mathrm{D}f_p(v_l)\right), \tag{3.61}$$

for all $v_1, \dots, v_l \in \mathrm{Vect}(M)$. Note that the map $f: M \to N$ does not need to be a diffeomorphism. In the case where $f: M \to N$ is a diffeomorphism, this definition agrees with the pullback of a (k, l)-tensor field for $k = 0$ defined earlier.

15. See Lemma 3.28.

We consider the map

$$\begin{aligned} \mathrm{d}\colon \Omega^0(M) &\to \Omega^1(M), \\ f &\mapsto \mathrm{d}f \end{aligned}$$

that is defined by the operation of $\mathrm{d}f$ on vector fields

$$\begin{aligned} \mathrm{d}f\colon \mathrm{Vect}(M) &\to C^\infty(M), \\ v &\mapsto \mathrm{d}f(v) := v[f]. \end{aligned} \tag{3.62}$$

We want to extend the map d to differential forms of arbitrary degree such that we get a sequence of maps

$$\Omega^0(M) \xrightarrow{\mathrm{d}} \Omega^1(M) \xrightarrow{\mathrm{d}} \Omega^2(M) \xrightarrow{\mathrm{d}} \cdots.$$

Theorem 3.29 (Gallot et al. 1990, Theorem 1.119, p. 43)**.** Let M be a differentiable manifold. For any $l \in \mathbb{N}$, there exists a unique local operator d from $\Omega^l(M)$ to $\Omega^{l+1}(M)$, called the **exterior derivative** such that

ED 1. for $l = 0$, the map $\mathrm{d}\colon \Omega^0(M) \to \Omega^1(M)$ is defined by (3.62),

ED 2. it holds that $\mathrm{d}^2 = \mathrm{d} \circ \mathrm{d} = 0$,

ED 3. for $\alpha \in \Omega^l(M)$ and $\beta \in \Omega^\star(M)$

$$\mathrm{d}(\alpha \wedge \beta) = \mathrm{d}\alpha \wedge \beta + (-1)^l \alpha \wedge \mathrm{d}\beta.$$

Let us come back to pullbacks of differential forms. The exterior derivative has the important feature that it commutes with all pullbacks. This property is referred to as naturality of the exterior derivative.

Proposition 3.30 (John M. Lee 2013, Proposition 14.26)**.** Let $f\colon M \to N$ be a smooth map between the differentiable manifolds M and N. Then for each $l \in \mathbb{N}$ the pullback $f_{\leftarrow}\colon \Omega^l(N) \to \Omega^l(M)$ commutes with d: for all $\omega \in \Omega^l(N)$

$$f_{\leftarrow}(\mathrm{d}\omega) = \mathrm{d}(f_{\leftarrow}\omega).$$

Moreover, pullbacks of differential forms have the following properties.

Proposition 3.31 (John M. Lee 2013, Lemma 14.16)**.** Suppose $f\colon M \to N$ to be a smooth map between the differentiable manifolds M and N, then the following properties hold:

(i) The map $f_{\leftarrow}\colon \Omega^l(N) \to \Omega^l(M)$ is linear over $\mathbb{R}$.

(ii) The pullback commutes with the wedge product, i.e.,

$$f_{\leftarrow}(\alpha \wedge \beta) = f_{\leftarrow}(\alpha) \wedge f_{\leftarrow}(\beta).$$

(iii) In any smooth chart,

$$f_{\leftarrow}\left(\sum_I^{\nearrow} \omega_I \, \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_l} \right) = \sum_I^{\nearrow} (\omega_I \circ f) \, \mathrm{d}\big(x^{i_1} \circ f\big) \wedge \cdots \wedge \mathrm{d}\big(x^{i_l} \circ f\big).$$

We say that a differential l-form $\omega \in \Omega^l(M)$ is **closed** if $\mathrm{d}\omega = 0$. It is called **exact** if it can be written as $\omega = \mathrm{d}\alpha$ with $\alpha \in \Omega^{l-1}(M)$. By ED 2, we know that $\mathrm{d} \circ \mathrm{d} = 0$ and, therefore, the condition of being closed is necessary for a differential form to be exact. In general, the converse is not true and the extent to which it fails is a topological property of the manifold. The study of this question leads to the *de Rham cohomology* for which we refer to Chapter 10 in Jeffrey M. Lee 2009. However, there is the following local result due to Poincaré which says that locally, every closed form is exact.

Lemma 3.32 (Poincaré lemma)**.** Let B be an open ball in $\mathbb{R}^n$ and let ω be a differential l-form on B with $l \geq 1$ such that $\mathrm{d}\omega = 0$. Then there exists a differential form α on B such that $\omega = \mathrm{d}\alpha$.

Lemma 3.32 corresponds to Theorem 4.1 in Lang 2001 to which we refer for the proof. The lemma is formulated for differential forms on an open ball in $\mathbb{R}^n$. However, it directly transfers to locally defined forms on an n-dimensional differentiable manifold M because of Proposition 3.30.

3.9. The Lie derivative

On page 71, we saw that (local) diffeomorphisms induce a pullback map on vector fields. From Section 3.6, we know that the local flow (3.55) of a vector field $v \in \mathrm{Vect}(M)$ provides the local diffeomorphism (3.56) from M to itself. The pullback that comes with this map can be used to relate tangent vectors from two separate tangent spaces at two distinct points p and q lying on the same integral curve of v as shown in Figure 3.12.

By Theorem 3.25, we know that vector fields can be applied as derivations to smooth real-valued functions on a manifold. The above considerations allow us to define a derivative for tensors of arbitrary type along a given vector field. We start with vector fields. The **Lie derivative**[16] $\mathfrak{L}_v w \in \mathrm{Vect}(M)$ **of**

16. It can be shown that $\mathfrak{L}_v w$ is indeed a smooth vector field (see John M. Lee 2013, Lemma 9.36, pp. 228–229).

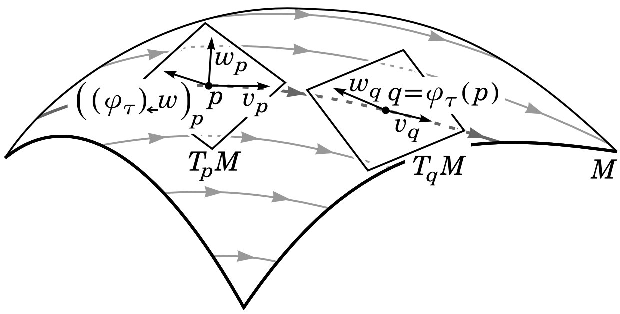


Figure 3.12.: Pullback map of vector fields induced by the flow of a vector field on a differentiable manifold M.

a vector field $w \in \mathrm{Vect}(M)$ along a vector field $v \in \mathrm{Vect}(M)$ is defined for all points $p \in M$ as

$$\mathfrak{L}_v w(p) := \frac{\mathrm{d}}{\mathrm{d}\tau}\bigg|_{\tau=0} \big((\varphi_\tau)_\leftarrow w\big)(p), \tag{3.63}$$

where $\varphi_\tau(p) = \gamma_p(\tau)$ is the flow of v through the point p with $\gamma_p(0) = p$.

Theorem 3.33 (John M. Lee 2013, Theorem 9.38)**.** If $v, w \in \mathrm{Vect}(M)$ are smooth vector fields on a differentiable manifold M, then it holds that

$$\mathfrak{L}_v w = [\![v, w]\!].$$

Theorem 3.34 (John M. Lee 2013, Theorem 9.44)**.** Let $v, w \in \mathrm{Vect}(M)$ be smooth vector fields on a differentiable manifold M with respective local flows $\varphi^v_{\tau_1}$ and $\varphi^w_{\tau_2}$. Then the vector fields commute if and only if their flows commute, i.e.,

$$[\![v, w]\!] = 0 \quad \Leftrightarrow \quad \varphi^v_{\tau_1} \circ \varphi^w_{\tau_2} = \varphi^w_{\tau_2} \circ \varphi^v_{\tau_1}.$$

Equation (3.63) defines the Lie derivative of a vector field. The extension of the definition to arbitrary tensor fields is straightforward. The **Lie derivative $\mathfrak{L}_v F$ of a (k, l)-tensor field** F on a differentiable manifold M along a vector field $v \in \mathrm{Vect}(M)$ is defined for all points $p \in M$ as

$$\mathfrak{L}_v F(p) := \frac{\mathrm{d}}{\mathrm{d}\tau}\bigg|_{\tau=0} \big((\varphi_\tau)_\leftarrow F\big)(p), \tag{3.64}$$

where $\varphi_\tau(p) = \gamma_p(\tau)$ is the flow of v through the point p with $\gamma_p(0) = p$.

The definition (3.64) includes the Lie derivative of a vector field w because a vector field is just a $(1, 0)$-tensor field. Moreover, it also comprises the derivative of a smooth function if the latter is interpreted as a tensor field of type $(0,0)$. The following proposition gathers some properties of the Lie-derivative (3.64). For the proof of these properties, we refer to Sections 5.3 and 5.4 of Abraham, Marsden, and Ratiu 1988.

Proposition 3.35. Let M be a differentiable manifold and $v \in \mathrm{Vect}(M)$ be a vector field on M. Suppose $f \in C^\infty(M)$ is a real-valued function on M (regarded as $(0,0)$-tensor field). Let $G \in \Gamma(\otimes^k_l TM)$ and $H \in \Gamma(\otimes^m_n TM)$ be a (k,l)- and a (m,n)-tensor field on M, respectively. Let $\alpha \in \Omega^k(M)$ be a differential k-form and $\beta \in \Omega^l(M)$ a differential l-form. Then the Lie derivative (3.64) has the following properties:

(i) $\mathfrak{L}_v f = v[f]$,

(ii) $\mathfrak{L}_v(fG) = \mathfrak{L}_v(f)G + f\mathfrak{L}_v G$,

(iii) $\mathfrak{L}_v(G \otimes H) = (\mathfrak{L}_v G) \otimes H + G \otimes \mathfrak{L}_v H$,

(iv) $\mathfrak{L}_v(\alpha \wedge \beta) = (\mathfrak{L}_v \alpha) \wedge \beta + \alpha \wedge \mathfrak{L}_v \beta$,

(v) Let $w_1, \dots, w_l \in \mathrm{Vect}(M)$ and $G \in \Gamma\big(\otimes^0_l TM\big)$, then

$$\begin{aligned}\mathfrak{L}_v\Big(G(w_1,\dots,w_l)\Big) = \Big(\mathfrak{L}_v G\Big)(w_1,\dots,w_l) \\ + G\Big(\mathfrak{L}_v w_1, w_2, \dots, w_l\Big) \\ + \dots + G\Big(w_1,\dots,w_{l-1},\mathfrak{L}_v w_l\Big).\end{aligned}$$

Let $\omega \in \Omega^k(M)$ and $v \in \mathrm{Vect}(M)$, then the **interior product** of ω and v is defined as the map

$$i_v \colon \Omega^k(M) \to \Omega^{k-1}(M), \;\; \omega \mapsto i_v\omega, \tag{3.65}$$

with

$$i_v\omega := \begin{cases} 0, & \text{if } k=0, \\ \omega(v, \underbrace{\cdot,\dots,\cdot}_{k-1}), & \text{if } k>0. \end{cases}$$

The interior product decreases the degree of a differential form by one, while the exterior derivative increases it by one. Let $\alpha \in \Omega^l(M)$, $\beta \in \Omega^\star(M)$ and $v \in \mathrm{Vect}(M)$, then the interior product has the property[17] that

$$i_v(\alpha \wedge \beta) = (i_v\alpha) \wedge \beta + (-1)^l \alpha \wedge (i_v\beta).$$

17. See Proposition 8.53 in Jeffrey M. Lee 2009.

Another common notation for the interior product reads $v \lrcorner \omega = i_v\omega$. The symbol $\lrcorner$ is particularly useful for inline formulas since it does not ruin the line spacing, while the i_v is more suitable to express concatenations of maps. With the interior product (3.65), the exterior derivative can be related to the Lie derivative[18] by **Cartan's magic formula**[19]

$$\mathfrak{L}_v = \mathrm{d} \circ i_v + i_v \circ \mathrm{d}\,. \tag{3.66}$$

3.10. Bilinear forms on the tangent spaces

In Section 2.7, we saw that a vector space can be equipped with a non-degenerate bilinear form. If we want to endow each tangent space T_pM of a differentiable manifold M with such a structure, then this can be realized by suitable tensor fields.

Symplectic Form

A **symplectic form** on a differentiable manifold M is a differential 2-form $\omega \in \Omega^2(M)$, that is closed and non-degenerate. A differential two-form on M is said to be **non-degenerate** if for all points $p \in M$ the evaluation of ω in a point $p \in M$ yields a bilinear form on the tangent space T_pM

$$\omega(p) : T_pM \times T_pM \to \mathbb{R} \tag{3.67}$$

that is non-degenerate according to condition (2.27). It can be readily verified that the bilinear forms (3.67) satisfy SF 1–4 for all $p \in M$ and therefore let the respective tangent space of M become symplectic vector spaces. A differentiable manifold M together with a symplectic form ω is referred to as **symplectic manifold** (M, ω).

The most prominent example of a symplectic manifold is given by the cotangent bundle T^*Q of a differentiable manifold Q (see Definition 3.20). The cotangent bundle T^*Q comes with a natural projection (3.31)

$$\pi_{T^*Q} : T^*Q \to Q,\ (p, \sigma_p) \mapsto p.$$

The differential of the natural projection

$$\mathrm{D}\pi_{T^*Q} : T(T^*Q) \to TQ$$

can be used to define the **canonical one-form** $\Theta \in \Omega^1(T^*Q)$ by requiring that

$$\Theta(w_{\sigma_p}) \stackrel{!}{=} \sigma_p\Big(\big(\mathrm{D}\pi_{T^*Q}\big)_{\sigma_p} w_{\sigma_p} \Big),$$

18. The definition of the Lie derivative from p. 81 holds in particular for differential forms.
19. For the proof of this formula, we refer to John M. Lee 2013, Theorem 14.35.

where $\sigma_p \in T_p^*Q$ and $w_{\sigma_p} \in T_{\sigma_p}(T^*Q)$. One can show[20] that the **canonical two-form** ω defined by

$$\omega := -\mathrm{d}\Theta$$

is a symplectic form such that the cotangent bundle (T^*Q, ω) is a symplectic manifold.

Let $\phi\colon Q \supseteq U \to \mathbb{R}^n,\ p \mapsto (q^1, \dots, q^n)$ be a chart of Q. The covector field $\sigma \in \Omega^1(Q)$ can be locally expressed as

$$\sigma = p_i \mathrm{d}q^i$$

where $\mathrm{d}q^1, \dots, \mathrm{d}q^n$ denote the dual coordinate fields (3.48) on Q induced by the chart (U, ϕ). This defines a chart of T^*Q as

$$\begin{aligned} &\Phi\colon T^*Q \supseteq \pi_{T^*Q}^{-1}(U) \to \mathbb{R}^{2n}, \\ &\quad \left(p, \sigma_p = p_i \mathrm{d}q_p^i\right) \mapsto \Phi\left(p, \sigma_p\right) := (q^1, \dots, q^n, p_1, \dots, p_n). \end{aligned} \tag{3.68}$$

This chart induces the dual coordinate fields $\mathrm{d}q^1, \dots, \mathrm{d}q^n, \mathrm{d}p^1, \dots, \mathrm{d}p^n$ on T^*Q. It follows by straightforward computation that $\Theta \in \Omega^1(T^*Q)$ is given by

$$\Theta = p_i \mathrm{d}q^i$$

and, consequently,

$$\omega = \mathrm{d}q^i \wedge \mathrm{d}p_i. \tag{3.69}$$

Coordinates of a symplectic manifold, in which the symplectic form ω takes the simple form (3.69) are called **canonical coordinates**. The existence of canonical coordinates is guaranteed by Darboux's theorem.[21]

Riemannian metric

Let M be a differentiable manifold and let $g \in \Gamma(\otimes^2 T^*M)$ be a covariant 2-tensor field with $g(p) = (p, g_p)$. If, for all $p \in M$, the tensor g_p is symmetric

$$g_p(u_p, v_p) = g_p(v_p, u_p), \text{ for all } u_p, v_p \in T_pM$$

and positive definite

$$g_p(u_p, u_p) > 0, \text{ for all } 0 \neq u_p \in T_pM,$$

then g is called a **Riemannian metric**. A differentiable manifold M together with a Riemannian metric is referred to as **Riemannian manifold**

20. See Abraham and Marsden 1987, Theorem 3.2.10.
21. See Theorem 22.13 in John M. Lee 2013.

(M, g). The operation of a Riemannian metric g as a covariant 2-tensor field is defined by equation (3.58) as

$$g(u,v)(p) = (u \cdot g \cdot v)(p) = u_p \cdot g_p \cdot v_p$$

for all $u, v \in \mathrm{Vect}(M)$ and all $p \in M$.

A Riemannian metric g on a manifold M endows each tangent space with the inner product[22] given by

$$\begin{aligned} \langle \cdot, \cdot \rangle_p \colon T_pM \times T_pM &\to \mathbb{R}, \\ (u_p, v_p) &\mapsto \langle u_p, v_p \rangle_p := g_p(u_p, v_p). \end{aligned} \tag{3.70}$$

Indeed, it can be easily seen that the map (3.70) fulfils axioms IP 1–4. Given a chart (U, ϕ), the Riemannian metric g can be expressed using the dual basis (3.48) as

$$g = g_{kl}\, \mathrm{d}x^k \otimes \mathrm{d}x^l.$$

Its application to vector fields $u, v \in \mathrm{Vect}(M)$ can be written using the local expressions $u = u^i \partial/\partial x^i$ and $v = v^j \partial/\partial x^j$ as

$$g(u,v) = u \cdot g \cdot v = g_{ij} u^i v^j.$$

Moreover, the Riemannian metric provides at each point $p \in M$ the linear map

$$\begin{aligned} g_p \cdot \colon T_pM &\to T_p^*M, \\ u_p &\mapsto g_p \cdot u_p, \end{aligned}$$

which is an isomorphism between the tangent and the cotangent space at the point p. Actually, the isomorphism does not only hold point by point but a Riemannian metric defines a vector bundle isomorphism over M

$$\begin{aligned} g \cdot \colon TM &\to T^*M, \\ (p, u_p) &\mapsto (p, g_p \cdot u_p). \end{aligned} \tag{3.71}$$

As a covariant 2-tensor field, a Riemannian metric defines a $C^\infty(M)$-multilinear map on vector fields by Lemma 3.28, which defines a linear, bijective map on sections

$$g \cdot \colon \mathrm{Vect}(M) \to \Omega^1(M).$$

For details we refer to John M. Lee 2013, pp. 341–343.

22. See Section 2.7.

3.11. The Frobenius theorem

This section gives a brief account of the Frobenius theorem that is one of the central theorems in the theory of differentiable manifolds. Our presentation closely follows Chapter 19 of John M. Lee 2013 and we omit proofs because a comprehensive presentation of the subject would go beyond the scope of this work. The brevity comes at the price that this section may be difficult to understand because, by omitting the proofs, we renounce to present a substantial amount of mathematical reasoning that underpins the subject. For a detailed presentation, we suggest John M. Lee 2013, Chapter 19, Spivak 1999a, Chapter 6, or Jeffrey M. Lee 2009, Chapter 11.

Let M be a differentiable manifold. A **distribution $\varDelta$ of rank l on M** is a subbundle of rank l of the tangent bundle TM. This means that for each point $p \in M$ an l-dimensional (vector) subspace $\varDelta_p \subseteq T_pM$ of the tangent space T_pM is given that smoothly depends on the point p.

A straightforward approach to define a distribution of rank l on a differentiable manifold M is to specify an l-dimensional (vector) subspace $\varDelta_p \subseteq T_pM$ at each point of M and to let

$$\varDelta \coloneqq \bigcup_{p\in M} \left(\{p\} \times \varDelta_p\right). \tag{3.72}$$

If each point of M has a neighbourhood U on which there are smooth vector fields $\mathfrak{b}_1, \dots, \mathfrak{b}_l \colon U \to TM$ such that for all $q \in U$ the tangent vectors $(\mathfrak{b}_1|_q, \dots, \mathfrak{b}_l|_q)$ form a basis of $\varDelta_q$, then (3.72) defines indeed a subbundle of TM according to Theorem 3.24 and Lemma 3.36. We say that the vector fields $\mathfrak{b}_1, \dots, \mathfrak{b}_l$ **(locally) span** the distribution.

Lemma 3.36 (John M. Lee 2013, Lemma 10.32)**.** Let $(E, \pi, M, \mathbb{R}^k)$ be a vector bundle of rank k over M, and suppose that for each $p \in M$ we are given an l-dimensional (vector) subspace $D_p \subseteq E_p$. Then

$$D \coloneqq \bigcup_{p\in M} \left(\{p\} \times D_p\right)$$

is a subbundle of E if and only if each point of M has a neighbourhood U on which there exist smooth local sections $\mathfrak{b}_1, \dots, \mathfrak{b}_l \colon U \to E$ such that $(\mathfrak{b}_1|_q, \dots, \mathfrak{b}_l|_q)$ forms a basis for D_q at each point $q \in U$.

Suppose $\varDelta \subseteq TM$ is a distribution. A non-empty immersed submanifold $N \subseteq M$ with inclusion map $\iota \colon N \hookrightarrow M$ for which

$$\mathrm{D}\iota_p(T_pN) = \varDelta_p, \quad \text{for all } p \in N \tag{3.73}$$

is called an **integral manifold of** $\varDelta$. Let u and v be smooth local sections of $\varDelta$, i.e., local vector fields $u\colon U \to \varDelta$ and $v\colon U \to \varDelta$ defined on an open subset $U \subseteq M$ such that $u_p, v_p \in \varDelta_p$ for each $p \in U$. We say that $\varDelta$ is **involutive** if given any pair of local sections u and v of $\varDelta$, their Lie bracket $[\![u, v]\!]$ is also a local section of $\varDelta$. Figure 3.13 shows two distributions of rank two on $\mathbb{R}^3$. The left one is involutive, the right one is not.

The following lemma simplifies the check whether a given distribution of rank l is involutive or not. It says that not every pair of smooth sections of the distribution has to be checked, but that it is sufficient to check only a set of local sections that span the distribution in a neighbourhood of each point.

Lemma 3.37 (John M. Lee 2013, Lemma 19.4)**.** Let $\varDelta \subseteq TM$ be a distribution on a differentiable manifold M. If in a neighbourhood U of every point of M there exist local sections $v_1, \dots, v_l\colon U \to \varDelta$ such that for all $q \in U$ the tangent vectors $(v_1|_q, \dots, v_l|_q)$ form a basis of $\varDelta_q$ and for which $[\![v_I, v_J]\!]$ is a section of $\varDelta$ for each $I, J = 1, \dots, l$, then $\varDelta$ is involutive.

Instead of characterizing a distribution of rank l on an n-dimensional differentiable manifold M by local vector fields, it can equivalently be defined using differential forms. If around each point of M there is an open subset $U \subseteq M$ and $n-l$ linearly independent differential one-forms $\alpha^1, \dots, \alpha^{n-l}$, then these one-forms define a distribution $\varDelta \subseteq TM$ of rank l by

$$\varDelta_q \coloneqq \ker \alpha^1\big|_q \cap \dots \cap \ker \alpha^{n-l}\big|_q, \tag{3.74}$$

for all $q \in U$. The one-forms $\alpha^1, \dots, \alpha^{n-l}$ are called **locally defining forms for** $\varDelta$.

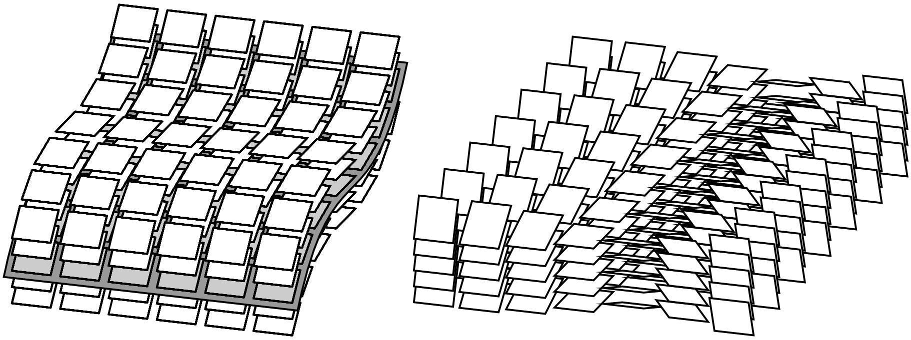
We say that an r-form $\beta \in \Omega^r(M)$ with $0 \le r \le n$ **annihilates the distribution** $\varDelta$ if

$$\beta(v_1, \dots, v_r) = 0$$

whenever $v_1, \dots, v_r$ are local sections of $\varDelta$. In the case $r = 0$, only the zero function annihilates $\varDelta$. Locally defining forms for $\varDelta$ annihilate the distribution by definition. Now, the criterion whether a given distribution is involutive or not can be formulated in terms of differential forms.

Theorem 3.38 (John M. Lee 2013, Theorem 19.7)**.** Suppose $\varDelta$ is a distribution. Then $\varDelta$ is involutive if and only if the following condition is satisfied: For each one-form $\alpha \in \Omega^1(U)$ defined on an open subset $U \subseteq M$ that annihilates the distribution, the form $d\alpha$ also annihilates the distribution on U.

As for the Lie bracket condition for involutivity, the condition from Theorem 3.38 needs only to be checked for a particular set of smooth defining



(a) An involtive distribution. (b) A non-involutive distribution.

Figure 3.13.: Two different distributions of rank two on $\mathbb{R}^3$.

forms in a neighbourhood of each point. So we get the following proposition as equivalent to Lemma 3.37 for differential forms.

Proposition 3.39 (John M. Lee 2013, Proposition 19.8)**.** Let Δ be a distribution of rank l on an n-dimensional differentiable manifold M and let $\alpha^1, \dots, \alpha^{n-l}$ be locally defining forms for Δ on an open subset $U \subseteq M$. Then the distribution Δ is involutive on U if and only if the forms $\mathrm{d}\alpha^1, \dots, \mathrm{d}\alpha^{n-l}$ annihilate Δ.

Let Δ be a distribution of rank l on an n-dimensional differentiable manifold M. We say that the distribution Δ is **integrable** if each point of M is contained in an integral manifold of Δ.

Proposition 3.40 (John M. Lee 2013, Proposition 19.3)**.** Every integrable distribution is involutive.

We know from (3.46), that a chart (U, ϕ) of M induces coordinate fields that locally span the tangent bundle TM. Now, we call a chart (U, ϕ) **flat for** Δ if $\phi(U)$ is a cube in $\mathbb{R}^n$, and if at points of U, the distribution Δ is spanned by the first l coordinate fields ${\partial}/{\partial x^1}, \dots, {\partial}/{\partial x^l}$. In any such chart, each slice of the form $x^{l+1} = c^{l+1}, \dots, x^n = c^n$ for constants $c^{l+1}, \dots, c^n$ is an integral manifold of Δ. This is the nicest possible local situation for integral manifolds. We say that a distribution $\Delta \subseteq TM$ is **completely integrable** if there exists a flat chart for Δ in a neighbourhood of each point of M. Obviously, every completely integrable distribution is integrable and therefore involutive by Proposition 3.40. In summary, this means that

$$\text{completely integrable} \;\Rightarrow\; \text{integrable} \;\Rightarrow\; \text{involutive}.$$

The Frobenius theorem (see John M. Lee 2013, Theorem 19.12) says that the implications are actually equivalences, such that

$$\text{completely integrable} \iff \text{integrable} \iff \text{involutive.}$$

Theorem 3.41 (Frobenius)**.** Every involutive distribution is completely integrable.

Considering all the maximal integral manifolds of an involutive distribution of rank l on a differentiable manifold M, we obtain a partition of M into l-dimensional submanifolds that "fit together" locally like the slices in a flat chart. To express more precisely what we mean by "fitting together," we need to extend our notion of a flat chart slightly. Let M be an n-dimensional differentiable manifold, and let $\mathfrak{F}$ be any collection of l-dimensional submanifolds of M. A chart (U, ϕ) of M is said to be **flat for** $\mathfrak{F}$ if $\phi(U)$ is a cube in $\mathbb{R}^n$, and each submanifold in $\mathfrak{F}$ intersects U in either the empty set or a countable union of l-dimensional slices of the form $x^{l+1} = c^{l+1}, \dots, x^n = c^n$. We define a **foliation of dimension** l **on** M to be a collection $\mathfrak{F}$ of disjoint, connected, nonempty, immersed l-dimensional submanifolds of M (called the **leaves of the foliation**), whose union is M, and such that in a neighbourhood of each point $p \in M$ there exists a flat chart for $\mathfrak{F}$. The involutive distribution shown in Figure 3.13a defines a two-dimensional foliation on $\mathbb{R}^3$. One of its leafs is depicted in dark grey.

There is a one-to-one correspondence between involutive distributions and foliations. One direction is expressed by the following proposition.

Proposition 3.42 (John M. Lee 2013, Proposition 19.19)**.** Let $\mathfrak{F}$ be a foliation on a differentiable manifold M. The collection of tangent spaces to the leaves of $\mathfrak{F}$ forms an involutive distribution on M.

The converse is provided by the global Frobenius theorem (see Theorem 19.21 in John M. Lee 2013).

Theorem 3.43 (Global Frobenius)**.** Let Δ be an involutive distribution on a differentiable manifold M. The collection of all maximal connected integral manifolds of Δ forms a foliation of M.

4 Finite-dimensional mechanical systems

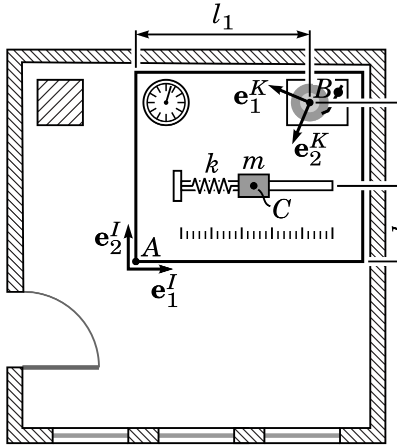
Theorien sind nicht verifizierbar;
aber sie können sich bewähren.
— Karl Popper

This chapter presents a physical theory for the description of mechanical systems with finitely many degrees of freedom. The presentation is based on Gallissot 1952, Godbillon 1969, Souriau 1970, Dombrowski et al. 1964a, Dombrowski et al. 1964b, Loos 1982 and Loos 1985. The first part of the presentation strongly follows Loos 1982.

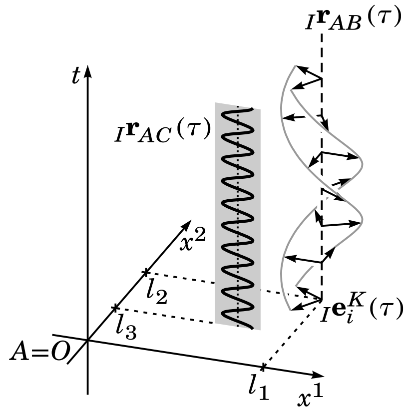
4.1. On axioms, postulates and the role of experiments

We try to achieve a clear distinction between the three fields: mathematics, mechanics (or physics in general), and experiments, that is, the observation of certain phenomena in the real world. So far, we have introduced mathematical terminology in the Chapters 2–3 and we hope that the reader digested it well. We use the language of differential geometry to formulate a physical theory that can describe finite-dimensional mechanical systems. In our presentation, we pay attention not to overload terminology in the sense that one designation may refer to multiple objects from different fields. An example for this common practice is given by the overloaded use of 'axiom'. In mathematics, axiom designates an unprovable statement that serves as the basis for mathematical reasoning. In physics, axiom is used to designate a physical law that is placed at the basis of a physical theory and whose validity relies on the fact that is has not (yet) been proved wrong by an experiment.[1] The notion of axiom in physics differs from its counterpart in mathematics by the fact that physical axioms are constantly put to the test by experiments, while the validity of axioms is meaningless in mathematics. In mechanics, for example, one speaks of Newton's axioms[2] of motion. However, in the domain of physics, we prefer the word 'postulate' to 'axiom'. We use axiom only in the context of mathematics.

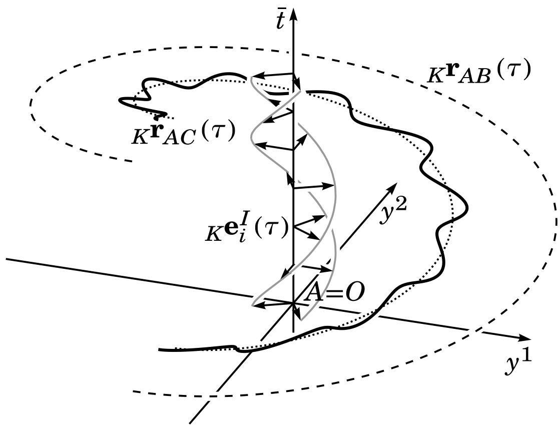
1. Following Popper 1935, an experiment cannot prove a physical theory to be true; but the latter can be falsified by an experiment.
2. See p. 19 of Newton 1729 — the English translation of Newton's Principia.



(a) Layout of the laboratory with the experimental setup.



(b) Motion of the oscillator expressed in the $(A, \mathbf{e}_1^I, \mathbf{e}_2^I)$-coordinate system defined by the lower left corner A of the clamping table and by the directions $\mathbf{e}_1^I$ and $\mathbf{e}_2^I$ defined by the edges of the table that join in A.



(c) Motion of the oscillator expressed in the $(A, \mathbf{e}_1^K, \mathbf{e}_2^K)$-coordinate system defined by the lower left corner A of the clamping table and by the body-fixed directions $\mathbf{e}_1^K$ and $\mathbf{e}_2^K$ rotating with the disc record.

Figure 4.1.: Example an experimental setup consisting of a mass-spring-oscillator mounted on a clamping table.

Concerning the question whether a certain physical theory is correct or not, the best we can do is to carry out experiments within the domain for which the theory is devised. The theory is considered to be valid as long as the predictions it provides are in accordance with the experimental observations. This means that in contrast to pure mathematics, in a physical theory the interpretation of the involved mathematical objects in regard to experiments needs to be specified.

In the following, we will study the illustrative example of Simon's[3] laboratory. It allows to reflect on the relation between an abstract mathematical model and its physical interpretation. Figure 4.1a shows the layout of the laboratory in which Simon studies the motion of an oscillator that consists of a block of mass m that is mounted on a rail such that it can move translationally with negligible friction. The mass is attached to a support by a spring with stiffness k. Simon can measure time using a chronometer and he is able to measure distances. He wants to study the motion of the oscillator. For this purpose, he initially pulls the mass m to the right from its equilibrium position and lets it go. At the moment he releases the mass with no velocity, Simon starts the chronometer to measure time. Since the oscillator has only one degree of freedom, its motion can be captured by observing the point C on the moving block. Simon decides to describe the position of C with respect to the clamping table on which the oscillator is mounted. He chooses the lower left corner (point A) as reference point and measures parallelly to the edges of the table that intersect in A (the I-frame in Figure 4.1a). At each time instant at which Simon measures, he determines two real numbers x^1 and x^2 such that the relative position of the point C with respect to A is given by

$$\mathbf{r}_{AC} = x^1 \mathbf{e}_1^I + x^2 \mathbf{e}_2^I. \tag{4.1}$$

Figure 4.1b shows a visualization of his measurements x^1 and x^2 with respect to the $(A, \mathbf{e}_1^I, \mathbf{e}_2^I)$-coordinate system. Instead of referring his position measurements to the I-frame, Simon could have used the K-frame fixed to the disc record he is playing while doing his measurements to express the relative position of the point C with respect to A as

$$\mathbf{r}_{AC} = y^1 \mathbf{e}_1^K + y^2 \mathbf{e}_2^K. \tag{4.2}$$

The coordinate systems $(A, \mathbf{e}_1^I, \mathbf{e}_2^I)$ and $(A, \mathbf{e}_1^K, \mathbf{e}_2^K)$ are related as follows:

$$\begin{aligned} \mathbf{e}_1^K &= \cos(\Omega t + \varphi_0)\mathbf{e}_1^I - \sin(\Omega t + \varphi_0)\mathbf{e}_2^I, \\ \mathbf{e}_2^K &= \sin(\Omega t + \varphi_0)\mathbf{e}_1^I + \cos(\Omega t + \varphi_0)\mathbf{e}_2^I, \end{aligned} \tag{4.3}$$

3. Simon was the most popular name during my time at the Institute for Nonlinear Mechanics.

where $\Omega = {}^{10}/_{9}\,\pi$ rad/s denotes the constant rotational speed of the record[4] and φ_0 is the angle describing the orientation of the I-frame with respect to the K-frame at the beginning of Simon's measurements (i.e., at time $t = 0$). By equation (4.3), it holds that

$$
\begin{aligned}
\bar{t} &= t, \\
y^1 &= x^1 \cos(\Omega t + \varphi_0) - x^2 \sin(\Omega t + \varphi_0), \\
y^2 &= x^1 \sin(\Omega t + \varphi_0) + x^2 \cos(\Omega t + \varphi_0).
\end{aligned}
\tag{4.4}
$$

Figure 4.1c depicts the corresponding results obtained with respect to the coordinate system $(A, \mathbf{e}_1^K, \mathbf{e}_2^K)$.

We observe from the example that time can only be measured *relatively* with respect to some chosen reference (the instant when Simon starts the chronometer). A first assumption underlying classical mechanics is that by our time measurements, we are able to decide whether two events happen at the same time or not irrespective of our motion as observer (in contrast to Einstein's theory of special relativity). Therefore, we can consider at each instant of time the set of synchronous events happening at that specific moment. In a laboratory, spatial measurements of distances and angles are realized between synchronous events (at the instant of measurement). These measurements allow the characterization of synchronous events relative to each other. It is not possible to relate events happening at different instants of time by spatial measurements. In a space-time context, the points A and C have to be seen as a collection of events such that for each time t (respectively $\bar{t}$), we get a pair A and C of synchronous events at time t (respectively $\bar{t}$). At a given time t, the corresponding pair of synchronous events A and C is related by the vector $\mathbf{r}_{AC}(t)$ from equation (4.1) or (4.2). Therefore, the tuples (x^1, x^2) and (y^1, y^2) are local coordinates on the space of synchronous events. The number t (respectively $\bar{t}$) obtained by reading the chronometer can be considered as an additional (independent) coordinate describing time.

Our objective is to mathematically describe finite-dimensional mechanical systems in order to predict their motion accurately with respect to experiments. For the example, this means that we want to derive equations which describe the evolution over time of the tuples (x^1, x^2) or (y^1, y^2), respectively. The resulting equations should be formulated independently of the choice of a particular set of coordinates. We start with a mathematical abstraction of the notion of space and time.

4. *Another One Bites the Dust* by *Queen* on a 16 inch, $33^1/_3$ rpm disc.

4.2. Space-time

In the example, we only considered two spatial dimensions instead of three because otherwise Figures 4.1b and 4.1c would have become four-dimensional and thus impossible to visualize. In general, however, we know that a spatial point in a laboratory can be described using three local coordinates (real numbers). If we want to keep track of time, we need to add a fourth coordinate. We model **space-time** as a four-dimensional smooth manifold $\mathcal{S}$. Points in $\mathcal{S}$ are referred to as **events**.

In mathematical terms, our observation that the relative time between two events can be measured means that a real number, the **time duration**, can be associated to each pair (p,q) of events $p,q \in \mathcal{S}$ (see Dombrowski et al. 1964a). This can be modelled by postulating the existence of a function

$$\varDelta\colon \mathcal{S}\times\mathcal{S} \to \mathbb{R},\ (p,q) \mapsto \varDelta(p,q) \tag{4.5}$$

which satisfies

$$\varDelta(p,q) + \varDelta(q,r) = \varDelta(p,r) \tag{4.6}$$

for all events $p,q,r \in \mathcal{S}$. If we choose an event $p \in \mathcal{S}$ as reference,[5] we can use (4.5) to define a temporal distance to all other events $q \in \mathcal{S}$, i.e., a map

$$t_p\colon \mathcal{S} \to \mathbb{R},\ q \mapsto t_p(q) := \varDelta(p,q). \tag{4.7}$$

The map (4.7) provides a (global) time coordinate. Its kernel

$$\ker t_p := \left\{q \in \mathcal{S} \mid t_p(q) = 0\right\}$$

is the set of all events happening at the same time as the event p. For each value of the time coordinate t, we have defined the space of synchronous events at t. By property (4.6), the time coordinates of two events q and r differ only by an additive constant. Therefore, the time coordinate (globally) defines the one-form $\vartheta := \mathrm{d}t$, where d denotes the exterior derivative, and the function (4.5) is given by

$$\varDelta(p,q) = \int_p^q \vartheta,$$

which is independent of the path of integration between p and q. We refer to Chapter 16 in John M. Lee 2013 for the theory about integration on manifolds.

Following Dombrowski et al. 1964a, we generalize the above considerations to finite-dimensional mechanical systems with n degrees of freedom.

5. For example, the event where Simon starts the chronometer.

Therefore, we consider an $(n+1)$-dimensional manifold M instead of the four-dimensional space-time $\mathcal{S}$. Rather than postulating the existence of an analogue function as (4.5) on M, we assume M to be endowed with a one-form ϑ which is closed and nowhere zero. We refer to the one-form ϑ as **time structure** on M. The time structure defines local **time functions** $t\colon M \supseteq U \to \mathbb{R}$ such that $\mathrm{d}t = \vartheta|_U$. Their existence around each point of M is guaranteed by the Poincaré lemma (see Lemma 3.32). The temporal distance of two events $p, q \in U$ is given by $t(q) - t(p)$. The requirement that ϑ is nowhere zero means that the local time functions have no stationary points such that time "passes". Note that the naive approach of defining the time function to be the first coordinate in a given local chart of M does not provide a definition that is invariant under changes of coordinates.

Let (U, ϕ) be a chart of M, such that

$$\phi\colon M \supseteq U \to \mathbb{R}^{n+1},\ p \mapsto \phi(p) = (x^0, \dots, x^n). \tag{4.8}$$

We say that the chart (4.8) is **adapted** to the time structure if $\vartheta|_U = \mathrm{d}x^0$. In this case, the coordinate x^0 is a local **time coordinate** and we will often use t instead of x^0 to denote it. In what follows, we will restrict our considerations to adapted charts. The existence of adapted charts is guaranteed by the existence of time functions and the fact that ϑ does not vanish. Therefore, the adapted charts provide an atlas of M. Let

$$\phi\colon M \supseteq U \to \mathbb{R}^{n+1},\ p \mapsto (x^0, \dots, x^n) \tag{4.9}$$

and

$$\psi\colon M \supseteq V \to \mathbb{R}^{n+1},\ p \mapsto (y^0, \dots, y^n)$$

be two adapted charts of M with $U \cap V \neq \emptyset$, then their coordinate change

$$\psi \circ \phi^{-1}\colon \phi(U \cap V) \to \psi(U \cap V)$$

is given by

$$\begin{aligned} y^0 &= x^0 + \text{const.}, \\ y^i &= \psi^i \circ \phi^{-1}(x^0, \dots, x^n),\ i = 1, \dots, n, \end{aligned}$$

where $\psi^i\colon V \to \mathbb{R}$ denotes the i-th coordinate function of the chart ψ. Coming back to Simon's oscillator, we observe that the coordinates (t, x^1, x^2) and $(\bar{t}, y^1, y^2)$ from Figures 4.1b and 4.1c can be interpreted as being provided by two adapted charts of a three-dimensional space-time manifold.

The $(n+1)$-dimensional manifold M is foliated[6] by the time structure. Indeed, the time structure can be used to single out spacelike tangent

6. See p. 89 for the definition of a foliation.

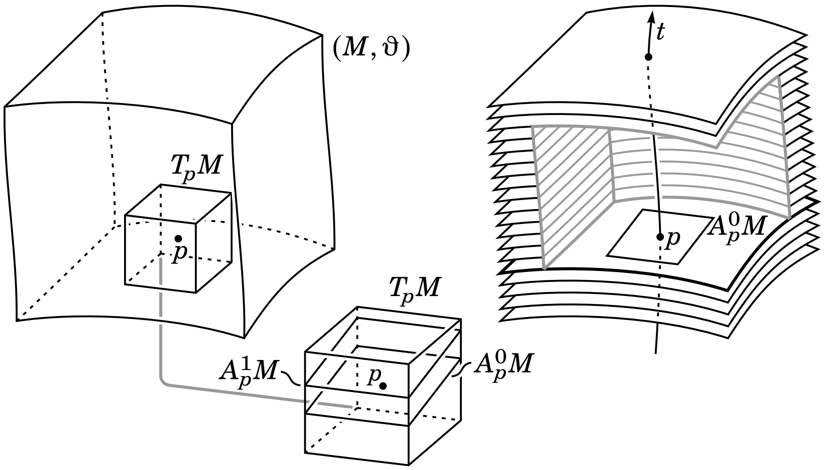


Figure 4.2.: Foliation of a $(2+1)$-dimensional manifold M by its time structure ϑ. At each point $p \in M$, the time structure defines the subspace A^0_pM and the *affine* subspace A^1_pM of the tangent space T_pM.

vectors, i.e., tangent vectors $v_p \in T_pM$ that do not have a component in time direction such that $\vartheta_p(v_p) = 0$. We introduce the **space of spacelike vectors in** $p \in M$ as

$$A^0_pM := \ker \vartheta_p = \left\{ v_p \in T_pM \mid \vartheta_p(v_p) = 0 \right\} \subset T_pM \tag{4.10}$$

and the corresponding subbundle of the tangent bundle TM to the generalized spacetime M as

$$A^0M := \bigcup_{p \in M} \left(\{p\} \times A^0_pM \right) \subset TM \tag{4.11}$$

and we call it the **spacelike bundle**. Indeed, it holds by Lemma 3.36 that A^0M is a subbundle of the vector bundle[7] TM because each chart (U, ϕ) of an adapted atlas of M induces the smooth local sections

$$\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n} : U \to TM$$

that provide a basis for A^0_qM at each $q \in U$. Therefore, A^0M is a distribution of rank n defined by the time structure ϑ. This distribution is involutive by Theorem 3.38 because $\mathrm{d}\vartheta = 0$ annihilates the distribution trivially.

7. The tangent bundle TM is a vector bundle according to Theorem 3.24.

By the Frobenius theorem (Theorem 3.41), the distribution (4.11) is completely integrable. Moreover, A^0M defines a foliation according to the global Frobenius theorem (Theorem 3.43). The leafs of this foliation are just the submanifolds with codimension[8] one of synchronous events that can be distinguished in classical mechanics. See Figure 4.2 for a visualization of the $(2+1)$-dimensional case.

At each point $p \in M$, the basis vectors

$$\left.\frac{\partial}{\partial x^1}\right|_p, \dots, \left.\frac{\partial}{\partial x^n}\right|_p \tag{4.12}$$

induced by an adapted chart such as (4.9) form a basis of A^0_pM. Note that any adapted chart induces such a basis of A^0_pM. We equip the bundle A^0M with a bundle metric

$$g = g_{ij}\,\mathrm{d}x^i \otimes \mathrm{d}x^j, \text{ with } g_{ij} = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right), \tag{4.13}$$

i.e., the tensor g_p is symmetric and positive definite for each $p \in M$. In Section 3.10, we saw that a Riemannian metric on a manifold endows the tangent spaces with an inner product. Now, if the fibres of a vector bundle are equipped with an inner product that smoothly depends on the point in the base manifold, one speaks of a **bundle metric**.[9] A bundle metric is the generalization of a Riemannian metric on a manifold to arbitrary vector bundles. Indeed, a Riemannian metric on a manifold is just a bundle metric on its tangent bundle. For this reason some authors[10] designate a bundle metric as *Riemannian metric*. We abstain from doing so since it might lead to confusion.

We remind the reader that by Einstein's summation convention a summation from 1 to n is understood in (4.13) over the repeated indices i and j that appear once as a lower and once as an upper index. The $\mathrm{d}x^k$ in (4.13) denote the dual vectors to the basis vectors $\partial/\partial x^l$ such that

$$\mathrm{d}x^k\left(\frac{\partial}{\partial x^l}\right) = \delta^k_l,$$

where δ^k_l denotes the Kronecker delta that equals one if $k = l$ and is zero else (see p. 29). The above construction can be summarized in the following definition.

8. See Section 3.3.
9. See Definition 1.8.11 in Jost 2008.
10. See Definition 6.42 in Jeffrey M. Lee 2009 or p. 308 in Spivak 1999a.

Definition 4.1 (Loos 1982, pp. 5–6)**.** An $(n+1)$-dimensional smooth manifold M with a time structure ϑ and a bundle metric g that endows the subspaces $A_p^0 M$ with an inner product for all $p \in M$ is called a **Galilean manifold** and it is denoted (M, ϑ, g).

4.3. State space and motion

In each point $p \in M$, the **affine space**[11] **of time-normalized vectors in** p (see Figure 4.2) is defined as

$$A_p^1 M := \left\{ v_p \in T_p M \mid \vartheta_p(v_p) = 1 \right\} \subset T_p M. \tag{4.14}$$

While $A_p^0 M$ is a vector subspace of $T_p M$, the set $A_p^1 M$ is an affine subspace of $T_p M$ such that the resulting bundle

$$A^1 M := \bigcup_{p \in M} \left(\{p\} \times A_p^1 M \right) \subset TM \tag{4.15}$$

is an affine subbundle of TM. The affine bundle (4.15) of time-normalized vectors is referred to as **state space**.[12]

The coordinate fields induced by an adapted chart $\phi \colon p \mapsto (x^0, \dots, x^n)$ (with $x^0 = t$) as the one from (4.9) can be used to express a time-normalized vector $v_p \in A_p^1 M$ as

$$v_p = \left.\frac{\partial}{\partial t}\right|_p + u^i \left.\frac{\partial}{\partial x^i}\right|_p. \tag{4.16}$$

Therefore, any adapted chart $\phi \colon M \supseteq U \to \mathbb{R}^{n+1}$ induces a corresponding **natural chart** of the state space $A^1 M$ as

$$\begin{aligned} \Phi \colon A^1 M \supseteq \pi^{-1}(U) &\to \mathbb{R}^{2n+1}, \\ (p, v_p) &\mapsto (t, x^1, \dots, x^n, u^1, \dots, u^n), \end{aligned} \tag{4.17}$$

where

$$\pi \colon A^1 M \to M, \;\; (p, v_p) \mapsto p \tag{4.18}$$

denotes the natural projection of the affine bundle $A^1 M$. Note that the state space $A^1 M$ is canonically endowed with the time structure

$$\hat{\vartheta} := \pi_{\leftarrow} \vartheta \tag{4.19}$$

11. See Section 2.9.
12. See Section 1.2 for a comment about alternative designations.

that is the pullback of the time structure of M with the natural projection (4.18). The natural chart (4.17) is an adapted chart with respect to the time structure (4.19) of A^1M because it holds that

$$\hat{\vartheta}|_{\pi^{-1}(U)} = \mathrm{d}t.$$

A curve

$$\gamma\colon \mathbb{R} \supseteq I \to M, \ \ \tau \mapsto \gamma(\tau) \tag{4.20}$$

in the Galilean manifold (M, ϑ, g) is just a smooth sequence of events. We call the curve (4.20) **time-parametrized** if $\vartheta(\dot{\gamma}) = 1$, where $\dot{\gamma}$ denotes the tangent field[13] along γ. The local time coordinate t increases monotonically along a time-parametrized curve because locally

$$1 = \vartheta\big(\dot{\gamma}\big) = \mathrm{d}t\big(\dot{\gamma}\big) = \dot{\gamma}[t] = \frac{\mathrm{d}}{\mathrm{d}\tau}(t \circ \gamma(\tau)). \tag{4.21}$$

Therefore, time does not decrease along a time-parametrized curve (see Figure 4.3). Condition (4.21) means the time-evolution along the motion is an affine function of the curve parameter τ, i.e.,

$$t \circ \gamma(\tau) = \tau + \tau_0,$$

where $\tau_0 \in \mathbb{R}$ is a constant. In words, a change in the parameter τ corresponds to the change in the time function t along the time-parametrized curve.

By its tangent field

$$\dot{\gamma}\colon I \to A^1M, \ \ \tau \mapsto \dot{\gamma}(\tau) = \big(\gamma(\tau), \dot{\gamma}_{\gamma(\tau)}\big) \tag{4.22}$$

a time-parametrized (w.r.t. ϑ) curve $\gamma\colon \mathbb{R} \supseteq I \to M$ defines a curve in the state space A^1M that consists of the curve $\gamma\colon I \to M$ together with the vector field (of time-normalized vectors) it induces along $\gamma(I)$. The curve (4.22) is a special type of curve in A^1M to which we refer as **second-order curve** or **motion** of a mechanical system.

To arrive at a coordinate-free characterization of second-order curves, we consider that an arbitrary curve

$$\beta\colon I \to A^1M, \ \ \tau \mapsto \beta(\tau)$$

through the state space A^1M defines, by projection, a curve

$$\alpha := \pi \circ \beta\colon I \to M \tag{4.23}$$

13. See Section 3.6.

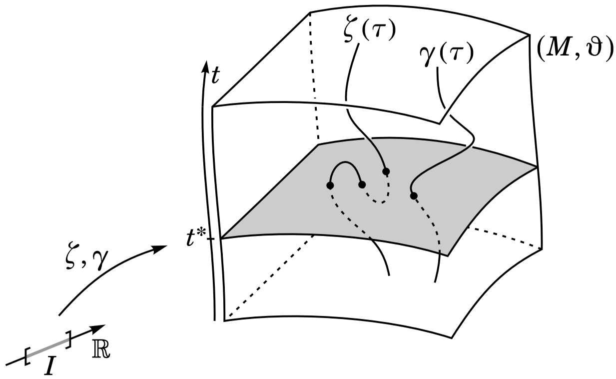


Figure 4.3.: While the curve $\gamma\colon \tau \mapsto \gamma(\tau)$ is time-parametrized, the curve $\zeta\colon \tau \mapsto \zeta(\tau)$ is *not* because $t \circ \zeta$ is not monotonically increasing as required by (4.21).

in the base manifold M. Now, the curve $\beta\colon I \to A^1M$ is a second-order curve if it is time-parametrized (with respect to $\hat{\vartheta}$) and satisfies the condition

$$\beta \overset{!}{=} \dot{\alpha} = (\pi \circ \beta)^{\dot{}} \tag{4.24}$$

such that the curve $\beta\colon I \to A^1M$ corresponds to the (time-normalized) tangent field along its (time-normalized) projection (4.23) onto the base manifold M. By condition (4.24) it follows that

$$\beta = \dot{\alpha}\colon I \to A^1M,\ \tau \mapsto \big(\alpha(\tau), \dot{\alpha}_{\alpha(\tau)}\big).$$

Condition (4.24) can be expressed in the local coordinates of the natural chart (4.17), as

$$\Phi \circ \beta(\tau) \overset{!}{=} \Phi\Big((\pi \circ \beta)^{\dot{}}(\tau)\Big) = \Phi\Big(\pi \circ \beta(\tau), (\pi \circ \beta)^{\dot{}}_{\pi \circ \beta(\tau)}\Big)$$

and, consequently,

$$(t(\tau), \mathbf{x}(\tau), \mathbf{u}(\tau)) \overset{!}{=} (t(\tau), \mathbf{x}(\tau), \dot{\mathbf{x}}(\tau)), \tag{4.25}$$

where the coordinates $x^1, \dots, x^n$ and $u^1, \dots, u^n$ are gathered as $\mathbb{R}^n$-tuples $\mathbf{x} \coloneqq (x^1, \dots, x^n)$ and $\mathbf{u} \coloneqq (u^1, \dots, u^n)$, respectively.

An arbitrary curve $\beta\colon I \to A^1M$ in the state space A^1M is the integral curve[14] of a vector field $X \in \mathrm{Vect}(W)$ defined on a neighbourhood $W \subseteq A^1M$ if

$$\dot{\beta}(\beta(\tau)) = X(\beta(\tau)). \tag{4.26}$$

14. See Section 3.6.

Equation (4.26) is just an ordinary differential equation in first-order form, which reads in the natural chart (4.17) as

$$\begin{aligned}\dot t(\tau) &= a(t(\tau),\mathbf{x}(\tau),\mathbf{u}(\tau)),\\ \dot{\mathbf{x}}(\tau) &= \mathbf{A}(t(\tau),\mathbf{x}(\tau),\mathbf{u}(\tau)),\\ \dot{\mathbf{u}}(\tau) &= \mathbf{B}(t(\tau),\mathbf{x}(\tau),\mathbf{u}(\tau)).\end{aligned} \tag{4.27}$$

In equation (4.27), aside from the local coordinates $x^1, \dots, x^n$ and $u^1, \dots, u^n$ we have also gathered the chart representations of the coefficient functions $a, A^1, \dots, A^n, B^1, \dots, B^n \in C^\infty(W)$ of the vector field

$$X = a\frac{\partial}{\partial t} + A^i\frac{\partial}{\partial x^i} + B^i\frac{\partial}{\partial u^i} \tag{4.28}$$

as tuples $\mathbf{A} := (A^1, \dots, A^n)$ and $\mathbf{B} := (B^1, \dots, B^n)$, respectively.

The integral curve $\beta\colon I \to A^1M$ is time-parametrized, i.e., $\hat\vartheta(\dot\beta) = 1$ if the vector field satisfies $\hat\vartheta(X) = 1$. This means that $a = 1$ in (4.27) and (4.28). If an integral curve $\beta\colon I \to A^1M$ of a local vector field $Z \in \mathrm{Vect}(W)$ with $W \subseteq A^1M$, i.e.,

$$\dot\beta(\beta(\tau)) = Z(\beta(\tau)) \tag{4.29}$$

should be a second-order curve, then the vector field Z cannot be arbitrary. First, the latter needs to be time-normalized such that

$$\hat\vartheta(Z) = 1. \tag{4.30}$$

Second, the vector field Z needs to obey the second-order condition

$$\mathrm{D}\pi \circ Z = \mathrm{id}_W. \tag{4.31}$$

Indeed, condition (4.24) together with (4.29) lead to

$$\beta = (\pi\circ\beta)^{\cdot} = \mathrm{D}\pi\circ\dot\beta = \mathrm{D}\pi\circ Z\circ\beta, \tag{4.32}$$

where $\mathrm{D}\pi\colon T(A^1M) \to TM$ denotes the differential[15] of the natural projection (4.18). Because condition (4.32) has to hold for arbitrary integral curves $\beta\colon I \to A^1M$, the second-order condition (4.31) follows.

A vector field $Z \in \mathrm{Vect}(W)$ on $W \subseteq A^1M$ that satisfies conditions (4.30) and (4.31) is called a **second-order (vector) field**. Second-order fields can be equivalently characterized using local coordinates by saying that a vector field $Z \in \mathrm{Vect}(W)$ on $W \subseteq A^1M$ is a second-order field if it can be expressed in every natural chart (4.17) with $W \cap \pi^{-1}(U) \neq \emptyset$ as

$$Z|_{W\cap\pi^{-1}(U)} = \frac{\partial}{\partial t} + u^i\frac{\partial}{\partial x^i} + Z^i\frac{\partial}{\partial u^i}, \tag{4.33}$$

15. See p. 71 for the definition.

with n smooth real-valued functions Z^i defined on $W \cap \pi^{-1}(U)$. It can be seen from the local expression (4.33), that second-order fields can only differ by the coefficients of their $\partial/\partial u^i$ part. Moreover, the differential equation (4.29) related to a second-order field is a second-order differential equation in first-order form

$$\begin{aligned} \dot{t}(\tau) &= 1, \\ \dot{\mathbf{x}}(\tau) &= \mathbf{u}(\tau), \\ \dot{\mathbf{u}}(\tau) &= \mathbf{B}(t(\tau), \mathbf{x}(\tau), \mathbf{u}(\tau)). \end{aligned} \tag{4.34}$$

The first equation of (4.34) can be solved to

$$t(\tau) := t \circ \beta(\tau) = \tau + \tau_0, \tag{4.35}$$

where $\tau_0 \in \mathbb{R}$ denotes again a constant. The second and third equation of (4.34) are equivalent to the second-order differential equation

$$\ddot{\mathbf{x}}(\tau) = \mathbf{B}(t(\tau), \mathbf{x}(\tau), \dot{\mathbf{x}}(\tau)).$$

In the study of finite-dimensional mechanical systems, we are interested in modelling the second-order field Z rather than its integral curves (4.29). Indeed, if a vector field is determined, all its integral curves are known for arbitrary initial conditions. Differential forms are particularly useful for the characterization of vector fields. With the time structure ϑ, we have already used a differential one-form to define the sets of spacelike (4.10) and of time-normalized vectors (4.14) on M, respectively. Furthermore, we have used the pullback $\hat{\vartheta}$ of the time structure ϑ on M to characterize time-normalized vector fields on A^1M (see equation (4.30)). From the local expression (4.33), we deduce a characterization of second-order fields using differential forms. We define the local one-forms $\theta^1, \dots, \theta^n \in \Omega^1(\pi^{-1}(U))$ as

$$\theta^i := \mathrm{d}x^i - u^i \mathrm{d}t, \qquad \text{with} \quad i = 1, \dots, n \tag{4.36}$$

and formulate the second-order condition as

$$Z \in \ker(\theta^1) \cap \dots \cap \ker(\theta^n) \quad \text{and} \quad \hat{\vartheta}(Z) = 1,$$

i.e., on $\pi^{-1}(U) \subseteq A^1M$ the vector field Z needs to be time-normalized and it has to lie in the distribution defined[16] by the differential one-forms (4.36). The remaining n free coefficients in the local representation of Z can be prescribed by requiring Z to lie in the distribution defined by the n one-forms

$$\lambda^i := \mathrm{d}u^i - Z^i \mathrm{d}t, \qquad \text{with} \quad i = 1, \dots, n.$$

16. The concept of defining forms of a distribution is treated on p. 87.

Vector and covector fields (or one-forms) on the state space A^1M are sections of the bundles $T(A^1M)$ and $T^*(A^1M)$, respectively. Therefore, we start by studying the geometric structure of these two vector bundles following Loos 1982.

4.4. Galilean manifolds and their related bundles

The differential of the natural projection (4.18), $D\pi\colon T(A^1M) \to TM$, defines the subbundle[17]

$$\mathrm{Ver}(A^1M) := \ker D\pi = \bigcup_{a \in A^1M} \big(\{a\} \times \ker D\pi_a\big) \tag{4.37}$$

of the tangent bundle $T(A^1M)$ that we call the **vertical bundle**. For any point $a \in A^1M$ the **space of vertical vectors in** a is given by

$$\mathrm{Ver}_a(A^1M) := \ker D\pi_a = \big\{ w \in T_a(A^1M) \mid D\pi_a(w) = 0 \big\}. \tag{4.38}$$

A section

$$V \in \Gamma\big(\mathrm{Ver}(A^1M)\big)$$

of the vertical bundle is called a **vertical vector field**. By definition (4.37), a vertical vector field is a vector field on A^1M which is π-related to the zero vector field on M. Let (U, ϕ) be an adapted chart of M and consider the corresponding natural chart (4.17) on the neighbourhood $\pi^{-1}(U)$ of A^1M. Then a vertical vector field V can be expressed on $\pi^{-1}(U)$ with respect to the coordinate fields induced by the natural chart as

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

Indeed, for points $a \in \pi^{-1}(U) \subseteq A^1M$, the vectors

$$\frac{\partial}{\partial u^1}\bigg|_a, \dots, \frac{\partial}{\partial u^n}\bigg|_a \tag{4.39}$$

provide a basis of $\mathrm{Ver}_a(A^1M)$ and, therefore, it holds that

$$\mathrm{Ver}_a(A^1M) = \mathrm{span}\left\{ \frac{\partial}{\partial u^1}\bigg|_a, \dots, \frac{\partial}{\partial u^n}\bigg|_a \right\}.$$

17. Indeed, by Theorem 3.24 the tangent bundles $T(A^1M)$ and TM are vector bundles. The pair of maps $(D\pi, \pi)$ provides a vector bundle homomorphism of constant rank between $T(A^1M)$ and TM because the projection is a surjective submersion. Therefore, the vertical bundle is a subbundle of $T(A^1M)$ by Proposition 3.23.

The vertical subbundle (4.37) naturally appears in the study of second-order fields because the difference of two second-order fields is always a vertical vector field as can be seen from the local expression (4.33) of a second-order field.

By the first isomorphism theorem for vector spaces,[18] it holds that

$$T_a(A^1M)/\mathrm{Ver}_a(A^1M) \cong T_{\pi(a)}M \tag{4.40}$$

for all $a \in A^1M$ because the natural projection $\pi\colon A^1M \to M$ is a surjective submersion. The space $\mathrm{Ver}_a(A^1M)$ is the tangent space at the point $a \in A^1M$ to A^1_pM with $p = \pi(a)$. The set A^1_pM is the affine hyperplane in T_pM defined by the equation $\vartheta_p(v) = 1$ for all $v \in T_pM$. Therefore, the tangent space $\mathrm{Ver}_a(A^1M)$ can be identified with $\ker \vartheta_p = A^0_pM$ (see equation (4.10)). This results in the pointwise isomorphism

$$\mathrm{Ver}_a(A^1M) \cong A^0_{\pi(a)}M \tag{4.41}$$

for all $a \in A^1M$. The isomorphism (4.41) can be locally expressed as

$$\left.\frac{\partial}{\partial u^i}\right|_a \mapsto \left.\frac{\partial}{\partial x^i}\right|_{\pi(a)} \tag{4.42}$$

using the basis vectors from (4.39) and (4.12). By the isomorphism (4.41), the bundle metric (4.13) on the bundle A^0M of spacelike vectors induces a bundle metric on the bundle $\mathrm{Ver}(A^1M)$ of vertical vectors that is defined as

$$\hat{g}_a\left(\left.\frac{\partial}{\partial u^i}\right|_a, \left.\frac{\partial}{\partial u^j}\right|_a\right) := g_{\pi(a)}\left(\left.\frac{\partial}{\partial x^i}\right|_{\pi(a)}, \left.\frac{\partial}{\partial x^j}\right|_{\pi(a)}\right), \tag{4.43}$$

for all $a \in A^1M$. According to equation (4.13), the bundle metric g on A^0M can be written as

$$g = g_{ij}\,\mathrm{d}x^i \otimes \mathrm{d}x^j$$

and by (4.43) it follows that

$$\hat{g} = g_{ij} \circ \pi\,\mathrm{d}u^i \otimes \mathrm{d}u^j. \tag{4.44}$$

We will often write g_{ij} instead of $g_{ij} \circ \pi$ for the coefficients in (4.44).

Let $w \in T_a(A^1M)$ be an arbitrary tangent vector at some point $a \in A^1_pM$, then

$$D\pi_a(w) - \vartheta_p(\mathrm{D}\pi_a(w))a \tag{4.45}$$

18. See Theorem 2.10 on p. 25.

is a spacelike vector at the point p. To see this, we need to check if (4.45) lies in the kernel of the time structure ϑ. Therefore, we calculate that

$$\vartheta_p\left(D\pi_a(w) - \vartheta_p(\mathrm{D}\pi_a(w))a\right) = \vartheta_p(D\pi_a(w)) - \vartheta_p(\mathrm{D}\pi_a(w))\vartheta_p(a) = 0$$

because $\vartheta_p(a) = 1$. Using the isomorphism (4.41), we can define a vector bundle homomorphism over A^1M (see p. 66 for the definition)

$$\mu\colon T(A^1M) \to \mathrm{Ver}(A^1M), \tag{4.46}$$

which we call the **vertical homomorphism** of the state space A^1M. Its local expression with respect to the natural chart (4.17) is given by

$$\mu|_{\pi^{-1}(U)} = \frac{\partial}{\partial u^i} \otimes \theta^i = \frac{\partial}{\partial u^i} \otimes (\mathrm{d}x^i - u^i \mathrm{d}t),$$

where the θ^i are the one-forms from (4.36). Apparently, the map (4.46) is surjective and it holds that $\mu(V) = 0$ for all local sections V of $\mathrm{Ver}(A^1M)$ and $\mu(Z) = 0$ for all second-order fields Z.

A theory for time-independent mechanical systems can be formulated on the tangent bundle of a time-independent configuration manifold. We refer to Godbillon 1969 for such a presentation. Godbillon uses a similar homomorphism as (4.46) that canonically exists on the double tangent bundle[19] of any differentiable manifold. It is known as the vertical endomorphism of the double tangent bundle (see Godbillon 1969, Chapter X or Morandi et al. 1990, Section 2). It is clear that (4.46) defines an endomorphism of the bundle $T(A^1M)$ when considered as map

$$\mu\colon T(A^1M) \to T(A^1M).$$

There is no canonically defined 'horizontal' subbundle

$$\mathrm{Hor}(A^1M) := \bigcup_{a\in A^1M} \left(\{a\} \times \mathrm{Hor}_a(A^1M)\right) \subset T(A^1M) \tag{4.47}$$

that would complement the vertical bundle $\mathrm{Ver}(A^1M)$ such that the tangent bundle $T(A^1M)$ would split as

$$\begin{aligned} T(A^1M) &= \mathrm{Hor}(A^1M) \oplus \mathrm{Ver}(A^1M) \\ &:= \bigcup_{a\in A^1M} \left(\{a\} \times \left(\mathrm{Hor}_a(A^1M) \oplus \mathrm{Ver}_a(A^1M)\right)\right). \end{aligned}$$

19. The double tangent bundle of a manifold is the tangent bundle of its tangent bundle.

The definition of a horizontal subbundle (4.47) allows to write the tangent space at each point $a \in A^1M$ as the direct sum

$$T_a(A^1M) = \mathrm{Hor}_a(A^1M) \oplus \mathrm{Ver}_a(A^1M).$$

In the study of tangent bundles (and double tangent bundles), it is well-known that the choice of a *particular* second-order field induces such a splitting.[20] There are several ways to define the horizontal bundle that results from the selection of a second-order field $Z \in \mathrm{Vect}(A^1M)$. A straightforward approach is to define n horizontal basis fields

$$H_i := \frac{\partial}{\partial x^i} + \frac{1}{2}\frac{\partial}{\partial u^i}[Z^j]\frac{\partial}{\partial u^j} \tag{4.48}$$

for each natural chart $(\pi^{-1}(U), \Phi)$ that is induced by an atlas of M. We drop the somewhat clumsy notation with square brackets for the application of a vector field on a real-valued function. Instead of (4.48), we write

$$H_i := \frac{\partial}{\partial x^i} + \frac{1}{2}\frac{\partial Z^j}{\partial u^i}\frac{\partial}{\partial u^j}. \tag{4.49}$$

The coefficients Z^j in (4.49) denote the defining coefficient functions in the local expression of the second-order field

$$Z = \frac{\partial}{\partial t} + u^j\frac{\partial}{\partial x^j} + Z^j\frac{\partial}{\partial u^j}.$$

Next, we define

$$\mathrm{Hor}_a(A^1M) := \mathrm{span}\left\{Z|_a, H_1|_a, \dots, H_n|_a\right\}$$

and use Lemma 3.36 to argue that (4.47) defines a smooth subbundle.

However, there is a coordinate-free alternative to this argumentation that makes use of Proposition 3.23. Following Loos 1985, p. 280, we consider the vector bundle homomorphism over A^1M

$$\eta\colon T(A^1M) \to T(A^1M) \tag{4.50}$$

that is defined as

$$\eta(X) := \tfrac{1}{2}\Big([\![Z, \mu\cdot X]\!] - \mu\cdot[\![Z,X]\!] + X - \hat{\vartheta}(X)Z\Big)$$

for all $X \in \mathrm{Vect}(A^1M)$. Indeed, one easily verifies that

$$\eta(fX + gY) = f\eta(X) + g\eta(Y)$$

20. See for example Yano et al. 1973 or Morandi et al. 1990.

for all $f,g \in C^\infty(A^1M)$ and all $X,Y \in \mathrm{Vect}(A^1M)$ such that (4.50) defines a vector bundle homomorphism over A^1M. The local coordinate expression of η reads

$$\eta = \frac{\partial}{\partial u^i} \otimes \upeta^i = \frac{\partial}{\partial u^i} \otimes \left(\mathrm{d}u^i - Z^i \mathrm{d}t - \frac{1}{2}\frac{\partial Z^i}{\partial u^j}(\mathrm{d}x^j - u^j \mathrm{d}t) \right),$$

where we have introduced the one-forms

$$\upeta^i = \mathrm{d}u^i - Z^i \mathrm{d}t - \frac{1}{2}\frac{\partial Z^i}{\partial u^j}(\mathrm{d}x^j - u^j \mathrm{d}t).$$

The one-forms $\mathrm{d}t, \uptheta^1, \dots, \uptheta^n, \upeta^1, \dots, \upeta^n$ (see equation (4.36)) are just the dual fields to the vector fields $Z, H_1, \dots, H_n, \partial/\partial u^1, \dots, \partial/\partial u^n$ on A^1M, i.e.,

$$\mathrm{d}t(Z) = 1, \quad \mathrm{d}t(H_i) = 0, \quad \mathrm{d}t\left(\frac{\partial}{\partial u^i}\right) = 0, \tag{4.51}$$

$$\uptheta^j(Z) = 0, \quad \uptheta^j(H_i) = \delta^j_i, \quad \uptheta^j\left(\frac{\partial}{\partial u^i}\right) = 0,$$

$$\upeta^j(Z) = 0, \quad \upeta^j(H_i) = 0, \quad \upeta^j\left(\frac{\partial}{\partial u^i}\right) = \delta^j_i. \tag{4.52}$$

Note that the $\mathrm{d}t$ is just the chart representation of the time structure $\hat{\vartheta}$ (i.e., locally $\hat{\vartheta} = \mathrm{d}t$). It follows from (4.52), that

$$\mathrm{Hor}(A^1M) = \ker \eta$$

and that

$$\eta|_{\mathrm{Ver}(A^1M)} = \mathrm{id}_{\mathrm{Ver}(A^1M)}.$$

Hence, it holds that $\eta \circ \eta = \eta$ such that η is a projection onto $\mathrm{Ver}(A^1M)$ and consequently

$$T(A^1M) = \ker \eta \oplus \mathrm{Ver}(A^1M) = \mathrm{Hor}(A^1M) \oplus \mathrm{Ver}(A^1M).$$

One can easily convince oneself, that $\ker \mu \cap \ker \eta$ is a line bundle that is spanned by the second-order field Z.

4.5. Basic and semi-basic differential forms

We have already seen that differential forms can be used to characterize vector fields. There are two types of differential forms that will reveal useful in the definition of forces, the so-called basic and semi-basic forms.

The natural projection $\pi: A^1M \to M$ is a surjective submersion. As such it defines an injection (see Section 3.8)

$$\pi_{\leftarrow}: \Omega^\star(M) \to \Omega^\star(A^1M)$$

of the differential forms on M to those on A^1M. These forms on A^1M that are given by $\operatorname{im} \pi_{\leftarrow} \subset \Omega^\star(A^1M)$ are called **basic** differential forms. These forms are said to be *basic* because they result from pulling differential forms on the base manifold M back to A^1M.

A differential l-form ω on A^1M is called **semi-basic** if $\omega(V_1, \dots, V_l) = 0$ as soon as one of the vector fields V_i is vertical,[21] i.e., if the interior product[22] $V \lrcorner \omega = 0$ for any vertical vector field V. An equivalent statement is that the local representation of ω with respect to the dual basis induced by the natural chart (4.17) does not contain terms in $\mathrm{d}u^1, \dots, \mathrm{d}u^n$, while the chart representations of the coefficients of the basis vectors $\mathrm{d}t$ and $\mathrm{d}x^1, \dots, \mathrm{d}x^n$ may depend on t, $\mathbf{x}$ and $\mathbf{u}$. Note that basic forms are semi-basic.

The vertical homomorphism (4.46) allows us to define a differentiation operation[23] on differential forms on A^1M with the property that the subalgebra of semi-basic forms is closed under its operation. First, we define the derivation

$$\mathcal{D}_\mu: \Omega^\star(A^1M) \to \Omega^\star(A^1M)$$

using the vertical homomorphism as

$$\big(\mathcal{D}_\mu \omega\big)\big(V_1, \dots, V_l\big) := \sum_{i=1}^{l} \omega\big(V_1, \dots, \mu(V_i), \dots, V_l\big). \tag{4.53}$$

The operator $\mathcal{D}_\mu$ is linear, does not alter the degree of the form and satisfies:

$$\begin{aligned}
&\mathcal{D}_\mu f = 0, \qquad (f \text{ smooth function on } A^1M)\\
&\mathcal{D}_\mu(\alpha \wedge \beta) = (\mathcal{D}_\mu \alpha) \wedge \beta + \alpha \wedge (\mathcal{D}_\mu \beta),\\
&\mathcal{D}_\mu(\mathrm{d}x^i) = \mathcal{D}_\mu(\mathrm{d}t) = 0,\\
&\mathcal{D}_\mu(\mathrm{d}u^i) = \mathrm{d}x^i - u^i \mathrm{d}t.
\end{aligned}$$

21. See p. 104 for the definition.
22. See p. 82 for the definition.
23. The vertical homomorphism is an example of a vector-valued differential form. It holds in general that vector-valued differential forms come along with certain derivations. We refer to Frölicher et al. 1956 for the general theory. We will make use of the differential concomitant of the vertical homomorphism (4.46) while in *time-independent* mechanics, the differential associate of the vertical endomorphism of the double tangent bundle is of interest (see Godbillon 1969, Chapters X and XI as well as Morandi et al. 1990, Section 2). See also Klein 1963.

Using (4.53) and the exterior derivative d, we define the linear operator

$$\boldsymbol{\partial} := \mathcal{D}_\mu \circ \mathrm{d} - \mathrm{d} \circ \mathcal{D}_\mu .$$

The operator $\boldsymbol{\partial}$ increases the degree of a form by one and obeys the following rules:

$$\begin{aligned}
&\boldsymbol{\partial} f = \frac{\partial f}{\partial u^i}(\mathrm{d}x^i - u^i \mathrm{d}t), && (f \text{ smooth function on } A^1M) \\
&\boldsymbol{\partial}(\alpha \wedge \beta) = \boldsymbol{\partial}\alpha \wedge \beta + (-1)^l \alpha \wedge \boldsymbol{\partial}\beta, && (\alpha \text{ is } l\text{-form on } A^1M) \\
&\boldsymbol{\partial}(\mathrm{d}x^i) = \boldsymbol{\partial}(\mathrm{d}t) = 0, && \\
&\boldsymbol{\partial}(\mathrm{d}u^i) = \mathrm{d}u^i \wedge \mathrm{d}t. &&
\end{aligned} \tag{4.54}$$

Moreover, it holds that

$$\boldsymbol{\partial} \circ \mathrm{d} = -\mathrm{d} \circ \boldsymbol{\partial} \tag{4.55}$$

because of $\mathrm{d}^2 = 0$. However, $\boldsymbol{\partial}^2 \neq 0$ but

$$\boldsymbol{\partial}^2 \omega = \hat{\vartheta} \wedge \boldsymbol{\partial}\omega, \tag{4.56}$$

where $\hat{\vartheta}$ denotes the time structure on A^1M. From the rules (4.54), it becomes obvious that $\boldsymbol{\partial}$ maps semi-basic forms to semi-basic forms.

Let Z be a second-order field and let ω be a semi-basic l-form. Then $Z \lrcorner \omega$ is a semi-basic $(l-1)$-form that is independent on the specific choice of Z. Indeed, if Z' denotes another second-order field, then it holds that $Z' = Z + V$ where V is a vertical vector field and consequently $V \lrcorner \omega = 0$ because ω is semi-basic. The following formula holds

$$\boldsymbol{\partial}(Z \lrcorner \omega) + Z \lrcorner \boldsymbol{\partial}\omega + \hat{\vartheta} \wedge (Z \lrcorner \omega) = l\omega. \tag{4.57}$$

To prove (4.57), one considers that the left-hand and the right-hand side represent derivations of the algebra of semi-basic differential forms that agree on the semi-basic zero-forms (smooth functions) and on the semi-basic one-forms and, therefore, are equal.

4.6. Action form of a second-order field

The projection $\eta\colon T(A^1M) \to T(A^1M)$ defines a surjective vector bundle homomorphism over A^1M when considered as map $\eta\colon T(A^1M) \to \mathrm{Ver}(A^1M)$ onto its image, which is the vertical subbundle $\mathrm{Ver}(A^1M)$. We denote this map again by η.

As it was suggested by Loos 1982, p. 20, the surjective vector bundle homomorphisms $\eta\colon T(A^1M) \to \mathrm{Ver}(A^1M)$ and $\mu\colon T(A^1M) \to \mathrm{Ver}(A^1M)$

together with the bundle metric (4.43) can be used to define a differential two-form Ω on A^1M as

$$\Omega(X,Y) := \hat{g}\big(\eta(X), \mu(Y)\big) - \hat{g}\big(\eta(Y), \mu(X)\big), \tag{4.58}$$

for all $X, Y \in \mathrm{Vect}(A^1M)$. Because η depends on the choice of a second-order field Z, we call Ω the **action form of** Z. The local expression of the action form (4.58) reads

$$\begin{aligned} \Omega|_{\pi^{-1}(U)} &= g_{ij}\, \eta^i \wedge \theta^j \\ &= g_{ij}\Big(\mathrm{d}u^i - Z^i \mathrm{d}t - \frac{1}{2}\frac{\partial Z^i}{\partial u^k}(\mathrm{d}x^k - u^k \mathrm{d}t)\Big) \wedge (\mathrm{d}x^j - u^j \mathrm{d}t). \end{aligned} \tag{4.59}$$

The properties of the action form (4.58) can be summarized in a theorem.

Theorem 4.2 (Loos 1982, p. 21)**.** Let (M, ϑ, g) be a Galilean manifold and let Z be a second-order field on its state space A^1M, the corresponding action form Ω has the following properties:

(i) Ω vanishes on $\ker \mu$. In particular, $\Omega|_{A^1_pM} = 0$ for any $p \in M$.

(ii) Ω determines the metric g by

$$g_{ij} = \Omega\Big(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial x^j}\Big).$$

(iii) Ω determines the second-order field Z. In fact, Z is the only vector field on A^1M for which it holds that

$$Z \lrcorner \Omega = 0, \;\; \hat{\vartheta}(Z) = 1.$$

Proof. Property (i) is clear by definition (4.58) and because

$$T_a(A^1_pM) = \mathrm{Ver}_a(A^1M) \subset \ker \mu_a.$$

Property (ii) follows directly from the local expression (4.59). Finally, property (iii) remains to be shown. With the interior product, a two-form Ω defines the map between vector and covector fields that is given by

$$\hat{f} : X \mapsto X \lrcorner \Omega. \tag{4.60}$$

The mapping (4.60) is a vector bundle homomorphism over A^1M between $T(A^1M)$ and $T^*(A^1M)$. First, we show that (4.60) has constant rank[24] $2n$.

24. The rank of a vector bundle homomorphism is defined on p. 66.

For this we observe that the $2n$ one-forms ${\partial}/{\partial u^i} \lrcorner \Omega$, ${\partial}/{\partial x^i} \lrcorner \Omega$ are linearly independent because by (i) and (ii) it follows from $a^i\,{\partial}/{\partial u^i} \lrcorner \Omega + b^i\,{\partial}/{\partial x^i} \lrcorner \Omega = 0$ that

$$\begin{aligned} 0 &= \left(a^i \frac{\partial}{\partial u^i} \lrcorner \Omega\right)\left(\frac{\partial}{\partial u^j}\right) + b^i \left(\frac{\partial}{\partial x^i} \lrcorner \Omega\right)\left(\frac{\partial}{\partial u^j}\right) \\ &= a^i\,\Omega\left(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial u^j}\right) + b^i\,\Omega\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial u^j}\right) = 0 - b^i g_{ij}, \end{aligned}$$

i.e., that $b^i = 0$, and thereby that

$$0 = \left(a^i \frac{\partial}{\partial u^i} \lrcorner \Omega\right)\left(\frac{\partial}{\partial x^j}\right) = a^i\,\Omega\left(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial x^j}\right) = a^i g_{ij},$$

i.e., that $a^i = 0$. Therefore, the homomorphism (4.60) has rank $2n$ because it cannot have full rank by property (i). For reasons of brevity, we also say that Ω has rank $2n$. Consequently,

$$\ker \Omega := \ker \hat{f} = \left\{(a, X_a) \in T(A^1M) \mid X_a \lrcorner \Omega_a = 0\right\},$$

where $X_a \in T_a(A^1M)$ and $\Omega(a) = (a, \Omega_a)$ for all $a \in A^1M$, is a line bundle. By equation (4.58), $\ker \eta \cap \ker \mu \subseteq \ker \Omega$ and equality follows for dimensional reasons. Because the vector field Z is the uniquely defined intersection of $\ker \eta \cap \ker \mu$ with $\hat{\vartheta}(Z) = 1$, the assertion follows. □

We want to establish a one-to-one correspondence between second-order fields and their action forms. The question is under which conditions a given differential two-form is the action form of a second-order field. It is not sufficient to require properties (i)–(iii) from Theorem 4.2. Indeed, there are other[25] two-forms than Ω from (4.58) that satisfy these properties. Consider for example the locally defined two-form

$$\Omega' = g_{ij}(\mathrm{d}u^i - Z^i \mathrm{d}t) \wedge (\mathrm{d}x^j - u^j \mathrm{d}t).$$

Similar forms can be found in Gallissot 1952, p. 153, and Souriau 1970, p. 132 (respectively on p. 129 of Souriau 1997). Souriau suggested to impose the condition

$$\mathrm{d}\Omega \overset{!}{=} 0 \tag{4.61}$$

in order to get rid of the ambiguity in the choice of Ω. Souriau refers to condition (4.61) as *Maxwell's principle* and he mentions[26] that this closure condition imposes restrictions on the second-order field Z. Loos[27] suggests

25. This is pointed out by Loos 1985, pp. 281–282.
26. See p. 143 of Souriau 1970 or p. 139 of the translation Souriau 1997.
27. See Loos 1982, p. 24 or Loos 1985, p. 281.

to use the assignment $Z \mapsto \Omega$ given by equation (4.58) together with the weaker closure condition

$$\eth\Omega \overset{!}{=} 0. \tag{4.62}$$

Condition (4.62) resolves the ambiguity without imposing restrictions on Z. One readily checks that the action form (4.58) does indeed satisfy (4.62) using the local expression (4.59) and the rules (4.54). With the closure condition (4.62) we can formulate the following theorem that establishes a bijective relation between action forms and second-order fields.

Theorem 4.3 (Loos 1982, p. 24)**.** Let (M, ϑ) be a manifold with time structure. A two-form Ω on A^1M is the action form of a second-order field Z if and only if it satisfies the following conditions:

(i) Ω vanishes on $\ker \mu$, i.e.,

$$\Omega(X, Y) = 0$$

for all X, Y with $\mu(X) = \mu(Y) = 0$.

(ii) Ω induces a bundle metric g on A^0M, i.e., the matrix

$$g_{ij} = \Omega\left(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial x^j}\right)$$

is symmetric and positive definite for all charts.

(iii) $\eth\Omega = 0$.

The second-order field Z is the only vector field on A^1M for which holds

$$Z \lrcorner \Omega = 0, \; \hat{\vartheta}(Z) = 1.$$

Proof. The necessity of conditions (i) and (ii) follows from Theorem 4.2. Direct calculation with (4.59) and the rules (4.54) shows that $\eth\Omega = 0$. To prove sufficiency, assume that the conditions (i), (ii), and (iii) are satisfied. We know from the proof of Theorem 4.2, that conditions (i) and (ii) imply that the rank of Ω is $2n$. Let $a \in A^1M$ and $0 \neq v \in \ker \Omega_a$. Then $v \in \ker \mu_a$ because $\ker \mu_a \subseteq (\ker \mu_a)^\perp$ (orthogonal complement with respect to the bilinear form Ω_a as defined in Section 2.7) by (i). Moreover, $\dim \ker \mu_a = n+1$ and so it follows by Proposition 2.13 that

$$\begin{aligned}\dim\, (\ker \mu_a)^\perp &= \dim T_a(A^1M) + \dim \ker \mu_a \cap (T_aA)^\perp - \dim \ker \mu_a \\ &= (2n+1) + 1 - (n+1) = n+1.\end{aligned}$$

and, consequently,

$$\ker \mu_a = (\ker \mu_a)^{\perp} \supseteq \left(T_a(A^1M)\right)^{\perp} = \ker \Omega_a.$$

By (ii), v cannot be a vertical vector, i.e., $v \notin \mathrm{Ver}_a(A^1M)$. A vertical vector can be expressed as a linear combination of $\partial/\partial u^i|_a$. But Ω is forbidden to vanish on vectors $\partial/\partial u^i|_a$ by (ii). Since $\ker \mu_a$ is spanned by $\mathrm{Ver}_a(A^1M)$ and $\partial/\partial t|_a + u^i(a)\,\partial/\partial x^i|_a$, there exists a unique element Z_a in $\ker \Omega_a$ of the form

$$Z_a = \left.\frac{\partial}{\partial t}\right|_a + u^i(a)\left.\frac{\partial}{\partial x^i}\right|_a + Z^i(a)\left.\frac{\partial}{\partial u^i}\right|_a.$$

The element Z is the second-order field on A^1M that is uniquely characterized by $Z \lrcorner \Omega = 0$ and $\hat{\vartheta}(Z) = 1$.

It remains to be shown that Ω is the action form of the second-order field Z defined by (4.58). For this, we consider the basis Z, $\partial/\partial x^i$, $\partial/\partial u^i$ and its dual one-forms $\mathrm{d}t$, $\theta^i = \mathrm{d}x^i - u^i\mathrm{d}t$, $\lambda^i = \mathrm{d}u^i - Z^i\mathrm{d}t$ induced by an adapted chart. By (i) and (ii), Ω has the form

$$\begin{aligned}
\Omega = {} & \Omega\left(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial x^j}\right)\left(\mathrm{d}u^i - Z^i\mathrm{d}t\right) \wedge \left(\mathrm{d}x^j - u^j\mathrm{d}t\right) \\
& + \frac{1}{2}\Omega\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\left(\mathrm{d}x^i - u^i\mathrm{d}t\right) \wedge \left(\mathrm{d}x^j - u^j\mathrm{d}t\right) \\
= {} & g_{ij}\left(\mathrm{d}u^i - Z^i\mathrm{d}t\right) \wedge \left(\mathrm{d}x^j - u^j\mathrm{d}t\right) \\
& + \frac{1}{2}\Omega\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\left(\mathrm{d}x^i - u^i\mathrm{d}t\right) \wedge \left(\mathrm{d}x^j - u^j\mathrm{d}t\right).
\end{aligned}$$

Direct calculation of $\boldsymbol{\partial}\Omega = 0$ shows that

$$\Omega\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \frac{1}{2}\left(g_{ik}\frac{\partial Z^k}{\partial u^j} - g_{jk}\frac{\partial Z^k}{\partial u^i}\right)$$

and, by (4.59), proves the assertion. □

4.7. Forces

In the previous section, we saw that to any second-order field Z_1 on the state space A^1M an action form Ω_1 can be uniquely associated and vice-versa. Moreover, we know that if we consider another second-order field Z_2 with action form Ω_2, then it can only differ from Z_1 by a vertical vector field, i.e., it holds that

$$Z_2 = Z_1 + V,$$

where V is a smooth section of the vertical bundle $\mathrm{Ver}(A^1M)$. Let us consider the differential two-form Φ by which the action forms Ω_1 and Ω_2 differ, i.e.,

$$\Omega_2 = \Omega_1 + \Phi. \tag{4.63}$$

It is clear that $\boldsymbol{\partial}\Phi = 0$ because $\boldsymbol{\partial}\Omega_1 = \boldsymbol{\partial}\Omega_2 = 0$.

In terms of the coordinate fields induced by a natural chart, the two second-order fields Z_1 and Z_2 can be written as

$$Z_1 = \frac{\partial}{\partial t} + u^i \frac{\partial}{\partial x^i} + Z_1^i \frac{\partial}{\partial u^i} \quad \text{and} \quad Z_2 = \frac{\partial}{\partial t} + u^i \frac{\partial}{\partial x^i} + Z_2^i \frac{\partial}{\partial u^i}, \tag{4.64}$$

respectively. Using (4.59) and (4.64), the two-form Φ is given as

$$\begin{aligned} \Phi = \Omega_2 - \Omega_1 &= g_{ij}(Z_2^i - Z_1^i)\,\mathrm{d}x^j \wedge \mathrm{d}t \\ &\quad + \frac{1}{2} g_{ij}\left(\frac{\partial Z_2^i}{\partial u^k} - \frac{\partial Z_1^i}{\partial u^k}\right)(\mathrm{d}x^j - u^j \mathrm{d}t) \wedge (\mathrm{d}x^k - u^k \mathrm{d}t) \\ &= g_{ij}(Z_2^i - Z_1^i)\,\mathrm{d}x^j \wedge \mathrm{d}t \\ &\quad + \frac{1}{2}\frac{\partial}{\partial u^k}\big(g_{ij}(Z_2^i - Z_1^i)\big)(\mathrm{d}x^j - u^j \mathrm{d}t) \wedge (\mathrm{d}x^k - u^k \mathrm{d}t), \end{aligned} \tag{4.65}$$

where the last equality uses that the coefficients $g_{ij} = g_{ij} \circ \pi$ are independent of $u^1, \dots, u^n$. The local expression shows that the two-form Φ is semi-basic (see Section 4.5).

Let us assume that we are given Ω_1 and Φ in equation (4.63). We observe that to any semi-basic and $\boldsymbol{\partial}$-closed two-form Φ we get a different action form Ω_2. In particular, it holds for $\Phi = 0$ that $\Omega_1 = \Omega_2$ and, therefore, that $Z_1 = Z_2$. The semi-basic differential two-forms Φ with $\boldsymbol{\partial}\Phi = 0$ have the character of a force because the choice of an other Φ implies a different action form Ω_2 and thereby a different vector field Z_2 defining the motion. We refer to semi-basic differential two-forms Φ with $\boldsymbol{\partial}\Phi = 0$ as **force two-forms**.

The coefficients of (4.65) depend on the difference $g_{ij}(Z_2^i - Z_1^i)$. This observation lets us come to another view on forces. By equation (4.44), the bundle metric $\hat{g}$ on the vertical bundle $\mathrm{Ver}(A^1M)$ has the local expression

$$\hat{g} = g_{ij}\, \mathrm{d}u^i \otimes \mathrm{d}u^j.$$

By the coordinate expression (4.64), the vertical vector field V locally reads

$$V = (Z_2^i - Z_1^i)\frac{\partial}{\partial u^i} = V^i \frac{\partial}{\partial u^i},$$

where $V^i := Z_2^i - Z_1^i$. The metric $\hat{g}$ defines a vector bundle isomorphism over A^1M

$$\begin{aligned} \hat{g}\cdot : \mathrm{Ver}(A^1M) &\to \mathrm{Ver}^*(A^1M), \\ V = V^k \frac{\partial}{\partial u^k} &\mapsto \hat{g}\cdot V = g_{ij} V^j \mathrm{d}u^i \end{aligned}$$

between the vertical bundle $\mathrm{Ver}(A^1M)$ and its dual bundle[28] $\mathrm{Ver}^*(A^1M)$. Analogously to a Riemannian metric (see Section 3.10), the bundle metric $\hat{g}$ also induces a linear bijective map on sections that we denote by the same symbol, i.e., we write

$$\hat{g}\cdot : \Gamma\Big(\mathrm{Ver}(A^1M)\Big) \to \Gamma\Big(\mathrm{Ver}^*(A^1M)\Big). \tag{4.66}$$

The map (4.66) establishes a one-to-one correspondence between smooth sections of the bundle $\mathrm{Ver}^*(A^1M)$ and the set of differences between second-order fields, the vertical vector fields. Therefore, we call a smooth section of the bundle $\mathrm{Ver}^*(A^1M)$ a **force**. As a section of $\mathrm{Ver}^*(A^1M)$, a force is a linear form on $\mathrm{Ver}(A^1M)$, i.e., a linear form

$$F\colon \mathrm{Ver}(A^1M) \to \mathbb{R} \tag{4.67}$$

that induces a $C^\infty(A^1M)$-linear map

$$F\colon \Gamma\Big(\mathrm{Ver}(A^1M)\Big) \to C^\infty(A^1M)$$

on the space of vertical vector fields.

If we consider that the Galilean metric models the mass of a finite-dimensional mechanical system and if we interpret vertical vector fields as (relative) accelerations, then with

$$\hat{g}\cdot V = F \tag{4.68}$$

we are facing Newton's second law that says "mass×(relative) acceleration = force". The following theorem establishes a bijective relation between forces (4.67) and force two-forms (4.65).

Theorem 4.4 (Loos 1982, p. 32)**.** The following formulae define bijections between

(i) the forces, i.e., the linear forms $F\colon \mathrm{Ver}(A^1M) \to \mathbb{R}$,

(ii) the semi-basic one-forms φ with $Z \lrcorner\, \varphi = 0$,

28. Equation (4.38) defines a vector space $\mathrm{Ver}_a(A^1M)$ at each point $a \in A^1M$. In analogy to the notation (3.27) used for the cotangent space, we denote its dual space by ${\mathrm{Ver}^*}_a(A^1M) := (\mathrm{Ver}_a(A^1M))^*$. Finally, we define the bundle

$$\mathrm{Ver}^*(A^1M) := \bigcup_{a\in A^1M} \Big(\{a\}\times {\mathrm{Ver}^*}_a(A^1M)\Big).$$

(iii) the force two-forms, i.e., the semi-basic two-forms Φ with $\partial\Phi = 0$:

$$\begin{aligned}\varphi &= F \circ \mu,\\ \varphi &= -Z \lrcorner \Phi,\\ \Phi &= -\tfrac{1}{2}(\partial\varphi + \hat{\vartheta} \wedge \varphi).\end{aligned}$$

In local coordinates, it holds that

$$F = F_i \mathrm{d}u^i, \tag{4.69}$$

$$\varphi = F_i(\mathrm{d}x^i - u^i \mathrm{d}t), \tag{4.70}$$

$$\Phi = F_i \mathrm{d}x^i \wedge \mathrm{d}t + \frac{1}{2}\frac{\partial F_i}{\partial u^j}(\mathrm{d}x^i - u^i \mathrm{d}t) \wedge (\mathrm{d}x^j - u^j \mathrm{d}t). \tag{4.71}$$

Proof. Because of $\mu\big(\mathrm{Ver}(A^1M)\big) = \mu(Z) = 0$, $F \circ \mu$ is a semi-basic one-form with $Z \lrcorner (F \circ \mu) = 0$. Conversely, a semi-basic one-form φ with $Z \lrcorner \varphi = 0$ vanishes on $\ker\mu$ and, therefore, defines a linear form on

$$T(A^1M)/\ker\mu \cong \mathrm{Ver}(A^1M). \tag{4.72}$$

The isomorphism (4.72) follows by Theorem 2.10 and equation (4.46). This proves the bijection between (i) and (ii). According to the properties (4.54) and (4.56), it holds that

$$\begin{aligned}\partial(\partial\varphi + \hat{\vartheta} \wedge \varphi) &= \partial\partial\varphi + \partial\hat{\vartheta} \wedge \varphi - \hat{\vartheta} \wedge \partial\varphi\\ &= \hat{\vartheta} \wedge \partial\varphi - \hat{\vartheta} \wedge \partial\varphi = 0\end{aligned}$$

and $Z \lrcorner (-Z \lrcorner \Phi) = -\Phi(Z, Z) = 0$. Finally, it holds that

$$\begin{aligned}-Z \lrcorner \Big(-\tfrac{1}{2}(\partial\varphi + \hat{\vartheta} \wedge \varphi)\Big) &= \tfrac{1}{2}\Big(Z \lrcorner \partial\varphi + (Z \lrcorner \hat{\vartheta}) \wedge \varphi - \hat{\vartheta} \wedge (Z \lrcorner \varphi)\Big)\\ &= \tfrac{1}{2}(Z \lrcorner \partial\varphi + \varphi) = \varphi,\end{aligned}$$

by the rule (4.57) applied to $\omega = \varphi$ and

$$-\tfrac{1}{2}\Big(\partial(-Z \lrcorner \Phi) + \hat{\vartheta} \wedge (-Z \lrcorner \Phi)\Big) = \tfrac{1}{2}\Big(\partial(-Z \lrcorner \Phi) + \hat{\vartheta} \wedge (Z \lrcorner \Phi)\Big) = \Phi,$$

again by (4.57) applied to $\omega = \Phi$. This proves the assertion. The coordinate expressions (4.69) to (4.71) follow by straightforward computation. □

Theorem 4.5 (Loos 1982, p. 25). Let Ω denote the action form of a mechanical system, let Z be its related second-order field and let F be a force. By the one-to-one correspondence (4.68), F is associated to a vertical vector field V. Moreover, F can be uniquely related to a force two-form Φ by Theorem 4.4. It then holds that the vector field $Z' = Z + V$ is the second-order field related to the action form $\Omega' = \Omega + \Phi$.

Proof. One easily verifies that $\Omega' = \Omega + \Phi$ is an action form, i.e., that it respects the properties from Theorem 4.3. Furthermore, one observes that Ω' and Ω induce the same Galilean metric g. It remains to be shown that $Z' \lrcorner \Omega' = (Z+V) \lrcorner (\Omega+\Phi) = 0$. Because $Z \lrcorner \Omega = 0$ and $V \lrcorner \Omega = 0$, it holds that

$$Z' \lrcorner \Omega' = (Z+V) \lrcorner (\Omega+\Phi) = V \lrcorner \Omega + Z \lrcorner \Phi.$$

By definition (4.58), it holds for $V \lrcorner \Omega$ that

$$\big(V \lrcorner \Omega\big)(Y) = \hat{g}\big(\eta(V), \mu(Y)\big) - \hat{g}\big(\eta(Y), \mu(V)\big) = \hat{g}\big(V, \mu(Y)\big) = F \circ \mu(Y),$$

where we used the properties $\mu(V) = 0$ and $\eta(V) = V$ of the vector bundle homomorphisms μ and η. The last equality follows by equation (4.68). By Theorem 4.4, it holds that $Z \lrcorner \Phi = -F \circ \mu$. Thus, it follows that

$$Z' \lrcorner \Omega' = V \lrcorner \Omega + Z \lrcorner \Phi = 0.$$

□

4.8. Modelling inertia — the kinetic energy

We learned from the example of Simon's oscillator at the beginning of this chapter that motion can only be characterized with respect to a certain reference event and relatively to chosen reference directions. Let (M, ϑ, g) be a Galilean manifold. We define a **reference field** to be a time-normalized vector field R defined on a neighbourhood U_R of M, i.e.,

$$R : M \supseteq U_R \to A^1 M$$

with $\pi \circ R = \mathrm{id}_M$. In Section 4.3, we defined the motion of a mechanical system to be a second-order curve $\dot{\gamma} : I \to A^1 M$, i.e., a curve which has the special form

$$\dot{\gamma} : I \to A^1 M, \ \tau \mapsto \big(\gamma(\tau), \dot{\gamma}_{\gamma(\tau)}\big), \tag{4.73}$$

where $\gamma : I \to M$ denotes a time-parametrized curve in the Galilean manifold (M, ϑ, g).

We define the **relative velocity** of the motion (4.73) at time $t(\tau) = t^*$ with respect to the reference field R as the spacelike vector

$$\dot{\gamma}_{\gamma(\tau)} - R_{\gamma(\tau)} \in A^0_{\gamma(\tau)} M,$$

where

$$R(\gamma(\tau)) = \big(\gamma(\tau), R_{\gamma(\tau)}\big).$$

The tangent vector $\dot\gamma_{\gamma(\tau)}$ is an element of the affine space

$$A^1_{\gamma(\tau)}M \subset T_{\gamma(\tau)}M.$$

It is only after choosing the reference element $R_{\gamma(\tau)}$ in the affine space that we can associate $\dot\gamma_{\gamma(\tau)}$ with an element of a vector space. Because the Galilean metric endows this vector space with an inner product, the notions of length and angle are available for relative velocities.

For an adapted chart $\phi\colon M \supseteq U \to \mathbb{R}^{n+1},\ p \mapsto \phi(p) = (x^0, \dots, x^n)$, the reference field $R = \partial/\partial x^0$ is said to be the **resting field** induced by the chart. In the other direction, we call the chart (U, ϕ) **the resting chart** of some given reference field $R = \partial/\partial y^0 + R^i\,\partial/\partial y^i$ that is defined with respect to the local coordinates $(y^0, \dots, y^n)$ of an adapted chart if its expression with respect to the coordinates induced by the chart (U, ϕ) takes the simple form $R = \partial/\partial x^0$.

Let us come back to the example of Simon's oscillator. Figure 4.1b visualizes the relative motion of the oscillator with respect to the clamping table. The curve ${}_I\mathbf{r}_{AC}$ can be regarded as the chart representation of a curve γ in a three-dimensional Galilean manifold (M, ϑ, g), i.e.,

$$\left(t, {}_I\mathbf{r}_{AC}\right) = \phi \circ \gamma\colon \mathbb{R} \supseteq I \to \mathbb{R}^3,$$

where $\phi\colon M \to \mathbb{R}^3,\ p \mapsto (t, x^1, x^2)$ is the (global) chart that corresponds to the $(A, \mathbf{e}^I_1, \mathbf{e}^I_2)$-system. Consequently, the tangent field along γ can be locally expressed as

$$\dot\gamma(\tau) = \left(\gamma(\tau), \left.\frac{\partial}{\partial t}\right|_{\gamma(\tau)} + \dot x^1(\tau)\left.\frac{\partial}{\partial x^1}\right|_{\gamma(\tau)} + \dot x^2(\tau)\left.\frac{\partial}{\partial x^2}\right|_{\gamma(\tau)}\right).$$

The resting field induced by the chart (M, ϕ) is given by $R = \partial/\partial t$.

Figure 4.1c shows the motion of the oscillator in the coordinates $(\bar t, y^1, y^2)$ corresponding to the reference system $(A, \mathbf{e}^K_1, \mathbf{e}^K_2)$. The resulting curve ${}_K\mathbf{r}_{AC}$ can be interpreted as chart representation of the same curve γ with respect to an other (global) chart $\psi\colon M \to \mathbb{R}^3,\ p \mapsto (\bar t, y^1, y^2)$ that is defined by the reference system $(A, \mathbf{e}^K_1, \mathbf{e}^K_2)$, i.e.,

$$\left(\bar t, {}_K\mathbf{r}_{AC}\right) = \psi \circ \gamma\colon \mathbb{R} \supseteq I \to \mathbb{R}^3.$$

Again, the tangent field along γ can be locally expressed with respect to the chart $\psi\colon M \to \mathbb{R}^3$ as

$$\dot\gamma(\tau) = \left(\gamma(\tau), \left.\frac{\partial}{\partial \bar t}\right|_{\gamma(\tau)} + \dot y^1(\tau)\left.\frac{\partial}{\partial y^1}\right|_{\gamma(\tau)} + \dot y^2(\tau)\left.\frac{\partial}{\partial y^2}\right|_{\gamma(\tau)}\right).$$

The chart (M, ψ) induces the resting field $\tilde R = \partial/\partial \bar t$.

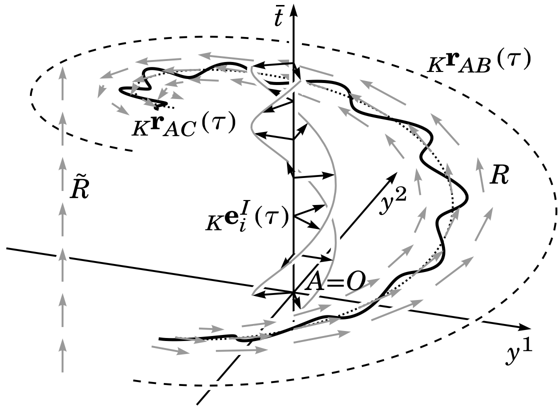


Figure 4.4.: Visualization of the reference fields R and $\tilde{R}$ with respect to the chart (M,ψ), respectively the $(A,\mathbf{e}_1^K,\mathbf{e}_2^K)$ reference system from Figure 4.1a.

However, in his study Simon has decided to study the motion of the oscillator with respect to the clamping table (not with respect to the vinyl disc). Therefore, Simon uses the reference field R and not $\tilde{R}$. The transformation rules (3.21) together with the coordinate transformations (4.4) allow him to express the reference field R with respect to the coordinate fields induced by the chart ψ as

$$R = \frac{\partial}{\partial t} = \frac{\partial}{\partial \bar{t}} - \Omega y^2 \frac{\partial}{\partial y^1} + \Omega y^1 \frac{\partial}{\partial y^2}.$$

The reference fields R and $\tilde{R}$ are depicted in Figure 4.4.

We have already mentioned in Section 4.7 that the mass of a mechanical system is modelled by the Galilean metric g that comes with the system's Galilean manifold (M,ϑ,g). Consequently, we define the **kinetic energy** with respect to the reference field $R\colon M \supseteq U_R \to A^1M$ as the function

$$\begin{aligned} T_R\colon \pi^{-1}(U_R) &\to \mathbb{R}, \\ (p,v_p) &\mapsto \tfrac{1}{2}\, g_p\big(v_p - R_p, v_p - R_p\big), \end{aligned} \tag{4.74}$$

with $v_p \in A_p^1 M$ and $R(p) = (p,R_p)$. The kinetic energy of the motion (4.73) with respect to the reference field R is then given by

$$T_R\big(\dot{\gamma}(\tau)\big) = \tfrac{1}{2}\, g_{\gamma(\tau)}\big(\dot{\gamma}_{\gamma(\tau)} - R_{\gamma(\tau)}, \dot{\gamma}_{\gamma(\tau)} - R_{\gamma(\tau)}\big).$$

Let (U,ϕ) be an adapted chart of M and let us assume for simplicity that $U \subseteq U_R$. Let $R = \partial/\partial t + R^i \partial/\partial x^i$ be an arbitrary reference field and

$v = \partial/\partial t + u^i \partial/\partial x^i$ a time-normalized vector field on M. Then the kinetic energy (4.74) locally reads

$$T_R = \underbrace{\tfrac{1}{2}g_{ij}u^iu^j}_{T_{R,2}} \underbrace{-g_{ij}u^iR^j}_{T_{R,1}} + \underbrace{\tfrac{1}{2}g_{ij}R^iR^j}_{T_{R,0}}, \tag{4.75}$$

where we used the local expression of the metric (4.13) and the symmetry of g, i.e., that $g_{ij} = g_{ji}$. By equation (4.75), the kinetic energy is the sum of

$$T_{R,2} \coloneqq \tfrac{1}{2}g_{ij}u^iu^j, \quad T_{R,1} \coloneqq -g_{ij}u^iR^j, \quad \text{and } T_{R,0} \coloneqq \tfrac{1}{2}g_{ij}R^iR^j.$$

The number in the subscript describes the respective degree of positive homogeneity[29] of each term with respect to $\mathbf{u} = (u^1, \dots, u^n)$. In the special case where R is a resting field (i.e., $R = \partial/\partial t$), the local expression of the kinetic energy (4.75) reduces to

$$T_R(p, v_p) = \tfrac{1}{2}g_{ij}u^iu^j.$$

Postulate 4.6. Let (M, ϑ, g) be the Galilean manifold of a finite-dimensional mechanical system. The force-free motion of the mechanical system with respect to the reference field R is defined by the action form

$$\Omega_R \coloneqq \mathrm{d}\big(T_R\hat{\vartheta} + \boldsymbol{\partial} T_R\big), \tag{4.76}$$

where T_R denotes the kinetic energy (4.74) relative to R.

The postulate is motivated by the following proposition.

Proposition 4.7 (Loos 1982, p. 35)**.** The differential two-form Ω_R defined by (4.76) is indeed an action form that induces a bundle metric g on A^0M. The difference between an (arbitrary) action form Ω and Ω_R is a force two-form $\Phi_R \coloneqq \Omega - \Omega_R$.

Proof. To check that (4.76) defines an action form, we have to check the properties (i) to (iii) from Theorem 4.3. According to the rules (4.54), (4.55), and (4.56) of $\boldsymbol{\partial}$, it holds that

$$\boldsymbol{\partial}\Omega_R = -\mathrm{d}\big(\boldsymbol{\partial} T_R \wedge \hat{\vartheta} + \boldsymbol{\partial}\boldsymbol{\partial} T_R\big) = -\mathrm{d}\big(\boldsymbol{\partial} T_R \wedge \hat{\vartheta} + \hat{\vartheta} \wedge \boldsymbol{\partial} T_R\big) = 0,$$

which shows that Ω_R enjoys property (iii). To establish properties (i) and (ii), we consider the local expression (4.75) of the kinetic energy and we calculate using the rules (4.54) that

$$\begin{aligned} T_R\hat{\vartheta} + \boldsymbol{\partial} T_R &= T_R\,\mathrm{d}t + \frac{\partial T_R}{\partial u^i}\big(\mathrm{d}x^i - u^i\mathrm{d}t\big) \\ &= -\big(T_{R,2} - T_{R,0}\big)\mathrm{d}t + p_i\,\mathrm{d}x^i, \end{aligned} \tag{4.77}$$

29. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be differentiable. The function f is called **positively homogeneous** of degree k if $f(\alpha\mathbf{x}) = \alpha^k f(\mathbf{x})$ for all $\alpha \in \mathbb{R}_0^+$ and all $\mathbf{x} \in \mathbb{R}^n$.

where we defined $p_i := \partial T_R / \partial u^i$ and we used the positive homogeneity of the terms $T_{R,2}$, $T_{R,1}$, and $T_{R,0}$ that compose the kinetic energy (4.75). By Euler's theorem, we know that a function $f\colon \mathbb{R}^n \to \mathbb{R}$ that is homogeneous of degree k satisfies

$$x^i \frac{\partial f}{\partial x^i} = kf.$$

Consequently, it holds that

$$p_i u^i = 2T_{R,2} + T_{R,1}.$$

The exterior derivative of (4.77) can be written in terms of the basis covector fields $\mathrm{d}t, \mathrm{d}x^1, \dots, \mathrm{d}x^n, \mathrm{d}p_1, \dots, \mathrm{d}p_n$ as

$$\Omega_R = \left(\mathrm{d}p_i + \frac{\partial K}{\partial x^i}\mathrm{d}t\right) \wedge \left(\mathrm{d}x^i - u^i \mathrm{d}t\right), \tag{4.78}$$

with $K := T_{R,2} - T_{R,0}$. It is clear from the expression (4.78) that Ω_R vanishes on $\ker \mu$. Moreover, it follows that

$$\Omega_R\left(\frac{\partial}{\partial u^j}, \frac{\partial}{\partial x^i}\right) = \mathrm{d}p_i\left(\frac{\partial}{\partial u^j}\right) = \frac{\partial p_i}{\partial u^j} = \frac{\partial^2 T_R}{\partial u^j \partial u^i} = g_{ij}.$$

A comparison of the local expression (4.78) with (4.59) shows that $\Omega - \Omega_R$ is semi-basic. The assertion that $\Phi_R = \Omega - \Omega_R$ is a force two-form follows because $\boldsymbol{\partial}\left(\Omega - \Omega_R\right) = \boldsymbol{\partial}\Omega - \boldsymbol{\partial}\Omega_R = 0$. □

Postulate 4.6 defines the action form Ω_R that describes the motion with respect to the field R of a mechanical system which is not subjected to forces (i.e., $\Phi_R = 0$). Proposition 4.7 tells us that a given arbitrary action form Ω decomposes as $\Omega = \Omega_R + \Phi_R$. If $\Phi_R = 0$, then we say that Ω defines a force-free motion with respect to the reference field R. We say that Φ_R is **the force two-form** that appears **with respect to the reference field** R.

4.9. Classification of forces

4.9.1. Inertia forces

Let R and $\tilde{R}$ be two reference fields and Ω be a given action form. By Proposition 4.7, we know that Ω decomposes as

$$\Omega = \Omega_R + \Phi_R = \Omega_{\tilde{R}} + \Phi_{\tilde{R}}$$

and, therefore,

$$\Omega_{\tilde{R}} - \Omega_R = \Phi_R - \Phi_{\tilde{R}}.$$

Being the difference of two forces Φ_R and $\Phi_{\tilde{R}}$, the two-form $\Omega_{\tilde{R}} - \Omega_R$ is a force two-form that we call

$$\Psi_{R,\tilde{R}} := \Omega_{\tilde{R}} - \Omega_R \tag{4.79}$$

the **inertia force two-form** between the reference fields R and $\tilde{R}$. The force Φ_R with respect to R is composed of the force $\Phi_{\tilde{R}}$ with respect to $\tilde{R}$ and of the inertia force two-form (4.79). As force two-form, the latter is a kinematic quantity because it does not depend on the motion Z. It depends only on the Galilean manifold (M, ϑ, g) and on the reference fields R and $\tilde{R}$, as can be seen from the following considerations.

Because the inertia force two-form (4.79) is given by the difference of two exact two-forms, it is exact. This means that there exists a one-form $\hat{\alpha}_{R,\tilde{R}}$ such that

$$\Psi_{R,\tilde{R}} = \Omega_{\tilde{R}} - \Omega_R = \mathrm{d}\hat{\alpha}_{R,\tilde{R}}.$$

By definition (4.76) and the linearity of the differential operators d and $\boldsymbol{\partial}$, we know that

$$\hat{\alpha}_{R,\tilde{R}} = \left(T_{\tilde{R}} - T_R\right)\hat{\vartheta} + \boldsymbol{\partial}\left(T_{\tilde{R}} - T_R\right).$$

Let the reference field R be defined on U_R such that $R = \partial/\partial t + R^i \partial/\partial x^i$ and let the field $\tilde{R}$ be given by $\tilde{R} = \partial/\partial t + \tilde{R}^i \partial/\partial x^i$ on $U_{\tilde{R}}$ such that $U_R \cap U_{\tilde{R}} \neq \emptyset$. Equations (4.75) and (4.77) lead to

$$\begin{aligned}\hat{\alpha}_{R,\tilde{R}} &= \left(T_{\tilde{R}} - T_R\right)\hat{\vartheta} + \boldsymbol{\partial}\left(T_{\tilde{R}} - T_R\right) \\ &= \tfrac{1}{2} g_{ij}\left(\tilde{R}^i \tilde{R}^j - R^i R^j\right)\mathrm{d}t + g_{ij}\left(R^j - \tilde{R}^j\right)\mathrm{d}x^i.\end{aligned} \tag{4.80}$$

on $\pi^{-1}(U_R) \cap \pi^{-1}(U_{\tilde{R}})$. The local expression (4.80) reveals that $\hat{\alpha}_{R,\tilde{R}}$ is a basic form. This motivates the following alternative definition of the one-form $\hat{\alpha}_{R,\tilde{R}}$. Indeed, let $\alpha_{R,\tilde{R}}$ be the one-form on $U_R \cap U_{\tilde{R}}$ defined by requiring

$$\alpha_{R,\tilde{R}}(R) \overset{!}{=} \tfrac{1}{2} g(R - \tilde{R}, R - \tilde{R}),$$

$$\alpha_{R,\tilde{R}}(v) \overset{!}{=} g(v, R - \tilde{R})$$

for all spacelike vector fields v, i.e., for all local sections of the spacelike bundle $A^0 M$. Then the one-form $\hat{\alpha}_{R,\tilde{R}}$ is the pullback of $\alpha_{R,\tilde{R}}$ with the natural projection, i.e.,

$$\hat{\alpha}_{R,\tilde{R}} = \pi_{\leftarrow}\left(\alpha_{R,\tilde{R}}\right).$$

This shows that the inertia force two-form (4.79) for the reference fields R and $\tilde{R}$ does indeed depend only on the Galilean manifold (M, ϑ, g) and, therefore, is a kinematic quantity.

If in classical mechanics the motion of a particle is studied with respect to a non-inertial frame of reference,[30] additional force effects appear in the equations of motions. These forces that result from the use of a non-inertial frame of reference instead of an inertial one are referred to as fictitious, apparent or as inertia forces. Two examples are the Coriolis force and the centrifugal force. In our presentation, these forces are provided by the inertia force two-form (4.79).

4.9.2. Potential forces

We say that a force F_R is a **potential force** if the related (see Theorem 4.4) force two-form Φ_R is closed, i.e., if

$$\mathrm{d}\Phi_R = 0. \tag{4.81}$$

According to the Poincaré lemma (see Lemma 3.32) there exists a neighbourhood $W \subseteq A^1M$ and a one-form β defined on W such that

$$\Phi_R\big|_W = \mathrm{d}\beta_R. \tag{4.82}$$

The closedness (exactness) of the force two-form implies the closedness (exactness) of the action form. Indeed, with Proposition 4.7, we saw that an action form is the sum of an exact form (see equation (4.76)) and the force two-form. Therefore, it makes sense to speak of a **closed (exact) mechanical system** if the force two-form is closed (exact).

Let us venture into the calculation of the coordinate expression of a closed two-form Φ_R by applying condition (4.81) to expression (4.71) such that

$$\begin{aligned}
0 &\overset{!}{=} \mathrm{d}\Phi_R \\
&= \mathrm{d}\left[\left(-F_i + \tfrac{1}{2}u^j\left(\frac{\partial F_i}{\partial u^j} - \frac{\partial F_j}{\partial u^i}\right)\right)\mathrm{d}t \wedge \mathrm{d}x^i + \tfrac{1}{2}\frac{\partial F_i}{\partial u^j}\mathrm{d}x^i \wedge \mathrm{d}x^j\right] \\
&= \left(-\frac{\partial F_i}{\partial x^k} + \tfrac{1}{2}\frac{\partial^2 F_i}{\partial t\,\partial u^k} + \tfrac{1}{2}u^j\left(\frac{\partial^2 F_i}{\partial x^k\,\partial u^j} - \frac{\partial^2 F_j}{\partial x^k\,\partial u^i}\right)\right)\mathrm{d}t \wedge \mathrm{d}x^i \wedge \mathrm{d}x^k \\
&\quad + \left(-\frac{\partial F_i}{\partial u^k} + \tfrac{1}{2}\left(\frac{\partial F_i}{\partial u^k} - \frac{\partial F_k}{\partial u^i}\right) + \tfrac{1}{2}u^j\left(\frac{\partial^2 F_i}{\partial u^k\,\partial u^j} - \frac{\partial^2 F_j}{\partial u^k\,\partial u^i}\right)\right)\mathrm{d}t \wedge \mathrm{d}x^i \wedge \mathrm{d}u^k \\
&\quad + \tfrac{1}{2}\frac{\partial^2 F_i}{\partial x^k\,\partial u^j}\mathrm{d}x^i \wedge \mathrm{d}x^j \wedge \mathrm{d}x^k + \tfrac{1}{2}\frac{\partial^2 F_i}{\partial u^k\,\partial u^j}\mathrm{d}x^i \wedge \mathrm{d}x^j \wedge \mathrm{d}u^k,
\end{aligned}$$

where we dropped the R by writing F instead of F_R for notational convenience. The above condition leads to restrictions on the real-valued functions F_i.

30. See Section 39 in Landau and E. M. Lifshitz 1969 or Sections IV.4–5 in Lanczos 1952.

The last term disappears if and only if

$$F_i(t,\mathbf{x},\mathbf{u}) = E_i(t,\mathbf{x}) + B_{ij}(t,\mathbf{x})u^j. \tag{4.83}$$

The vanishing of the $\mathrm{d}t\wedge\mathrm{d}x^i\wedge\mathrm{d}u^k$-term requires that

$$B_{ij} = -B_{ji}.$$

The annihilation of the $\mathrm{d}x^i\wedge\mathrm{d}x^j\wedge\mathrm{d}x^k$-term leads to

$$\sum_{\text{cyclic}} \frac{\partial B_{ij}}{\partial x^k} = 0. \tag{4.84}$$

Finally, the annihilation of the first term imposes that

$$\frac{\partial B_{ij}}{\partial t} = \frac{\partial E_i}{\partial x^j} - \frac{\partial E_j}{\partial x^i} \tag{4.85}$$

Consequently, a closed two-form Φ_R has the local form

$$\Phi_R = E_i \mathrm{d}x^i \wedge \mathrm{d}t + \tfrac{1}{2}B_{ij}\mathrm{d}x^i \wedge \mathrm{d}x^j. \tag{4.86}$$

The suggestive use of the letters B and E lets us identify (4.84) and (4.85) as a generalized version of Maxwell's equations. We see from the local expression (4.86) that a closed force two-form is basic. Therefore, as a one-form β_R satisfying (4.82), we consider the locally defined basic one-form

$$\begin{aligned}\beta_R &= -V_R(t,\mathbf{x})\,\mathrm{d}t + A_i^R(t,\mathbf{x})\,\mathrm{d}x^i \\ &= \big(-V_R(t,\mathbf{x}) + A_i^R(t,\mathbf{x})u^i\big)\mathrm{d}t + A_i^R(t,\mathbf{x})\,(\mathrm{d}x^i - u^i\mathrm{d}t) \\ &= \big(-V_R + A_i^R u^i\big)\mathrm{d}t + \partial\big(-V_R + A_i^R u^i\big).\end{aligned} \tag{4.87}$$

The second and third equality follow by telescopic expansion and by the rules (4.54), respectively. In the last line, we omitted the function arguments for brevity. In the context of a charged particle moving in an electromagnetic field the function V_R is known as scalar potential[31] of the field and the $\mathbb{R}^3$-tuple (A_1^R, A_2^R, A_3^R) is said to be its vector potential.[31] With the one-form (4.87) it holds that

$$E_i = -\left(\frac{\partial V_R}{\partial x^i} + \frac{\partial A_i^R}{\partial t}\right), \qquad B_{ij} = 2\frac{\partial A_j^R}{\partial x^i}.$$

It is important to notice that the Poincaré lemma guarantees the existence of a one-form β_R and *not* its uniqueness. Indeed, two one-forms β_R and β_R'

31. See p. 45 in Landau and E. Lifshitz 1971.

that differ by the differential $\mathrm{d}f$ of a function $f = f(t, \mathbf{x})$ lead to the same force two-form (4.82) because $\mathrm{d} \circ \mathrm{d} = 0$. With

$$\mathrm{d}f = \frac{\partial f}{\partial t}\mathrm{d}t + \frac{\partial f}{\partial x^i}\mathrm{d}x^i,$$

this implies that the coefficient functions of β_R from (4.87) are related to those of

$$\beta_R' = \beta_R + \mathrm{d}f = -V'\mathrm{d}t + A_i'\mathrm{d}x^i$$

by

$$V' = V - \frac{\partial f}{\partial t}, \quad \text{and} \quad A_i' = A_i + \frac{\partial f}{\partial x^i}, \tag{4.88}$$

without changing the resulting force two-form Φ_R. Note that we dropped the letter R in equation (4.88) for notational convenience. The invariance property (4.88) of the coefficient functions of the one-form β_R is known as **gauge invariance**.[32]

In classical mechanics (no electromagnetism), one assumes $B_{ij} = 0$ such that the coefficient functions (4.83) are independent of $u^1, \ldots, u^n$. In this case, the closed force two-form (4.86) reduces to

$$\Phi_R = E_i \mathrm{d}x^i \wedge \mathrm{d}t.$$

Accordingly, the one-form β_R from (4.87) reduces to

$$\beta_R = -V_R(t, \mathbf{x})\,\mathrm{d}t + A_i^R(t)\,\mathrm{d}x^i.$$

Because of the gauge invariance (4.88), we can add a differential $\mathrm{d}f$ without changing the resulting force two-form. We choose $f(t, \mathbf{x}) = -A_i^R(t)x^i$ such that

$$\beta_R' = \beta_R + \mathrm{d}f = \left(-V_R - \frac{\mathrm{d}A_i^R}{\mathrm{d}t}x^i \right)\mathrm{d}t =: -V_R'\mathrm{d}t.$$

This proves that in classical mechanics the force two-form of a potential force can be derived from a one-form

$$\beta_R' = -V_R'(t, \mathbf{x})\,\mathrm{d}t \tag{4.89}$$

without loss of generality. The coefficient function V_R' in (4.89) is known as **potential energy** with respect to the reference field R. In what follows, we will consider one-forms of the form (4.87) because they comprise the form (4.89) used in classical mechanics.

32. See Section 18 in Landau and E. Lifshitz 1971.

4.9.3. Nonpotential forces

The previous considerations allow us to split a given force two-form

$$\Phi_R = \Phi_R^{\mathrm{p}} + \Phi_R^{\mathrm{np}}$$

into a part $\Phi_R^{\mathrm{p}} = \mathrm{d}\beta_R$ that is defined by a one-form (4.87) and the remaining part Φ_R^{np} which we will refer to as **nonpotential force** two-form. By comparing equations (4.76) and (4.87), it is clear that the sum $\Omega_R + \Phi_R^{\mathrm{p}}$ can be written as

$$\begin{aligned}\Omega_R + \Phi_R^{\mathrm{p}} = \Omega_R + \mathrm{d}\beta_R &= \mathrm{d}\Big[\big(T_R - V_R + A_i^R u^i\big)\hat{\vartheta} + \partial\big(T_R - V_R + A_i^R u^i\big)\Big] \\ &= \mathrm{d}\big(L_R\hat{\vartheta} + \partial L_R\big),\end{aligned}$$

with $L_R \coloneqq T_R - V_R + A_i^R u^i$. This brings us to the following generalization of Postulate 4.6.

> **Postulate 4.8.** Let (M, ϑ, g) be the Galilean manifold of a finite-dimensional mechanical system. The motion of the mechanical system with respect to the reference field R under the influence of the force $\Phi_R = \Phi_R^{\mathrm{p}} + \Phi_R^{\mathrm{np}}$ is defined[33] by the action form
>
> $$\Omega \coloneqq \Omega_R^{\mathrm{p}} + \Phi_R^{\mathrm{np}} \tag{4.90}$$
>
> where
>
> $$\Omega_R^{\mathrm{p}} \coloneqq \Omega_R + \Phi_R^{\mathrm{p}} = \Omega_R + \mathrm{d}\beta_R = \mathrm{d}\big(L_R\hat{\vartheta} + \partial L_R\big), \tag{4.91}$$
>
> with the **Lagrangian**
>
> $$L_R \coloneqq T_R - V_R + A_i^R u^i \tag{4.92}$$
>
> that is defined with respect to the reference field R.

The one-form defined by the Lagrangian as

$$\omega_R \coloneqq L_R\hat{\vartheta} + \partial L_R \tag{4.93}$$

that is used in equation (4.91) is referred to as **Cartan one-form**. The Cartan one-form determines the Lagrangian by

$$L_R = X \lrcorner\, \omega_R, \tag{4.94}$$

33. The motion of a mechanical system is an integral curve of the second-order field associated with the action form Ω of the mechanical system. The second-order field is uniquely defined according to Theorem 4.3.

where X is the unique[34] vector field determined by

$$X \lrcorner \Omega = 0 \quad \text{and} \quad \hat{\vartheta}(X) = 1. \tag{4.95}$$

In the local coordinates induced on the neighbourhood $\pi^{-1}(U) \subseteq A^1M$ by the natural chart (4.17), the Cartan one-form reads

$$\omega_R = L_R \,\mathrm{d}t + \frac{\partial L_R}{\partial u^i}(\mathrm{d}x^i - u^i \mathrm{d}t), \tag{4.96}$$

where we will suppress the letter R on many occasions for ease of notation.

The exterior derivative establishes an assignment $\omega \mapsto \Omega^{\mathrm{p}}$ between Cartan one-forms and the set of action forms (i.e., differential two-forms having the properties specified by Theorem 4.3). This map is not surjective because action forms that are not closed cannot be reached. The assignment is not injective either. Indeed, if two Cartan one-forms ω_1 and ω_2 define the same action form

$$\Omega = \mathrm{d}\omega_1 = \mathrm{d}\omega_2,$$

they may still differ by the differential of a function $f \in C^\infty(A^1M)$, i.e.,

$$\omega_2 - \omega_1 = \mathrm{d}f \tag{4.97}$$

because $\mathrm{d} \circ \mathrm{d} = 0$. Since the difference $\omega_2 - \omega_1$ is semi-basic, the function f in (4.97) needs to satisfy

$$f = \pi_{\leftarrow} g = g \circ \pi$$

for some function $g \in C^\infty(M)$. The non-uniqueness of the Cartan one-form transfers to the Lagrangians by equation (4.94). The Lagrangians L_1 and L_2 defining the Cartan one-forms ω_1 and ω_2 can locally differ by

$$L_2 - L_1 = Z \lrcorner \mathrm{d}f = \frac{\partial f}{\partial t} + u^i \frac{\partial f}{\partial x^i} \tag{4.98}$$

and still define the same action form Ω. If the difference (4.98) is evaluated along a second-order curve $\dot{\gamma}$, then we get by (4.25) that

$$\begin{aligned}\big(L_2 - L_1\big)(\dot{\gamma}(\tau)) &= \left(\frac{\partial}{\partial t}[f] + u^i \frac{\partial}{\partial x^i}[f]\right) \circ \dot{\gamma}(\tau) \\ &= \left(\frac{\partial f \circ \Phi^{-1}}{\partial t} + u^i(\tau)\frac{\partial f \circ \Phi^{-1}}{\partial x^i}\right)\big(t(\tau), \mathbf{x}(\tau), \dot{\mathbf{x}}(\tau)\big),\end{aligned}$$

34. According to Theorem 4.2(iii), the vector field X obeying (4.95) is indeed uniquely determined.

from which we retrieve the classical statement[35] that (the chart representations of) the Lagrangians $\mathrm{L}_1 := L_1 \circ \Phi^{-1}(t, \mathbf{x}, \mathbf{u})$ and $\mathrm{L}_2 := L_2 \circ \Phi^{-1}(t, \mathbf{x}, \mathbf{u})$ may differ by the total derivative with respect to t of a function $\mathrm{f}(t, \mathbf{x})$, which depends on time t and the positions $\mathbf{x}$.

The following considerations allow us to pin down the form of Lagrangians. First, a Lagrangian needs to define the bundle metric by

$$g_{ij} = \Omega\left(\frac{\partial}{\partial u^i}, \frac{\partial}{\partial x^j}\right) = \frac{\partial^2 L}{\partial u^i \partial u^j},$$

according to Theorem 4.3(ii). Because it holds for the bundle metric that $\partial g_{ij} / \partial u^k = 0$, the Lagrangian needs to satisfy

$$\frac{\partial^3 L}{\partial u^k \partial u^i \partial u^j} = 0,$$

and it therefore has the local form

$$L = \tfrac{1}{2} g_{ij} u^i u^j + a_i u^i + a_0 \tag{4.99}$$

on $\pi^{-1}(U)$, with coefficients $a_0, \dots, a_n$ that do not depend on $u^1, \dots, u^n$. This means that there are functions $\bar{a}_0, \dots, \bar{a}_n : U \to \mathbb{R}$ defined on the neighbourhood U such that

$$a_\alpha := \pi_{\leftarrow}(\bar{a}_\alpha), = \bar{a}_\alpha \circ \pi$$

with $\alpha = 0, \dots, n$. With the local expression (4.75) of the kinetic energy the Lagrangian (4.92) can be written as

$$L_R = T_R - V_R + A_i^R u^i = \tfrac{1}{2} g_{ij} u^i u^j + \left(A_i^R - g_{ij} R^j\right) u^i + \tfrac{1}{2} g_{ij} R^i R^j - V_R.$$

The comparison with (4.99) leads to the equalities

$$\begin{aligned} a_0 &= \tfrac{1}{2} g_{ij} R^i R^j - V_R, \\ a_i &= A_i - g_{ij} R^j, \end{aligned} \tag{4.100}$$

with $i = 1, \dots, n$.

The case of classical mechanics where the one-form β_R reduces to (4.89) can be studied by setting $A_i = 0$. It follows from equation (4.100) that the reference field R and the potential V_R can be determined from the coefficients $a_0, \dots, a_n$ of a given Lagrangian (4.99) as

$$R^i = -g^{ij} a_j \tag{4.101}$$

35. See Landau and E. M. Lifshitz 1969 p. 4.

and

$$V_R = \tfrac{1}{2} g_{ij} R^i R^j - a_0.$$

The coefficients g^{ij} in equation (4.101) are given by the inverse matrix to the coefficient matrix of the Galilean metric g,

$$\left[g^{ij}\right] = \left[g_{ij}\right]^{-1},$$

such that $g^{ij} g_{jk} = \delta^i_k$.

4.10. Lagrangian and Hamiltonian mechanics

As a differentiable manifold, the state space A^1M allows a local description using charts. This means that Postulate 4.8 can be expressed with respect to different sets of local coordinates. We show that Lagrange's, Hamel's as well as Hamilton's equations are different coordinate representations of Postulate 4.8.

4.10.1. Lagrange's equations of the second kind

We start by using the local coordinates provided by the natural chart (4.17). By equation (4.91) and the rules (4.54), the action form Ω^{p}_R is locally given by

$$\Omega^{\mathrm{p}}_R = \mathrm{d}L \wedge \mathrm{d}t + \mathrm{d}\left(\frac{\partial L}{\partial u^i}\right) \wedge (\mathrm{d}x^i - u^i \mathrm{d}t) - \frac{\partial L}{\partial u^i} \mathrm{d}u^i \wedge \mathrm{d}t \tag{4.102}$$

and, by equation (4.71), the nonpotential force Φ^{np}_R in (4.90) reads

$$\Phi^{\mathrm{np}}_R = F_i \, \mathrm{d}x^i \wedge \mathrm{d}t + \frac{1}{2} \frac{\partial F_i}{\partial u^j} (\mathrm{d}x^i - u^i \mathrm{d}t) \wedge (\mathrm{d}x^j - u^j \mathrm{d}t). \tag{4.103}$$

Note that we lightened the notation by suppressing the reference field R when writing the Lagrangian in (4.102).

By Theorem 4.2(iii), we know that the action form $\Omega = \Omega^{\mathrm{p}}_R + \Phi^{\mathrm{np}}_R$ determines the vector field $X \in \mathrm{Vect}(A^1M)$ that describes the motion by

$$\hat{\vartheta}(X) = 1, \tag{4.104}$$

$$X \lrcorner \Omega = 0. \tag{4.105}$$

Condition (4.104) requires the vector field X to be time-normalized such that it can be locally written as

$$X = \frac{\partial}{\partial t} + A^i \frac{\partial}{\partial x^i} + B^i \frac{\partial}{\partial u^i}, \tag{4.106}$$

where the coefficients A^i and B^i with $i = 1, \dots, n$ are smooth real-valued functions defined on the neighbourhood $\pi^{-1}(U) \subseteq A^1 M$. Condition (4.105) can be rewritten as

$$0 = X \lrcorner \Omega = X \lrcorner \Omega_R^{\mathrm{p}} + X \lrcorner \Phi_R^{\mathrm{np}}. \tag{4.107}$$

Using equations (4.102), (4.103), (4.106), property (i) from Proposition 3.35 as well as the definition (3.62) of the exterior derivative of a real-valued function, we compute both terms separately as

$$\begin{aligned}
X \lrcorner \Omega_R^{\mathrm{p}} &= \mathfrak{L}_X L \,\mathrm{d}t - \mathrm{d}L + \mathfrak{L}_X\left(\frac{\partial L}{\partial u^i}\right)\left(\mathrm{d}x^i - u^i \mathrm{d}t\right) - \mathrm{d}\left(\frac{\partial L}{\partial u^i}\right)\left(A^i - u^i\right) \\
&\qquad - B^i \frac{\partial L}{\partial u^i}\mathrm{d}t + \frac{\partial L}{\partial u^i}\mathrm{d}u^i \\
&= \left[\mathfrak{L}_X L - \frac{\partial L}{\partial t} - u^i \mathfrak{L}_X\left(\frac{\partial L}{\partial u^i}\right) - \frac{\partial^2 L}{\partial t \partial u^i}\left(A^i - u^i\right) - B^i \frac{\partial L}{\partial u^i}\right]\mathrm{d}t \\
&\quad + \left[\mathfrak{L}_X\left(\frac{\partial L}{\partial u^i}\right) - \frac{\partial L}{\partial x^i} - \frac{\partial^2 L}{\partial x^i \partial u^j}\left(A^j - u^j\right)\right]\mathrm{d}x^i \\
&\quad + \frac{\partial^2 L}{\partial u^i \partial u^j}\left(u^j - A^j\right)\mathrm{d}u^i
\end{aligned} \tag{4.108}$$

and

$$\begin{aligned}
X \lrcorner \Phi_R^{\mathrm{np}} &= \left[A^i F_i - \tfrac{1}{2} u^i \left(A^j - u^j\right)\left(\frac{\partial F_j}{\partial u^i} - \frac{\partial F_i}{\partial u^j}\right)\right]\mathrm{d}t \\
&\qquad + \left[-F_i + \tfrac{1}{2}\left(A^j - u^j\right)\left(\frac{\partial F_j}{\partial u^i} - \frac{\partial F_i}{\partial u^j}\right)\right]\mathrm{d}x^i.
\end{aligned} \tag{4.109}$$

By equation (4.107), the sum of (4.108) and (4.109) has to vanish. In particular, the $\mathrm{d}u^i$-component of (4.108) must be zero. This implies that

$$A^j = u^j \quad \text{for } j = 1, \dots, n \tag{4.110}$$

because the matrix

$$\left[\frac{\partial^2 L}{\partial u^i \partial u^j}\right] = \left[g_{ij}\right]$$

is positive definite and thus has full rank. Equation (4.110) requires the vector field X to be a second-order field as we saw in (4.33). The annihilation of the $\mathrm{d}x^i$-part of the sum (4.107) together with (4.110) leads to **Lagrange's equations** of the second kind

$$\mathfrak{L}_X\left(\frac{\partial L}{\partial u^i}\right) - \frac{\partial L}{\partial x^i} = F_i. \tag{4.111}$$

Let us consider a time-parametrized integral curve $\beta\colon I \to A^1M$, $\tau \mapsto \beta(\tau)$ of the vector field X defined by equation (4.111), i.e., a curve for which holds

$$\dot{\beta}(\tau) = X\big(\beta(\tau)\big). \tag{4.112}$$

Since X is a second-order field by (4.110), we know by equation (4.34) that

$$\dot{\mathbf{x}}(\tau) = \mathbf{u}(\tau). \tag{4.113}$$

The integral curve β defined by (4.112) needs to satisfy equation (4.111) such that

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\partial L}{\partial u^i}\circ\beta(\tau)\right) - \frac{\partial L}{\partial x^i}\circ\beta(\tau) = F_i\circ\beta(\tau) \tag{4.114}$$

by definition (3.22) of the Lie derivative (see p. 60). By the definition of vector fields as derivations on $C^\infty(A^1M)$ (see equations (3.13) and (3.43)) together with equation (4.113), we recognize (4.114) as Lagrange's equations of the second kind in their classical form[36]

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\partial \mathrm{L}}{\partial u^i}\big(t(\tau), \mathbf{x}(\tau), \dot{\mathbf{x}}(\tau)\big)\right) - \frac{\partial \mathrm{L}}{\partial x^i}\big(t(\tau), \mathbf{x}(\tau), \dot{\mathbf{x}}(\tau)\big) = \mathrm{F}_i(\,\dots\,),$$

where the upright letters $\mathrm{L} \coloneqq L\circ\Phi^{-1}$ and $\mathrm{F}_i \coloneqq F_i\circ\Phi^{-1}$ denote the representations of the Lagrangian L and of the coefficient functions $F_1, \dots, F_n$ with respect to the natural chart (4.17). We abridged the function arguments on the right-hand side by dots to avoid a line break. The function $t(\tau) = t\circ\beta(\tau)$ from (4.35) is an affine function of the parameter τ because of equations (4.104) and (4.112).

4.10.2. Hamel's equations

In the previous section, we showed that Postulate 4.8 directly leads to Lagrange's equations of the second kind when the involved objects are expressed in the natural chart (4.17). By equation (4.113), the representation of the motion with respect to the local coordinates of the natural chart (4.17) satisfies

$$\dot{\mathbf{x}}(\tau) = \mathbf{u}(\tau),$$

which is just the local expression of the second-order condition (4.24) as we saw with (4.25). This means that the velocity coordinates of the motion

$$u^1(\tau) = \Phi^{n+1}\circ\beta(\tau),\ \dots,\ u^n(\tau) = \Phi^{2n}\circ\beta(\tau)$$

36. See Lagrange 1780, p. 25 or Landau and E. M. Lifshitz 1969, p. 3.

provided by the natural chart correspond to the derivatives with respect to τ of the position coordinates

$$x^1(\tau) = \Phi^1 \circ \beta(\tau),\ \dots,\ x^n(\tau) = \Phi^n \circ \beta(\tau).$$

Now, for some applications, e.g., the formulation of constraints (see Section 4.12), it may be beneficial to describe the motion using a different set of velocity parameters, i.e., to consider another chart of the state space than the natural chart. In this section, we will derive the equations of motion that appear in place of Lagrange's equations (4.111) if we consider such an alternative chart in which the velocity coordinates $u^1, \dots, u^n$ are replaced by a set of parameters $v^1, \dots, v^n$ that result from the $u^1, \dots, u^n$ by an affine transformation (see equation (4.127)).

The generalized velocities $u^1, \dots, u^n$ were introduced in (4.16) as the parameters that uniquely define a time-normalized vector $v_p \in A^1_p M$ at some point $p \in U \subseteq M$

$$v_p = \left.\frac{\partial}{\partial t}\right|_p + u^i \left.\frac{\partial}{\partial x^i}\right|_p. \tag{4.115}$$

The basis vectors of T_pM that are used in (4.115) are those induced by an adapted ($x^0 = t$) chart

$$\phi\colon M \supseteq U \to \mathbb{R}^{n+1},\ p \mapsto (x^0, \dots, x^n) \tag{4.116}$$

of the Galilean manifold (M, ϑ, g). We take a step back and consider an *arbitrary* tangent vector $w_p \in T_pM$ at some point $p \in M$ given by

$$w_p = u^\nu \left.\frac{\partial}{\partial x^\nu}\right|_p, \tag{4.117}$$

where $\nu = 0, \dots, n$. In what follows, we use lowercase Greek letters to denote indices that range from 0 to n, while lowercase Latin letters stand for indices between 1 and n. The tangent space T_pM is a vector space of dimension $n+1$.[37] It is reasonable to study the set of velocity coordinates that originate from the $u^0, \dots, u^n$ by a linear transformation (see Section 2.3). We define a set of new basis vectors

$$\left.\mathfrak{b}_\sigma\right|_p := B^\nu_\sigma \left.\frac{\partial}{\partial x^\nu}\right|_p, \tag{4.118}$$

where $\sigma = 0, \dots, n$ and the coefficients B^ν_σ form a regular $(n+1)$-by-$(n+1)$ matrix. The tangent vector w_p from (4.117) can be represented as

$$w_p = v^\sigma \left.\mathfrak{b}_\sigma\right|_p = B^\nu_\sigma v^\sigma \left.\frac{\partial}{\partial x^\nu}\right|_p,$$

37. See Theorem 3.15 for the proof.

with

$$u^\nu = B^\nu_\sigma v^\sigma. \tag{4.119}$$

We assume that the coefficients B^ν_σ smoothly depend on the point, i.e., that they are smooth real-valued functions defined on the manifold M

$$B^\nu_\sigma : M \to \mathbb{R},\ p \mapsto B^\nu_\sigma(p),$$

which satisfy

$$\det\left[B^\nu_\sigma(p)\right] \neq 0 \tag{4.120}$$

for all $p \in M$. In the context of principal bundles the basis $(\mathfrak{b}_0|_p, \dots, \mathfrak{b}_n|_p)$ could be defined as a section of the so-called bundle of frames of M. However, we abstain from introducing the theory of principal bundles and refer the interested reader to Chapter 8 in Spivak 1999b.

With assumption (4.120), the point-by-point requirement (4.118) defines local basis fields

$$\mathfrak{b}_\sigma : M \supseteq U \to TM,\ p \mapsto \mathfrak{b}_\sigma(p) = \left(p, \mathfrak{b}_\sigma|_p\right).$$

The fact that the chart (4.116) is adapted implies for its induced basis vectors $\partial/\partial x^0, \dots, \partial/\partial x^n$ that

$$\vartheta\left(\frac{\partial}{\partial x^0}\right) = \mathrm{d}t\left(\frac{\partial}{\partial t}\right) = 1$$

and that

$$\vartheta\left(\frac{\partial}{\partial x^i}\right) = \mathrm{d}t\left(\frac{\partial}{\partial x^i}\right) = 0,$$

with $i = 1, \dots, n$. The analogue requirement for the basis fields $(\mathfrak{b}_0, \dots, \mathfrak{b}_n)$ reads

$$\vartheta(\mathfrak{b}_0) \stackrel{!}{=} 1 \quad \text{and} \quad \vartheta(\mathfrak{b}_i) \stackrel{!}{=} 0, \quad \text{for } i = 1, \dots, n \tag{4.121}$$

and imposes the restrictions

$$B^0_0 = 1 \quad \text{and} \quad B^0_i = 0, \quad \text{for } i = 1, \dots, n \tag{4.122}$$

on the $(n+1)^2$ coefficients functions $B^\nu_\sigma : M \to \mathbb{R}$. Therefore, the linear transformation (4.118) is restricted to

$$\begin{aligned} \mathfrak{b}_0 &= \frac{\partial}{\partial x^0} + B^i_0 \frac{\partial}{\partial x^i} = \frac{\partial}{\partial t} + b^i \frac{\partial}{\partial x^i}, \\ \mathfrak{b}_i &= B^j_i \frac{\partial}{\partial x^j} \end{aligned} \tag{4.123}$$

if it should be adapted to the time structure. Note that we introduced $b^i := B^i_0$ in the first equation for notational convenience. The coefficient functions B^ν_σ from (4.123) can be written in matrix form as

$$\left[B^\nu_\sigma\right] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ b^1 & B^1_1 & \cdots & B^1_n \\ \vdots & \vdots & \ddots & \vdots \\ b^n & B^n_1 & \cdots & B^n_n \end{bmatrix}.$$

By using the rule for expansion of a determinant[38] according to the first row, we obtain for the regularity condition that

$$\det\left[B^\nu_\sigma\right] \neq 0 \Leftrightarrow \det\left[B^i_j\right] \neq 0. \tag{4.124}$$

Therefore, the linear coordinate changes on the tangent space of M that are compatible with the time structure are given by choosing n^2 coefficient functions $B^i_j\colon M \to \mathbb{R}$ defining a regular n-by-n matrix for each point $p \in M$ and by n functions $b^i\colon M \to \mathbb{R}$. We have already observed that the state space A^1M is an affine subbundle (see Section 4.3) of the tangent bundle TM and, therefore, it is not surprising that the transformations (4.123) compatible with the time structure are affine transformations.

It can be readily observed from the transformation rule (4.119) that the restrictions (4.122) imply that $u^0 = v^0$. The basis fields (4.123) complying with the time structure can be used to express a time-normalized $v_p \in A^1_pM$ as

$$v_p = \mathfrak{b}_0|_p + v^i\,\mathfrak{b}_i|_p. \tag{4.125}$$

The adapted chart (4.116) together with the basis fields $\mathfrak{b}_0, \dots, \mathfrak{b}_n$ define a chart on the state space A^1M by

$$\begin{aligned} \Psi\colon A^1M \supseteq \pi^{-1}(U) &\to \mathbb{R}^{2n+1}, \\ (p, v_p) &\mapsto (\bar{t}, \bar{x}^1, \dots, \bar{x}^n, v^1, \dots, v^n), \end{aligned} \tag{4.126}$$

where the coordinates $v^1, \dots, v^n$ are the coefficients from (4.125) that satisfy

$$u^i = B^i_j v^j + b^i \tag{4.127}$$

because of the transformation rules (4.123). The first $n+1$ coordinates $\bar{t}, \bar{x}^1, \dots, \bar{x}^n$ are provided by the adapted chart (4.116) of the base manifold M. Therefore, they are equal to the coordinates $t, x^1, \dots, x^n$ provided by the natural chart (4.17). The bars on t and the x^i shall allow to distinguish both

38. See Theorem 2.4 in Lang 2004.

sets of coordinates in calculations. Please be aware that $\bar{t} = t$ and $\bar{x}^i = x^i$ imply $\mathrm{d}\bar{t} = \mathrm{d}t$ and $\mathrm{d}\bar{x}^i = \mathrm{d}x^i$, but that in general $\partial/\partial\bar{t} \neq \partial/\partial t$ and $\partial/\partial\bar{x}^i \neq \partial/\partial x^i$. By the regularity condition (4.124), the coefficient matrix $[B^i_j]$ is invertible and we denote its inverse matrix by $[A^j_k]$ such that

$$B^i_j A^j_k = \delta^i_k.$$

Using the newly defined coefficient functions A^j_k, the coordinate change from (4.127) can be rewritten as

$$v^i = A^i_j (u^j - b^j). \tag{4.128}$$

With equations (4.127) and (4.128), the change of coordinates between the natural chart (4.17) and the chart (4.126) is given by

$$\begin{aligned} t &= \bar{t}, \\ x^i &= \bar{x}^i, \\ u^i &= B^i_j v^j + b^i \end{aligned} \tag{4.129}$$

and

$$\begin{aligned} \bar{t} &= t, \\ \bar{x}^i &= x^i, \\ v^i &= A^i_j (u^j - b^j), \end{aligned} \tag{4.130}$$

respectively. The chart (4.126) induces the basis fields

$$\frac{\partial}{\partial \bar{t}}, \frac{\partial}{\partial \bar{x}^1}, \dots, \frac{\partial}{\partial \bar{x}^n}, \frac{\partial}{\partial v^1}, \dots, \frac{\partial}{\partial v^n},$$

which are local vector fields on $\pi^{-1}(U) \subseteq A^1M$. From the chart representation of the natural projection $\pi\colon A^1M \to M$ (see equation (4.18))

$$\begin{aligned} &\phi \circ \pi \circ \Psi^{-1}\colon \mathbb{R}^{2n+1} \supseteq \Psi\big(\pi^{-1}(U)\big) \to \phi(U) \subseteq \mathbb{R}^{n+1}, \\ &\qquad (\bar{t}, \bar{x}^1, \dots, \bar{x}^n, v^1, \dots, v^n) \mapsto (t, x^1, \dots, x^n) = (\bar{t}, \bar{x}^1, \dots, \bar{x}^n), \end{aligned}$$

we can see that for all points $a \in \pi^{-1}(U)$ the vector fields

$$\frac{\partial}{\partial v^1}, \dots, \frac{\partial}{\partial v^n},$$

provide a basis of $\mathrm{Ver}_a(A^1M) = \ker \mathrm{D}\pi_a \subset T_a(A^1M)$ and, therefore,

$$\mathrm{Ver}_a(A^1M) = \operatorname{span}\left\{ \frac{\partial}{\partial v^1}\bigg|_a, \dots, \frac{\partial}{\partial v^n}\bigg|_a \right\}.$$

Moreover, it holds that

$$T_a(A^1M) = \operatorname{span}\left\{ \frac{\partial}{\partial \bar{t}}\bigg|_a, \frac{\partial}{\partial \bar{x}^1}\bigg|_a, \dots, \frac{\partial}{\partial \bar{x}^n}\bigg|_a \right\} \oplus \mathrm{Ver}_a(A^1M)$$

and by the isomorphism (2.16), we know that

$$\operatorname{span}\left\{ \frac{\partial}{\partial \bar{t}}\bigg|_a, \frac{\partial}{\partial \bar{x}^1}\bigg|_a, \dots, \frac{\partial}{\partial \bar{x}^n}\bigg|_a \right\} \cong T_a(A^1M)/\mathrm{Ver}_a(A^1M).$$

Finally, using the isomorphism (4.40), we conclude that

$$\operatorname{span}\left\{ \frac{\partial}{\partial \bar{t}}\bigg|_a, \frac{\partial}{\partial \bar{x}^1}\bigg|_a, \dots, \frac{\partial}{\partial \bar{x}^n}\bigg|_a \right\} \cong T_{\pi(a)}M. \tag{4.131}$$

Since

$$T_{\pi(a)}M = \operatorname{span}\left\{ \frac{\partial}{\partial t}\bigg|_{\pi(a)}, \frac{\partial}{\partial x^1}\bigg|_{\pi(a)}, \dots, \frac{\partial}{\partial x^n}\bigg|_{\pi(a)} \right\},$$

the isomorphism (4.131) can be expressed as

$$\frac{\partial}{\partial \bar{x}^\nu}\bigg|_a \mapsto \frac{\partial}{\partial x^\nu}\bigg|_{\pi(a)}. \tag{4.132}$$

After the identification of these bases, equation (4.123) defines the basis vectors

$$\begin{aligned} \mathfrak{b}'_0|_a &:= \frac{\partial}{\partial \bar{t}}\bigg|_a + b^i \circ \pi(a) \frac{\partial}{\partial \bar{x}^i}\bigg|_a, \\ \mathfrak{b}'_i|_a &= B^j_i \circ \pi(a) \frac{\partial}{\partial \bar{x}^j}\bigg|_a \end{aligned} \tag{4.133}$$

on the neighbourhood $\pi^{-1}(U) \subseteq A^1M$. It can be easily verified that the $\mathfrak{b}'_\nu$ are compatible (see equation (4.121)) with the time structure $\hat{\vartheta}$ of the state space (see equation (4.19)). We will drop the prime in the notation of the basis vectors (4.133) because the distinction remains possible via the point of evaluation similar to the basis vectors in (4.132). Moreover, we refrain from denoting the coefficient functions in (4.133) by $\hat{B}^j_i = B^j_i \circ \pi$ and $\hat{b}^i = b^i \circ \pi$, as we did for the coefficients of $\hat{g}$ in (4.44). Because this notation would be to heavy, we use the same symbol for both types of coefficient functions $B^j_i\colon M \supseteq U \to \mathbb{R}^{n+1}$ and $B^j_i = B^j_i \circ \pi\colon \pi^{-1}(U) \to \mathbb{R}^{2n+1}$. The vectors

$$\mathfrak{b}_0|_a, \dots, \mathfrak{b}_n|_a, \frac{\partial}{\partial v^1}\bigg|_a, \dots, \frac{\partial}{\partial v^n}\bigg|_a \tag{4.134}$$

provide a basis of $T_a(A^1M)$ for all points $a \in \pi^{-1}(U) \subseteq A^1M$.

Let $(\mathfrak{b}^0, \dots, \mathfrak{b}^n, \mathrm{d}v^1, \dots, \mathrm{d}v^n)$ denote the local covector fields that are dual to the basis fields (4.134) such that

$$\mathfrak{b}^\nu\Big(\mathfrak{b}_\sigma\Big) \stackrel{!}{=} \delta^\nu_\sigma, \qquad \mathrm{d}v^i\Big(\mathfrak{b}_\sigma\Big) \stackrel{!}{=} 0$$
$$\mathfrak{b}^\nu\Big(\frac{\partial}{\partial v^j}\Big) \stackrel{!}{=} 0, \qquad \mathrm{d}v^i\Big(\frac{\partial}{\partial v^j}\Big) \stackrel{!}{=} \delta^i_j,$$

with $\nu, \sigma = 0, \dots, n$ and $i, j = 1, \dots, n$. For calculations in local coordinates it is helpful to express the covector fields $\mathfrak{b}^0, \dots, \mathfrak{b}^n, \mathrm{d}v^1, \dots, \mathrm{d}v^n$ in terms of the dual coordinate fields $\mathrm{d}\bar{t}, \mathrm{d}\bar{x}^1, \dots, \mathrm{d}\bar{x}^n, \mathrm{d}v^1, \dots, \mathrm{d}v^n$ and $\mathrm{d}t, \mathrm{d}x^1, \dots, \mathrm{d}x^n$, $\mathrm{d}u^1, \dots, \mathrm{d}u^n$ induced by the chart (4.126) and the natural chart (4.17), respectively. Straightforward computation leads to

$$\begin{aligned}
\mathfrak{b}^0 &= \mathrm{d}\bar{t} = \mathrm{d}t,\\
\mathfrak{b}^i &= A^i_j(\mathrm{d}\bar{x}^j - b^j\mathrm{d}\bar{t}) = A^i_j(\mathrm{d}x^j - b^j\mathrm{d}t),\\
\mathrm{d}v^i &= \left(\frac{\partial A^i_j}{\partial t}(u^j - b^j) - A^i_j\frac{\partial b^j}{\partial t}\right)\mathrm{d}t + \left(\frac{\partial A^i_j}{\partial x^k}(u^j - b^j) - A^i_j\frac{\partial b^j}{\partial x^k}\right)\mathrm{d}x^k + A^i_j\mathrm{d}u^j,
\end{aligned} \tag{4.135}$$

where the last line is obtained by taking the exterior derivative of the coordinate functions (4.128). Using the relations (4.135) and the exterior derivative of the expression (4.127), it follows that

$$\begin{aligned}
\mathrm{d}t &= \mathrm{d}\bar{t} = \mathfrak{b}^0,\\
\mathrm{d}x^i &= \mathrm{d}\bar{x}^i = B^i_j\mathfrak{b}^j + b^i\mathfrak{b}^0,\\
\mathrm{d}u^i &= \left[v^j\left(\frac{\partial B^i_j}{\partial \bar{t}} + b^k\frac{\partial B^i_j}{\partial \bar{x}^k}\right) + \frac{\partial b^i}{\partial \bar{t}} + b^k\frac{\partial b^i}{\partial \bar{x}^k}\right]\mathfrak{b}^0 + B^k_l\left(v^j\frac{\partial B^i_j}{\partial \bar{x}^k} + \frac{\partial b^i}{\partial \bar{x}^k}\right)\mathfrak{b}^l + B^i_j\mathrm{d}v^j.
\end{aligned} \tag{4.136}$$

By Theorem 4.2(iii), we know that the action form $\Omega = \Omega^{\mathrm{p}}_R + \Phi^{\mathrm{np}}_R$ determines the vector field $X \in \mathrm{Vect}(A^1M)$ that describes the motion by

$$\hat{\vartheta}(X) = 1, \tag{4.137}$$
$$X \lrcorner \Omega = 0. \tag{4.138}$$

Equation (4.137) requires the vector field X to be time-normalized. Because of $\mathfrak{b}^0 = \mathrm{d}t$ (see equation (4.135)), it locally holds that $\hat{\vartheta} = \mathfrak{b}^0$ and therefore the time-normalized vector field X (satisfying (4.137)) can be locally written as

$$X = \mathfrak{b}_0 + C^i\mathfrak{b}_i + D^i\frac{\partial}{\partial v^i}. \tag{4.139}$$

This time-normalized vector field is then uniquely determined by condition (4.138) that can be written as

$$0 = X \lrcorner \Omega = X \lrcorner \Omega^{\mathrm{p}}_R + X \lrcorner \Phi^{\mathrm{np}}_R, \tag{4.140}$$

where $\Omega^{\mathrm{p}}_R = \mathrm{d}\omega_R$ with $\omega_R = L_R\hat{\vartheta} + \boldsymbol{\partial} L_R$ according to Postulate 4.8. As before, we drop the letter R that stands for the reference field wherever it becomes a notational impediment. First, we inspect the term $X \lrcorner \Omega^{\mathrm{p}}_R = X \lrcorner \mathrm{d}\omega_R$ in (4.140). For this we express

$$\omega_R = L\hat{\vartheta} + \boldsymbol{\partial} L = L\,\mathrm{d}t + \frac{\partial L}{\partial u^i}(\mathrm{d}x^i - u^i\mathrm{d}t)$$

in terms of the coordinates $\bar{t}, \bar{x}^1, \dots, \bar{x}^n$ and the covector fields $\mathfrak{b}^0, \dots, \mathfrak{b}^n$ as

$$\omega_R = L\,\mathfrak{b}^0 + \frac{\partial L}{\partial v^i}(\mathfrak{b}^i - v^i\mathfrak{b}^0), \tag{4.141}$$

where we used equations (4.129), (4.136) and that

$$\frac{\partial}{\partial v^i} = \frac{\partial}{\partial v^i}[t]\frac{\partial}{\partial t} + \frac{\partial}{\partial v^i}[x^j]\frac{\partial}{\partial x^j} + \frac{\partial}{\partial v^i}[u^j]\frac{\partial}{\partial u^j} = B^j_i\frac{\partial}{\partial u^j}. \tag{4.142}$$

Taking the exterior derivative of the Cartan one-form (4.141) yields

$$\mathrm{d}\omega_R = \mathrm{d}L \wedge \mathfrak{b}^0 + \mathrm{d}\Big(\frac{\partial L}{\partial v^i}\Big) \wedge \Big(\mathfrak{b}^i - v^i\mathfrak{b}^0\Big) + \frac{\partial L}{\partial v^i}\Big(\mathrm{d}\mathfrak{b}^i - \mathrm{d}v^i \wedge \mathfrak{b}^0\Big) \tag{4.143}$$

and this leads to

$$\begin{aligned} X \lrcorner \Omega^{\mathrm{p}}_R &= X \lrcorner \mathrm{d}\omega_R \\ &= i_X\Big(\mathrm{d}L \wedge \mathfrak{b}^0\Big) + i_X \circ \mathrm{d}\Big(\frac{\partial L}{\partial v^i}\Big)\Big(\mathfrak{b}^i - v^i\mathfrak{b}^0\Big) \\ &\quad - \mathrm{d}\Big(\frac{\partial L}{\partial v^i}\Big) i_X\Big(\mathfrak{b}^i - v^i\mathfrak{b}^0\Big) + \frac{\partial L}{\partial v^i}\Big(i_X \circ \mathrm{d}\mathfrak{b}^i - i_X\Big(\mathrm{d}v^i \wedge \mathfrak{b}^0\Big)\Big). \end{aligned} \tag{4.144}$$

We digest expression (4.144) term by term. The first one can be readily expressed as

$$i_X\Big(\mathrm{d}L \wedge \mathfrak{b}^0\Big) = \mathfrak{L}_X L\,\mathfrak{b}^0 - \mathrm{d}L = \Big(\mathfrak{L}_X L - \mathfrak{b}_0[L]\Big)\mathfrak{b}^0 - \mathfrak{b}_i[L]\mathfrak{b}^i - \frac{\partial L}{\partial v^i}\mathrm{d}v^i$$

using the definition (3.62) of the exterior derivative of a real-valued function. By the same argument,

$$i_X \circ \mathrm{d}\Big(\frac{\partial L}{\partial v^i}\Big) = \mathfrak{L}_X\Big(\frac{\partial L}{\partial v^i}\Big)$$

and the second term can be rewritten as

$$i_X \circ \mathrm{d}\Big(\frac{\partial L}{\partial v^i}\Big)\Big(\mathfrak{b}^i - v^i\mathfrak{b}^0\Big) = \mathfrak{L}_X\Big(\frac{\partial L}{\partial v^i}\Big)\Big(\mathfrak{b}^i - v^i\mathfrak{b}^0\Big).$$

With the local expression (4.139) of the vector field X, the third term of equation (4.144) becomes

$$-\mathrm{d}\Big(\frac{\partial L}{\partial v^i}\Big) i_X\big(\mathfrak{b}^i - v^i\mathfrak{b}^0\big) = -\mathrm{d}\Big(\frac{\partial L}{\partial v^i}\Big)\big(C^i - v^i\big),$$

with

$$\mathrm{d}\Big(\frac{\partial L}{\partial v^i}\Big) = \mathfrak{b}_0\Big[\frac{\partial L}{\partial v^i}\Big]\mathfrak{b}^0 + \mathfrak{b}_j\Big[\frac{\partial L}{\partial v^i}\Big]\mathfrak{b}^j + \frac{\partial}{\partial v^j}\Big[\frac{\partial L}{\partial v^i}\Big]\mathrm{d}v^j.$$

The final term

$$\frac{\partial L}{\partial v^i}\Big(i_X \circ \mathrm{d}\mathfrak{b}^i - i_X\big(\mathrm{d}v^i \wedge \mathfrak{b}^0\big)\Big)$$

needs a bit more care. By equation (4.135), we know that

$$\mathfrak{b}^i = A^i_j(\mathrm{d}x^j - b^j\mathrm{d}t)$$

and consequently

$$\begin{aligned}\mathrm{d}\mathfrak{b}^i &= \mathrm{d}\big(A^i_j(\mathrm{d}x^j - b^j\mathrm{d}t)\big)\\ &= \mathrm{d}A^i_j \wedge (\mathrm{d}x^j - b^j\mathrm{d}t) - A^i_j\,\mathrm{d}b^j \wedge \mathrm{d}t\\ &= B^j_k\,\mathrm{d}A^i_j \wedge \mathfrak{b}^k - A^i_j\,\mathrm{d}b^j \wedge \mathfrak{b}^0,\end{aligned}$$

where in the last line we made use of (4.136). Therefore, the interior product $i_X \circ \mathrm{d}\mathfrak{b}^i$ can be expressed using the local expression (4.139) of the vector field X as

$$\begin{aligned}i_X \circ \mathrm{d}\mathfrak{b}^i &= i_X\big(B^j_k\,\mathrm{d}A^i_j \wedge \mathfrak{b}^k - A^i_j\,\mathrm{d}b^j \wedge \mathfrak{b}^0\big)\\ &= B^j_k\,\mathfrak{L}_X A^i_j\,\mathfrak{b}^k - B^j_k C^k\big(\mathfrak{b}_0[A^i_j]\mathfrak{b}^0 + \mathfrak{b}_l[A^i_j]\mathfrak{b}^l\big)\\ &\qquad - A^i_j\big(\mathfrak{L}_X b^j - \mathfrak{b}_0[b^j]\big)\mathfrak{b}^0 + A^i_j\mathfrak{b}_k[b^j]\mathfrak{b}^k\\ &= -\Big(B^j_k C^k\mathfrak{b}_0[A^i_j] + A^i_j\big(\mathfrak{L}_X b^j - \mathfrak{b}_0[b^j]\big)\Big)\mathfrak{b}^0\\ &\qquad + \Big(B^j_k\,\mathfrak{L}_X A^i_j - B^j_l C^l\mathfrak{b}_k[A^i_j] + A^i_j\mathfrak{b}_k[b^j]\Big)\mathfrak{b}^k.\end{aligned} \tag{4.145}$$

The interior product $-i_X\big(\mathrm{d}v^i \wedge \mathfrak{b}^0\big)$ lets us recover our breath. Indeed, using the local expression (4.139) of the vector field X, it readily follows that

$$-i_X\big(\mathrm{d}v^i \wedge \mathfrak{b}^0\big) = -D^i\mathfrak{b}^0 + \mathrm{d}v^i,$$

such that

$$\begin{aligned}i_X \circ \mathrm{d}\mathfrak{b}^i - i_X\big(\mathrm{d}v^i \wedge \mathfrak{b}^0\big) &= -\Big(B^j_k C^k\mathfrak{b}_0[A^i_j] + A^i_j\big(\mathfrak{L}_X b^j - \mathfrak{b}_0[b^j]\big) + D^i\Big)\mathfrak{b}^0\\ &\qquad + \Big(B^j_k\,\mathfrak{L}_X A^i_j - B^j_l C^l\mathfrak{b}_k[A^i_j] + A^i_j\mathfrak{b}_k[b^j]\Big)\mathfrak{b}^k + \mathrm{d}v^i.\end{aligned}$$

Finally, we can gather our results and restate (4.144) with respect to the basis covector fields $\mathfrak{b}^0, \dots, \mathfrak{b}^n, \mathrm{d}v^1, \dots, \mathrm{d}v^n$ as

$$\begin{aligned} X \lrcorner \Omega_R^{\mathrm{p}} = \Bigg[& \mathfrak{L}_X L - \mathfrak{b}_0[L] - v^i \mathfrak{L}_X\Big(\frac{\partial L}{\partial v^i}\Big) - \mathfrak{b}_0\Big[\frac{\partial L}{\partial v^i}\Big](C^i - v^i) \\ & - \frac{\partial L}{\partial v^i}\Big(B^j_k C^k \mathfrak{b}_0[A^i_j] + A^i_j\big(\mathfrak{L}_X b^j - \mathfrak{b}_0[b^j]\big) + D^i\Big)\Bigg]\mathfrak{b}^0 \\ & + \Bigg[\mathfrak{L}_X\Big(\frac{\partial L}{\partial v^k}\Big) - \mathfrak{b}_k[L] - \mathfrak{b}_k\Big[\frac{\partial L}{\partial v^i}\Big](C^i - v^i) \\ & + \frac{\partial L}{\partial v^i}\Big(B^j_k \mathfrak{L}_X A^i_j - B^j_l C^l \mathfrak{b}_k[A^i_j] + A^i_j \mathfrak{b}_k[b^j]\Big)\Bigg]\mathfrak{b}^k \\ & + \Bigg[\frac{\partial^2 L}{\partial v^j \partial v^i}\big(v^i - C^i\big)\Bigg]\mathrm{d}v^j. \end{aligned} \tag{4.146}$$

With (4.146), we have derived the first term of the sum (4.140). In order to express the second term $X \lrcorner \Phi_R^{\mathrm{np}}$, we start by expressing the two-form Φ_R^{np} modelling the nonpotential forces with respect to the basis covector fields $\mathfrak{b}^0, \dots, \mathfrak{b}^n, \mathrm{d}v^1, \dots, \mathrm{d}v^n$ using the coordinate transformation rules (4.129), (4.136) and (4.142) as

$$\begin{aligned} \Phi_R^{\mathrm{np}} &= F_i\, \mathrm{d}x^i \wedge \mathrm{d}t + \frac{1}{2}\frac{\partial F_i}{\partial u^j}\big(\mathrm{d}x^i - u^i \mathrm{d}t\big) \wedge \big(\mathrm{d}x^j - u^j \mathrm{d}t\big) \\ &= B^i_j F_i\, \mathfrak{b}^j \wedge \mathfrak{b}^0 + \frac{1}{2} B^i_j \frac{\partial F_i}{\partial v^k}\big(\mathfrak{b}^j - v^j \mathfrak{b}^0\big) \wedge \big(\mathfrak{b}^k - v^k \mathfrak{b}^0\big) \\ &= \bar{F}_j\, \mathfrak{b}^j \wedge \mathfrak{b}^0 + \frac{1}{2}\frac{\partial \bar{F}_j}{\partial v^k}\big(\mathfrak{b}^j - v^j \mathfrak{b}^0\big) \wedge \big(\mathfrak{b}^k - v^k \mathfrak{b}^0\big), \end{aligned} \tag{4.147}$$

where we have introduced $\bar{F}_j \coloneqq B^i_j F_i$. Consequently, we obtain

$$\begin{aligned} X \lrcorner \Phi_R^{\mathrm{np}} = & \Bigg[C^j \bar{F}_j - \tfrac{1}{2} v^j \big(C^k - v^k\big)\Big(\frac{\partial \bar{F}_k}{\partial v^j} - \frac{\partial \bar{F}_j}{\partial v^k}\Big)\Bigg]\mathfrak{b}^0 \\ & + \Bigg[-\bar{F}_j + \tfrac{1}{2}\big(C^k - v^k\big)\Big(\frac{\partial \bar{F}_k}{\partial v^j} - \frac{\partial \bar{F}_j}{\partial v^k}\Big)\Bigg]\mathfrak{b}^j. \end{aligned} \tag{4.148}$$

We immediately notice the apparent similarity with the one-form (4.109).

By equation (4.140), the sum of (4.146) and (4.148) needs to vanish, i.e., each component with respect to the basis $\mathfrak{b}^0, \dots, \mathfrak{b}^n, \mathrm{d}v^1, \dots, \mathrm{d}v^n$ needs to be zero separately. From the components in $\mathrm{d}v^1, \dots, \mathrm{d}v^n$ it follows that

$$C^i = v^i \tag{4.149}$$

for $i = 1, \dots, n$, because the matrix

$$\frac{\partial^2 L}{\partial v^j \partial v^i} = B^l_j g_{lk} B^k_i$$

is positive definite and, therefore, has full rank. With the kinematic condition (4.149), the annihilation of the $\mathfrak{b}^1, \dots, \mathfrak{b}^n$ components of the sum (4.140) leads to **Hamel's equations**

$$\mathfrak{L}_X\left(\frac{\partial L}{\partial v^k}\right) - \mathfrak{b}_k[L] + \frac{\partial L}{\partial v^i}\left(B^j_k \mathfrak{L}_X A^i_j - B^j_l v^l \mathfrak{b}_k[A^i_j] + A^i_j \mathfrak{b}_k[b^j]\right) = \bar{F}_k. \quad (4.150)$$

For the case of a linear transformation $b^i = 0$ depending only on the generalized coordinates $x^1, \dots, x^n$ and *not* on $x^0 = t$, i.e.,

$$B^j_i \circ \phi^{-1}(t, x^1, \dots, x^n) = B^j_i \circ \phi^{-1}(x^1, \dots, x^n),$$

the resulting equations were studied in Hamel 1904a, § 5. In the context of a geometric treatment of time-independent mechanics, Bloch 2015, Section 3.8 refers to these equations as *Hamel's equations*. The author decided to follow this suggestion.

However, the difficulties that arise in naming equations should not go unmentioned. On the one hand, there is the observation that in our geometric formulation of the mechanics of finite-dimensional mechanical systems (Postulate 4.8) Lagrange's equations of the second kind (4.111) and Hamel's equations are "just" different chart representations of the same geometric objects. With this view, one would rather avoid different names. On the other hand, by scientific probity, equations should be attributed to the scientist that derived them first. However, this question is by far not easy because there are always fine nuances in the presentations. In the field of technical mechanics, Bremer 1988, p. 47 refers to Hamel's equations in their variational form as *Hamel–Boltzmann equations*. Hamel[39] refers to the equations as *Lagrange–Euler equations* since they include Lagrange's equations of the second kind as well as Euler's equations describing the rotation of a rigid body. The author thinks that this designation might easily lend to confusion with the Euler–Lagrange equations that appear as the stationarity condition in the calculus of variations. The study of Volterra 1898, led the author and his coworkers suggest the designation as *Volterra–Hamel–Boltzmann equations* (see Winandy et al. 2018). However, this name is not only somewhat clumsy but it ignores the work of Voronets 1901, whose contribution is recognized by Hamel 1904b, footnote on p. 424. We refer to Chapter III of Neĭmark et al. 1972 for a presentation of the contribution of the different authors.

39. See pp. 480–481 of Hamel 1949.

4.10.3. Hamilton's equations

We introduce the coordinates

$$p_i := \frac{\partial L_R}{\partial u^i} = g_{ij}(u^j - R^j) + A_i^R, \tag{4.151}$$

to which we refer as **generalized momentum** coordinates. The last equality follows by equations (4.92) and (4.75) together with $\partial V_R / \partial u^i = 0$. The full rank of the matrix $[g_{ij}]$ guarantees that the relation (4.151) can be resolved for $u^1, \dots, u^n$ as

$$u^i = g^{ij}(p_j - A_j^R) + R^i, \tag{4.152}$$

where $g^{ik} g_{kj} = \delta^i_j$. Therefore, the coordinate functions

$$\begin{aligned}\tilde{t} &= t,\\ \tilde{x}^i &= x^i,\\ p_i &= g_{ij}(u^j - R^j) + A_i^R\end{aligned}$$

with $i = 1, \dots, n$ define the chart

$$\begin{aligned}\tilde{\Phi}\colon A^1 M \supseteq \pi^{-1}(U) &\to \mathbb{R}^{2n+1},\\ (p, v_p) &\mapsto (\tilde{t}, \tilde{x}^1, \dots, \tilde{x}^n, p_1, \dots, p_n)\end{aligned} \tag{4.153}$$

on $\pi^{-1}(U)$. The chart (4.153) results from the natural chart (4.17) by using the coordinates $p_1, \dots, p_n$ instead of $u^1, \dots, u^n$ for the representation of time-normalized vectors. We refer to $(\tilde{t}, \tilde{x}^1, \dots, \tilde{x}^n, p_1, \dots, p_n)$ as **canonical coordinates**.[40] As before, the tildes on t and the x^i allow the distinction between the canonical coordinates and those provided by the natural chart (4.17).

It is by representing the objects from Postulate 4.8 in canonical coordinates that we obtain Hamilton's equations. We rewrite the Cartan one-form (4.96) as

$$\omega_R = \left(L_R - u^i \frac{\partial L_R}{\partial u^i}\right) \mathrm{d}t + \frac{\partial L_R}{\partial u^i} \mathrm{d}x^i = -H_R \,\mathrm{d}\tilde{t} + p_i \,\mathrm{d}\tilde{x}^i, \tag{4.154}$$

where we used that $\mathrm{d}t = \mathrm{d}\tilde{t}$ and $\mathrm{d}x^i = \mathrm{d}\tilde{x}^i$ for $i = 1, \dots, n$. Moreover, we defined

40. These coordinates are by no means canonically defined since they depend on the choice of a reference field R. Physically, the quantities $p_1, \dots, p_n$ are generalized momenta. In the context of time-independent mechanics playing on the cotangent bundle T^*Q of a time-independent configuration manifold Q, the position and generalized momentum coordinates provided by the chart (3.68) are *canonical* coordinates according to the Darboux theorem (see p. 84). Moreover, Hamilton's equations are also referred to as *canonical equations* (see Landau and E. M. Lifshitz 1969, p. 132). So we use the adjective canonical because of tradition.

the **Hamiltonian** as

$$H_R : A^1M \supseteq \pi^{-1}(U) \to \mathbb{R},\ a \mapsto H_R(a) := -\left(L_R - u^i \frac{\partial L_R}{\partial u^i}\right)(a),$$

which is a *local* function defined on the neighbourhood $\pi^{-1}(U)$. It follows by Euler's theorem[41] that

$$H_R = u^i \frac{\partial L_R}{\partial u^i} - L_R = T_{R,2} - T_{R,0} + V_R, \tag{4.155}$$

where the last equality uses (4.75) and (4.92). We compute Ω_R^{p} by taking the exterior derivative of the Cartan one-form (4.154) as

$$\begin{aligned}\Omega_R^{\mathrm{p}} = \mathrm{d}\omega_R &= -\mathrm{d}H \wedge \mathrm{d}\tilde{t} + \mathrm{d}p_i \wedge \mathrm{d}\tilde{x}^i \\ &= -\frac{\partial H}{\partial \tilde{x}^i}\mathrm{d}\tilde{x}^i \wedge \mathrm{d}\tilde{t} - \frac{\partial H}{\partial p_i}\mathrm{d}p_i \wedge \mathrm{d}\tilde{t} + \mathrm{d}p_i \wedge \mathrm{d}\tilde{x}^i,\end{aligned} \tag{4.156}$$

where we stuck to our policy of dropping the letter R whenever it is hindering.

We use (3.47) to express the basis vectors $\partial/\partial u^i$ induced by the natural chart (4.17) with respect to the basis vectors induced by the chart (4.153) as

$$\frac{\partial}{\partial u^i} = \frac{\partial}{\partial u^i}[\tilde{t}]\frac{\partial}{\partial \tilde{t}} + \frac{\partial}{\partial u^i}[\tilde{x}^j]\frac{\partial}{\partial \tilde{x}^j} + \frac{\partial}{\partial u^i}[p_j]\frac{\partial}{\partial p_j} = g_{ji}\frac{\partial}{\partial p_j}.$$

With this relation and equation (4.152), the nonpotential force Φ_R^{np} adopts the local form

$$\begin{aligned}\Phi_R^{\mathrm{np}} = F_i\,\mathrm{d}\tilde{x}^i \wedge \mathrm{d}\tilde{t} + \frac{1}{2} g_{rj}\frac{\partial F_i}{\partial p_r}\Big[\mathrm{d}\tilde{x}^i - \Big(g^{ik}(p_k - A_k^R) + R^i\Big)\mathrm{d}\tilde{t}\Big]& \\ \wedge \Big[\mathrm{d}\tilde{x}^j - \Big(g^{jl}(p_l - A_l^R) + R^j\Big)\mathrm{d}\tilde{t}\Big]&\end{aligned}$$

because $\mathrm{d}t = \mathrm{d}\tilde{t}$ and $\mathrm{d}x^i = \mathrm{d}\tilde{x}^i$ for $i = 1, \dots, n$.

The time-normalized field X from (4.106) that describes the motion can be locally expressed as

$$X = \frac{\partial}{\partial \tilde{t}} + A^i \frac{\partial}{\partial \tilde{x}^i} + E_i \frac{\partial}{\partial p_i}, \tag{4.157}$$

where we adopted the convention that a lower index appearing in the denominator is considered to be an upper index. By Theorem 4.2(iii) the time-normalized vector field X is uniquely determined by

$$0 = X \lrcorner \Omega = X \lrcorner \Omega_R^{\mathrm{p}} + X \lrcorner \Phi_R^{\mathrm{np}},$$

41. See the proof of Proposition 4.7.

where $\Omega = \Omega_R^{\mathrm{p}} + \Phi_R^{\mathrm{np}}$ is the action form from Postulate 4.8. As before, we compute $X \lrcorner \Omega_R^{\mathrm{p}}$ and $X \lrcorner \Phi_R^{\mathrm{np}}$ separately. With the local expressions (4.156) and (4.157), we get

$$X \lrcorner \Omega_R^{\mathrm{p}} = -\left(A^i \frac{\partial H}{\partial \tilde{x}^i} + E_i \frac{\partial H}{\partial p_i}\right) \mathrm{d}\tilde{t} + \left(\frac{\partial H}{\partial \tilde{x}^i} + E_i\right) \mathrm{d}\tilde{x}^i + \left(\frac{\partial H}{\partial p_i} - A^i\right) \mathrm{d}p_i \quad (4.158)$$

and

$$\begin{aligned} X \lrcorner \Phi_R^{\mathrm{np}} = \Big[& A^i F_i - \tfrac{1}{2}\left(g^{ik}(p_k - A_k^R) + R^i\right)\left(A^j - g^{jl}(p_l - A_l^R) - R^j\right) \\ & \cdot \left(g_{ri}\frac{\partial F_j}{\partial p_r} - g_{rj}\frac{\partial F_i}{\partial p_r}\right)\Big] \mathrm{d}\tilde{t} \qquad (4.159) \\ & + \left[-F_i + \tfrac{1}{2}\left(A^j - g^{jl}(p_l - A_l^R) - R^j\right)\left(g_{ri}\frac{\partial F_j}{\partial p_r} - g_{rj}\frac{\partial F_i}{\partial p_r}\right)\right] \mathrm{d}\tilde{x}^i. \end{aligned}$$

The one-form $X \lrcorner \Omega = X \lrcorner \Omega_R^{\mathrm{p}} + X \lrcorner \Phi_R^{\mathrm{np}}$ is zero if each component vanishes. Since the coefficient of the $\mathrm{d}p_i$-component in (4.158) has to vanish, it follows that

$$A^i = \frac{\partial H}{\partial p_i} \overset{(3.13)}{=} \frac{\partial H \circ \tilde{\Phi}^{-1}}{\partial p_i} \circ \tilde{\Phi} = g^{ij}(p_j - A_j^R) + R^i \qquad (4.160)$$

because by (4.155), (4.75), and (4.152)

$$H \circ \tilde{\Phi}^{-1}\left(\tilde{t}, \tilde{\mathbf{x}}, \tilde{\mathbf{p}}\right) = \tfrac{1}{2} g^{ij}(p_i - A_i^R)(p_j - A_j^R) + R^j(p_j - A_j^R) + V.$$

With equation (4.160), the annihilation of the $\mathrm{d}\tilde{x}^i$-component of $X \lrcorner \Omega$ implies that

$$E_i = -\frac{\partial H}{\partial \tilde{x}^i} + F_i. \qquad (4.161)$$

Finally, it is by considering a time-parametrized integral curve (see equation (4.26)) of the vector field X determined by equations (4.160) and (4.161) that we obtain **Hamilton's equations**[42]

$$\begin{aligned} \dot{\tilde{x}}^i(\tau) &= \frac{\partial \mathrm{H}}{\partial p_i}\left(\tilde{t}(\tau), \tilde{\mathbf{x}}(\tau), \mathbf{p}(\tau)\right), \\ \dot{p}_i(\tau) &= -\frac{\partial \mathrm{H}}{\partial \tilde{x}^i}\left(\tilde{t}(\tau), \tilde{\mathbf{x}}(\tau), \mathbf{p}(\tau)\right) + \mathrm{F}_i\left(\tilde{t}(\tau), \tilde{\mathbf{x}}(\tau), \mathbf{p}(\tau)\right), \end{aligned} \qquad (4.162)$$

where we used upright letters $\mathrm{H} := H \circ \tilde{\Phi}^{-1}$ and $\mathrm{F}_i := F_i \circ \tilde{\Phi}^{-1}$ to denote the representation of the Hamiltonian H and of the functions F_i with respect to the chart (4.153). Note that the time function $\tilde{t}(\tau) = \tilde{t} \circ \beta(\tau)$ evaluated along the motion is affine in τ because β is an integral curve of a time-normalized vector field.

42. See p. 132 of Landau and E. M. Lifshitz 1969.

4.11. The variational approach

In Section 4.2, we saw that the time structure ϑ of a Galilean manifold (M, ϑ, g) allows to single out spacelike tangent vectors to M (see equation (4.10)). The same can be done with tangent vectors to the state space A^1M that is endowed with the time structure $\hat{\vartheta} = \pi^*\vartheta$ induced by ϑ. At each point $a \in A^1M$, the space of spacelike vectors on A^1M is the set

$$A^0_a(A^1M) := \ker \hat{\vartheta}_a = \left\{ z \in T_a(A^1M) \mid \hat{\vartheta}_a(z) = 0 \right\} \subset T_a(A^1M).$$

Similar to (4.11), we define the corresponding subbundle of the tangent bundle $T(A^1M)$ of the state space A^1M as

$$A^0(A^1M) := \ker \hat{\vartheta} = \bigcup_{a \in A^1M} A^0_a(A^1M) \subset T(A^1M).$$

A **virtual displacement field** Y is a smooth section of this bundle, i.e.,

$$Y \colon A^1M \to A^0(A^1M), \ a \mapsto Y(a),$$

with $\bar{\pi} \circ Y = \mathrm{id}_{A^1M}$ and where $\bar{\pi} \colon T(A^1M) \to A^1M$ denotes the natural projection of the tangent bundle $T(A^1M)$.

In Section 4.10, we introduced different sets of local coordinates that are related to an adapted chart (U, ϕ) of the Galilean manifold (M, ϑ, g). These coordinates can be used for the local expression of the virtual displacement field Y on the neighbourhood $\pi^{-1}(U) \subseteq A^1M$ as

$$\begin{aligned} Y &= \delta x^i \frac{\partial}{\partial x^i} + \delta u^i \frac{\partial}{\partial u^i} \\ &= \delta s^i \mathfrak{b}_i + \delta v^i \frac{\partial}{\partial v^i} \\ &= \delta \tilde{x}^i \frac{\partial}{\partial \tilde{x}^i} + \delta p_i \frac{\partial}{\partial p_i}. \end{aligned} \tag{4.163}$$

where the respective coefficients are smooth real-valued functions defined on the neighbourhood $\pi^{-1}(U)$. The choice of the peculiar notation will become clear in the following.

With equations (4.106), (4.139) and (4.157), we have seen that a time-normalized vector field X can be locally written as

$$\begin{aligned} X &= \frac{\partial}{\partial t} + A^i \frac{\partial}{\partial x^i} + B^i \frac{\partial}{\partial u^i} \\ &= \mathfrak{b}_0 + C^i \mathfrak{b}_i + D^i \frac{\partial}{\partial v^i} \\ &= \frac{\partial}{\partial \tilde{t}} + A^i \frac{\partial}{\partial \tilde{x}^i} + E_i \frac{\partial}{\partial p_i}. \end{aligned}$$

According to Theorem 4.2(iii), the requirement

$$X \lrcorner \Omega \stackrel{!}{=} 0, \tag{4.164}$$

for a given action form Ω uniquely determines a time-normalized vector field X. Condition (4.164) is equivalent to demanding that

$$\bar{Y} \lrcorner (X \lrcorner \Omega) = \Omega(X, \bar{Y}) \stackrel{!}{=} 0, \tag{4.165}$$

for all $\bar{Y} \in \mathrm{Vect}(A^1M)$. Equation (4.165) is referred to as **variational form** of condition (4.164). In Section 4.10.1, the vanishing of the $\mathrm{d}x^i$- and $\mathrm{d}u^i$-components of the one-form $X \lrcorner \Omega$ was sufficient to derive Lagrange's equations of the second kind. In Section 4.10.2 the annihilation of the $\mathfrak{b}^i$- and $\mathrm{d}v^i$-components of the form $X \lrcorner \Omega$ implied Hamel's equations. Finally, Hamilton's equations followed by exploiting (only) that the $\mathrm{d}\tilde{x}^i$- and the $\mathrm{d}p_i$-part of the form $X \lrcorner \Omega$ need to vanish. This means that, it is sufficient to exploit condition (4.165) for *spacelike* vector fields on A^1M, i.e., for virtual displacement fields $\bar{Y} = Y$. This observation is not astonishing since we know from Section 4.6 that the action form has rank $2n$ and that it is blind on the line bundle over A^1M spanned by the time-normalized vector field X that satisfies (4.164). Now, because of $\hat{\vartheta}(X) = 1$ (see equation (4.51)) and $\hat{\vartheta}(Y) = 0$, virtual displacement fields Y can be used to test the $2n$ complementary directions to $\mathrm{span}\{X\}$.

4.11.1. Variational families of curves

In its development, classical mechanics has been intimately related to the calculus of variations and variational principles such as Hamilton's principle.[43] The classical notion of a virtual displacement has to be seen as a vector field along the curve that is to be varied. By adopting this perspective we are able to make the link with classical results such as the principle of virtual work, the two central equations of Lagrange and Hamel as well as Hamilton's principle.

We start with a smooth two-parameter map

$$\begin{aligned} \kappa \colon\ ]-r, r[\times I &\to A^1M, \\ (\varepsilon, \tau) &\mapsto \kappa(\varepsilon, \tau), \end{aligned} \tag{4.166}$$

that we assume to be an injective immersion for some $r > 0$. As such, the map (4.166) defines a two-dimensional submanifold that is parametrized

43. Section 4.11.3 is concerned with Hamilton's principle.

by ε and τ. Moreover, the map (4.166) defines the two curves (immersed one-dimensional submanifolds)

$$\begin{aligned} \kappa_\varepsilon\colon I &\to A^1 M, \\ \tau &\mapsto \kappa_\varepsilon(\tau) := \kappa(\varepsilon, \tau) \end{aligned} \tag{4.167}$$

and

$$\begin{aligned} \kappa_\tau\colon\,]-r, r[\, &\to A^1 M, \\ \varepsilon &\mapsto \kappa_\tau(\varepsilon) := \kappa(\varepsilon, \tau) \end{aligned} \tag{4.168}$$

in the state space $A^1 M$. The curve (4.167) results from (4.166) by considering fixed values of the parameter ε, while the curve (4.168) is engendered by (4.166) in keeping τ fixed. We call the set of all curves (4.167) a **variational family of curves**. The map (4.166) is required to contain the time-parametrized curve

$$\beta\colon I \to A^1 M \tag{4.169}$$

such that

$$\beta(\tau) \stackrel{!}{=} \kappa_0(\tau) = \kappa(0, \tau) \tag{4.170}$$

for all $\tau \in I$. The chart representation[44] of the map (4.166) with respect to the coordinates of the natural chart (4.17) has the form

$$\Phi \circ \kappa = \Big(t(\varepsilon, \tau), \mathbf{x}(\varepsilon, \tau), \mathbf{u}(\varepsilon, \tau)\Big). \tag{4.171}$$

For simplicity we assume $\operatorname{im} \kappa \subseteq \pi^{-1}(U)$ such that we do not need to split the map κ into components lying in the different neighbourhoods $\pi^{-1}(U_\alpha)$ provided by an atlas (U_α, ϕ_α) of the manifold M. Technically, this can be realized by considering the restrictions

$$\kappa|_{U_\alpha}\colon \kappa^{-1}(U_\alpha) \cap \Big(]-r, r[\times I\Big) \to A^1 M.$$

for all charts (U_α, ϕ_α).

The chart representation of the curve (4.169) is $\Phi \circ \beta = (t(\tau), \mathbf{x}(\tau), \mathbf{u}(\tau))$. By equations (4.170) and (4.171), it holds that

$$t(0, \tau) = t(\tau), \qquad \mathbf{x}(0, \tau) = \mathbf{x}(\tau), \qquad \mathbf{u}(0, \tau) = \mathbf{u}(\tau).$$

Classically, one considers variations of the curve β that keep time fixed such that the chart representation (4.171) reduces to

$$\Phi \circ \kappa = \Big(t(\tau), \mathbf{x}(\varepsilon, \tau), \mathbf{u}(\varepsilon, \tau)\Big). \tag{4.172}$$

44. The chart representation of a map is defined in equation (3.3).

A coordinate-free definition of the restricted variational family from (4.171) is given by

$$\hat{\vartheta}(\delta\kappa_\tau) \stackrel{!}{=} 0, \tag{4.173}$$

where $\delta\kappa_\tau$ denotes the tangent field along the curve κ_τ from (4.168). We have used a δ to denote differentiation with respect to the curve parameter ε, while in equation (3.50) we denoted the derivative with a dot because there the curve parameter is τ. Condition (4.173) requires the tangent field along the curve κ_τ to be spacelike. Indeed, by equation (3.47), the local representation of the tangent field $\delta\kappa_\tau$ is given by

$$\delta\kappa_\tau(\varepsilon) = \delta\kappa_\tau[t](p)\left.\frac{\partial}{\partial t}\right|_p + \delta\kappa_\tau[x^i](p)\left.\frac{\partial}{\partial x^i}\right|_p + \delta\kappa_\tau[u^i](p)\left.\frac{\partial}{\partial u^i}\right|_p, \tag{4.174}$$

with $p = \kappa_\tau(\varepsilon)$ and condition (4.173) can be locally written as

$$0 \stackrel{!}{=} \hat{\vartheta}(\delta\kappa_\tau) = \mathrm{d}t_{\kappa_\tau(\varepsilon)}\left(\delta\kappa_\tau|_{\kappa_\tau(\varepsilon)}\right) = \delta\kappa_\tau|_{\kappa_\tau(\varepsilon)}[t] = \left.\frac{\mathrm{d}}{\mathrm{d}\bar{\varepsilon}}\right|_{\bar{\varepsilon}=\varepsilon} t\circ\kappa_\tau(\bar{\varepsilon}).$$

Accordingly, the chart representation (4.171) of κ is indeed restricted to the form (4.172) by condition (4.173). The local description of the tangent field $\delta\kappa_\tau$ from (4.174) reduces to the one of a virtual displacement field, i.e.,

$$\delta\kappa_\tau(\varepsilon) = \delta x^i(p)\left.\frac{\partial}{\partial x^i}\right|_p + \delta u^i(p)\left.\frac{\partial}{\partial u^i}\right|_p, \tag{4.175}$$

with $p = \kappa_\tau(\varepsilon)$ and where we have introduced the short-hand notations

$$\begin{aligned}
\delta x^i(p) &= \left.\frac{\mathrm{d}}{\mathrm{d}\bar{\varepsilon}}\right|_{\bar{\varepsilon}=\varepsilon} x^i\circ\kappa_\tau(\bar{\varepsilon}) = \left.\frac{\mathrm{d}}{\mathrm{d}\bar{\varepsilon}}\right|_{\bar{\varepsilon}=\varepsilon} x^i(\bar{\varepsilon}),\\
\delta u^i(p) &= \left.\frac{\mathrm{d}}{\mathrm{d}\bar{\varepsilon}}\right|_{\bar{\varepsilon}=\varepsilon} u^i\circ\kappa_\tau(\bar{\varepsilon}) = \left.\frac{\mathrm{d}}{\mathrm{d}\bar{\varepsilon}}\right|_{\bar{\varepsilon}=\varepsilon} u^i(\bar{\varepsilon})
\end{aligned}$$

relying on (4.172). In an analogue way, we use a dot to denote the tangent field along the curve $\kappa_\varepsilon(\tau)$ as

$$\dot{\kappa}_\varepsilon(\tau) = \left.\frac{\partial}{\partial t}\right|_p + \dot{x}^i(p)\left.\frac{\partial}{\partial x^i}\right|_p + \dot{u}^i(p)\left.\frac{\partial}{\partial u^i}\right|_p,$$

with $p = \kappa_\varepsilon(\tau)$. The two-parameter map κ defines the two local vector fields

$$X\colon \kappa\big(]-r, r[\times \operatorname{int} I\big) \to T(A^1M)$$

and

$$Y\colon \kappa\big(]-r, r[\times \operatorname{int} I\big) \to T(A^1M)$$

by the respective differentiation with respect to the parameters τ and ε, i.e.,

$$X(\kappa(\varepsilon,\tau)) := \dot{\kappa}_\varepsilon(\tau) \tag{4.176}$$

and

$$Y(\kappa(\varepsilon,\tau)) := \delta\kappa_\tau(\varepsilon). \tag{4.177}$$

In equation (4.176), the right-hand side denotes the time-normalized tangent vector that is defined at the point $\kappa(\varepsilon,\tau)$ by the curve $\kappa_\varepsilon\colon I \to A^1M$. Because of condition (4.173), the vector field Y is a virtual displacement field.

The local vector fields

$$X = \frac{\partial}{\partial t} + \dot{x}^i \frac{\partial}{\partial x^i} + \dot{u}^i \frac{\partial}{\partial u^i}$$

and

$$Y = \delta x^i \frac{\partial}{\partial x^i} + \delta u^i \frac{\partial}{\partial u^i} \tag{4.178}$$

defined by equations (4.176) and (4.177) commute, i.e., their Lie bracket[45] vanishes. To see this we consider a smooth function $f \in C^\infty(A^1M)$ and we compute

$$\begin{aligned}[\![X,Y]\!][f](\kappa(\varepsilon,\tau)) &= X\big[Y[f]\big](\kappa(\varepsilon,\tau)) - Y\big[X[f]\big](\kappa(\varepsilon,\tau)) \\ &= \frac{\partial^2(f\circ\kappa)}{\partial\tau\partial\varepsilon}(\varepsilon,\tau) - \frac{\partial^2(f\circ\kappa)}{\partial\varepsilon\partial\tau}(\varepsilon,\tau) = 0\end{aligned}$$

using equations (3.22), (4.176), and (4.177). Therefore, the vanishing of their Lie bracket is a necessary condition for two vector fields to be induced by a variational family of curves. Moreover, by construction, the vector field X is time-normalized and Y is a virtual displacement field by condition (4.173). According to the Frobenius theorem from Section 3.11, the vector fields X and Y span an involutive distribution of which $\operatorname{im}\kappa \subset A^1M$ is the integral manifold. The curves (4.167) and (4.168) are integral curves of the vector fields X and Y, respectively. Let

$$\varphi\colon\,]-r,r[\, \times \operatorname{im}\kappa \to A^1M$$

denote the flow[46] of the vector field Y. Then it holds by definition of the vector field Y that

$$\kappa(\varepsilon+\varepsilon',\tau) = \varphi\big(\varepsilon, \kappa(\varepsilon',\tau)\big)$$

45. See equation (3.45).
46. See Section 3.6 and in particular equations (3.55) and (3.56).

for all values $\tau \in I$ and $\varepsilon, \varepsilon' \in]-r, r[$ for which $\varepsilon + \varepsilon' \in]-r, r[$. In particular, for $\varepsilon' = 0$ this means that

$$\kappa(\varepsilon, \tau) = \varphi\big(\varepsilon, \kappa(0, \tau)\big) = \varphi\big(\varepsilon, \beta(\tau)\big) \tag{4.179}$$

and

$$\kappa_\varepsilon(\tau) = \varphi_\varepsilon \circ \beta(\tau).$$

Therefore, we can see a variational family of curves as being the result of dragging the time-parametrized curve β using the flow of a vector field Y. If the vector field Y is a virtual displacement field (i.e., if Y is spacelike), then the resulting variational family contains only time-parametrized curves because the curve β is time-parametrized.

We refer to a variational family in which the curves (4.167) are not only time-parametrized but second-order curves as a **variational family of second-order curves**. By condition (4.24), this means that

$$\kappa_\varepsilon = (\pi \circ \kappa_\varepsilon)^{\boldsymbol{\cdot}}$$

and, therefore, the curves can be constructed from a two-parameter map

$$\begin{aligned} \alpha \colon\,]-r, r[\times I &\to M, \\ (\varepsilon, \tau) &\mapsto \alpha(\varepsilon, \tau) \end{aligned}$$

for which we assume that the variational family

$$\begin{aligned} \alpha_\varepsilon \colon I &\to M, \\ \tau &\mapsto \alpha_\varepsilon(\tau) \coloneqq \alpha(\varepsilon, \tau). \end{aligned} \tag{4.180}$$

it induces in the Galilean manifold (M, ϑ, g) consists of time-parametrized curves. As before, the curve β from (4.169) that we now assume to be a second-order curve should be contained in the variational family such that

$$\alpha(0, \tau) \overset{!}{=} \pi \circ \beta(\tau)$$

for all $\tau \in I$. The two parameter map α defines the curve

$$\begin{aligned} \alpha_\tau \colon\,]-r, r[&\to M, \\ \varepsilon &\mapsto \alpha_\tau(\varepsilon) \coloneqq \alpha(\varepsilon, \tau) \end{aligned}$$

such that the restriction to variations that keep time fixed can be formulated as

$$\vartheta(\delta\alpha_\tau) \overset{!}{=} 0,$$

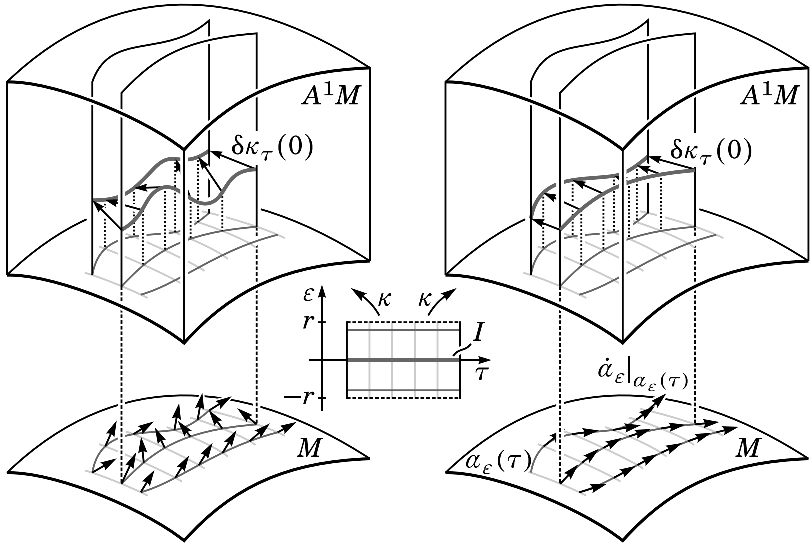


Figure 4.5.: The left side of the picture shows a family of time-parametrized curves in A^1M, while the right side depicts a family of second-order curves in A^1M that is induced by a family α_ε of time-parametrized curves in M.

which is an analogue condition to (4.173). As we did in Section 4.3, we denote by

$$\begin{aligned} &\dot\alpha_\varepsilon \colon I \to A^1M, \\ &\quad \tau \mapsto \dot\alpha_\varepsilon(\tau) = \Big(\alpha_\varepsilon(\tau), \dot\alpha_\varepsilon|_{\alpha_\varepsilon(\tau)}\Big) \end{aligned}$$

the second-order curve given by the tangent field along the curve (4.180) for a fixed value of ε. Finally, we obtain a variational family of second-order curves by considering the two-parameter map that is defined as

$$\begin{aligned} &\kappa\colon\ ]-r,r[\times I \to A^1M, \\ &\qquad (\varepsilon,\tau) \mapsto \kappa(\varepsilon,\tau) \coloneqq \dot\alpha_\varepsilon(\tau). \end{aligned} \tag{4.181}$$

For of a variational family of second-order curves, the vector field X from equation (4.176) is a second-order field, i.e., $\dot x^i = u^i$ such that

$$X = \frac{\partial}{\partial t} + u^i \frac{\partial}{\partial x^i} + \dot u^i \frac{\partial}{\partial u^i}$$

because its integral curves $\kappa_\varepsilon = \dot\alpha_\varepsilon(\tau)$ are second-order curves by assumption. The vector field Y is spacelike and it needs to satisfy $[\![X, Y]\!] = 0$. Therefore, it has the local form

$$Y = \delta x^i \frac{\partial}{\partial x^i} + \mathfrak{L}_X \delta x^i \frac{\partial}{\partial u^i}. \tag{4.182}$$

In the view of equation (4.179), this means that in order to guarantee that a second-order curve β is dragged to a second-order curve $\varphi_\varepsilon \circ \beta$ the generating vector field Y of the flow φ_ε needs to have the form (4.182). Figure 4.5 shows an attempt to visualize both types of variational families.

4.11.2. Virtual work and the central equation

We saw that it is sufficient to exploit the variational form (4.165) using virtual displacement fields only. Actually, even less is required if we consider from the beginning that the vector field X that is determined by

$$X \lrcorner \Omega = 0, \quad \hat{\vartheta}(X) = 1.$$

is a second-order field. The real-valued function given by the variational form $\Omega(X,Y)$, where the vector field

$$\begin{aligned} X &= \frac{\partial}{\partial t} + u^i \frac{\partial}{\partial x^i} + B^i \frac{\partial}{\partial u^i} \\ &= \mathfrak{b}_0 + v^i \mathfrak{b}_i + D^i \frac{\partial}{\partial v^i} \\ &= \frac{\partial}{\partial \tilde{t}} + \left(g^{ij}(p_j - A_j^R) + R^i \right) \frac{\partial}{\partial \tilde{x}^i} + E_i \frac{\partial}{\partial p_i}. \end{aligned}$$

is a second-order field and where Y is a virtual displacement field, is known as **virtual work**.

Using the local descriptions of the virtual displacement field Y from equation (4.163), we obtain the following expressions for the virtual work

$$\begin{aligned} \Omega(X,Y) &= \left[\mathfrak{L}_X \left(\frac{\partial L}{\partial u^i} \right) - \frac{\partial L}{\partial x^i} - F_i \right] \delta x^i \\ &= \Bigg[\mathfrak{L}_X \left(\frac{\partial L}{\partial v^k} \right) - \mathfrak{b}_k [L] - \bar{F}_k \\ &\quad + \frac{\partial L}{\partial v^i} \left(B_k^j \mathfrak{L}_X A_j^i - B_l^j v^l \mathfrak{b}_k [A_j^i] + A_j^i \mathfrak{b}_k [b^j] \right) \Bigg] \delta s^k \\ &= \left[E_i + \frac{\partial H}{\partial \tilde{x}^i} - F_i \right] \delta \tilde{x}^i, \end{aligned} \tag{4.183}$$

where we made use of our previous results (4.108), (4.109), (4.146), (4.148), (4.158), and (4.159).

A well-known tool for the formulation of mechanical theories is the so-called principle of virtual work which says that the virtual work vanishes for all virtual displacements. The principle of virtual work may be chosen as postulate for the formulation of a physical theory for finite-dimensional mechanical systems, i.e., one could postulate the virtual work (4.183) and require it to vanish for all virtual displacements δx^i, δs^k, respectively for all $\delta\tilde{x}^i$. In our presentation that is based on Postulate 4.8 the observation that the virtual work vanishes is a theorem rather than a principle. Nevertheless, we speak of the principle of virtual work in order to relate our presentation to classical approaches. The observation that Postulate 4.8 generalizes the principle of virtual work can be found in the book by Souriau, where he writes that the virtual work is a truncated[47] version of the action form Ω.

We pursue our endeavour to establish a firm link to the classical results by deriving Lagrange's central equation and Hamel's generalized version of it. We consider the virtual work that results from feeding the action form Ω from Postulate 4.8 with a second-order field X and a virtual displacement field Y

$$\Omega(X,Y) = \Omega^{\mathrm{p}}_R(X,Y) + \Phi^{\mathrm{np}}_R(X,Y) = \mathrm{d}\omega_R(X,Y) + \Phi^{\mathrm{np}}_R(X,Y), \tag{4.184}$$

where

$$\mathrm{d}\omega_R = \mathrm{d}L\wedge\mathrm{d}t + \mathrm{d}\left(\frac{\partial L}{\partial u^i}\right)\wedge(\mathrm{d}x^i - u^i\mathrm{d}t) - \frac{\partial L}{\partial u^i}\mathrm{d}u^i\wedge\mathrm{d}t$$

and

$$\Phi^{\mathrm{np}}_R = F_i\mathrm{d}x^i\wedge\mathrm{d}t + \frac{1}{2}\frac{\partial F_i}{\partial u^j}(\mathrm{d}x^i - u^i\mathrm{d}t)\wedge(\mathrm{d}x^j - u^j\mathrm{d}t),$$

as we know from (4.102) and (4.103). Postulate 4.8 requires the virtual work (4.184) to vanish. Considering that $X\lrcorner\,(\mathrm{d}x^i - u^i\mathrm{d}t) = 0$ because X is a second-order field and that $Y\lrcorner\,\mathrm{d}t = 0$ since Y is a virtual displacement field, it follows that

$$0 \overset{!}{=} \Omega(X,Y) = -\mathfrak{L}_Y L + \mathfrak{L}_X\left(\frac{\partial L}{\partial u^i}\right)\delta x^i + \frac{\partial L}{\partial u^i}\delta u^i - F_i\delta x^i, \tag{4.185}$$

for all coefficient functions $\delta x^i, \delta u^i \in C^\infty(\pi^{-1}(U))$. Note that we used the local expression (4.178) of the virtual displacement field Y to derive equation (4.185). Using the product rule for the differentiation of real-valued

47. Footnote 2 on p. XVII of Souriau 1970 reads: *Le principe des travaux virtuels classique n'est qu'une forme tronquée de cette formule (XI).* In Souriau 1997, the English translation of the book, one finds in footnote 6 on p. XX the infelicitous translation that: *The classical principle of virtual work is only an abbreviated form of formula (XI).*

functions by vector fields, the principle of virtual work (4.185) can be rewritten as

$$0 \overset{!}{=} \mathfrak{L}_X\left(\frac{\partial L}{\partial u^i}\delta x^i\right) - \mathfrak{L}_Y L - F_i\delta x^i + \frac{\partial L}{\partial u^i}\left(\delta u^i - \mathfrak{L}_X\delta x^i\right), \tag{4.186}$$

for all $\delta x^i, \delta u^i \in C^\infty(\pi^{-1}(U))$. Equation (4.186) reduces to

$$0 \overset{!}{=} \mathfrak{L}_X\left(\frac{\partial L}{\partial u^i}\delta x^i\right) - \mathfrak{L}_Y L - F_i\delta x^i, \tag{4.187}$$

for all $\delta x^i, \delta u^i \in C^\infty(\pi^{-1}(U))$ if one considers only virtual displacement fields of the form (4.182).

The virtual work expression from (4.187) is invariant under coordinate changes from $t, x^1, \dots, x^n, u^1, \dots, u^n$ with coordinate fields

$$\frac{\partial}{\partial t}, \frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial u^1}, \dots, \frac{\partial}{\partial u^n},$$

to $\bar{t}, \bar{x}^1, \dots, \bar{x}^n, v^1, \dots, v^n$ together with the basis vector fields

$$\mathfrak{b}_0, \dots, \mathfrak{b}_n, \frac{\partial}{\partial v^1}, \dots, \frac{\partial}{\partial v^n}.$$

To see this, we need to derive the transformation rules between these basis vector fields. The following consideration allows us to make use of the transformation rules between the dual bases $\mathrm{d}t, \mathrm{d}x^1, \dots, \mathrm{d}x^n, \mathrm{d}u^1, \dots, \mathrm{d}u^n$ and $\mathfrak{b}^0, \dots \mathfrak{b}^n, \mathrm{d}v^1, \dots, \mathrm{d}v^n$ that we know from Section 4.10.2. Let w be a vector field on an n-dimensional differentiable manifold M. Then w can be expressed with respect to given basis vector fields $\mathfrak{e}_1, \dots, \mathfrak{e}_n$ as

$$w = \mathfrak{e}^i(w)\,\mathfrak{e}_i,$$

where $\mathfrak{e}^1, \dots, \mathfrak{e}^n$ denote the corresponding dual fields of $\mathfrak{e}_1, \dots, \mathfrak{e}_n$ defined by $\mathfrak{e}^i(\mathfrak{e}_j) = \delta^i_j$. Consequently, it holds for the virtual displacement field

$$Y = \delta x^i \frac{\partial}{\partial x^i} + \delta u^i \frac{\partial}{\partial u^i} = \delta s^i \mathfrak{b}_i + \delta v^i \frac{\partial}{\partial v^i}$$

from (4.163) that

$$\delta s^i = \mathfrak{b}^i(Y) = Y \lrcorner \left(A^i_j(\mathrm{d}x^j - b^j \mathrm{d}t)\right) = A^i_j \delta x^j \tag{4.188}$$

and

$$\delta v^i = \mathrm{d}v^i(Y) = \left(\frac{\partial A^i_j}{\partial x^k}(u^j - b^j) - A^i_j \frac{\partial b^j}{\partial x^k}\right)\delta x^k + A^i_j \delta u^j, \tag{4.189}$$

where we used the transformation rules from (4.135). Note that equation (4.189) corresponds exactly to what we would obtain if we would carry out the variation, i.e., apply δ as derivative with respect to ε on the coordinate change

$$v^i(\varepsilon,\tau) = A^i_j(t(\tau),\mathbf{x}(\varepsilon,\tau))\Big(u^j(\varepsilon,\tau) - b^j(t(\tau),\mathbf{x}(\varepsilon,\tau))\Big),$$

from (4.130). We know that

$$\frac{\partial}{\partial v^i} = B^j_i \frac{\partial}{\partial u^j}$$

from (4.142) and it follows by equation (4.147) together with (4.188) that

$$F_i \delta x^i = \bar{F}_i \delta s^i.$$

Therefore, the principle of virtual work (4.187) can be rewritten as

$$0 \stackrel{!}{=} \mathfrak{L}_X\Big(\frac{\partial L}{\partial v^i}\delta s^i\Big) - \mathfrak{L}_Y L - \bar{F}_i \delta s^i, \tag{4.190}$$

for all smooth functions δs^i and δv^i defined on the neighbourhood $\pi^{-1}(U)$. Equation (4.190) is known as **Lagrange's central equation**.[48]

To emphasize its significance, we apply the product rule to the first term and we express the second term using the local expression (4.163) of Y as

$$0 \stackrel{!}{=} \left(\mathfrak{L}_X\Big(\frac{\partial L}{\partial v^i}\Big) - \mathfrak{b}_i[L] - \bar{F}_i\right)\delta s^i + \frac{\partial L}{\partial v^i}\big(\mathfrak{L}_X \delta s^i - \delta v^i\big), \tag{4.191}$$

for all $\delta s^i, \delta v^i \in C^\infty(\pi^{-1}(U))$. Equation (4.191) can be found in Bremer 1988, p. 47. It directly leads to Hamel's equations (4.150) if we consider the δs^i given by (4.188) and use the δv^i from (4.189) together with the assumption that the vector fields X and Y are induced by a variational family of second-order curves (4.181) (i.e., $\delta u^i = \mathfrak{L}_X \delta x^i$ by equation (4.182)). Note that we have already used the latter assumption in the derivation of (4.191) at the transition from (4.186) to (4.187). We observe that starting from Lagrange's central equation the derivation of Hamel's equations is not difficult. In the classical view, Lagrange's equations are considered to be a special case of (4.191) with $\delta s^i = \delta x^i$ and $\delta v^i = \delta u^i$.

In order to derive Hamel's generalized central equation that does not rely on the assumption $\delta u^i = \mathfrak{L}_X \delta x^i$, we solve (4.189) for $A^i_j \delta u^j$ such that

$$A^i_j \delta u^j = \delta v^i - \left(\frac{\partial A^i_j}{\partial x^k} B^j_l v^l - A^i_j \frac{\partial b^j}{\partial x^k}\right) B^k_r \delta s^r,$$

48. See Hamel 1904a, p. 15, Bremer 1988, p. 47, or Bremer 2008, p. 21.

where we used that $u^j - b^j = B^j_l v^l$ and that $\delta x^k = B^k_r \delta s^r$. Moreover, we calculate

$$\mathfrak{L}_X \delta x^i = \mathfrak{L}_X (B^i_j \delta s^j) = B^i_j \mathfrak{L}_X \delta s^j + \delta s^j \mathfrak{L}_X B^i_j$$

such that we can rewrite the principle of virtual work (4.186) as

$$\begin{aligned} 0 \overset{!}{=} \mathfrak{L}_X\left(\frac{\partial L}{\partial v^i}\delta s^i\right) - \mathfrak{L}_Y L - \bar{F}_i \delta s^i + \frac{\partial L}{\partial v^i}\left(\delta v^i - \mathfrak{L}_X \delta s^i\right) \\ + \frac{\partial L}{\partial v^i}\left(\mathfrak{L}_X A^i_k - \left(\frac{\partial A^i_j}{\partial x^k} B^j_l v^l - A^i_j \frac{\partial b^j}{\partial x^k}\right)\right) B^k_r \delta s^r, \end{aligned} \tag{4.192}$$

for all $\delta s^i, \delta v^i \in C^\infty(\pi^{-1}(U))$. Equation (4.192) is known as **Hamel's generalized central equation**.[49]

Instead of carrying out a coordinate transformation of (4.186), Hamel's generalized central equation can also be derived using the results from Section 4.10.2. We evaluate the virtual work (4.184) using the local expressions (4.143) and (4.147) as

$$\Omega(X,Y) = -\mathfrak{L}_Y L + \mathfrak{L}_X\left(\frac{\partial L}{\partial v^i}\right)\delta s^i + \frac{\partial L}{\partial v^i}\left(\mathrm{d}\mathfrak{b}^i(X,Y) - \mathrm{d}v^i \wedge \mathfrak{b}^0\right) - \bar{F}_i \delta s^i.$$

The term $\mathrm{d}\mathfrak{b}^i(X,Y)$ can be evaluated using (4.145) with the second-order condition $C^l = v^l$. Using the product rule of differentiation, the principle of virtual work can be stated as

$$\begin{aligned} 0 \overset{!}{=} \mathfrak{L}_X\left(\frac{\partial L}{\partial v^i}\delta s^i\right) - \mathfrak{L}_Y L + \frac{\partial L}{\partial v^i}\left(\delta v^i - \mathfrak{L}_X \delta s^i\right) - \bar{F}_i \delta s^i \\ + \frac{\partial L}{\partial v^i}\left(B^j_r \mathfrak{L}_X A^i_j - B^j_l v^l \mathfrak{b}_r[A^i_j] + A^i_j \mathfrak{b}_r[b^j]\right)\delta s^r, \end{aligned} \tag{4.193}$$

for all $\delta s^i, \delta v^i \in C^\infty(\pi^{-1}(U))$. Equation (4.193) is just Hamel's generalized central equation (4.192) because by (4.133)

$$\mathfrak{b}_r[A^i_j] = B^k_r \frac{\partial}{\partial \bar{x}^k}[A^i_j] = B^k_r \frac{\partial}{\partial x^k}[A^i_j]$$

and

$$\mathfrak{b}_r[b^j] = B^k_r \frac{\partial}{\partial \bar{x}^k}[b^j] = B^k_r \frac{\partial}{\partial x^k}[b^j]$$

because $A^i_j = A^i_j \circ \pi$ and $b^j = b^j \circ \pi$ do not depend on the coordinates $u^1, \dots, u^n$ respectively $v^1, \dots, v^n$.

Hamel's generalized central equation allows to derive Hamel's equation without the restriction $\delta u^i = \mathfrak{L}_X \delta x^i$ to variational families of second-order

49. See Hamel 1904b, p. 424.

curves. The third term of expression (4.193) drops out if we apply the product rule of differentiation to its first term and if we express its second term using the local expression (4.163) of Y. The remaining part can then be recognized as variational form of Hamel's equations.

4.11.3. Hamilton's principle

We saw in the previous section that the explicit consideration of the kinematics (i.e., of the second-order condition) allows to arrive at the notion of virtual work. In this section, we will see that the action integral on which Hamilton's principle[50] is based in classical texts appears as the restriction to second-order curves of a more general action integral. Moreover, we will show that for exact mechanical systems, Hamilton's principle becomes a theorem of Postulate 4.8.

On page 124, we saw that the action form of a mechanical system with Galilean manifold (M, ϑ, g) is exact if the mechanical system is only subjected to potential forces. By Postulate 4.8, the action form of an exact mechanical system can be defined using a Lagrangian and its related Cartan one-form (4.93). Using the Cartan one-form $\omega = L\hat{\vartheta} + \boldsymbol{\partial} L$ from (4.93), we define the **action**[51] of a mechanical system as the functional on time-parametrized curves $\beta\colon I \to A^1M$ that is given by the integral[52]

$$A[\beta] := \int_{\beta(I)} \iota_{\leftarrow}\omega, \tag{4.194}$$

where we dropped the letter R denoting the reference field. In the integral expression (4.194),

$$\iota\colon \beta(I) \hookrightarrow A^1M \tag{4.195}$$

denotes the inclusion map[53] of the subset $\beta(I)$ of A^1M. The set $\beta(I) \subset A^1M$ is an immersed submanifold of A^1M. Indeed, the map $\beta\colon I \to A^1M$ is an injective immersion because of equation (4.35) and since the tangent vector of a time-parametrized curve does never vanish. The action (4.194) can be rewritten as

$$\begin{aligned} A[\beta] &= \int_I \beta_{\leftarrow}\omega \\ &= \int_I \big(\beta_{\leftarrow}\omega\big)(\partial/\partial\tau)\,\mathrm{d}\tau \\ &= \int_I \omega_{\beta(\tau)}\big(\mathrm{D}\beta_\tau(\partial/\partial\tau)\big)\,\mathrm{d}\tau, \end{aligned} \tag{4.196}$$

50. See p. 2 of Landau and E. M. Lifshitz 1969.
51. See Section 8.3 of Loos 1982.
52. We refer to Chapter 16 in John M. Lee 2013 for the theory about integration on manifolds.
53. See footnote on p. 63.

where we used the definition of the integral, equation (3.49) and the definition (3.61) of the pullback of the one-form ω with the map $\beta\colon I \to A^1M$. For the natural chart (4.17), it holds that $\Phi \circ \beta = (t(\tau), \mathbf{x}(\tau), \mathbf{u}(\tau))$, such that

$$\mathrm{D}\beta_\tau({\partial}/{\partial\tau}) = \frac{\partial}{\partial t}\bigg|_{\beta(\tau)} + \dot{x}^i(\tau)\frac{\partial}{\partial x^i}\bigg|_{\beta(\tau)} + \dot{u}^i(\tau)\frac{\partial}{\partial u^i}\bigg|_{\beta(\tau)}$$

and

$$\omega = L\,\mathrm{d}t + \frac{\partial L}{\partial u^j}(\mathrm{d}x^j - u^j\mathrm{d}t).$$

Therefore, the integrand in (4.196) is given by

$$\omega_{\beta(\tau)}\big(\mathrm{D}\beta_\tau({\partial}/{\partial\tau})\big) = L\big(\beta(\tau)\big) + \frac{\partial L}{\partial u^j}\bigg|_{\beta(\tau)}\big(\dot{x}^j(\tau) - u^j(\tau)\big). \tag{4.197}$$

If the action (4.194) is evaluated on a second-order curve $\dot{\gamma}\colon I \to A^1M$ (see equation (4.22)), it reads

$$A[\dot{\gamma}] = \int_I L(\dot{\gamma}(\tau))\mathrm{d}\tau = \int_I \mathrm{L}(t(\tau), \mathbf{x}(\tau), \dot{\mathbf{x}}(\tau))\mathrm{d}\tau, \tag{4.198}$$

because the second term in (4.197) vanishes along second-order curves. Note that, we used again the upright $\mathrm{L} := L \circ \Phi^{-1}$ introduced on p. 132 to denote the chart representation of the Lagrangian $L\colon A^1M \to \mathbb{R}$. The right-hand side of (4.198) is the action integral as it can be found in classical texts.[54] The classical version of **Hamilton's principle** states that between two events p and q with $t(q) > t(p)$ the motion of a mechanical system makes the integral (4.198) stationary. Classically, the stationarity is studied for variational families of second-order curves. The integral (4.194) can be found in Cartan 1922. That is why the one-form ω is referred to as Cartan one-form. The observation that (4.194) comprises the classical action (4.198) when evaluated along second-order curves can be found on page 17 of Cartan 1922.

In our context, this principle can be reformulated as follows: For exact mechanical systems a motion β (i.e., an integral curve of the vector field $X \in \mathrm{Vect}(A^1M)$ that is determined by $X \lrcorner \mathrm{d}\omega = 0$ and $\hat{\vartheta}(X) = 1$) relates two events p and q if it is an extremal of the variational problem defined by the action (4.194) for fixed *position* endpoints, which shall mean that the variational family is such that for $I = [\tau_0, \tau_1]$

$$\delta\kappa_{\tau_0}(\varepsilon = 0) \in \mathrm{Ver}_{\beta(\tau_0)}(A^1M) \tag{4.199}$$

54. See for example Landau and E. M. Lifshitz 1969, p. 2, or Hamel 1949, p. 235.

and

$$\delta\kappa_{\tau_1}(\varepsilon = 0) \in \mathrm{Ver}_{\beta(\tau_0)}(A^1M). \tag{4.200}$$

We can see from the local expression (4.175) that the conditions (4.199) and (4.200) mean that the position coordinates at the endpoints are kept fixed while the velocity parameters may be varied.

We define[55] the **first variation** of the action (4.194) as

$$\delta A[Y] := \frac{\partial}{\partial\varepsilon}\bigg|_{\varepsilon=0} \int_{\varphi_\varepsilon\circ\beta(I)} i_{\leftarrow}\omega,$$

where $i\colon \kappa_\varepsilon(I) \hookrightarrow A^1M$ denotes the inclusion map of the subset $\kappa_\varepsilon(I) \subset A^1M$ and φ is the flow of the vector field $Y \in \mathrm{Vect}(A^1M)$. The first variation can be rewritten as

$$\begin{aligned}
\delta A[Y] &= \tfrac{\partial}{\partial\varepsilon}\big|_{\varepsilon=0} \int_{\beta(I)} \iota_{\leftarrow}\big((\varphi_\varepsilon)_{\leftarrow}\omega\big) \\
&= \int_{\beta(I)} \tfrac{\partial}{\partial\varepsilon}\big|_{\varepsilon=0} \iota_{\leftarrow}\big((\varphi_\varepsilon)_{\leftarrow}\omega\big) \\
&= \int_{\beta(I)} \iota_{\leftarrow}\big(\mathfrak{L}_Y\omega\big),
\end{aligned} \tag{4.201}$$

where ι denotes the inclusion map (4.195). In equation (4.201), we used definition (3.64) of the Lie derivative and that differentiation and integration can be interchanged. Finally, by Cartan's magic formula (3.66), $\mathfrak{L}_Y\omega$ can be written as

$$\mathfrak{L}_Y\omega = i_Y \circ \mathrm{d}\omega + \mathrm{d}\circ i_Y\omega$$

such that the first variation (4.201) becomes

$$\begin{aligned}
\delta A[Y] &= \int_{\beta(I)} \iota_{\leftarrow}\big(i_Y\circ\mathrm{d}\omega + \mathrm{d}\circ i_Y\omega\big) \\
&= \int_{\beta(I)} \iota_{\leftarrow}(i_Y\circ\mathrm{d}\omega) + \int_{\partial\beta(I)} \mathrm{ι}_{\leftarrow}(i_Y\omega),
\end{aligned} \tag{4.202}$$

where we used Stoke's theorem.[56] In equation (4.202), $\partial\beta(I)$ denotes the boundary of the interval I, i.e., for $I = [\tau_0, \tau_1]$

$$\int_{\partial\beta(I)} \mathrm{ι}_{\leftarrow}(i_Y\omega) = \big[(i_Y\omega)\circ\beta\big]_{\tau_0}^{\tau_1} = (i_Y\omega)\circ\beta(\tau_1) - (i_Y\omega)\circ\beta(\tau_0). \tag{4.203}$$

Moreover, the upright iota denotes the inclusion map $\mathrm{ι}\colon \partial\beta(I) \hookrightarrow \beta(I)$ of the boundary $\partial\beta(I)$ of the one-dimensional manifold $\beta(I)$. For variations of the curve β with fixed position endpoints, the integral (4.203) vanishes because the Cartan one-form is semi-basic as can be easily seen from its

55. This definition follows Hermann 1988.
56. See Theorem 16.11 of John M. Lee 2013.

expression (4.96) in local coordinates and, therefore, it vanishes when it is evaluated on the vertical vectors (4.199) and (4.200) at the respective points $\beta(\tau_0)$ and $\beta(\tau_1)$. Consequently, the first variation (4.202) of the action (4.194) can be written as

$$\begin{aligned}\delta A[Y] &= \int_{\beta(I)} \iota^{\leftarrow}(i_Y \circ \mathrm{d}\omega) \\ &= \int_I \beta^{\leftarrow}(i_Y \circ \mathrm{d}\omega)\,(\partial/\partial\tau)\,\mathrm{d}\tau \\ &= \int_I \mathrm{d}\omega\big(Y, X\big) \circ \beta(\tau)\,\mathrm{d}\tau,\end{aligned} \tag{4.204}$$

where the vector field X satisfies $\dot{\beta} = X(\beta)$. We say that the first variation vanishes if $\delta A[Y] = 0$ for all vector fields $Y \in \mathrm{Vect}(A^1M)$. We see from (4.204) that Postulate 4.8 implies $\delta A[Y] = 0$ for all vector fields Y. The first variation (4.204) vanishes *in particular* for virtual displacement fields and for virtual displacement fields of the form (4.182). Thus, we have proved the following theorem.

Theorem 4.9 (Hamilton's principle)**.** Let $\Omega = \mathrm{d}\omega$ denote the action form of an exact mechanical system provided by Postulate 4.8. Then the integral curves $\dot{\beta} = X(\beta)$ of the vector field $X \in \mathrm{Vect}(A^1M)$ determined by $\hat{\vartheta}(X) = 1$ and $X \lrcorner \Omega = 0$ make the action (4.194) stationary.

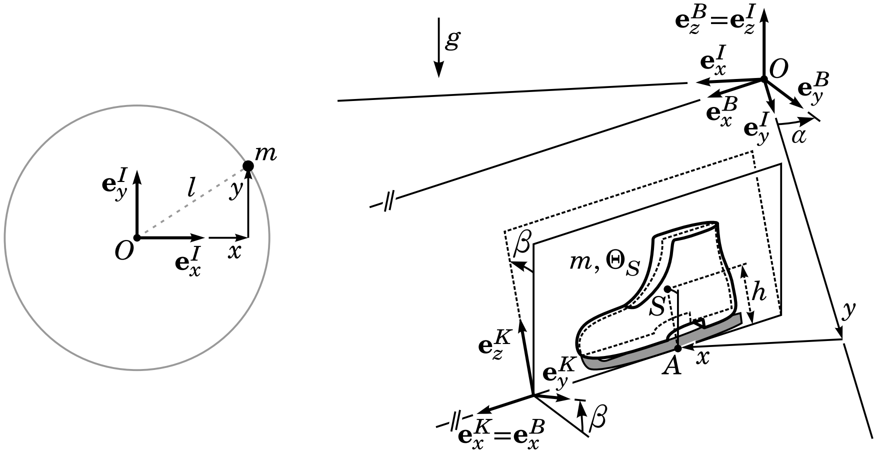
4.12. Constraints

In Section 4.2, we saw that the configuration of a mechanical system at a particular instant of time can be seen as a point in the Galilean manifold (M, ϑ, g) of the system (see Definition 4.1). This means that the manifold M defines the kinematics of the mechanical system under study. For technical applications, it is useful if one is able to subject an initially free system to additional constraints. Let us consider the example of a point mass moving freely in the plane (see Figure 4.6a) that should be restricted to keep a constant distance l to the point O. This restriction can be expressed in terms of the coordinates x and y as

$$g(x, y) := x^2 + y^2 - l^2 \overset{!}{=} 0. \tag{4.205}$$

Another example is the ice skate from Figure 4.6b whose position can be described by the coordinates $\mathbf{q} = (x, y, \alpha, \beta)$. The characteristic property of an ice skate is that its blade does only slide in longitudinal direction. Mathematically, this condition can be written as

$$h(\mathbf{q}, \dot{\mathbf{q}}) := -\dot{x} \sin\alpha + \dot{y} \cos\alpha \overset{!}{=} 0. \tag{4.206}$$



(a) A point mass that is constrained to move on a circle.

(b) An ice skate modelled as a rigid body that slides on a horizontal plane.

Figure 4.6.: Two examples of mechanical systems involving constraints.

This section deals with the description of finite-dimensional mechanical systems that are subjected to algebraic restrictions such as (4.205) or (4.206). Note that condition (4.205) constrains the position of the point mass, while (4.206) restricts the translational velocity of the ice skate. However, the restriction (4.205) can be brought to velocity level by differentiating it with respect to time, i.e.,

$$\frac{\mathrm{d}g}{\mathrm{d}t}(x, y, \dot{x}, \dot{y}) = 2x\dot{x} + 2y\dot{y} = 0. \tag{4.207}$$

We observe from (4.207) that both restrictions (4.205) and (4.206) are linear in the velocity parameters, when expressed on velocity level. Therefore, the following definitions seem reasonable.

Let (M, ϑ, g) with $\dim M = n+1$ be the Galilean manifold of a mechanical system with n degrees of freedom whose motion is given by the action form Ω. Let $\mathit{\Delta}$ be a distribution of rank $l < n+1$ on M. We say that $\mathit{\Delta}$ is **compatible with the time structure** ϑ if around each point p of M there is a neighbourhood $U \in M$ such that

$$A_q^1 M \cap \mathit{\Delta}_q \neq \emptyset \tag{4.208}$$

for all $q \in U$. The distribution given by the spacelike bundle A^0M is an example of a distribution that is *not* compatible with ϑ. Our aim is to impose restrictions on the spatial directions in which a given mechanical system can move. The compatibility requirement (4.208) avoids that we impose restrictions in time direction.

We define a **constraint** to be a distribution $\varDelta$ of rank l on M that is compatible with the time structure and that comes with a constraint force two-form Φ^{c} which guarantees that the vector field Z^{c} defined by

$$\begin{aligned} \hat{\vartheta}(Z^{\mathrm{c}}) &= 1, \\ Z^{\mathrm{c}} \lrcorner (\Omega + \Phi^{\mathrm{c}}) &= 0, \end{aligned} \tag{4.209}$$

satisfies

$$\mathrm{D}\pi_a(Z^{\mathrm{c}}_a) \in \varDelta_{\pi(a)} \tag{4.210}$$

for all $a \in W \subseteq A^1M$, where W denotes the neighbourhood on which the motion is studied. This means that the integral curves of Z^{c} lie in the distribution $\varDelta$. For a given constraint distribution of rank l, the force two-form Φ^{c} and thereby the vector field Z^{c} is by no means uniquely defined. For this, further restrictions need to be imposed.

One fruitful approach is to assume that the constraint forces are *ideal*. The principle of d'Alembert/Lagrange says that a constraint force is **ideal** if it does not produce any virtual work for compatible virtual displacements (see Glocker 2001, p. 48). The following considerations allow us to recast this principle for our purposes. A *compatible virtual displacement* means a spacelike vector $v_p \in A^0_pM$ that is compatible with the constraint distribution in the sense that $v_p \in \varDelta_p$ for all $p \in M$. Therefore, compatible virtual displacements are elements from the intersection

$$A^0_pM \cap \varDelta_p, \tag{4.211}$$

which is a (vector) subspace of A^0_pM. Consequently, a field of compatible virtual displacements is a smooth local section v of A^0M defined on an open subset U of M that satisfies $v_p \in \varDelta_p$ for all $p \in U$. Equation (4.67) defines forces as linear forms on the space of vertical vector fields. By the pointwise isomorphism (4.41), any local section $u\colon U \to A^0M$ defines a local section $\hat{u}\colon \pi^{-1}(U) \to \mathrm{Ver}(A^1M)$. This holds in particular for fields of compatible virtual displacements. In the coordinates of a chart (U, ϕ) of M and of its corresponding natural chart $(\pi^{-1}(U), \varPhi)$ of A^1M, these sections read

$$u = u^i \frac{\partial}{\partial x^i} : U \to A^0M$$

and

$$\hat{u} = u^i \circ \pi \frac{\partial}{\partial u^i} : \pi^{-1}(U) \to \mathrm{Ver}(A^1M),$$

respectively. According to (4.42), the coordinate fields are related as

$$\left.\frac{\partial}{\partial u^i}\right|_a \mapsto \left.\frac{\partial}{\partial x^i}\right|_{\pi(a)},$$

for all $a \in \pi^{-1}(U) \subseteq A^1M$. A constraint force

$$F^{\mathrm{c}} \in \Gamma\Big(\mathrm{Ver}^*(A^1M)\Big)$$

is said to be **ideal** if

$$F^{\mathrm{c}} \cdot \hat{v} = 0, \tag{4.212}$$

for all (local) sections v of the bundle of compatible virtual displacements $A^0M \cap \Delta$. The corresponding two-form Φ^{c} follows then by the bijective relation

$$\Phi^{\mathrm{c}} = -\tfrac{1}{2}\Big(\mathbf{\partial}(F^{\mathrm{c}} \circ \mu) + \hat{\vartheta} \wedge (F^{\mathrm{c}} \circ \mu)\Big)$$

established by Theorem 4.4. A constraint for which the constraint forces are assumed to be ideal is called an **ideal constraint**.

Let us study the situation in local coordinates. An adapted chart (U, ϕ) on M provides the coordinates $(t, x^1, \dots, x^n)$ and according to (3.74) a distribution of rank l on M can be locally defined by $n-l$ linearly independent differential one-forms

$$\alpha^\nu = H_i^\nu \mathrm{d}x^i + c^\nu \mathrm{d}t$$

with $\nu = 1, \dots, n-l$ as

$$\Delta_q := \ker \alpha^1\big|_q \cap \dots \cap \ker \alpha^{n-l}\big|_q,$$

for all $q \in U$. The compatibility condition (4.208) requires the sets

$$A_q^1M \cap \Delta_q = \left\{ \frac{\partial}{\partial t} + u^i \frac{\partial}{\partial x^i} \in T_qM \,\middle|\, H_i^1 u^i + c^1 = 0, \dots, H_i^{n-l} u^i + c^{n-l} = 0 \right\}$$

to be non-empty for all points $q \in U$. This is the case if the coefficient matrix $\mathbf{H}(q) = [H_i^\nu(q)]$ has rank $n-l$ for each point $q \in U$ because then the linear equation

$$\mathbf{Hu} + \mathbf{c} = \mathbf{0}$$

has at least one solution $\mathbf{u}$ for a given value of $\mathbf{c}$. This means that the differential one-forms $\alpha^1, \dots, \alpha^{n-l}$, *and* $\mathrm{d}t$ are linearly independent. The spaces of compatible virtual displacements (4.211) are then given by

$$A_q^0M \cap \Delta_q = \left\{ v^i \frac{\partial}{\partial x^i} \in T_qM \,\middle|\, H_i^1 v^i = 0, \dots, H_i^{n-l} v^i = 0 \right\},$$

for each $q \in U$, respectively by

$$\mathbf{H}\mathbf{v} = \mathbf{0} \tag{4.213}$$

in matrix notation. Under the isomorphism (4.42), equation (4.213) defines the subspace of vertical vectors on which an *ideal* constraint force needs to vanish in order to be ideal.

Indeed, by condition (4.212), at any point $a \in \pi^{-1}(U) \subseteq A^1M$, an ideal constraint force must satisfy

$$F^{\mathrm{c}} \cdot \hat{v}(a) = \left(F_i^{\mathrm{c}}(a)\,\mathrm{d}u_a^i\right) \cdot \left(v^i \circ \pi(a) \frac{\partial}{\partial u^i}\bigg|_a\right) = 0, \tag{4.214}$$

whenever the $v^i \circ \pi(a) = v^i(p)$ satisfy equation (4.213). If $\mathbf{F}^{\mathrm{c}}$ denotes the $\mathbb{R}^n$-tuple that gathers the coefficients $F_1^{\mathrm{c}}, \dots, F_n^{\mathrm{c}}$, then condition (4.214) can be expressed in matrix notation as

$$(\mathbf{F}^{\mathrm{c}})^{\mathrm{T}}\mathbf{v} = 0,$$

for all $\mathbf{v} \in \mathbb{R}^n$ that satisfy equation (4.213). In other words, condition (4.212) requires $\mathbf{F}^{\mathrm{c}}$ to lie in the annihilator space[57] of $\ker \mathbf{H}$, i.e.,

$$\mathbf{F}^{\mathrm{c}} \in (\ker \mathbf{H})^{\circ} = \left\{\mathbf{F} \in \mathbb{R}^n \mid \mathbf{F}^{\mathrm{T}}\mathbf{w} = 0, \text{ for all } \mathbf{w} \in \ker \mathbf{H}\right\}$$

in order to define an ideal constraint force.

The constraint force is ideal if and only if it has the form

$$F^{\mathrm{c}} = F_i^{\mathrm{c}}\,\mathrm{d}u^i = H_i^{\nu}\lambda_{\nu}\,\mathrm{d}u^i, \tag{4.215}$$

with $n-l$ coefficients $\lambda_1, \dots, \lambda_{n-l}$ (instead of n for an arbitrary force). Let $\boldsymbol{\lambda}$ denote the $\mathbb{R}^{n-l}$-tuple gathering the coefficients $\lambda_1, \dots, \lambda_{n-l}$, then (4.215) can be expressed as

$$\mathbf{F}^{\mathrm{c}} = \mathbf{H}^{\mathrm{T}}\boldsymbol{\lambda} \tag{4.216}$$

in matrix notation. First, we observe that tuples $\mathbf{F}^{\mathrm{c}}$ of the form (4.216) vanish on the directions defined by $\ker \mathbf{H}$ and, therefore, are elements of $(\ker \mathbf{H})^{\circ}$. We still need to show the reverse direction, i.e., that all elements from $(\ker \mathbf{H})^{\circ}$ can be written as $\mathbf{H}^{\mathrm{T}}\boldsymbol{\lambda}$ for some $\boldsymbol{\lambda} \in \mathbb{R}^{n-l}$. By the dimension formula (2.25), it holds that $\dim\,(\ker \mathbf{H})^{\circ} = \dim \mathbb{R}^n - \dim \ker \mathbf{H}$. From the rank-nullity theorem, we know that $\dim \ker \mathbf{H} = n - \dim \operatorname{im} \mathbf{H} = l$ because $\dim \operatorname{im} \mathbf{H} = \operatorname{rank} \mathbf{H} = n-l$. Therefore, $\dim\,(\ker \mathbf{H})^{\circ} = \dim \operatorname{im} \mathbf{H} = n-l$ which proves the assertion.

57. See equation (2.24) for the definition.

As the diligent reader might have expected, the results from Section 3.11 can be used to determine whether a constraint distribution is involutive or not. Constraints for which the constraint distribution is involutive are said to be **holonomic** constraints and those for which it is not involutive are referred to as **nonholonomic** constraints.

The defining one-forms corresponding to the respective restriction of the pendulum (4.205) and the ice skate (4.206)) are given by

$$\alpha_g = \mathrm{d}g = 2x\mathrm{d}x + 2y\mathrm{d}y \tag{4.217}$$

and

$$\alpha_h = -\sin\alpha\,\mathrm{d}x + \cos\alpha\,\mathrm{d}y, \tag{4.218}$$

respectively. We check the involutivity of both forms according to Proposition 3.39. As expected, the one-form α_g is involutive on the punctured plane $\mathbb{R}^2\backslash\{\mathbf{0}\}$ because $\mathrm{d}\alpha = \mathrm{d}\circ\mathrm{d}g = 0$. We need to exclude the origin since there the kernel of α_g is degenerate. Let us consider the chart $\psi\colon \mathbb{R}^2\backslash\{\mathbf{0}\} \to \mathbb{R}^2$, $p \mapsto \psi(p) = (\varphi, r)$ that describes the punctured plane using polar coordinates (in reversed order). The coordinate change between the (x, y)- and (φ, r)-coordinates is given by

$$\begin{aligned} x &= r\cos\varphi, \\ y &= r\sin\varphi. \end{aligned}$$

The chart $\psi\colon \mathbb{R}^2\backslash\{\mathbf{0}\} \to \mathbb{R}^2$ is flat[58] for the distribution defined by α_g. Indeed, it holds that

$$\alpha_g = 2r\mathrm{d}r = \mathrm{d}(r^2)$$

and, therefore, the distribution is spanned by the coordinate field $\partial/\partial\varphi$ (see Figure 4.7). In polar coordinates, the integral manifolds of the distribution defined by α_g are given by the slices $r = \bar{r}$ for constant values $\bar{r} > 0$.

Let us come to the one-form α_h from (4.218). Its exterior derivative is given by

$$\mathrm{d}\alpha_h = -\dot{\alpha}\cos\alpha\,\mathrm{d}t \wedge \mathrm{d}x - \dot{\alpha}\sin\alpha\,\mathrm{d}t \wedge \mathrm{d}y.$$

Let u, v be local sections of the distribution defined by α_h, then it locally holds that

$$\begin{aligned} u &= a_u^t\frac{\partial}{\partial t} + a_u^x\frac{\partial}{\partial x} + a_u^y\frac{\partial}{\partial y} + a_u^\alpha\frac{\partial}{\partial \alpha} + a_u^\beta\frac{\partial}{\partial \beta} \\ v &= a_v^t\frac{\partial}{\partial t} + a_v^x\frac{\partial}{\partial x} + a_v^y\frac{\partial}{\partial y} + a_v^\alpha\frac{\partial}{\partial \alpha} + a_v^\beta\frac{\partial}{\partial \beta} \end{aligned}$$

58. See p. 88 for the definition.

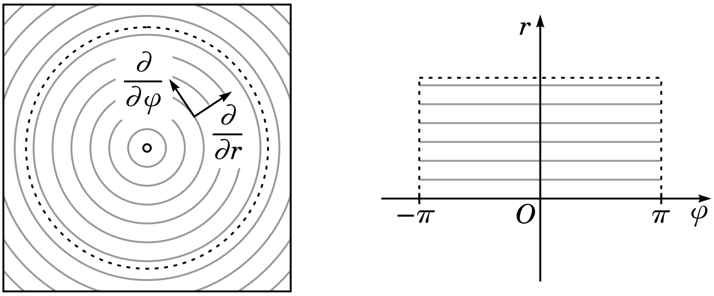


Figure 4.7.: Polar coordinates correspond to a flat chart of the distribution defined by the one-form α_g from (4.217) on the punctured plane $\mathbb{R}^2\backslash\{\mathbf{0}\}$.

with the restrictions

$$-a_u^x \sin\alpha + a_u^y \cos\alpha = 0,$$
$$-a_v^x \sin\alpha + a_v^y \cos\alpha = 0.$$

The evaluation of the involutivity condition from Proposition 3.39 for vector fields $u, v \in \ker \alpha_h$ yields

$$\begin{aligned} \mathrm{d}\alpha_h(u,v) &= -\dot\alpha \cos\alpha\,(a_u^t a_v^x - a_v^t a_u^x) - \dot\alpha \sin\alpha\,(a_u^t a_v^y - a_v^t a_u^y) \\ &= -a_u^t \dot\alpha\,(a_v^x \cos\alpha + a_v^y \sin\alpha) + a_v^t \dot\alpha\,(a_u^x \cos\alpha + a_u^y \sin\alpha), \end{aligned}$$

which is clearly non-zero such that α_h defines a nonholonomic constraint.

Ideal holonomic constraints reveal to be a special case of a more general situation. Let (M, ϑ, g) be a Galilean manifold and Ω be an action form on A^1M. Let $f\colon M' \to M$ be an immersion from a manifold M' to M. We say that f is **compatible** with the time structure if $\vartheta' := f_{\leftarrow}\vartheta$ defines a time structure on M'. In this case the Galilean metric g induces a Galilean metric $g' := f_{\leftarrow}(g)$, i.e.,

$$g'(u,v) = g\big(\mathrm{D}f(u), \mathrm{D}f(v)\big),$$

for all $u, v \in \Gamma(A^0M')$ and $Df\colon TM' \to TM$ maps the affine subbundle A^1M' to A^1M.

Theorem 4.10 (Loos 1982, p. 50)**.** The differential two-form $\Omega' := (\mathrm{D}f)_{\leftarrow}\Omega$ is an action form on A^1M' that induces the Galilean metric g'. If Ω is closed then Ω' is also closed. If $\mathrm{d}\Omega$ is basic then $\mathrm{d}\Omega'$ is also basic.

Proof. $\mathrm{D}(\mathrm{D}f)$ maps the bundle $\mathrm{Ver}(A^1M')$ to $\mathrm{Ver}(A^1M)$ and it commutes with the vertical homomorphisms μ' and μ, i.e.,

$$\mu \circ \mathrm{D}(\mathrm{D}f) = \mathrm{D}(\mathrm{D}f) \circ \mu'. \tag{4.219}$$

Furthermore, it holds that if $\sigma \in \Omega^\star(A^1M)$ is basic, respectively semi-basic, then its pullback $(\mathrm{D}f)_{\leftarrow}\sigma \in \Omega^\star(A^1M')$ is also basic, respectively semi-basic and the pullback $(\mathrm{D}f)_{\leftarrow}$ commutes with the differentiation operators d and ∂, which proves the assertion. □

Because of Theorem 4.10, we call the Galilean manifold (M', ϑ', g') together with the action form Ω' the **mechanical system on M' induced by f.** In the case of holonomic constraints, M' denotes the integral manifold of the distribution Δ of the constraint and f is the inclusion map $\iota : M' \hookrightarrow M$. Locally, M' can be described by $n-l$ equations

$$g^\nu(t, x^1, \dots, x^n) = 0,$$

where $\nu = 1, \dots, n-l$. The one-forms

$$\alpha^\nu := \mathrm{d}g^\nu = \frac{\partial g^\nu}{\partial x^i}\mathrm{d}x^i + \frac{\partial g^\nu}{\partial t}\mathrm{d}t$$

are defining forms of the distribution Δ. As we have seen, the compatibility with the time structure (4.208) requires the differentials $\mathrm{d}t, \mathrm{d}g^1, \dots, \mathrm{d}g^{n-l}$ to be linearly independent or, equivalently, the coefficient matrix

$$\left[\frac{\partial g^\nu}{\partial x^i}\right]$$

needs to have rank $n-l$. In this case, the time structure $\vartheta' = \iota_{\leftarrow}\vartheta$ corresponds to the restriction

$$\vartheta' = \iota_{\leftarrow}\vartheta = \vartheta|_{M'}.$$

The spacelike bundle A^0M' and state-space A^1M' of the integral manifold M' are given by

$$A^0M' = A^0M \cap TM' = \bigcup_{p \in M'} \Big(\{p\} \times \big(A^0_pM \cap T_pM'\big)\Big)$$

and

$$A^1M' = A^1M \cap TM' = \bigcup_{p \in M'} \Big(\{p\} \times \big(A^1_pM \cap T_pM'\big)\Big),$$

respectively. Note that by equation (3.73) it holds that

$$\mathrm{D}\iota_p\big(T_pM'\big) = \Delta_p,$$

for all $p \in M'$. The action form Ω' on the submanifold A^1M' is given by the restriction of Ω to the state-space A^1M', i.e.,

$$\Omega' = (\mathrm{D}\iota)_{\leftarrow}\Omega = \Omega|_{A^1M'}.$$

Let Z' denote the vector field on A^1M' that is uniquely determined by

$$\vartheta'(Z') = 1,\ Z' \lrcorner \Omega' = 0.$$

We will show that for a mechanical system that is subjected to an *ideal* holonomic constraint with integral manifold M', the vector field Z' corresponds to the restriction to A^1M' of the vector field Z^{c} determined by equations (4.209) and (4.210), i.e.,

$$Z' = Z^{\mathrm{c}}|_{A^1M'}.$$

Condition $Z' \lrcorner \Omega' = 0$ is equivalent to the variational form

$$\Omega'_a\left(Z'_a, Y'_a\right) = 0, \tag{4.220}$$

for all $Y'_a \in T_a(A^1M')$ and all $a \in A^1M'$. Since Ω' is the pullback of Ω with the differential of the inclusion map $\iota\colon M' \hookrightarrow M$, condition (4.220) can be written as

$$\Omega_a\left(\mathrm{D}(\mathrm{D}\iota)_a(Z'_a), \mathrm{D}(\mathrm{D}\iota)_a(Y'_a)\right) = 0, \tag{4.221}$$

for all $Y'_a \in T_a(A^1M')$ and all $a \in A^1M'$. Let Z denote the second-order field of the unconstrained system, i.e., the vector field that is uniquely defined by

$$\vartheta(Z) = 1,\ Z \lrcorner \Omega = 0.$$

We know from Section 4.7 that for all points $a \in A^1M'$

$$\mathrm{D}(\mathrm{D}\iota)_a(Z'_a) = Z_a + V_a,$$

for some vertical vector $V_a \in \mathrm{Ver}_a(A^1M)$ such that condition (4.221) yields

$$\Omega_a\left(Z_a + V_a, \mathrm{D}(\mathrm{D}\iota)_a(Y'_a)\right) = 0,$$

for all $Y'_a \in T_a(A^1M')$ and all $a \in A^1M'$. Using that $Z_a \lrcorner \Omega_a = 0$ it follows with the definition (4.58) of the action form that

$$\hat{g}\left(\eta_a(V_a), \mu_a \circ \mathrm{D}(\mathrm{D}\iota)_a(Y'_a)\right) - \hat{g}\left(\eta_a \circ \mathrm{D}(\mathrm{D}\iota)_a(Y'_a), \mu_a(V_a)\right) = 0$$

for all $Y'_a \in T_a(A^1M')$ and all $a \in A^1M'$. We saw on p. 106 that the vertical homomorphism μ_a vanishes on vertical vectors and on p. 108 that η_a is a projection onto $\mathrm{Ver}_a(A^1M)$. Consequently, it follows that

$$\hat{g}\left(V_a, \mu_a \circ \mathrm{D}(\mathrm{D}\iota)_a(Y'_a)\right) = \hat{g}\left(V_a, \mathrm{D}(\mathrm{D}\iota)_a \circ \mu'_a(Y'_a)\right) = 0,$$

for all $Y'_a \in T_a(A^1M')$ and all $a \in A^1M'$, where $\mu'\colon T(A^1M) \to \mathrm{Ver}_a(A^1M)$ denotes the vertical homomorphism on A^1M'. The second equality follows because $\mathrm{D}(\mathrm{D}\iota)$ commutes with μ according to (4.219). Since μ' is surjective, it holds that

$$\hat{g}\Big(V_a, \mathrm{D}(\mathrm{D}\iota)_a Y_a\Big) = F_a\Big(\mathrm{D}(\mathrm{D}\iota)_a Y_a\Big) = 0, \tag{4.222}$$

for all $Y_a \in \mathrm{Ver}_a(A^1M')$ and all $a \in A^1M'$, where we identified the constraint force $F_a := \hat{g} \cdot V_a$ according to Section 4.7. Because the differential

$$\mathrm{D}(\mathrm{D}\iota)|_{A^1M'} : A^1M' \to A^1M$$

maps $\mathrm{Ver}(A^1M')$ to $\mathrm{Ver}(A^1M)$, the differential $\mathrm{D}(\mathrm{D}\iota)_a$ in a is just the inclusion map of the vector subspace $\mathrm{Ver}_a(A^1M') \subset \mathrm{Ver}_a(A^1M)$. Thus, condition (4.222) agrees with (4.212) and the constraint force F_a is ideal. Condition (4.222), which can be rewritten as

$$\hat{g}\Big(\mathrm{D}(\mathrm{D}\iota)_a(Z'_a) - Z_a, \mathrm{D}(\mathrm{D}\iota)_a Y_a\Big) = 0,$$

for all $Y_a \in \mathrm{Ver}_a(A^1M')$ and all $a \in A^1M'$, uniquely defines the vector field Z'. To see this, we consider another $Z''_a \in T_a(A^1M')$ with $\mathrm{D}\pi_a(Z''_a) = a$ and

$$\hat{g}\Big(\mathrm{D}(\mathrm{D}\iota)_a(Z''_a) - Z_a, \mathrm{D}(\mathrm{D}\iota)_a Y_a\Big) = 0,$$

for all $Y_a \in \mathrm{Ver}_a(A^1M')$ and all $a \in A^1M'$. Since

$$\mathrm{D}(\mathrm{D}\iota)_a(Z''_a) - \mathrm{D}(\mathrm{D}\iota)_a(Z'_a) \in \mathrm{Ver}_a(A^1M'),$$

it follows that

$$\hat{g}\Big(\mathrm{D}(\mathrm{D}\iota)_a(Z''_a) - \mathrm{D}(\mathrm{D}\iota)_a(Z'_a), \mathrm{D}(\mathrm{D}\iota)_a Y_a\Big) = \hat{g}|_{A^1M'}\Big(Z''_a - Z'_a, Y_a\Big) = 0,$$

for all $Y_a \in \mathrm{Ver}_a(A^1M')$ and all $a \in A^1M'$. It follows that $Z'_a = Z''_a$ because the restriction of $\hat{g}$ to $\mathrm{Ver}(A^1M')$ is non-degenerate.

Conclusion 5

Science did not progress by that harmonious path, the illusion of which is easily created after the event.

— René Dugas

This thesis presents a description of finite-dimensional mechanical systems that may be explicitly time-dependent and include nonpotential forces. The language of contemporary differential geometry allowed us to define the involved physical quantities as coordinate-free objects.

Chapter 1 reveals the different viewpoints adopted in geometric and technical mechanics. The presented comparison of coordinate-free approaches for the description of finite-dimensional mechanical systems not only illustrates the different underlying assumptions but it highlights the restrictions they imply for the resulting physical theory.

Compared to Loos 1982, the structure of this presentation is reorganized such that forces take the centre stage. In the author's eyes, this choice brings the geometric presentation closer to more classical texts that are based on the principle of virtual work. The thesis bridges the divide between geometric and technical mechanics by establishing a firm link to classical results. First, Lagrange's, Hamel's and Hamilton's equations are put on an equal footing by showing that each set of equations can be derived from Postulate 4.8 by choosing a respective chart of the state space. While Lagrange's and Hamilton's equations can be found in Loos 1982, the author could not find the presented derivation of Hamel's equations in the literature. The explicit elaboration of the link between Postulate 4.8 and the principle of virtual work identified by Souriau led us to the respective central equation of Lagrange and Hamel. These equations, as suggested by their name, are at the heart of many classical works on the dynamics of finite-dimensional mechanical systems. To the author's knowledge no coordinate-free formulation of mechanics exists in which the connection with these classical results is made.

The study of Hamilton's principle allowed us to establish the link to the calculus of variations. Classically, Hamilton's principle postulates the stationarity of the action for variational families of second-order curves. We saw that by using the Cartan one-form, this principle can be "generalized" to an action for which the stationarity is postulated for variational families of arbitrary[1] time-parametrized curves. The principle of virtual work is an expression that "looks like" the stationarity condition of a variational principle. If the different virtual work contributions (the forces) can be obtained from a potential, then there exists a corresponding variational principle (the classical form of Hamilton's principle). In our geometric approach, the stationarity condition reads

$$X \lrcorner \Omega \stackrel{!}{=} 0$$

and there exists a corresponding variational principle (Hamilton's principle) if the action form Ω can be locally derived from a potential, i.e., if Ω is closed meaning that $\mathrm{d}\Omega = 0$. With the comparison of Postulate 4.8 and the principle of virtual work, we have pointed out the structural equivalence of both approaches.

Section 4.12 gives a coordinate-free definition of constraints as a distribution on M that is compatible with the time structure ϑ. This definition not only allows to distinguish between holonomic and nonholonomic constraints using the Frobenius theorem, but it shows that a theory built upon Postulate 4.8 may be used for the description of constrained mechanical systems.

Finally, this work makes part[2] of the results from Loos 1982 available in English and, thereby, may prevent this major contribution from falling into oblivion. Moreover, the unified reformulation of classical results around Postulate 4.8 allows a critical retrospective on the development of classical mechanics and the theory for finite-dimensional systems in particular. We consider two excerpts that underline the pioneering role played by Georg Hamel and Élie Cartan. On page 416 of Hamel 1904b, one can read:

> *So fruchtbringend nun auch die Verknüpfung der Mechanik mit der Variationsrechnung gewesen ist (Lagrange, Hamilton, Jacobi), so läßt sich doch nicht leugnen, daß die [obige] Auffassungsweise der virtuellen Verschiebungen einseitig ist und ihrer mechanischen Bedeutung nicht voll entspricht; daß sie namentlich der Anknüpfung weiterer, außerhalb des Gesichtskreises der Variationsrechnung liegender Beziehungen im Wege steht. Auch die merkwürdige, in der Literatur weitverbreitete Meinung, als ob das Wesen der allgemeinen Lagrangeschen Mechanik in den*

1. The adjective arbitrary means that there is no restriction to variational families of second-order curves.

2. The typescript Loos 1982 contains results that surpass the author's current mathematical competencies by far. An example of such a result is given by the proof of the statement that to any exact action form Ω there is a semi-basic one-form ω such that $\Omega = \mathrm{d}\omega$ (See pp. 61–62 in Loos 1982).

sogenannten Variationsprinzipien stecke, scheint mir in jenem Dogma ihre Wurzeln zu haben.

Hamel's statement can be translated into English as:

> *As fruitful as the connection of mechanics with the calculus of variations has been (Lagrange, Hamilton, Jacobi), it cannot be denied that the [above] interpretation of the virtual displacements is one-sided and does not fully correspond to their mechanical meaning; that it impedes, in particular, the connection to results that do not lie within the scope of variational calculus. Likewise it appears to me that the roots of the strange, widespread opinion in the literature that the essence of general Lagrangian mechanics can be found in the so-called variational principles lie in that dogma.*

With the results of Section 4.11 in mind, the extensive insight of Hamel becomes apparent. Élie Cartan writes on page 17 of Cartan 1922:

> *16. Nous avons vu que l'action élémentaire d'Hamilton pouvait s'obtenir en supposant que dans l'expression*
>
> $$\omega_\delta = \sum p_i \delta q_i - H\delta t,$$
>
> *on a*
>
> $$\delta q_i = q_i' \delta t.$$
>
> *Il est remarquable que les trajectoires d'un système matériel réalisent encore l'extremum de l'intégrale*
>
> $$W = \int_{t_0}^{t_1} \sum p_i \delta q_i - H\delta t,$$
>
> *en supposant simplement que les q_i et les q_i' sont des fonctions quelconques des t assujetties aux seules conditions que les q_i prennent des valeurs données à l'avance aux limites. On ne suppose donc plus, comme dans le principe d'Hamilton, que les q_i' soient les dérivées des q_i par rapport au temps. On peut même plus généralement supposer que les q_i, q_i' et t sont des fonctions d'un même paramètre u variant de 0 à 1, les quantités q_i et t prenant aux limites des valeurs données.*

A translation into English of the previous excerpt is given as:

> *16. We have seen that the elementary action of Hamilton can be obtained from the expression*
>
> $$\omega_\delta = \sum p_i \delta q_i - H\delta t,$$
>
> *by supposing that*
>
> $$\delta q_i = q_i' \delta t.$$

> *It is remarkable that the trajectories of a material system still realize an extremum of the integral*
>
> $$W = \int_{t_0}^{t_1} \sum p_i \delta q_i - H\delta t,$$
>
> *by simply supposing that the q_i and q_i' are arbitrary functions of t subjected to the sole conditions that the q_i take values given in advance at the limits. Thus, one does not suppose anymore, as with Hamilton's principle, that the q_i' are the derivatives of the q_i with respect to time. One can even more generally suppose that the q_i, q_i' and t are functions of a common parameter u running from 0 to 1, the quantities q_i and t taking given values at the limits.*

This observation is precisely what we have worked out in Section 4.11 for the Cartan one-form (4.93). Élie Cartan's *elementary action of Hamilton* corresponds to the local expression (4.154) of the Cartan one-form in canonical coordinates. The above two extracts give an impression of the far-sightedness of Hamel and Cartan.

Élie Cartan is considered to be the father of differential forms. It is astonishing to see that almost a century after Cartan 1922 differential forms have not found their way into classical mechanics. The literature survey about the central position that is still given to the classical version of Hamilton's principle in modern textbooks about finite-dimensional mechanical systems is left to the reader.

Compared to Georg Hamel and Élie Cartan, Ottmar Loos had the language of contemporary differential geometry at his disposal. Indeed, we saw in Section 4.7, that it is by giving a coordinate-free definition of the action form (4.58) that Loos was able to arrive at his definition of forces. Hence, in the case of finite-dimensional mechanical systems, Loos is able to give a final answer to the fundamental question about the definition of forces. This key result directly relies on the methods of global differential geometry. The same holds for the characterization of action forms (see Theorem 4.3). Indeed, it is hard to see how one could come up with the closure condition $\partial\Omega = 0$ using only local coordinates and index calculus. Loos' contributions to mechanics confirm once more that there is a firm link between mechanics and mathematics.

We already observed in Section 1.4 that the literature on the mathematical foundations of the physical description of finite-dimensional mechanical systems is not in an as good condition as it might be expected. A pessimist could object that many recently published textbooks have not even digested the knowledge from Hamel's and Cartan's era.

We have seen that Postulate 4.8 can be used as foundation of a coordinate-free physical theory dealing with finite-dimensional mechanical systems.

We demonstrated that classical principles such as the principle of virtual work and Hamilton's principle seamlessly fit in the coordinate-free picture as theorems. Therefore, an obvious continuation of this work lies in the integration of existing results as theorems of Postulate 4.8. From our discussion of Hamilton's principle in Section 4.11.3, we have learned that this process is more demanding than a mere reorganization of existing results. Indeed, concerning Hamilton's principle, we saw that it is the coordinate-free definition of a "generalized" action that lets the classical version of Hamilton's principle dovetail with the geometric description. Other coordinate-free considerations reveal to be less expedient. The classical action (4.198) can for example be interpreted as a real-valued function on the infinite-dimensional manifold of curves in M as suggested in Chapter 8 of Marsden et al. 1999. However, with the adoption of this view, we leave the (finite-dimensional) application area of differential geometry. This observation clearly underlines that the embedding of existing results into a coordinate-free formulation of mechanics is an intellectually demanding process.

A practically relevant class of finite-dimensional mechanical systems is given by multibody systems, which consist of rigid bodies interrelated by ideal constraints and by scalar force laws such as springs and dampers. A coordinate-free presentation of this subject from engineering mechanics would be a further contribution to the desired rapprochement of geometric and technical mechanics.

In view of teaching, the focus is shifted towards a didactic treatment of the subject. While geometric mechanics dissects the mathematical concepts underlying the description of finite-dimensional mechanical systems, technical mechanics has the objective to impart theoretical knowledge in view of applications. Because technical mechanics prescribes an economic use of mathematical concepts, the adaptation of a geometric description for engineering purposes is identified as an important but particularly challenging task.

Calculus on $\mathbb{R}^n$ A

We briefly summarize matrix notations that make calculus on $\mathbb{R}^n$ more comfortable. An element $\mathbf{x} \in \mathbb{R}^n$ is an n-tuple $\mathbf{x} = (x^1, \dots, x^n)$ of real numbers. To ease calculations, it may be re-interpreted as column vector

$$\mathbf{x} = \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix}.$$

The corresponding row vector is denoted by

$$\mathbf{x}^{\mathrm{T}} = \begin{bmatrix} x^1 & \cdots & x^n \end{bmatrix}.$$

For a vector-valued function $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{x} \mapsto \mathbf{f}(\mathbf{x})$, we define

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} := \begin{bmatrix} \dfrac{\partial f^1}{\partial \mathbf{x}} \\ \vdots \\ \dfrac{\partial f^m}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f^1}{\partial x^1} & \cdots & \dfrac{\partial f^1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f^m}{\partial x^1} & \cdots & \dfrac{\partial f^m}{\partial x^n} \end{bmatrix}. \tag{A.1}$$

In the case $m = 1$ of a real-valued function $f\colon \mathbb{R}^n \to \mathbb{R}$, $\mathbf{x} \mapsto f(\mathbf{x})$, equation (A.1) reduces to the row vector of length n

$$\frac{\partial f}{\partial \mathbf{x}} := \begin{bmatrix} \frac{\partial f}{\partial x^1} & \cdots & \frac{\partial f}{\partial x^n} \end{bmatrix}.$$

For a real-valued function $L\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $(\mathbf{x}, \mathbf{u}) \mapsto L(\mathbf{x}, \mathbf{u})$, we define the second-derivatives as the m-by-n matrix

$$\frac{\partial^2 L}{\partial \mathbf{x} \partial \mathbf{u}} := \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial L}{\partial \mathbf{u}}^{\mathrm{T}} \right)$$

and the m-by-m matrix

$$\frac{\partial^2 L}{\partial \mathbf{u} \partial \mathbf{u}} := \frac{\partial}{\partial \mathbf{u}} \left(\frac{\partial L}{\partial \mathbf{u}}^{\mathrm{T}} \right).$$

Bibliography

Abraham, R. and J. E. Marsden (1987). *Foundations of Mechanics*. 2nd ed. Addison-Wesley.

Abraham, R., J. E. Marsden, and T. S. Ratiu (1988). *Manifolds, Tensor Analysis, and Applications*. 2nd ed. Vol. 75. Applied Mathematical Sciences. Springer.

Arnold, V. I. (1989). *Mathematical Methods of Classical Mechanics*. 2nd ed. Vol. 60. Graduate Texts in Mathematics. Springer.

Artin, M. (2011). *Algebra*. 2nd ed. Prentice Hall.

Aubin, T. (2001). *A Course in Differential Geometry*. Vol. 27. Graduate Studies in Mathematics. American Mathematical Society.

Bishop, R. L. and S. I. Goldberg (1980). *Tensor Analysis on Manifolds*. Dover Publications.

Bloch, A. M. (2015). *Nonholonomic Mechanics and Control*. 2nd ed. Vol. 24. Interdisciplinary Applied Mathematics. Springer.

Bravetti, A., H. Cruz, and D. Tapias (2017). "Contact Hamiltonian mechanics". In: *Annals of Physics* 376, pp. 17–39.

Bremer, H. (1988). *Dynamik und Regelung mechanischer Systeme*. Vol. 67. Leitfäden der angewandten Mathematik und Mechanik. Teubner.

– (2008). *Elastic Multibody Dynamics*. Vol. 35. Intelligent Systems, Control, and Automation: Science and Engineering. Springer.

Bullo, F. and A. D. Lewis (2004). *Geometric Control of Mechanical Systems*. Vol. 49. Texts in Applied Mathematics. Springer.

Cartan, E. (1922). *Leçons sur les invariants intégraux*. Hermann.

Cortés, V. and A. S. Haupt (2017). *Mathematical Methods of Classical Physics*. Springer Briefs in Physics. Springer.

Crampin, M. (1983). "Tangent bundle geometry for Lagrangian dynamics". In: *Journal of Physics A: Mathematical and General* 16, pp. 3755–3772.

Crampin, M. and F. A. E. Pirani (1987). *Applicable Differential Geometry*. Vol. 59. London Mathematical Society Lecture Note Series. Cambridge University Press.

Dombrowski, H. D. and K. Horneffer (1964a). "Das verallgemeinerte Galileische Relativitätsprinzip". In: *Nachrichten der Akademie der Wissenschaften in Göttingen: Mathematisch-Physikalische Klasse* 17, pp. 233–246.

– (1964b). "Die Differentialgeometrie des Galileischen Relativitätsprinzips". In: *Mathematische Zeitschrift* 86, pp. 291–311.

Fischer, G. (2010). *Lineare Algebra*. 17th ed. vieweg studium – Grundkurs Mathematik. Vieweg+Teubner.

Frölicher, A. and A. Nijenhuis (1956). "Theory of vector-valued differential forms: Part I. Derivations in the graded ring of differential forms". In: *Indagationes Mathematicae (Proceedings)* 59, pp. 338–359.

Gallissot, F. (1952). "Les formes extérieures en mécanique". In: *Annales de l'institut Fourier* 4, pp. 145–297.

Gallot, S., D. Hulin, and J. Lafontaine (1990). *Riemannian Geometry*. 2nd ed. Universitext. Springer.

Glocker, Ch. (2001). *Set-Valued Force Laws*. Vol. 1. Lecture Notes in Applied Mechanics. Springer.

Godbillon, C. (1969). *Géométrie différentielle et mécanique analytique*. Collection Méthodes. Hermann.

Golubitsky, M. and V. Guillemin (1973). *Stable Mappings and Their Singularities*. Graduate Texts in Mathematics 14. Springer.

Hall, B. (2015). *Lie Groups, Lie Algebras, and Representations*. Vol. 222. Graduate Texts in Mathematics. Springer.

Hamel, G. (1904a). "Die Lagrange-Eulerschen Gleichungen der Mechanik". In: *Zeitschrift für Mathematik und Physik* 50.1–2, pp. 1–57.

– (1904b). "Über die virtuellen Verschiebungen in der Mechanik". In: *Mathematische Annalen* 59.3, pp. 416–434.

– (1949). *Theoretische Mechanik*. Vol. 57. Grundlehren der mathematischen Wissenschaften. Springer.

Hermann, R. (1988). "Differential form methods in the theory of variational systems and Lagrangian field theories". In: *Acta Applicandae Mathematicae* 12, pp. 35–78.

Hoffman, K. and R. Kunze (1971). *Linear Algebra*. 2nd ed. Prentice-Hall.

Hornfeck, B. (1969). *Algebra*. De Gruyter Lehrbuch. De Gruyter.

Jost, J. (2008). *Riemannian Geometry and Geometric Analysis*. 5th ed. Universitext. Springer.

Klein, J. (1962). "Espaces variationnels et mécanique". In: *Annales de l'institut Fourier* 12, pp. 1–124.

– (1963). "Les systèmes dynamiques abstraits". In: *Annales de l'institut Fourier* 13.2, pp. 191–202.

Kühnel, W. (2011). *Matrizen und Lie-Gruppen*. 1st ed. Vieweg+Teubner.

Lagrange, J.-L. (1780). "Théorie de la libration de la lune". In: *Nouveaux Mémoires de l'Académie royale des Sciences et Belles-Lettres de Berlin*.

– (1788). *Méchanique Analitique*. Veuve Desaint.

Lanczos, Cornelius (1952). *The Variational Principles of Mechanics*. Vol. 4. Mathematical Expositions. University of Toronto Press.

Landau, L. D. and E. Lifshitz (1971). *The Classical Theory of Fields*. 3rd ed. Vol. 2. Course of Theoretical Physics. Pergamon Press.

Landau, L. D. and E. M. Lifshitz (1969). *Mechanics*. 2nd ed. Vol. 1. Course of Theoretical Physics. Pergamon Press.

Lang, S. (2001). *Fundamentals of Differential Geometry*. Vol. 191. Graduate Texts in Mathematics. Springer.

– (2004). *Linear Algebra*. 3rd ed. Undergraduate Texts in Mathematics. Springer.

– (2005). *Undergraduate Algebra*. 3rd ed. Undergraduate Texts in Mathematics. Springer.

Lee, Jeffrey M. (2009). *Manifolds and Differential Geometry*. Vol. 107. Graduate Studies in Mathematics. American Mathematical Society.

Lee, John M. (2013). *Introduction to Smooth Manifolds*. 2nd ed. Vol. 218. Graduate Texts in Mathematics. Springer.

Libermann, P. and C.-M. Marle (1987). *Symplectic Geometry and Analytical Mechanics*. Vol. 35. Mathematics and Its Applications. Springer.

Lichnerowicz, A. (1945). "Les espaces variationnels généralisés". In: *Annales scientifiques de l'École Normale Supérieure* 62, pp. 339–384.

Loos, O. (1982). *Analytische Mechanik*. Seminarausarbeitung, Institut für Mathematik, Universität Innsbruck.

– (1985). "Automorphism groups of classical mechanical systems". In: *Monatshefte für Mathematik* 100.4, pp. 277–292.

Marsden, J. E. and T. S. Ratiu (1999). *Introduction to Mechanics and Symmetry*. 2nd ed. Vol. 17. Texts in Applied Mathematics. Springer.

Morandi, G. et al. (1990). "The inverse problem in the calculus of variations and the geometry of the tangent bundle". In: *Physics Reports* 188.3–4, pp. 147–284.

Munkres, J. R. (2000). *Topology*. 2nd ed. Prentice Hall.

Neĭmark, J. I. and N. A. Fufaev (1972). *Dynamics of Nonholonomic Systems*. Vol. 33. Translations of Mathematical Monographs. American Mathematical Society.

Newton, I. (1729). *The Mathematical Principles of Natural Philosophy*. Printed for Benjamin Motte.

Oliva, W. M. (2002). *Geometric Mechanics*. Vol. 1798. Lecture Notes in Mathematics. Springer.

Papastavridis, J. G. (2014). *Analytical Mechanics*. World Scientific.

Päsler, M. (1968). *Prinzipe der Mechanik*. De Gruyter.

Popper, K. (1935). *Logik der Forschung*. Vol. 9. Schriften zur wissenschaftlichen Weltauffassung. Springer.

Roberson, R. E. and R. Schwertassek (1988). *Dynamics of Multibody Systems*. Springer.

Roman, S. (2008). *Advanced Linear Algebra*. 3rd ed. Vol. 135. Graduate Texts in Mathematics. Springer.

Scheck, F. (2007). *Theoretische Physik 1: Mechanik*. 8th ed. Springer-Lehrbuch. Springer.

Souriau, J.-M. (1970). *Structure des Systèmes Dynamiques*. Dunod.

– (1997). *Structure of Dynamical Systems*. Vol. 149. Progress in Mathematics. Springer.

Spivak, M. (1999a). *A Comprehensive Introduction to Differential Geometry*. 3rd ed. Vol. 1. Publish or Perish.

– (1999b). *A Comprehensive Introduction to Differential Geometry*. 3rd ed. Vol. 2. Publish or Perish.

– (1999c). *A Comprehensive Introduction to Differential Geometry*. 3rd ed. Vol. 3. Publish or Perish.

– (1999d). *A Comprehensive Introduction to Differential Geometry*. 3rd ed. Vol. 4. Publish or Perish.

– (1999e). *A Comprehensive Introduction to Differential Geometry*. 3rd ed. Vol. 5. Publish or Perish.

Talman, R. (2007). *Geometric Mechanics*. 2nd ed. Wiley.

Truesdell, C. (1984). “Suppesian Stews (1980/81)”. In: *An Idiot’s Fugitive Essays on Science*. Springer, pp. 503–579.

Volterra, V. (1898). “Sopra una classe di equazioni dinamiche”. In: *Atti della Reale Accademia delle scienze di Torino* XXXIII, pp. 451–475.

Voronets, P. V. (1901). “Sur les équations du mouvement pour les systèmes non holonômes”. In: *Matematicheskii Sbornik* 22.4, pp. 659–686.

Winandy, T., G. Capobianco, and S. R. Eugster (2018). “A geometric view on the kinematics of finite-dimensional mechanical systems”. In: *Proceedings in Applied Mathematics and Mechanics* 18.1, pp. 1–2.

Yano, K. and S. Ishihara (1973). *Tangent and Cotangent Bundles*. Vol. 16. Pure and Applied Mathematics. Marcel Dekker.

This work was typeset using LuaTeX. The text font is TeX Gyre Schola 10pt. As mathematical font the related font TeX Gyre Schola Math is used. Because the derivative dots of the TeX Gyre Schola Math font are very fine, the original font was modified using FontForge. Additionally some kerning needed to be done, especially in typesetting mathematical formulas. The figures are produced with CorelDRAW for which the author programmed a plug-in to create editable LuaTeX-formulas.

The readability of a monograph strongly depends on its format and typesetting. At many European universities dissertations must be submitted in the format DIN A4. Common formats of publication are DIN A5 and 170×240 mm. These two formats correspond to an approximate scaling to 70% and 80% of DIN A4, respectively. Moreover, 170×240 mm does more or less correspond to the size of contemporary monographs by publishers such as Birkhäuser, Springer, or Wiley.

The larger font size that results from a scaling to 170×240 mm increases the readability of a work compared to the format DIN A5. Knowing that the format 170×240 mm is an accepted publication format for doctoral theses at renowned German universities such as the Karlsruhe Institute of Technology and the University of Siegen, the author officially asked to publish his work in 170×240 mm instead of DIN A5. The request was rejected by the PhD board (Promotionsausschuss) of the Faculty 7 on the (only) basis that the existing bookshelves are adapted to DIN A5. The reader may excuse the comment that this argument appears somewhat contradictory in times where universities and funding organizations strive for the immediate dissemination of scientific results via all available communication media. In this regard, it makes a difference whether a research monograph is agreeable to read or not. Fortunately, the kind reader may print the digital version of this work in any format he desires.