Lecture notes on
“Geometric Dynamics and Symmetry”

Bob Rink*

May 9, 2012

Contents

*VU University Amsterdam, Department of Mathematics, De Boelelaan 1081a, 1081 HV Amsterdam, e-mail `b.w.rink@vu.nl`

Introduction

Classical mechanics has had an enormous influence on mathematics. The development of differential calculus by Newton for instance was mainly inspired by his desire to understand the motion of the planets. But there are also deep connections between classical mechanics and differential geometry, topology, dynamical systems theory and the calculus of variations, to name a few.

Because mechanics is such a broad and ill-defined topic, I found it hard to decide what should be included in a first course aimed mainly at MSc and graduate students in Mathematics. Of course, I have omitted many parts of the theory. Nevertheless, I have tried to gather a representative mix of topics, ranging from the calculus of variations to bifurcation theory and from symplectic geometry to fluid mechanics. Without claiming to be complete in any way, I have tried to stress a geometric view towards mechanics, that I believe helps understanding its dynamics. I hope that this point of view will inspire the reader to further explore the connections between the seemingly so different topics that constitute the subject of "Geometric dynamics and symmetry".

This is a one-semester course, consisting of two times 45 minutes of lectures per week. Additionally, and perhaps more importantly, students are to prepare exercises from these lecture notes that they are invited to present during a one-hour exercise session following the lectures. The final grade will be decided upon during an oral exam after the course, and will mainly depend on the degree of "active participation and problem solving skills" during these exercise sessions.

1 Mechanical systems

The most important principle in classical mechanics is the property that a mechanical system can be given an arbitrary initial position and velocity, but that these then determine the behaviour of the system completely.

Let us see what this means. Let $q : I \to U, t \mapsto q(t)$ be a C^2 curve in an open subset $U \subset \mathbb{R}^n$, defined for t in an open interval $I \subset \mathbb{R}$. We assume that $q(t)$ describes the position or "configuration" of a mechanical system at time t. The velocities of the mechanical system are then given by the derivatives

$$\dot{q}_j(t) := \frac{dq_j(t)}{dt} \in \mathbb{R}\ ,\ j = 1, \ldots, n\ .$$

The above main principle of classical mechanics then implies that the accelerations at time t,

$$\frac{d\dot{q}_j(t)}{dt} = \frac{d^2q_j(t)}{dt^2}\ ,$$

are determined by the positions and velocities at time t, that is

$$\frac{d\dot{q}_j(t)}{dt} = a_j\left(t, q(t), \dot{q}(t)\right)\ ,$$

for certain functions $a_j : I \times U \times \mathbb{R}^n \to \mathbb{R}$. The exact form of the functions a_j depends on the physical properties of the mechanical system under consideration.

The ordinary differential equations

$$\begin{aligned} \frac{dq_j}{dt} &= \dot{q}_j\ , \\ \frac{d\dot{q}_j}{dt} &= a_j(t, q, \dot{q})\ . \end{aligned} \tag{1.1}$$

are defined on the "phase space" $U \times \mathbb{R}^n$ and are called the "equations of motion" for the mechanical system under consideration. Under the mild condition that the a_j are locally Lipschitz continuous, the existence and uniqueness theory for ordinary differential equations indeed guarantees that $q(t)$ is determined by these equations of motion once the initial position $q(t_0)$ and velocity $\dot{q}(t_0)$ are given.

1.1 Two classical examples

A simple but famous example of a classical mechanical system was found by Galilei. He discovered experimentally that falling objects accelerate constantly towards the earth. In other words, Galilei discovered that falling objects describe a curve $q : t \mapsto q(t) \in \mathbb{R}^3$ that satisfies equations (1.1) with

$$a_1(t, q, \dot{q}) = a_2(t, q, \dot{q}) = 0\ ,\ a_3(t, q, \dot{q}) = -g\ . \tag{1.2}$$

The constant $g \in (0, \infty)$ is called the gravitational constant and it too can be determined experimentally: in Pisa it is approximately equal to $9,80\ m/s^2$. Integrating the equations $\frac{d\dot{q}_1}{dt} = \frac{d\dot{q}_2}{dt} = 0, \frac{d\dot{q}_3}{dt} = -g$ gives that $\dot{q}_1(t) = \dot{q}_1(0), \dot{q}_2(t) = \dot{q}_2(0), \dot{q}_3(t) = \dot{q}_3(0) - gt$. Integrating also $\frac{dq_i}{dt} = \dot{q}_i$, we then find that

$$q_1(t) = q_1(0) + \dot{q}_1(0)t, q_2(t) = q_2(0) + \dot{q}_2(0)t, q_3(t) = q_3(0) + \dot{q}_3(0)t - \frac{1}{2}gt^2 \ .$$

Thus, one can even explicitly solve Galiei's equations of motion.

A more complicated but equally famous mechanical system is the Kepler system that describes the motion of a planet around the sun, given by a curve $t \mapsto q(t) \in \mathbb{R}^2$. Based on observations by the astronomer Tycho Brahe, Kepler formulated the following principles for this motion, now known as *Kepler's laws*:

1. A planet moves on an ellipse in $\mathbb{R}^2$, with the sun in one of its focal points. Call this point $q_s \in \mathbb{R}^2$.
2. The area of the domain bounded by the line segment from q_s to $q(t_0)$, the orbit of the planet and the line segment from q_s to $q(t)$ is proportional to $t - t_0$.
3. The square of the period of a planetary orbit divided by the third power of the length of the long axis of its elliptic orbit is the same for every planet.

Newton proved that Kepler's laws are in fact equivalent to the equations of motion

$$m\frac{d^2q}{dt^2} = -\frac{mMG}{||q - q_s||^3}(q - q_s) \ , \tag{1.3}$$

in which m is the mass of the planet, M is the mass of the sun and G is a universal gravitational constant that is the same for every planet. Having at our disposal the techniques of modern calculus, the proof of Newton's theorem is quite elementary, but we will not present it here. In fact, it is well-known that the solutions to equation (1.3) describe a *conic section*, i.e. a circle, ellipse, parabola or hyperbola.

The right hand side of equation (1.3) is called the *force* acting on the planet, so equation (1.3) illustrates *Newton's first law* that says that the mass of a body times its acceleration is equal to the force acting on the body. Newton's *second law* asserts that the total force acting on a body is equal to the sum of the forces that are acting on it. So for instance, if $t \mapsto q^{(i)}(t) \in \mathbb{R}^2$ $(i = 1, \ldots, N)$ describe the positions of a collection of N planets moving in $\mathbb{R}^2$, each with their own mass m_i, then the equations of motion for these planets are given by

$$m_i\frac{d^2q^{(i)}}{dt^2} = \sum_{j \neq i} \phi_{ji}(q) \ , \tag{1.4}$$

where

$$\phi_{ji}(q) := -\frac{m_i m_j G}{||q^{(i)} - q^{(j)}||^3}(q^{(i)} - q^{(j)}) \tag{1.5}$$

is the force that planet j exerts on planet i. We also see here an illustration of Newton's *third law*, which says that the force that one body exerts on another body is equal to minus the force that this other body exerts on the first body. For the motion of the planets this is expressed by the fact that $\phi_{ji} = -\phi_{ij}$ as is clear from formula (1.5).

1.2 One degree of freedom mechanical systems

Let us study a mechanical system with one degree of freedom. In this case, $t \mapsto q(t), I \to J \subset \mathbb{R}$ describes a curve in an open interval J. Let us assume that it satisfies a second order differential equation of the form

$$m\frac{d^2q}{dt^2} = \phi(q) \ . \tag{1.6}$$

The force $\phi : J \to \mathbb{R}$ is a continuous function, of which we have assumed for simplicity that it depends on the position only, and $m > 0$ is the mass of this mechanical system. Writing again $\dot{q} := \frac{dq}{dt}$, the above second order equation is equivalent to the first order system

$$\begin{aligned} \frac{dq}{dt} &= \dot{q} \ , \\ m\frac{d\dot{q}}{dt} &= \phi(q) \ , \end{aligned} \tag{1.7}$$

and the phase space of this first order system is $J \times \mathbb{R}$. Now let $V : J \to \mathbb{R}$ be a primitive of $-\phi$, i.e. $\frac{dV(q)}{dq} = -\phi(q)$. Then the crucial remark is that the function of position and velocity

$$E(q, \dot{q}) := \frac{1}{2}m\dot{q}^2 + V(q)$$

is a constant of motion, namely if $t \mapsto (q(t), \dot{q}(t))$ is a solution to equations (1.7), then :

$$\frac{d}{dt}E(q(t), \dot{q}(t)) = \frac{\partial E(q, \dot{q})}{\partial q}\dot{q} + \frac{\partial E(q, \dot{q})}{\partial \dot{q}}\frac{\phi(q)}{m} = -\phi(q)\dot{q} + m\dot{q}\frac{\phi(q)}{m} = 0 \ .$$

The existence of this constant of motion is a direct consequence of our assumption that ϕ be a function of q only. Moreover, it is of great help for drawing the phase portrait of (1.7). For instance: we call a point $(q, \dot{q}) \in J \times \mathbb{R}$ a singular point of E if $\frac{\partial E(q,\dot{q})}{\partial q} = -\phi(q) = 0$ and $\frac{\partial E(q,\dot{q})}{\partial \dot{q}} = m\dot{q} = 0$. Then it is clear that the set of singular points of E is equal to the set of equilibrium points of the system of differential equations (1.7) and that all singular points are contained in the line $\dot{q} = 0$.

If $(q_0, \dot{q}_0)$ is not singular and $E(q_0, \dot{q}_0) = h \in \mathbb{R}$, then the level set $E^{-1}(h)$ is a smooth curve, that is a one-dimensional C^1 submanifold of $J \times \mathbb{R}$, near $(q_0, \dot{q}_0)$. If $\dot{q}_0 \neq 0$, it is actually the graph of one of the functions

$$\dot{q}(q) = \pm\sqrt{\frac{2}{m}\left(h - V(q)\right)} \ . \tag{1.8}$$

Because E is a constant of motion, the integral curve of system (1.7) with initial condition $q(0) = q_0, \dot{q}(0) = \dot{q}_0$ is contained in the union of these graphs.

In fact, we can say a little more. Assume that $\dot{q}_0 > 0$, i.e. choose the +-sign in equation (1.8). The situation $\dot{q}_0 < 0$ is similar. Then using that for solutions to (1.7) the equality $\dot{q} = \frac{dq}{dt}$ holds, equation (1.8) tells us that the unique solution $t \mapsto q(t)$ of equation (1.6) with initial conditions $q(t_0) = q_0$, $\left.\frac{dq(t)}{dt}\right|_{t=t_0} = \dot{q}_0$, actually satisfies the *first order* differential equation

$$\frac{dq}{dt} = \sqrt{\frac{2}{m}\left(h - V(q)\right)} \ .$$

Now if on the interval $[q_1, q_2]$ containing q_0, the right hand side of this equation is nonzero, and hence positive, then q is a strictly increasing function of t on this interval and by the inverse-function theorem, t is a strictly increasing function of q, with $t(q_0) = t_0$ and $\frac{dt(q)}{dq} = 1/\left.\frac{dq(t)}{dt}\right|_{t=t(q)}$. Integration of this equality then gives that for $q \in [q_1, q_2]$,

$$t(q) - t_0 = \int_{q_0}^{q} \frac{dq}{\sqrt{\frac{2}{m}\left(h - V(q)\right)}} \ .$$

This formula tells us how much time it takes to reach a $q \in [q_1, q_2]$ from q_0. We may be able to solve it for $q = q(t)$ to find the solution explicitly.

Remark 1.1 (Linear friction) Including in equation (1.6) a linear friction term leads to the equation

$$m\frac{d^2q}{dt^2} = \phi(q) - c\frac{dq}{dt} \ ,$$

where $c > 0$ is called the friction constant. For solutions to this equation one computes that

$$\frac{d}{dt}E\left(q(t), \frac{dq(t)}{dt}\right) = -c\left(\frac{dq(t)}{dt}\right)^2 \ ,$$

which means that E decreases along solutions of the differential equation, unless $q(t)$ is constant. We say that in this case, E is a *Lyapunov function.*

1.3 More degrees of freedom

A mechanical system with several degrees of freedom is defined by Newton's system of second order ordinary differential equations

$$m_i\frac{d^2q_i}{dt^2} = \phi_i(q) \ , \ i = 1, \ldots, n \ . \tag{1.9}$$

Now $q \in U \subset \mathbb{R}^n$ is an element of an open subset of $\mathbb{R}^n$ and the continuous functions $\phi_i : U \to \mathbb{R}$ are the components of the force acting on q. Again, we assumed that they only depend on q. Defining again the velocities $\dot{q}_i := \frac{dq_i}{dt}$, this system of n second order equations is equivalent to the system of $2n$ first order equations on $U \times \mathbb{R}^n$

$$\begin{aligned} \frac{dq_i}{dt} &= \dot{q}_i \\ m_i \frac{d\dot{q}_i}{dt} &= \phi_i(q) \ , \ i = 1, \ldots, n \ . \end{aligned} \tag{1.10}$$

Inspired by the previous section, we may want to assume that there exists a C^1 function $V : U \to \mathbb{R}$ with the property that

$$\phi_i(q) = -\frac{\partial V(q)}{\partial q_i} \ .$$

If this is the case, then we call V the potential of the force ϕ and ϕ is called a conservative force. In general, the requirement that a force be conservative is very restrictive. For instance, if the ϕ_i are C^1, then V is C^2 and because $\frac{\partial^2 V}{\partial q_i \partial q_j} = \frac{\partial^2 V}{\partial q_j \partial q_i}$, the requirement that $\phi_i = -\frac{\partial V}{\partial q_i}$ for all i leads to the conclusion that

$$\frac{\partial \phi_i}{\partial q_j} = \frac{\partial \phi_j}{\partial q_i} \ , \ i, j = 1, \ldots, n \ . \tag{1.11}$$

It turns out the condition (1.11) is sufficient for the existence of an open subset U_{q_0} near each $q_0 \in U$ and a C^2 function $V_{q_0} : U_{q_0} \to \mathbb{R}$ such that $\phi_i = -\frac{\partial V_{q_0}}{\partial q_i}$ on U_{q_0}. The existence of such a function on the entire U can only be guaranteed under strong topological conditions on U, for instance that it be starshaped or, more generally, simply connected. In fact, the very question when a force is conservative was the motive for Poincaré to introduce the subject of topology.

If ϕ is conservative, then the function of positions and velocities

$$E(q, \dot{q}) := \sum_{i=1}^{n} \frac{1}{2} m_i \dot{q}_i^2 + V(q)$$

is again a constant of motion for the system (1.10), because along solutions

$$\sum_{i=1}^{n} \left(\frac{\partial E(q, \dot{q})}{\partial q_i} \dot{q}_i + \frac{\partial E(q, \dot{q})}{\partial \dot{q}_i} \frac{\phi_i(q)}{m_i} \right) = \sum_{i=1}^{n} \left(-\phi_i(q) \dot{q}_i + m_i \dot{q}_i \frac{\phi_i(q)}{m_i} \right) = 0 \ .$$

The singular points of E correspond to the equilibrium points of (1.10). At regular points, the level set of E is a submanifold of $U \times \mathbb{R}^n$ of dimension $2n - 1$.

Remark 1.2 (Energy) The function $V = V(q)$ is called the *potential energy* of a conservative mechanical system. The function

$$T(\dot{q}) := \sum_{i=1}^{n} \frac{1}{2} m_i \dot{q}_i^2 \tag{1.12}$$

is called *kinetic energy* and the function $E(q, \dot{q}) = T(\dot{q}) + V(q)$ is called the *total energy* of the mechanical system. The concept of conservation of total energy is of course quite fundamental in classical physics.

1.4 Exercises

Exercise 1.1 (Galilei's laws with friction) *If we include in Galilei's model for falling objects a friction force $\phi(q, \dot{q}) = -c\dot{q}$ that acts in the direction of minus the velocity $\dot{q}$ and is proportional to $||\dot{q}||$, then his equations become*

$$\frac{dq_1}{dt} = \dot{q}_1 \ , \frac{dq_2}{dt} = \dot{q}_2, \frac{dq_3}{dt} = \dot{q}_3, m\frac{d\dot{q}_1}{dt} = -c\dot{q}_1, m\frac{d\dot{q}_2}{dt} = -c\dot{q}_2, m\frac{d\dot{q}_3}{dt} = -mg - c\dot{q}_3 \ .$$

$c > 0$ *is called the friction constant. Solve these equations and show that* $\lim_{t\to\infty}(\dot{q}_1, \dot{q}_2, \dot{q}_3) = (0, 0, -\frac{mg}{c})$*. Hence we observe that in the presence of friction, a falling object does not accelerate without bound, but instead approaches a limiting speed.*

Exercise 1.2 (Charged particle) *The motion of a charged test particle in an electric and magnetic field is given by a curve $q(t) \in \mathbb{R}^3$ satisfying*

$$m\frac{d^2q}{dt^2} = e\left(E + \frac{dq}{dt} \times B\right) \ ,$$

where $E = E(q, t) \in \mathbb{R}^3$ and $B = B(q, t) \in \mathbb{R}^3$ are vector functions, called the electric and magnetic field respectively and $e \in \mathbb{R}$ is called the charge of the particle. Show that $\langle \frac{d^2q}{dt^2}, B\rangle = 0$ if $E = 0$. Under the assumption that $E = 0$ and $B = (b, 0, 0) \in \mathbb{R}^3$ is constant, solve the equations of motion.

Exercise 1.3 *Show that Galilei's equations (1.2) are conservative. Give T, V and E. Apart from E, can you point out other constants of motion?*

Exercise 1.4 *According to Newton's laws, the motion of a system of N planets is given by*

$$m_i\frac{d^2q^{(i)}}{dt^2} = \sum_{j\neq i} \phi_{ji}(q) \ ,$$

where ϕ_{ji} is given in equation (1.5). Let $M := \sum_{i=1}^N m_i$ be the total mass of the planets and define the center of mass as

$$z(t) := \frac{1}{M}\sum_{i=1}^N m_i q^{(i)}(t) \ .$$

Prove that $z(t)$ is a linear function of t. Show that the components of

$$\mu := \frac{dz}{dt} = \frac{1}{M}\sum_{i=1}^N m_i \dot{q}^{(i)}$$

are constants of motion.

Exercise 1.5 (Harmonic oscillator) *A one-dimensional harmonic oscillator is described by the one degree of freedom mechanical system*

$$m\frac{d^2q}{dt^2} = \phi(q) \ ,$$

with a force given by Hooke's law: $\phi(q) = -kq, \mathbb{R} \to \mathbb{R}$. *The constant* $k > 0$ *is sometimes called the spring constant of the oscillator. Give the constant of motion for the harmonic oscillator, compute the equilibrium points and draw the phase portrait. Can you also solve the equations of motion explicitly?*

Exercise 1.6 *For the equations* $\frac{d^2q}{dt^2} = -\sin q$ *(mathematical pendulum) and* $\frac{d^2q}{dt^2} = q^3 - q$ *(Duffing oscillator), give the constant of motion, compute the equilibrium points and draw the phase portrait.*

Exercise 1.7 *If we define, for* $i = 1, \ldots, N$ *and* $k = 1, 2$, *the total force acting on the* k*-th coordinate of planet* i, *by* $\phi_k^{(i)}(q) = \sum_{j \neq i} (\phi_{ji}(q))_k$, *then the equations of motion (1.4) can be written as*

$$m_i \frac{d^2 q_k^{(i)}}{dt^2} = \phi_k^{(i)}(q) \ .$$

Show that

$$\phi_k^{(i)}(q) = -\frac{\partial V(q)}{\partial q_k^{(i)}} \ ,$$

with

$$V(q) := -G \sum_{i<j} \frac{m_i m_j}{||q^{(i)} - q^{(j)}||} \ .$$

This shows that the forces acting on the planets are conservative. Give the constant of motion.

2 Lagrangian mechanics

Lagrange showed that Newton's equations (1.9) are defined in a coordinate-invariant way. More precisely, he formulated Newton's equation in such a way that they behave well under coordinate transformations. Lagrange's construction will be the main topic of this section.

2.1 New position variables

One may wonder what happens to the equations of motion (1.9) for $q(t) \in U \subset \mathbb{R}^n$ if we make an arbitrary change of the position variables, that is if assume that $q(t) = \Phi(t, Q(t))$ for $Q(t) \in \tilde{U} \subset \mathbb{R}^m$. The reason we ask this question is that, once we know how an arbitrary Φ changes the equations of motion, we might be able to make a clever choice of Φ that changes the equations (1.9) for $q(t)$ into much simpler equations for $Q(t)$.

If Φ is C^2, then a twice differentiable curve $t \mapsto Q(t)$ in $\tilde{U} \subset \mathbb{R}^m$ defined on some open time-interval $I \subset \mathbb{R}$ is transformed by Φ into a twice differentiable curve

$$q(t) := \Phi(t, Q(t))$$

in $U \subset \mathbb{R}^n$. Differentiation of $q(t)$ gives that

$$\frac{dq_i(t)}{dt} = \frac{\partial \Phi_i(t, Q(t))}{\partial t} + \sum_{j=1}^{m} \frac{\partial \Phi_i(t, Q(t))}{\partial Q_j} \frac{dQ_j(t)}{dt} \ ,$$

i.e. $\frac{dq}{dt} = \frac{\partial \Phi}{\partial t} + \frac{\partial \Phi}{\partial Q}\frac{dQ}{dt}$, which is a nice *affine* transformation formula. Differentiating this once more we nevertheless find that even if Φ does not depend explicitly on t, the second order derivative $\frac{d^2 q}{dt^2}$ in general is not simply the image of $\frac{d^2 Q}{dt^2}$ under the linear map $\frac{\partial \Phi}{\partial Q}$.

Lagrange realised that one can take another approach, which might at first seem a little artificial. The first thing he remarked is that

$$m_i \frac{d^2 q_i(t)}{dt^2} = \frac{d}{dt}\left(\left. \frac{\partial T(\dot{q})}{\partial \dot{q}_i} \right|_{\dot{q} = \frac{dq(t)}{dt}} \right) \ , \tag{2.1}$$

where T is the kinetic energy defined in (1.12). The next remark is that one can express the kinetic energy $T\left(\frac{dq(t)}{dt}\right)$ explicitly as a function of t, $Q(t)$ and $\frac{dQ(t)}{dt}$, by defining the function $\tilde{T} : I \times \tilde{U} \times \mathbb{R}^m \to \mathbb{R}$ by

$$\tilde{T}(t, Q, \dot{Q}) = T(\dot{q}) \ ,$$

in which

$$\dot{q} = \frac{\partial \Phi(t, Q)}{\partial t} + \frac{\partial \Phi(t, Q)}{\partial Q} \cdot \dot{Q} \ .$$

In other words, $\tilde{T}$ is defined by $\tilde{T}(t, Q, \dot{Q}) = T\left(\frac{\partial \Phi(t,Q)}{\partial t} + \frac{\partial \Phi(t,Q)}{\partial Q} \cdot \dot{Q}\right)$. This definition is such that $T\left(\frac{dq(t)}{dt}\right) = \tilde{T}\left(t, Q(t), \frac{dQ(t)}{dt}\right)$ if $q(t) = \Phi(t, Q(t))$. Note that the transformed kinetic energy $\tilde{T}(t,Q,\dot{Q})$ in general will depend explicitly on the time t and the new position variable Q. Now Lagrange's discovery was that not $\frac{d}{dt}\left(\frac{\partial \tilde{T}(t,Q,\dot{Q})}{\partial \dot{Q}_j}\Big|_{Q=Q(t),\dot{Q}=\frac{dQ(t)}{dt}}\right)$, but the quantity $\frac{d}{dt}\left(\frac{\partial \tilde{T}(t,Q,\dot{Q})}{\partial \dot{Q}_j}\Big|_{Q=Q(t),\dot{Q}=\frac{dQ(t)}{dt}}\right) - \frac{\partial \tilde{T}(t,Q,\dot{Q})}{\partial Q_j}\Big|_{Q=Q(t),\dot{Q}=\frac{dQ(t)}{dt}}$ depends linearly on $\frac{d}{dt}\left(\frac{\partial T(\dot{q})}{\partial \dot{q}_i}\Big|_{\dot{q}=\frac{dq(t)}{dt}}\right)$. In fact, we have the following quite general result:

Theorem 2.1 *Let $\Phi : I \times \tilde{U} \to U$ be a C^2 map, $t \mapsto Q(t)$ a C^2 curve in $\tilde{U}$ and $q(t) = \Phi(t, Q(t))$ the corresponding curve in U. Let $L : I \times U \times \mathbb{R}^n \to \mathbb{R}$ be a C^2 function of $(t, q, \dot{q})$ and let $\tilde{L} : I \times \tilde{U} \times \mathbb{R}^m \to \mathbb{R}$ be the corresponding function of $(t, Q, \dot{Q})$ defined by*

$$\tilde{L}(t,Q,\dot{Q}) = L\left(t, \Phi(t,Q), \frac{\partial \Phi(t,Q)}{\partial t} + \frac{\partial \Phi(t,Q)}{\partial Q}\cdot \dot{Q}\right) . \tag{2.2}$$

Furthermore, define the continuous functions $[L]_i^q$ on I by

$$[L]_i^q(t) := \frac{d}{dt}\left(\frac{\partial L(t,q,\dot{q})}{\partial \dot{q}_i}\Big|_{\dot{q}=\frac{dq(t)}{dt}}\right) - \frac{\partial L(t,q,\dot{q})}{\partial q_i}\Big|_{\dot{q}=\frac{dq(t)}{dt}} , \ i = 1,\ldots,n , \tag{2.3}$$

and similarly $[\tilde{L}]_j^Q$:

$$[\tilde{L}]_j^Q(t) := \frac{d}{dt}\left(\frac{\partial \tilde{L}(t,Q,\dot{Q})}{\partial \dot{Q}_j}\Bigg|_{\dot{Q}=\frac{dQ(t)}{dt}}\right) - \frac{\partial \tilde{L}(t,Q,\dot{Q})}{\partial Q_j}\Bigg|_{\dot{Q}=\frac{dQ(t)}{dt}} , \ j = 1,\ldots,m . \tag{2.4}$$

Then we have the following transformation formula:

$$[\tilde{L}]_j^Q(t) = \sum_{i=1}^{n} [L]_i^q(t) \cdot \frac{\partial \Phi_i(t,Q)}{\partial Q_j}\Big|_{Q=Q(t)} , \ j = 1,\ldots,m . \tag{2.5}$$

One can prove this theorem by a very long direct computation, using the relations $q = \Phi(t,Q)$ and $\frac{dQ}{dt} = \frac{\partial \Phi(t,Q)}{\partial t} + \frac{\partial \Phi(t,Q)}{\partial Q}\frac{dq}{dt}$ and differentiating the identity (2.2) with respect to Q_j and $\dot{Q}_j$. But one can also prove Theorem 2.1 in a surprisingly different way, as we shall see in the next section.

The conclusion of Theorem 2.1 is that $[\tilde{L}]^Q$ depends linearly on $[L]^q$, namely $[\tilde{L}]^Q = [L]^q \cdot \frac{\partial \Phi}{\partial Q}$, viewing $[L]^q(t)$ and $[\tilde{L}]^Q(t)$ as row vectors. Note that this is *not* the same as the transformation formula $\frac{dq}{dt} = \frac{\partial \Phi}{\partial t} + \frac{\partial \Phi}{\partial Q}\cdot\frac{dQ}{dt}$ for the velocity vectors, even if Φ is independent of t. The classical terminology is that $[L]^q$ transforms *covariantly* and $\frac{dq}{dt}$ transforms *contravariantly.*

2.2 A variational proof of Theorem 2.1

This section contains a "variational" proof of Theorem 2.1. It will become clear what this means. I myself think that this proof is quite surprising.

We start with letting $q : I \to U \subset \mathbb{R}^n$ be a C^2 curve in U and $L : I \times U \times \mathbb{R}^n \to \mathbb{R}$ a C^2 Lagrangian function. Then we can integrate L over compact pieces of the curve $t \mapsto q(t)$. Hence we fix some $a, b \in I, a < b$ and define the *action* of L along q restricted to $[a, b]$ as

$$A(q) := \int_a^b L\left(t, q(t), \frac{dq(t)}{dt}\right) dt \ .$$

The next step is to consider small perturbations of the curve $t \mapsto q(t)$, depending on an auxiliary parameter ε. That is we consider C^2 maps

$$\tilde{q} : I \times (-\varepsilon_0, \varepsilon_0) \to U \ ,$$

with the property that $\tilde{q}(t, 0) = q(t)$. For an $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$ close to 0, the curve $q_\varepsilon : t \mapsto \tilde{q}(t, \varepsilon)$ lies "close" to the curve $t \mapsto q(t)$, and hence such a map $\tilde{q}$ is called a *variation* of the curve $t \mapsto q(t)$. Now the action can be viewed as a function $\varepsilon \mapsto A(q_\varepsilon)$ on $(-\varepsilon_0, \varepsilon_0)$. Due to the theorem for interchanging differentiation and integration, this function is itself differentiable and differentiation gives:

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} A(q_\varepsilon) = \int_a^b \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} L\left(t, q_\varepsilon(t), \frac{dq_\varepsilon(t)}{dt}\right) dt =$$
$$\int_a^b \sum_{i=1}^n \left(\left.\frac{\partial L(t,q,\dot{q})}{\partial q_i}\right|_{\substack{q\,=\,q(t)\\ \dot{q}\,=\,\frac{dq(t)}{dt}}} \left.\frac{\partial \tilde{q}_i(t,\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} + \left.\frac{\partial L(t,q,\dot{q})}{\partial \dot{q}_i}\right|_{\substack{q\,=\,q(t)\\ \dot{q}\,=\,\frac{dq(t)}{dt}}} \left.\frac{\partial^2 \tilde{q}_i(t,\varepsilon)}{\partial t \partial \varepsilon}\right|_{\varepsilon=0} \right) dt \ .$$

Partial integration with respect to t of the second term in this expression now gives that this is equal to

$$-\int_a^b \sum_{i=1}^n [L]_i^q(t) \left.\frac{\partial \tilde{q}_i(t,\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} dt + \text{ “boundary terms” } .$$

If we consider only variations $\tilde{q}$ of q with fixed endpoints, that is $\tilde{q}(a, \varepsilon) = \tilde{q}(a, 0) = q(a)$ and $\tilde{q}(b, \varepsilon) = \tilde{q}(b, 0) = q(b)$, then $\frac{\partial \tilde{q}(a,\varepsilon)}{\partial \varepsilon} = \frac{\partial \tilde{q}(b,\varepsilon)}{\partial \varepsilon} = 0$, whence the boundary terms disappear. From now on we will assume that our variations have fixed endpoints.

If $q(t) = \Phi(t, Q(t))$ with $\Phi : I \times \tilde{U} \to U$, and $\tilde{Q} : I \times (-\varepsilon_0, \varepsilon_0) \to \tilde{U}$ is a variation of $Q(t)$ with fixed endpoints, consisting of curves $Q_\varepsilon(t) := \tilde{Q}(t, \varepsilon)$, then $\tilde{q}(t, \varepsilon) := \Phi(t, \tilde{Q}(t, \varepsilon))$ is a variation of $t \mapsto q(t)$ with fixed endpoints. The definition (2.2) of $\tilde{L}$ implies that $\tilde{L}\left(t, Q_\varepsilon(t), \frac{dQ_\varepsilon(t)}{dt}\right) = L\left(t, q_\varepsilon(t), \frac{dq_\varepsilon(t)}{dt}\right)$, so that

$$A(q_\varepsilon) = \tilde{A}(Q_\varepsilon) := \int_a^b \tilde{L}\left(t, Q_\varepsilon(t), \frac{dQ_\varepsilon(t)}{dt}\right) dt \ .$$

Thus, differentiation of the identity $A(q_\varepsilon) = \tilde{A}(Q_\varepsilon)$ with respect to ε at $\varepsilon = 0$ gives that

$$\int_a^b \sum_{i=1}^n [L]_i^q(t) \left.\frac{\partial \tilde{q}_i(t,\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} dt = \int_a^b \sum_{j=1}^m [\tilde{L}]_j^Q(t) \left.\frac{\partial \tilde{Q}_j(t,\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} dt \ . \tag{2.6}$$

On the other hand, differentiation of $\tilde{q}_i(t,\varepsilon) = \Phi_i(t, \tilde{Q}(t,\varepsilon))$ gives that

$$\left.\frac{\partial \tilde{q}_i(t,\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} = \sum_{j=1}^m \left.\frac{\partial \Phi_i(t,Q)}{\partial Q_j}\right|_{Q=Q(t)} \left.\frac{\partial \tilde{Q}_j(t,\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} ,$$

and we conclude that

$$\int_a^b \sum_{j=1}^m \left([\tilde{L}]_j^Q(t) - \sum_{i=1}^n [L]_i^q(t) \left.\frac{\partial \Phi_i(t,Q)}{\partial Q_j}\right|_{Q=Q(t)} \right) \left.\frac{\partial \tilde{Q}_j(t,\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0} dt = 0 \ , \tag{2.7}$$

for every C^2 variation $\tilde{Q}$ of Q.

Finally, suppose that $[\tilde{L}]_j^Q(t) \neq \sum_{i=1}^n [L]_i^q(t) \left.\frac{\partial \Phi_i(t,Q)}{\partial Q_j}\right|_{Q=Q(t)}$ for some $1 \leq j \leq m$ and at some $t = t^*$ with $a < t^* < b$. Then because of continuity, inequality holds in an interval $[t^* - \delta, t^* + \delta]$. If we now choose a variation $\tilde{Q}$ of Q with $\tilde{Q}_k(t,\varepsilon) = Q_k(t)$ for $k \neq j$ and $\tilde{Q}_j(t,\varepsilon) = Q_j(t) + \varepsilon\chi(t)$, with χ some nonzero C^2 function of fixed sign and with compact support in $[t^* - \delta, t^* + \delta]$, then for this variation, formula (2.7) is not true: a contradiction. This proves Theorem 2.1.

2.3 Euler-Lagrange equations

For a given C^2 function $L : I \times U \times \mathbb{R}^n \to \mathbb{R}$, the equations of motion

$$[L]^q = 0$$

for the curve $t \mapsto q(t)$ in U are called the *Euler-Lagrange equations* for L and L is called the *Lagrangian function* for the equations $[L]^q = 0$.

A particular consequence of Theorem 2.1 is that if $t \mapsto q(t)$ solves the Euler-Lagrange equations for L and $q(t) = \Phi(t, Q(t))$, then $t \mapsto Q(t)$ automatically solves the Euler-Lagrange equations for $\tilde{L}$. If for all $t \in I$, $\Phi(t, \cdot)$ is a diffeomorphism, that is if $\frac{\partial \Phi}{\partial Q}$ is invertible, then the reverse statement is also true. This expresses that Euler-Lagrange equations for $q(t)$ are defined coordinate-invariantly, which is hardly surprising because they are equivalent to a *variational principle.*

More precisely, the proof of Theorem 2.1 shows that $[L]^q = 0$ if and only if

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} A(\tilde{q}(\cdot,\varepsilon)) = \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \int_a^b L\left(t, \tilde{q}(t,\varepsilon), \frac{\partial \tilde{q}(t,\varepsilon)}{\partial t}\right) dt = 0$$

for all $a, b \in I$ and all C^2 maps $\tilde{q} : I \times (-\varepsilon_0, \varepsilon_0) \to U$ with the property that

1. $\tilde{q}$ is a variation of q, i.e. $\tilde{q}(t,0) = q(t)$.

2. $\tilde{q}$ has fixed endpoints, i.e. $\tilde{q}(a,\varepsilon) = q(a)$ and $\tilde{q}(b,\varepsilon) = q(b)$.

In short, we say that *the action A is stationary at the curve $t \mapsto q(t)$ for variations with fixed endpoints*. Some authors express this by writing

$$\delta A(q) = \delta \int_a^b L\left(t, q(t), \frac{dq(t)}{dt}\right) dt = 0 \ ,$$

thus expressing that the "derivative" of A at the curve $t \mapsto q(t)$ is zero. We will not use this notation.

The statement that $t \mapsto q(t)$ is a solution of the equations $[L]^q = 0$ if and only if it is stationary for $A(q) = \int_a^b L(t, q(t), \frac{dq(t)}{dt})dt$ for variations with fixed endpoints is known as *Hamilton's principle.*

Let us study the Euler-Lagrange equations for L in some more detail now. Written out explicitly, they read $\frac{dq_i}{dt} = \dot{q}_i$ and

$$\frac{\partial^2 L(t,q,\dot{q})}{\partial t \partial \dot{q}_i} + \sum_{j=1}^n \left(\frac{\partial^2 L(t,q,\dot{q})}{\partial q_j \partial \dot{q}_i} \dot{q}_j + \frac{\partial^2 L(t,q,\dot{q})}{\partial \dot{q}_j \partial \dot{q}_i} \frac{d\dot{q}_j}{dt} \right) - \frac{\partial L(t,q,\dot{q})}{\partial q_i} = 0 \ , \ i = 1, \ldots, n \ . \tag{2.8}$$

If the second order derivative matrix $\left.\frac{\partial^2 L(t,q,\dot{q})}{\partial \dot{q}^2}\right|_{\substack{q \,=\, q(t) \\ \dot{q} \,=\, \frac{dq(t)}{dt}}}$ is invertible, then we call L a *non-degenerate Lagrangian* at $(t, q, \dot{q})$. Nondegeneracy implies that near $(t, q, \dot{q})$ we can rewrite the Euler-Lagrange equations explicitly as a system $\frac{dq_i}{dt} = \dot{q}_i$, $\frac{d\dot{q}_i}{dt} = \tilde{\phi}_i(t, q, \dot{q})$, and in particular we then have local existence and uniqueness of the solutions to the Euler-Lagrange equations. The requirement that $\frac{\partial^2 L}{\partial \dot{q}^2}$ is invertible is called the *Legendre condition.*

The final remark in this section is that if, for a C^2 Lagrangian function $L = L(t, q, \dot{q})$, we define the function $h = \frac{\partial L}{\partial \dot{q}} \cdot \dot{q} - L$, that is

$$h(t,q,\dot{q}) := \sum_{i=1}^n \frac{\partial L(t,q,\dot{q})}{\partial \dot{q}_i} \dot{q}_i - L(t,q,\dot{q}) \ , \tag{2.9}$$

then it is easy to compute that

$$\frac{d}{dt} h\left(t, q(t), \frac{dq(t)}{dt}\right) = - \left.\frac{\partial L(t,q,\dot{q})}{\partial t}\right|_{\substack{q \,=\, q(t) \\ \dot{q} \,=\, \frac{dq(t)}{dt}}} + \sum_{i=1}^n [L]_i^q(t) \frac{dq_i(t)}{dt} \ . \tag{2.10}$$

Corollary 2.2 *If $t \mapsto q(t)$ is a solution to the Euler-Lagrange equations $[L]^q = 0$ and L does not explicitly depend on time, then $h := \frac{\partial L}{\partial \dot{q}} \cdot \dot{q} - L$ is constant along $t \mapsto q(t)$.*

2.4 Natural mechanical systems

Let us now return to Newton's equations of motion in conservative form,

$$m_i \frac{d^2 q_i}{dt} = -\frac{\partial V(q)}{\partial q_i} \ , \tag{2.11}$$

for which we defined the kinetic energy $T(\dot q) = \sum_{i=1}^n \frac{1}{2} m_i \dot q_i^2$, the potential energy $V(q)$ and the total energy $E = T + V$. The interesting remark is that if we also define the *Lagrangian* function

$$L : U \times \mathbb{R}^n \ , \ L(q, \dot q) = T(\dot q) - V(q) \ ,$$

of kinetic energy *minus* potential energy, then

$$[L]_i^q(t) = m_i \frac{d^2 q_i(t)}{dt^2} + \left.\frac{\partial V(q)}{\partial q_i}\right|_{q=q(t)} \ .$$

In other words, $t \mapsto q(t)$ solves Newton's equations of motion (2.11) if and only if $[T - V]^q = 0$.

Moreover, if $L(q, \dot q) = T(\dot q) - V(q)$, with $T(\dot q) = \sum_{i=1}^n \frac{1}{2} m_i \dot q_i^2$, then $\sum_{i=1}^n \frac{\partial L(q,\dot q)}{\partial \dot q_i} \dot q_i = \sum_{i=1}^n \frac{\partial T(\dot q)}{\partial \dot q_i} \dot q_i = 2T(\dot q)$, so that

$$h(q, \dot q) = T(\dot q) + V(q) = E(q, \dot q) \ ,$$

and we find that h is equal to our old constant of motion, the total energy E.

Remark 2.3 (The Euler equation) The partial differential equation

$$\sum_{i=1}^n \frac{\partial S(q, \dot q)}{\partial \dot q_i} \dot q_i = 2S(q, \dot q) \ , \tag{2.12}$$

that we encountered above is called the *differential equation of Euler*. Solutions to Euler's equation (2.12) are called homogeneous of degree 2 in the variable $\dot q$, because equation (2.12) is equivalent to the statement that $S(q, \lambda \dot q) = \lambda^2 S(q, \dot q)$ for all $\lambda > 0$, as is quite easy to check. Homogeneity of degree 2 implies that

$$S(q, \dot q) = \frac{1}{2} \sum_{i,j=1}^n \beta_{ij}(q) \dot q_i \dot q_j \ .$$

for a collection of real-valued functions $q \mapsto \beta_{ij}(q)$ that is symmetric in the sense that $\beta_{ij} = \beta_{ji}$ for all $1 \le i, j \le n$. This follows from Taylor expanding S with respect to $\dot q$ at $\dot q = 0$ up to order two and noting that the homogeneity implies that the remainder vanishes. In fact, $\beta_{ij}(q) = \left.\frac{\partial^2 S(q,\dot q)}{\partial \dot q_i \partial \dot q_j}\right|_{\dot q = 0}$. The functions $q \mapsto \beta_{ij}(q)$ are as smooth as S is.

S is a nondegenerate Lagrangian at $(q, \dot q)$ if and only if the matrix $\beta(q)$ with elements $\beta_{ij}(q)$ is invertible. If this is the case for every q, then we say that β defines a pseudo-Riemannian metric on U. The solution curves to the corresponding Euler-Lagrange equations $[S]^q = 0$ are called the geodesics of the metric. Riemannian geometry will be the topic of Chapter 4.

Some more terminology: Mechanical systems with a Lagrangian of the form $L(q, \dot{q}) = S(q, \dot{q}) - V(q)$, with S is homogeneous of degree 2 in $\dot{q}$, are sometimes called *natural* mechanical systems. S is called the kinetic energy or free energy of the natural system and V the potential energy. By the above remarks, the total energy $E := S + V$ of a natural mechanical system is conserved.

2.5 Lagrangian equations for continua

Using Hamilton's principle as a physical postulate, we can derive equations of motion for the evolution of continuous media such as gases, fluids or elastic solids.

A continuum is modeled by a map $\psi : X \to \mathbb{R}^n$, where $X \subset \mathbb{R}^m$ is an open subset of $\mathbb{R}^m$. X is called the reference configuration of the continuum, and for $x \in X$, $\psi(x)$ is the location in $\mathbb{R}^n$ of the element of the continuum with label x. A *motion* of the continuum is a smooth map $u : I \times X \to \mathbb{R}^n$, where $I \subset \mathbb{R}$ is an open time-interval. Then the map $u(t, \cdot) : X \to \mathbb{R}^n$ describes the configuration of the continuum at time t, whereas the curve $t \mapsto u(t, x), I \to \mathbb{R}^n$ describes the motion of the element of the continuum with label x.

If $\rho : X \to \mathbb{R}$ is a mass density function, then the kinetic energy of the continuum is obtained by integrating the kinetic energy density $\frac{1}{2}\rho(x)||\dot{u}(x)||^2$ over X, where we denoted $\dot{u}(x) = \frac{\partial u(t,x)}{\partial t}$:

$$T(\dot{u}) = \int_X \frac{1}{2}\rho(x)||\dot{u}(x)||^2 \, d_m x \, .$$

The potential energy of the continuum at time t depends on the function $x \mapsto u(t, x)$ and may depend on the (x-)derivatives of this function. It is usually obtained by integrating a potential energy density function. For instance if we assume that it costs energy to stretch the continuum but not to bend it, then we are saying that the potential energy density depends only on the values of x, $u(t, x)$ and the matrix of first derivatives $Du(t, x) := \left(\frac{\partial u_i(t,x)}{\partial x_j}\right)_{i,j}$, and not on higher order derivatives. That is

$$V(u(t, \cdot)) = \int_X W(x, u(t, x), Du(t, x)) d_m x \, ,$$

in which $W : X \times \mathbb{R}^n \times \mathbb{R}^{n\times m} \to \mathbb{R}$ is a smooth function. Let us denote the arguments of W by (x, u, B), where $x \in X, u \in \mathbb{R}^n, B \in \mathbb{R}^{n\times m}$.

Now Hamilton's principle means that $u : I \times X \to \mathbb{R}^n$ satisfies

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \int_a^b \int_X \frac{1}{2}\rho(x) \left|\left|\frac{\partial \tilde{u}(t, x, \varepsilon)}{\partial t}\right|\right|^2 - W(x, \tilde{u}(t, x, \varepsilon), D\tilde{u}(t, x, \varepsilon)) \, d_m x \, dt \; = 0$$

for all smooth variations $\tilde{u} : I \times X \times (-\varepsilon_0, \varepsilon_0) \to \mathbb{R}^n$ of u with fixed (temporal) endpoints. Let us choose the variations of the form $\tilde{u}(t, x, \varepsilon) = u(t, x) + \varepsilon\phi(t, x)$, where ϕ is smooth and has a compact support that is contained in $I \times X$. Bringing the derivative inside the integral we then obtain

$$\int_a^b \int_X \sum_i \rho(x) \frac{\partial u_i(t,x)}{\partial t} \frac{\partial \phi_i(t,x)}{\partial t} - \sum_i \left. \frac{\partial W(t,u,B)}{\partial u_i} \right|_{\substack{u\,=\,u(t,x)\\ B\,=\,Du(t,x)}} \cdot \phi_i(t,x)$$
$$- \sum_{i,j} \left. \frac{\partial W(t,u,B)}{\partial B_{ij}} \right|_{\substack{u\,=\,u(t,x)\\ B\,=\,Du(t,x)}} \cdot \frac{\partial \phi_i(t,x)}{\partial x_j} \; d_m x \; dt = 0 \; . \tag{2.13}$$

If u is smooth enough, integration by parts gives that this equals

$$\sum_i \int_a^b \int_X \phi_i(t,x) \left(-\rho(x) \frac{\partial^2 u_i(t,x)}{\partial t^2} - \left. \frac{\partial W(t,u,B)}{\partial u_i} \right|_{\substack{u=u(t,x)\\ B=Du(t,x)}} + \sum_j \frac{\partial}{\partial x_j} \left. \frac{\partial W(t,u,B)}{\partial B_{ij}} \right|_{\substack{u=u(t,x)\\ B=Du(t,x)}} \right) d_m x dt.$$

Thus we derived from Hamilton's principle the partial differential equations

$$\rho(x) \frac{\partial^2 u_i}{\partial t^2} + \frac{\partial W}{\partial u_i} - \sum_{j=1}^m \frac{\partial}{\partial x_j} \frac{\partial W}{\partial B_{ij}} = 0 \; , \; i = 1, \ldots, n \; . \tag{2.14}$$

Equations (2.14) form a system of n partial differential equations for the n functions u_i, depending explicitly on the second order derivatives of the u_i with respect to t and the first and second order derivatives of the u_i with respect to the x_j. If W had been a function of the second order derivatives $D^2 u$ as well, then the resulting partial differential equation for u would have included fourth-order x-derivatives of u, etc. The reader may want to try and derive these equations.

A C^2 solution of equations (2.14) automatically satisfies equation (2.13) for every C^2 function $\phi = \phi(t,x)$ with compact support in $I \times X$. If u is not C^2 but only C^1 then it is not sensible to write equations (2.14), but equation (2.13) is still meaningful. We say that u is a *weak solution* or *solution in the distributional sense* of equations (2.14) if it satisfies the equality (2.13) for all smooth and compactly supported ϕ. In this way the space of functions in which we can search for solutions is enlarged, and the chance of finding any is bigger.

2.6 Excursion: submanifolds of $\mathbb{R}^n$

In the next section, we will use the notion of a submanifold of Euclidean space. So let us refresh our memory by recalling the following definitions:

Definition 2.4 (Immersions, embeddings, submersions and diffeomorphisms) *Let $0 \leq d \leq n$ and $k \in \mathbb{N} \cup \{\infty\}$ and let $W \subset \mathbb{R}^d$ and $U, V \subset \mathbb{R}^n$ be nonempty open subsets.*

- *A C^k map $h : W \to \mathbb{R}^n$ is called a C^k* immersion *if for all $w \in W$, the derivative $D_w h : \mathbb{R}^d \to \mathbb{R}^n$ of h at w is an injective linear map.*
- *A C^k map $h : W \to \mathbb{R}^n$ is called a C^k* embedding *if it is an immersion and a homeomorphism onto its image.*

- *A C^k map $g : U \to \mathbb{R}^{n-d}$ is called a C^k* submersion *if for all $u \in U$, the derivative $D_u g : \mathbb{R}^n \to \mathbb{R}^{n-d}$ of g at u is a surjective linear map.*
- *A C^k map $\Phi : U \to V$ is called a C^k* diffeomorphism *if it is bijective and for all $u \in U$, the derivative $D_u\Phi : \mathbb{R}^n \to \mathbb{R}^n$ of Φ at u is a bijective linear map.*

Note that $\Phi : U \to V$ is a diffeomorphism if and only if it is both an embedding and a submersion.

With these definitions, it is possible to give a number of equivalent characterizations of a submanifold of $\mathbb{R}^n$. In fact, one has the following Theorem/Definition, that we present here without a proof:

Definition 2.5 (Submanifolds of $\mathbb{R}^n$) *Let $0 \le d \le n$, $k \in \mathbb{N} \cup \{\infty\}$ and let $Q \subset \mathbb{R}^n$ be a nonempty subset. We say that Q is a C^k* submanifold *of $\mathbb{R}^n$ of dimension d, if for every $q \in Q$, one of the following* equivalent *statements is true:*

- *There exists an open neighborhood U of q in $\mathbb{R}^n$, an open subset W in $\mathbb{R}^d$, and a C^k function $f : W \to \mathbb{R}^{n-d}$ such that $Q \cap U = \mathrm{graph}(f) = \{(w, f(w)) \mid w \in W\}$, after re-ordering the coordinates of $\mathbb{R}^n$ if necessary.*
- *There exists an open neighborhood U of q in $\mathbb{R}^n$, an open subset $W \subset \mathbb{R}^d$ and a C^k embedding $h : W \to \mathbb{R}^n$ such that $Q \cap U = h(W)$.*
- *There exists an open neighborhood U of q in $\mathbb{R}^n$ and a C^k submersion $g : U \to \mathbb{R}^{n-d}$ with the property that $Q \cap U = g^{-1}(\{0_{n-d}\})$.*
- *There exist an open neighborhood U of q in $\mathbb{R}^n$, an open subset V of $\mathbb{R}^n$ and a C^k diffeomorphism $\Phi : U \to V$ such that*

$$\Phi(Q \cap U) = V \cap (\mathbb{R}^d \times \{0_{n-d}\}) \ .$$

This shows that a d-dimensional submanifold of $\mathbb{R}^n$ locally is equal to the graph of a function, the image of an embedding and the zero-set of a submersion and moreover looks like an open subset of $\mathbb{R}^d$. Also, these four properties are equivalent.

An open subset X of $\mathbb{R}^n$ is a trivial example of an n-dimensional C^∞ submanifold of $\mathbb{R}^n$. The reader can check that it satisfies definition 2.5 with $U = V = W = X$, $f = g = 0$ and $h = \mathrm{id}_X$.

Let Q be a C^1 submanifold of $\mathbb{R}^n$ and $q \in Q$. Recall that one says that $v \in \mathbb{R}^n$ is a *tangent vector* to Q at q if there exist an open neighborhood I of 0 in $\mathbb{R}$ and a C^1 curve $\gamma : I \to Q \subset \mathbb{R}^n$ in Q, with the property that $\gamma(0) = q$ and $\gamma'(0) = v$. The collection of all tangent vectors to Q at q is a linear subspace of $\mathbb{R}^n$, called the *tangent space* to Q at q, and denoted T_qQ. Now we can recall the following result, again without a proof:

Proposition 2.6

- *Let $W \subset \mathbb{R}^d$ be open and let $h : W \to \mathbb{R}^n$ be a C^1 embedding. Then the tangent space $T_{h(w)}(h(W))$ to $h(W)$ at $h(w)$ is equal to the image of the derivative $D_w h : \mathbb{R}^d \to \mathbb{R}^n$ of h at w.*

- *Let $V \subset \mathbb{R}^n$ be open and let $g : V \to \mathbb{R}^{n-d}$ be a C^1 submersion. Then the tangent space $T_u(g^{-1}(\{0_{n-d}\}))$ to $g^{-1}(\{0_{n-d}\})$ at u is equal to the kernel of the derivative $D_u g : \mathbb{R}^n \to \mathbb{R}^{n-d}$ of g at u.*

2.7 Holonomic and nonholonomic constraints

Let us return now to finite dimensionsional Lagrangian mechanics. Let $L : U \times \mathbb{R}^n \to \mathbb{R}$ be a C^2 Lagrangian function on an open subset $U \subset \mathbb{R}^n$ and assume that $\Phi : \tilde{U} \to U$ is a C^2 *embedding* of an open subset $\tilde{U} \subset \mathbb{R}^m$ into U, with $m < n$. This implies that $C := \Phi(\tilde{U})$ is an m-dimensional C^2 submanifold of $\mathbb{R}^n$. Of course, Φ induces a Lagrangian $\tilde{L} : \tilde{U} \times \mathbb{R}^m \to \mathbb{R}$ which satisfies the relation $[\tilde{L}]^Q(t) = [L]^q(t)\, \frac{\partial \Phi(Q)}{\partial Q}\Big|_{Q=Q(t)}$.

Recall that the derivative $D_Q\Phi = \frac{\partial \Phi(Q)}{\partial Q}$ is everywhere injective and that the tangent space T_qC to C at the point $q = \Phi(Q)$ is equal to the image of $\frac{\partial \Phi(Q)}{\partial Q}$. This then leads us to conclude that

$$[\tilde{L}]^Q(t) = 0 \text{ if and only if } [L]^q(t) \cdot \dot{q} = 0 \text{ for all } \dot{q} \in T_{q(t)}C \ , \tag{2.15}$$

viewing $\dot{q} \in T_{q(t)}C$ as a column vector and $[L]^q(t)$ as a row vector, that is as a *linear form* on $T_{q(t)}C$.

Remark 2.7 Because $[\tilde{L}]^Q = 0$ if and only if $t \mapsto Q(t)$ is stationary for $\int_a^b \tilde{L}(Q(t), \frac{dQ(t)}{dt})dt$ for all variations of $t \mapsto Q(t)$ with fixed endpoints, it is clear that the statement that (2.15) holds for all t, is equivalent to the statement that $t \mapsto q(t)$ is stationary for the action $A = \int_a^b L(q(t), \frac{dq(t)}{dt})dt$ for variations with fixed endpoints *that lie entirely in* C. This can be considered an alternative version of Hamilton's principle.

If we assume that C is the zero-set of a smooth submersion $g : U \to \mathbb{R}^{n-m}$, then statement (2.15) is also equivalent to the existence of *Lagrange multipliers* $\lambda_j = \lambda_j(t), j = 1, \ldots, n-m$ for which

$$[L]^q(t) = \sum_{j=1}^{n-m} \lambda_j(t) dg_j(q(t)) \ , \ q(t) \in C \ , \tag{2.16}$$

because $T_{q(t)}C$ is also equal to the kernel of $D_{q(t)}g$. Here we have used the notation $dg_j(q) := D_q g_j$, which is customary for the derivative of a real-valued function.

Solutions $q(t)$ to equations (2.16) are of course not necessarily solutions to the Euler-Lagrange equations $[L]^q = 0$, but they do have the property that they lie entirely in the submanifold C. In some applications we definitely want $q(t)$ to lie in a certain submanifold of $C \subset \mathbb{R}^n$, that we then call the *constraint manifold* of the mechanical system. It is a modeling *assumption* that equations (2.16) are the physically correct equations of motion for the constrained mechanical system with Lagrangian L. This assumption would have to be checked experimentally though.

The right hand side of equation (2.16) can be interpreted as a reaction force that keeps the curve $q(t)$ on the constraint manifold C. Indeed, because the dg_j are "perpendicular" to the level sets of g, these reaction forces do not insert any energy into the mechanical system. We say that they do not do any "work". This claim will be made precise in Exercise 2.7.

The situation becomes more complicated if we do not only impose constraints on the positions but also on the velocities. Then the constraint manifold becomes a submanifold $C \subset U \times \mathbb{R}^n$ and we require that $(q(t), \frac{dq(t)}{dt}) \in C$ for all t.

If for each $q \in U$, the intersection of C with the "fiber" $\{q\} \times \mathbb{R}^n$ is a linear subspace C_q of fixed dimension r with $1 \leq r \leq n$, then C is called a smooth distribution in U of rank r. A distribution can for instance be specified by defining for each $q \in U$ a collection of independent linear maps

$$\alpha_j(q) : \mathbb{R}^n \to \mathbb{R} \ , \ j = 1, \ldots, n-r \ ,$$

such that $C_q = \cap_{j=1}^{n-r} \ker \alpha_j(q)$. We call the α_j *linear velocity constraints* or simply constraints. The requirement that $(q(t), \frac{dq(t)}{dt}) \in C$ then becomes

$$\alpha_j(q(t)) \cdot \frac{dq(t)}{dt} = 0 \ , \ j = 1, \ldots, n-r \ . \tag{2.17}$$

If $g : U \to \mathbb{R}^{n-r}$ is a smooth submersion, then the dg_j define a distribution that has the special property that it is *integrable*: a curve that satisfies (2.17) for $\alpha_j = dg_j$, automatically has the property that $\frac{d}{dt} g_j(q(t)) = 0$ for all t, and hence such a curve lies in a level set of g. Thus the constraints dg_j are really only a constraint on the positions. Such constraints are called *holonomic*.

We call the α_j a system of *nonholonomic* constraints at $q_0 \in U$ if there does not exist an open neighborhood U_{q_0} of q_0 in U and functions g_j on U_{q_0} with the property that $\alpha_j = dg_j$ on U_{q_0}. In view of (2.16) it is quite natural to assume that the equations of motion for a mechanical system with Lagrangian L and nonholonomic constraints α_j are then given by the so-called *Vakonomic equations*

$$[L]^q(t) = \sum_{j=1}^{n-r} \lambda_j(t) \alpha_j(q(t)) \ , \ \alpha_j(q) \cdot \frac{dq}{dt} = 0 \ , j = 1, \ldots, n-r \ . \tag{2.18}$$

Rolling objects are examples of mechanical systems with nonholonomic constraints. Perhaps the most famous nonholonomic mechanical system is the *rattleback*, which is famous for its preference to rotate in one direction only.

2.8 Exercises

Exercise 2.1 *Let $U \subset \mathbb{R}^n$ and $\tilde{U} \subset \mathbb{R}^m$ be open subsets and let $T : U \times \mathbb{R}^n \to \mathbb{R}$ and $V : U \to \mathbb{R}$ be C^2 functions, with $T(q, \dot{q}) = \sum_{i,j=1}^n \frac{1}{2} \beta_{ij}(q) \dot{q}_i \dot{q}_j$ with $\beta_{ij}(q) = \beta_{ji}(q)$ and let*

$L(q,\dot q) = T(\dot q) - V(q)$. *Let* $\Phi : \tilde U \to U$ *be a* C^2 *map and let* $\tilde L$ *be defined by* $\tilde L(Q,\dot Q) = L(\Phi(Q), \frac{\partial \Phi(Q)}{\partial Q}\cdot \dot Q)$. *Show that*

$$\tilde L(Q,\dot Q) = \frac{1}{2}\sum_{i,j=1}^{m} \frac{1}{2}\tilde\beta_{ij}(Q)\dot Q_i \dot Q_j - V(\Phi(Q)) \ ,$$

in which

$$\tilde\beta_{ij}(\Phi(Q)) = \sum_{k,l=1}^{n} \beta_{kl}(Q)\frac{\partial \Phi_k(Q)}{\partial Q_i}\frac{\partial \Phi_l(Q)}{\partial Q_j} \ .$$

Prove that $\tilde\beta_{ij}(Q) = \tilde\beta_{ji}(Q)$.

Exercise 2.2 (Rotating coordinates) *Let* $U = \tilde U = \mathbb{R}^2$ *and let* $\Phi : \mathbb{R}\times\mathbb{R}^2 \to \mathbb{R}^2$ *be a one-parameter family of rotations:* $q = \Phi(t,Q) = e^{t\sigma J}\cdot Q$, *where* σ *is a real constant and*

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \ .$$

- *If* $T(\dot q) = \frac{1}{2}m\langle \dot q, \dot q\rangle$ *is the kinetic energy of a free particle, compute the transformed kinetic energy* $\tilde T(t,Q,\dot Q)$. *Show that it does not depend on* t *explicitly so that* $\tilde T = \tilde T(Q,\dot Q)$.
- *Show that*

$$[\tilde T]^Q(t) = m\left(\frac{d^2Q(t)}{dt^2} + 2\sigma J\frac{dQ(t)}{dt} - \sigma^2 Q(t)\right) \ .$$

- *Let* $V = V(q)$ *be a rotation-symmetric potential energy function. Show that* $\tilde V$ *is independent of* t *and write down the Euler-Lagrange equations for* $\tilde L(Q,\dot Q) = \tilde T(Q,\dot Q) - \tilde V(Q)$.

Exercise 2.3 *We model a one-dimensional string by letting the reference and target configuration be one-dimensional:* $X = J \subset \mathbb{R}$ *is an open interval and for a smooth* $u : I\times J \to \mathbb{R}$, $u(t,\cdot)$ *describes the configuration of the string at time* t. *For a smooth mass density* $\rho : J \to \mathbb{R}$, *and velocity function* $\dot u : J \to \mathbb{R}$, *define the kinetic energy*

$$T(\dot u) = \int_J \frac{1}{2}\rho(x)\dot u(x)^2 dx \ .$$

Denoting $u^{(j)}(x) := \frac{\partial^j u(x)}{\partial x^j}$, *assume that the potential energy is given by*

$$V(u) = \int_J W(x, u(x), u^{(1)}(x), \ldots, u^{(k)}(x))dx \ ,$$

where W *is a smooth potential energy density function* $W : \mathbb{R}^{k+2} \to \mathbb{R}$. *Use Hamilton's principle to derive for* u *the partial differential equation*

$$\rho(x)\frac{\partial^2 u(t,x)}{\partial t^2} + \sum_{j=0}^{k}(-1)^j \frac{\partial^j}{\partial x^j}\left.\frac{\partial W(x,u,\ldots,u^{(k)})}{\partial u^{(j)}}\right|_{u=u(t,x),\ldots,u^{(k)}=u^{(k)}(t,x)} = 0 \ .$$

If $W = W(u^{(1)})$ depends on the amount of "stretching" of the string only, we say that the string is purely elastic. For an elastic string derive the equation

$$\rho(x)\frac{\partial^2 u(t,x)}{\partial t^2} = \frac{\partial^2 u(t,x)}{\partial x^2} W'\left(\frac{\partial u(t,x)}{\partial x}\right) \ .$$

Exercise 2.4 (Minimal surfaces) *Let $X \subset \mathbb{R}^2$ be a bounded open subset with a C^2 boundary ∂X. Let $h : \overline{X} \to \mathbb{R}$ be a continuous function that is C^2 on X. Then the graph of h,* $\mathrm{graph}(h) := \{(x, h(x)) \mid x \in X\}$ *is a smooth surface with a finite area that equals*

$$S(h) := \int_X || \begin{pmatrix} 1 \\ 0 \\ \frac{\partial h(x)}{\partial x_1} \end{pmatrix} \times \begin{pmatrix} 0 \\ 1 \\ \frac{\partial h(x)}{\partial x_2} \end{pmatrix} || \, d_2x = \int_X \sqrt{1 + \left(\frac{\partial h(x)}{\partial x_1}\right)^2 + \left(\frac{\partial h(x)}{\partial x_2}\right)^2} \, d_2x \ .$$

We say that the graph of h is a minimal surface over X if $S(g) \geq S(h)$ for all continuous functions $g : \overline{X} \to \mathbb{R}$ that are C^2 on X and equal to h on ∂X. In particular this implies that $S(h + \varepsilon\phi) \geq S(h)$ if ϕ is some C^2 function on X with compact support contained in X. Prove that this implies that on X, h satisfies the partial differential equation

$$\left(1 + h_{x_1}^2 + h_{x_2}^2\right)(h_{x_1x_1} + h_{x_2x_2}) = h_{x_1}^2 h_{x_1x_1} + 2h_{x_1}h_{x_2}h_{x_1x_2} + h_{x_2}^2 h_{x_2x_2} \ .$$

Here we used the shorthand notation $h_\alpha := \frac{\partial h}{\partial \alpha}$. This equation is called the minimal surface equation.

Exercise 2.5 (The Skate) *The configuration of a* skate *consists of its position $(x, y) \in \mathbb{R}^2$, together with an angle $\phi \in \mathbb{R}/2\pi\mathbb{Z}$ that determines the direction in which the skate is pointing. The skate can only move in this particular direction: otherwise it would damage the ice.*

Try to formulate the above restriction as a constraint in the phase space of configurations and velocities of the skate. Argue that this constraint is nonholonomic. Write down the kinetic energy of the skate, as well as its Vakonomic equations of motion.

Exercise 2.6 (Shock waves) *A particular example of the one-dimensional string equations that arise in Exercise 2.3 is the wave equation*

$$u_{tt} = u_{xx} \ .$$

The goal of this exercise is to show that certain nondifferentiable functions may qualify as weak solutions to the wave equation.

Formulate the wave equation in its weak form.

Next, let $t \mapsto X(t)$ be any C^1 function and define $\Gamma := \{(x,t) \in \mathbb{R}^2 \,|\, x = X(t)\}$. If $v : \mathbb{R}^2 \backslash \Gamma \to \mathbb{R}$ is a continuous function, then we denote by v^-, v^+ and $[v] : \Gamma \to \mathbb{R}$ its left and right limits at Γ and its jump *over Γ:*

$$v^-(x,t) := \lim_{x \nearrow X(t)} v(x,t), \ v^+(x,t) := \lim_{x \searrow X(t)} v(x,t) \ \text{and} \ [v](x,t) := v^+(x,t) - v^-(x,t) \ ,$$

for $(x,t) \in \Gamma$ and if these limits exist.

Assume now that $u : \mathbb{R}^2 \to \mathbb{R}$ is a continuous function that is C^2 except possibly on Γ, but that $u_t^\pm, u_x^\pm, u_{tt}^\pm, u_{xx}^\pm$ and $u_{tx}^\pm$ exist. Show that u is a weak solution to the wave equation, precisely when

$$[u_t]\frac{dX}{dt} = [u_x] \text{ for all } (x,t) \in \Gamma \ .$$

This is the famous Rankine-Hugoniot jump condition. It relates the speed of a shock wave to the values of u_x and u_t before and after the passage of the shock wave.

Exercise 2.7 *Consider the Lagrangian system with holonomic constraints of equation (2.16). Assume that L is independent of t. Let us verify the claim that the right hand side of equation (2.16) does not do any work:*

- *Using (2.10), prove that $h := \frac{\partial L}{\partial \dot{q}} \cdot \dot{q} - L$ is constant along every curve $t \mapsto q(t)$ in the constraint manifold $C \subset U$, if and only if for every curve $t \mapsto q(t)$ in the constraint manifold there exist Lagrange multipliers $\lambda_j(t)$ for which the constraint equations (2.16) hold.*

- *Similarly, prove that h is constant along every curve $t \mapsto q(t)$ satisfying the nonholonomic constraints (2.17), if and only if for every curve $t \mapsto q(t)$ that satisfy these constraints, there exist Lagrange multipliers $\lambda_j(t)$ for which the Vakonomic equations (2.18) hold.*

Recall that if $L = T - V$, then $h = E = T + V$. Hence we have proved that in natural systems with holonomic or nonholonomic constraints, total energy is preserved.

3 Differential geometry

Until now, in most examples the position or configuration of a mechanical system was determined by n coordinates $q_i \in \mathbb{R}$, that is by an element q of an open subset $U \subset \mathbb{R}^n$. We saw in Section 2.7 that this is not always a natural assumption: for a constrained mechanical system, the allowed configurations lie in a submanifold $C \subset \mathbb{R}^n$ with a dimension lower than n. This C is then called a *constraint manifold*.

Another example is the *rigid body* that we will study in Chapter 6. Its configuration is a rotation matrix. The collection of all rotation matrices is called the special orthogonal group $SO(3,\mathbb{R})$, which is a compact subset of $\mathbb{R}^{3\times 3} \cong \mathbb{R}^9$. Since there are essentially 3 possible axes of rotation in $\mathbb{R}^3$, $SO(3,\mathbb{R})$ locally "looks like" an open subset of $\mathbb{R}^3$.

In this chapter, we will generalize the notion of a submanifold of $\mathbb{R}^n$ and we will make precise what we mean when we say that a topological space locally looks like an open subset of $\mathbb{R}^d$.

3.1 The language of manifolds

We will start by giving the "abstract" definition of a *differentiable manifold*. Submanifolds of $\mathbb{R}^n$ will turn out to be examples of such abstract manifolds.

Definition 3.1 (Differentiable manifolds) *Let $n \in \mathbb{N} \cup \{0\}$ and $k \in \mathbb{N} \cup \{0, \infty\}$. An n-dimensional C^k manifold is a topological Hausdorff space Q together with a collection $\mathcal{A}$ of homeomorphisms $\kappa : Q_\kappa \to U_\kappa$ from an open subset $Q_\kappa \subset Q$ to an open subset $U_\kappa \subset \mathbb{R}^n$ that satisfy*

- *For each $q \in Q$ there is at least one κ with $q \in Q_\kappa$.*
- *If $\kappa : Q_\kappa \to U_\kappa$ and $\lambda : Q_\lambda \to U_\lambda$ are two homeomorphisms with $Q_\kappa \cap Q_\lambda \neq \emptyset$, then $\lambda \circ \kappa^{-1} : \kappa(Q_\kappa \cap Q_\lambda) \to \lambda(Q_\kappa \cap Q_\lambda)$ is a C^k map of open subsets of $\mathbb{R}^n$.*

The homeomorphisms $\kappa : Q_\kappa \to U_\kappa$ are called *coordinate charts* for Q and the collection $\mathcal{A}$ of all coordinate charts for Q is called the *atlas* of Q. The maps $\lambda \circ \kappa^{-1}$ are called *recoordinatizations*. The definition implies that the recoordinatizations are in fact diffeomorphisms.

These definitions should be quite familiar to the reader. In fact, the main point to remember about manifolds is that they locally "look like an open subset of $\mathbb{R}^n$".

Not surprisingly, the C^k submanifolds of $\mathbb{R}^n$ of dimension d that we defined in Section 2.6 are examples of d-dimensional C^k manifolds:

Proposition 3.2 (Submanifolds of $\mathbb{R}^n$ are manifolds) *If $Q \subset \mathbb{R}^n$ is a d-dimensional C^k submanifold of $\mathbb{R}^n$ in the sense of definition 2.5, then Q is a d-dimensional C^k manifold in the sense of definition 3.1.*

Proof: This is simple: by definition 2.5, $Q \subset \mathbb{R}^n$ is a C^k submanifold of $\mathbb{R}^n$ of dimension d, if and only if for every point $q \in Q$ there exists an open neighborhood U of q in $\mathbb{R}^n$

and a C^k diffeomorphism $\Phi : U \to \Phi(U) \subset \mathbb{R}^n$ with the property that $\Phi(U \cap Q) = \Phi(U) \cap \left(\mathbb{R}^d \times \{0_{n-d}\}\right)$. Clearly, the composition of Φ with the projection onto the first d coordinates then is a coordinate chart for Q near q. Varying q over all elements in Q then produces an atlas for Q that satisfies definition 3.1. $\square$

3.2 The tangent space and bundle

Recall that we introduced the tangent space to a submanifold of $\mathbb{R}^n$ in Section 2.6. In this section, we will generalize this notion to abstract manifolds.

If $\kappa : Q_\kappa \to U_\kappa$ is some coordinate chart for Q with $q \in Q_\kappa$, it is convenient to denote by $q_\kappa \in U_\kappa$ the point $q \in Q$ in the chart κ, that is

$$q_\kappa := \kappa(q) \in U_\kappa \subset \mathbb{R}^n \ .$$

Suppose now that $\gamma : I \to Q$ is a C^1 curve in Q with $\gamma(0) = q$. Then $\kappa \circ \gamma$ is a curve in U_κ with $(\kappa \circ \gamma)(0) = q_\kappa$. The velocity of γ at time $t = 0$ in the chart κ is given by

$$\dot{q}_\kappa := (\kappa \circ \gamma)'(0) \in \mathbb{R}^n \ .$$

If now $\lambda : Q_\lambda \to U_\lambda$ is some other coordinate chart for Q with $q \in Q_\lambda$, then differentiation of $(\lambda \circ \gamma) = (\lambda \circ \kappa^{-1}) \circ (\kappa \circ \gamma)$ with respect to t at $t = 0$ gives that

$$\dot{q}_\lambda = D_{q_\kappa}(\lambda \circ \kappa^{-1}) \cdot \dot{q}_\kappa \ . \tag{3.1}$$

Here, $D_{q_\kappa}(\lambda \circ \kappa^{-1}) : \mathbb{R}^n \to \mathbb{R}^n$ denotes the derivative of $\lambda \circ \kappa^{-1}$ at the point q_κ. This observation inspires the following definitions:

Definition 3.3 (Tangent space and tangent bundle) *Let $k \geq 1$ and let Q be a C^k manifold of dimension n. Let $q \in Q$. Denote by $\mathcal{A}_q$ the set of charts around q, i.e.*

$$\mathcal{A}_q = \{\kappa \in \mathcal{A} \mid q \in Q_\kappa \} \ .$$

A tangent vector *to Q at q is a map*

$$\dot{q} : \mathcal{A}_q \to \mathbb{R}^n, \kappa \mapsto \dot{q}_\kappa$$

which satisfies

$$\dot{q}_\lambda = D_{q_\kappa}(\lambda \circ \kappa^{-1}) \cdot \dot{q}_\kappa \ ,$$

for any $\kappa, \lambda \in \mathcal{A}_q$. The collection of all tangent vectors to Q at q is called the tangent space *to Q at q and is denoted T_qQ. The disjoint union*

$$TQ := \bigcup_{q \in Q} T_qQ$$

is called the tangent bundle of Q.

Indeed, for a C^1 curve $\gamma : I \to Q$ with $\gamma(0) = q$, the transformation formula (3.1) says that the map

$$\gamma'(0) : \kappa \mapsto (\kappa \circ \gamma)'(0)$$

is an element of T_qQ.

On the other hand, for a given $\dot{q} \in T_qQ$, the C^1 curve $\gamma : t \mapsto \kappa^{-1}(q_\kappa + t\dot{q}_\kappa)$ in Q satisfies $\gamma'(0) = \dot{q}$. Thus, T_qQ is exactly equal to the set of all possible velocity vectors of C^1 curves through the point q. This shows that our definition of tangent space is the correct generalisation of the definition of the tangent space to a submanifold of $\mathbb{R}^n$ given in Section 2.6.

One can define addition and scalar multiplication on T_qQ in an "obvious" manner, by letting, for $\dot{q}^{(1)}, \dot{q}^{(2)} \in T_qQ$ and $\alpha \in \mathbb{R}$,

$$\dot{q}^{(1)} + \dot{q}^{(2)} : \kappa \mapsto \dot{q}^{(1)}_\kappa + \dot{q}^{(2)}_\kappa \ ,$$
$$\alpha\dot{q}^{(1)} : \kappa \mapsto \alpha\dot{q}_\kappa \ .$$

This makes T_qQ a linear space. In fact, it is not hard to check that the map

$$D_q\kappa : \dot{q} \mapsto \dot{q}_\kappa \ , T_qQ \to \mathbb{R}^n$$

is a linear isomorphism from T_qQ to $\mathbb{R}^n$, so T_qQ has dimension n.

Moreover, TQ has a natural atlas, consisting of the charts

$$D\kappa : \bigcup_{q \in Q_\kappa} T_qQ \to U_\kappa \times \mathbb{R}^n \ , \dot{q} \in T_qQ \mapsto (\kappa(q), D_q\kappa \cdot \dot{q}) = (q_\kappa, \dot{q}_\kappa) \ ,$$

with $\kappa \in \mathcal{A}$. In fact, the recoordinatisations for this induced atlas are given by

$$D\lambda \circ (D\kappa)^{-1} : \kappa(Q_\kappa \cap Q_\lambda) \times \mathbb{R}^n \to \lambda(Q_\kappa \cap Q_\lambda) \times \mathbb{R}^n \ , \ (q_\kappa, \dot{q}_\kappa) \mapsto (q_\lambda, \dot{q}_\lambda),$$

in which $q_\lambda = (\lambda \circ \kappa^{-1})(q_\kappa)$ and $\dot{q}_\lambda = D_{q_\kappa}(\lambda \circ \kappa^{-1}) \cdot \dot{q}_\kappa$. This shows that TQ is a manifold of differentiability class C^{k-1} and dimension $2n$.

3.3 The cotangent bundle

If V is a linear space, then we denote by V^* the space of real-valued linear functions on V:

$$V^* := \{\alpha : V \to \mathbb{R} \mid \alpha \text{ is a linear map } \} \ .$$

V^* is also called the *dual* of V. The elements of V^* are called *linear forms* on V.

We usually think of an element of $(\mathbb{R}^n)^*$ as a row vector of length n, which makes it isomorphic to $\mathbb{R}^n$.

If Q is a C^k manifold of dimension n, then the dual $(T_qQ)^*$ of the tangent space to Q at q is called the *co-tangent space* to q at Q and its elements are called *co-vectors*. The disjoint union

$$T^*Q := \bigcup_{q \in Q} (T_qQ)^*$$

is called the *cotangent bundle* of Q. As was the case for the tangent bundle, a chart $\kappa : Q_\kappa \to U_\kappa$ for Q induces a chart for T^*Q, namely the inverse of the mapping

$$D^*\kappa : U_\kappa \times (\mathbb{R}^n)^* \to \bigcup_{q\in Q_\kappa} (T_qQ)^* \ , \ (q_\kappa, p_\kappa) \mapsto p_\kappa \circ D_{\kappa(q_\kappa)}\kappa \in (T_{\kappa(q_\kappa)}Q)^* \ .$$

In this way we produce an atlas for T^*Q with which it becomes a C^{k-1} manifold of dimension $2n$: if $\lambda : Q_\lambda \to U_\lambda$ is another chart for Q, then

$$(D^*\lambda)^{-1} \circ D^*\kappa : \kappa(Q_\kappa \cap Q_\lambda) \times (\mathbb{R}^n)^* \to \lambda(Q_\kappa \cap Q_\lambda) \times (\mathbb{R}^n)^* \ , \ (q_\kappa, p_\kappa) \mapsto (q_\lambda, p_\lambda) \ ,$$

with $q_\lambda = (\lambda \circ \kappa^{-1})(q_\kappa)$ and $p_\kappa = p_\lambda \circ D_{q_\kappa}(\lambda \circ \kappa^{-1})$. We say that cotangent vectors transform covariantly, which should be compared with the contravariant transformation formula $\dot q_\lambda = D_{q_\kappa}(\lambda \circ \kappa^{-1}) \cdot \dot q_\kappa$.

3.4 A coordinate invariant interpretation of Lagrange's variables

In this section, I will show how the square bracket $[L]^\delta(t)$ can be interpreted in a coordinate invariant way. In particular, this will lead to a coordinate invariant formulation of the Euler-Lagrange equations $[L]^\delta = 0$.

So we start with a C^2 manifold Q and a C^2 Lagrangian function $L : TQ \to \mathbb{R}$ on the tangent bundle. When $\kappa : Q_\kappa \to U_\kappa \subset \mathbb{R}^n$ is a coordinate chart for Q, then we denote by

$$L_\kappa := L \circ (D\kappa)^{-1} \ : \ U_\kappa \times \mathbb{R}^n \to \mathbb{R}$$

the Lagrangian in the coordinates induced by κ.

Assume now that $\delta : I \to Q$ is a C^2 curve in Q. Let $t \in I$ and suppose that $\delta(t) \in Q_\kappa$. We then defined the functions

$$[L_\kappa]_i^{\kappa\circ\delta}(t) := \frac{d}{dt}\left(\left.\frac{\partial L_\kappa(q_\kappa, \dot q_\kappa)}{\partial(\dot q_\kappa)_i}\right|_{\dot q_\kappa = (\kappa\circ\delta)'(t)} \right) - \left.\frac{\partial L_\kappa(q_\kappa, \dot q_\kappa)}{\partial(q_\kappa)_i}\right|_{\dot q_\kappa = (\kappa\circ\delta)'(t)} \quad \text{for } i = 1, \ldots, n \ .$$

Suppose now that $\lambda : Q_\lambda \to U_\lambda$ is another chart for Q, with $\delta(t) \in Q_\lambda$. Then the definition of L_κ and L_λ implies that $L_\kappa = L_\lambda \circ (D\lambda \circ (D\kappa)^{-1})$, that is

$$L_\kappa(q_\kappa . \dot q_\kappa) = L_\lambda((\lambda \circ \kappa^{-1})(q_\kappa), D_{q_\kappa}(\lambda \circ \kappa^{-1}) \cdot \dot q_\kappa) \ .$$

But this means that we can apply Theorem 2.1 with $U = U_\lambda, \tilde U = U_\kappa, \Phi = \lambda \circ \kappa^{-1}$, $L = L_\lambda$, $\tilde L = L_\kappa$, $Q(t) = (\kappa \circ \gamma)(t)$ and $q(t) = (\lambda \circ \gamma)(t)$. The conclusion is that

$$[L_\kappa]^{\kappa\circ\delta}(t) = [L_\lambda]^{\lambda\circ\delta}(t) \cdot D_{(\kappa\circ\delta)(t)}(\lambda \circ \kappa^{-1}) \ ,$$

viewing $[L_\kappa]^{\kappa\circ\delta}(t)$ and $[L_\lambda]^{\lambda\circ\delta}(t)$ as a row vectors.

The important remark now is that if $\dot q \in T_{\delta(t)}Q$ is an arbitrary tangent vector, taking the value $\dot q_\kappa \in \mathbb{R}^n$ in the chart κ, then $\dot q_\lambda = D_{(\lambda\circ\delta)(t)}(\lambda \circ \kappa^{-1}) \cdot \dot q_\kappa$. But this implies that

$$[L_\kappa]^{\kappa\circ\delta}(t) \cdot \dot q_\kappa = [L_\lambda]^{\lambda\circ\delta}(t) \cdot \dot q_\lambda \ .$$

In other words, we have a uniquely defined linear map

$$[L]^\delta(t) : T_{\delta(t)}Q \to \mathbb{R} \ , \ \dot{q} \mapsto [L_\kappa]^{\kappa\circ\delta}(t) \cdot \dot{q}_\kappa \ ,$$

and this definition is independent of κ. It thus turns out that $[L]^\delta(t)$ can be interpreted as an element of $(T_{\delta(t)}Q)^*$, and $t \mapsto [L]^\delta(t)$ as a curve in T^*Q.

3.5 The tangent map

In this section, we define what we mean by the derivative of maps between manifolds.

We begin by defining what we mean by a C^k map of C^k manifolds. Let Q and $\tilde{Q}$ be C^k manifolds of dimensions n and m respectively. Then a map $\Phi : Q \to \tilde{Q}$ is called C^l with $l \leq k$ if it is C^l in coordinate charts, that is if for all coordinate charts $\kappa : Q_\kappa \to U_\kappa$ for Q and $\lambda : \tilde{Q}_\lambda \to \tilde{U}_\lambda$ for $\tilde{Q}$,

$$\lambda \circ \Phi \circ \kappa^{-1} : U_\kappa \cap \kappa(\Phi^{-1}(\tilde{Q}_\lambda)) \to \tilde{U}_\lambda$$

is C^l as a map from an open subset of $\mathbb{R}^n$ to an open subset of $\mathbb{R}^m$.

Now, let $\Phi : Q \to \tilde{Q}$ be a C^1 map between C^1 manifolds. If $\gamma : I \to Q$ is a curve in Q with $\gamma(0) = q$ and velocity $\gamma'(0) \in T_qQ$, then Φ transports the curve γ to the curve $(\Phi \circ \gamma) : I \to \tilde{Q}$ in $\tilde{Q}$ with $(\Phi \circ \gamma)(0) = \Phi(q)$ and $(\Phi \circ \gamma)'(0) \in T_{\Phi(q)}\tilde{Q}$.

Proposition 3.4 *Let Q, $\tilde{Q}$ be C^1 manifolds and $\Phi : Q \to \tilde{Q}$ a C^1 map. Then there is a uniquely defined linear map*

$$T_q\Phi : T_qQ \to T_{\Phi(q)}\tilde{Q} \ ,$$

with the property that

$$(\Phi \circ \gamma)'(0) = T_q\Phi \cdot \gamma'(0)$$

for all C^1 curves γ in Q with $\gamma(0) = q$. $T_q\Phi$ is called the tangent map *of Φ at q and satisfies*

$$D_{\Phi(q)}\lambda \circ T_q\Phi \circ (D_q\kappa)^{-1} = D_{q_\kappa}(\lambda \circ \Phi \circ \kappa^{-1}) \tag{3.2}$$

for any pair of charts κ for Q and λ for $\tilde{Q}$.

Proof: Let $\kappa : Q_\kappa \to U_\kappa$ and $\lambda : \tilde{Q}_\lambda \to \tilde{U}_\lambda$ be charts for Q and $\tilde{Q}$ respectively. Then differentiation of

$$\lambda \circ \Phi \circ \gamma = (\lambda \circ \Phi \circ \kappa^{-1}) \circ (\kappa \circ \gamma)$$

at the point q_κ gives that

$$D_{\Phi(q)}\lambda \cdot (\Phi \circ \gamma)'(0) = D_{q_\kappa}(\lambda \circ \Phi \circ \kappa^{-1}) \cdot D_q\kappa \cdot \gamma'(0) \ .$$

As any element of T_qQ is the tangent vector of some curve and $D_{\Phi(q)}\lambda$ and $D_q\kappa$ are linear isomorphisms, this proves the result. □

Note that formula (3.2) exactly says that "the map $T_q\Phi$ in induced coordinates" is equal to the derivative of "the map Φ in coordinates".

The map

$$T\Phi : TQ \to T\tilde{Q} \ , \dot{q} \in T_qQ \mapsto T_q\Phi \cdot \dot{q}$$

is called the *tangent map* of Φ. By definition, $T\Phi$ sends T_qQ to $T_{\Phi(q)}\tilde{Q}$. In local coordinates, $T\Phi$ is given by

$$D\lambda \circ T\Phi \circ (D\kappa)^{-1} : (q_\kappa, \dot{q}_\kappa) \mapsto ((\lambda \circ \Phi \circ \kappa^{-1})(q_\kappa), D_{q_\kappa}(\lambda \circ \Phi \circ \kappa^{-1}) \cdot \dot{q}_\kappa) \ .$$

This proves that $T\Phi$ is C^{l-1} when Φ is C^l.

Finally, we remark without proof that when $\Phi : Q \to \tilde{Q}$ and $\Psi : \tilde{Q} \to \bar{Q}$ are C^1 maps, then the composition $\Psi \circ \Phi : Q \to \bar{Q}$ is also C^1 and we have the *chain rule*:

$$T_q(\Psi \circ \Phi) = T_{\Phi(q)}\Psi \circ T_q\Phi \ ,$$

that is $T(\Psi \circ \Phi) = T\Psi \circ T\Phi$.

Remark 3.5 (A coordinate invariant formulation of Lagrange's theorem) Let Q and $\tilde{Q}$ be C^2 manifolds, $\gamma : I \to \tilde{Q}$ a C^2 curve in $\tilde{Q}$, $L : TQ \to \mathbb{R}$ a C^2 Lagrangian function on TQ and $\Phi : \tilde{Q} \to Q$ a C^2 mapping. Then $\Phi \circ \gamma : I \to Q$ is a curve in Q and $L \circ T\Phi : T\tilde{Q} \to \mathbb{R}$ is a Lagrangian on $T\tilde{Q}$. The coordinate invariant version of Theorem 2.1 now says that the identity

$$[L \circ T\Phi]^\gamma(t) = [L]^{\Phi\circ\gamma}(t) \circ T_{\gamma(t)}\Phi$$

holds for linear forms on $T_{\gamma(t)}\tilde{Q}$. The proof is an exercise in using definitions.

3.6 Submanifolds and Whitney embedding

With the tangent map at our disposal, we can straightforwardly define what it means for a smooth map of smooth manifolds to be an immersion, embedding, submersion or diffeomorphism, analogous to the definitions in Section 2.6.

Definition 3.6 *Let P, Q and R be C^k manifolds of finite dimensions.*

- *A C^k map $h : P \to Q$ is called a C^k* immersion *if for all $p \in P$, the tangent map $T_ph : T_pP \to T_{h(p)}Q$ is an injective linear map.*
- *A C^k map $h : P \to Q$ is called a C^k* embedding *if it is an immersion and a homeomorphism onto its image.*
- *A C^k map $g : Q \to R$ is called a C^k* submersion *if for all $q \in Q$, the tangent map $T_qg : T_qQ \to T_{g(q)}R$ is a surjective linear map.*
- *A C^k map $\Phi : P \to Q$ is called a C^k* diffeomorphism *if it is bijective and, for all $p \in P$, the tangent map $T_p\Phi : T_pP \to T_{\Phi(p)}Q$ is a bijective linear map.*

At the risk of boring the reader, we now also introduce the notion of a submanifold of a differentiable manifold. Using Definition/Theorem 2.5 and a local coordinate system for Q, it is easy to prove the following Definition/Theorem:

Definition 3.7 *Let $n \in \mathbb{N} \cup \{0\}$, $k \in \mathbb{N} \cup \{\infty\}$ and $0 \leq d \leq n$. Let Q be a C^k manifold of dimension n and let $P \subset Q$ be a nonempty subset. We say that P is a C^k submanifold of Q of dimension d if for every $p \in P$ one of the following* equivalent *statements is true:*

- *There exists an open neighborhood U of p in Q, an open subset $W \subset \mathbb{R}^d$ and a C^k embedding $h : W \to Q$ such that $P \cap U = h(W)$.*
- *There exists an open neighborhood U of p in Q and a C^k submersion $g : U \to \mathbb{R}^{n-d}$ with the property that $P \cap U = g^{-1}(\{0_{n-d}\})$.*
- *There exist an open neighborhood U of p in Q, an open subset V of $\mathbb{R}^n$ and a C^k diffeomorphism $\Phi : U \to V$ such that*
$$\Phi(P \cap U) = V \cap (\mathbb{R}^d \times \{0_{n-d}\}) \ .$$

It also follows easily that if P is a d-dimensional C^k submanifold of a C^k manifold Q in the sense of definition 3.7, then P is a d-dimensional C^k manifold in the sense of definition 3.1, exactly as one would expect.

Let us finish this section with flagging a deeper result, the famous *Whitney embedding theorem*. This theorem says that every abstract n-dimensional C^k manifold Q that satisfies a mild topological condition, can be embedded in some $\mathbb{R}^N$.

Theorem 3.8 (Whitney embedding) *Let Q be an n-dimensional C^k manifold that is the countable union of compact subsets. Then there exists a C^k embedding $h : Q \to \mathbb{R}^{2n}$ of Q into $\mathbb{R}^{2n}$. In particular, Q is diffeomorphic to the submanifold $h(Q) \subset \mathbb{R}^{2n}$.*

If a topological space is the countable union of compact subsets, it is called *second-countable*. Second-countability is a rather weak condition. For instance every subset of $\mathbb{R}^n$, with the induced topology, is second-countable, because $\mathbb{R}^n$ is itself a countable union of compact subsets. This also shows that second-contability is an obvious necessary condition for a manifold to be diffeomorphic to a submanifold of $\mathbb{R}^{2n}$.

The importance of Whitney's embedding theorem is that it allows us to think of all second-countable differentiable manifolds as submanifolds of Euclidean space. This is in particular convenient if you don't like the abstract definition 3.1.

The proof of Whitney's theorem is unfortunately much too complicated to discuss here.

3.7 Vector fields

A vector field assigns in a smooth way to each $q \in Q$ a tangent vector in T_qQ. In other words, a vector field is a C^1 map v from Q to TQ with the property that $v(q) \in T_qQ$ for all $q \in Q$. We can say this in another way by defining the *bundle projection*

$$\pi_{TQ} : TQ \to Q \ , \dot{q} \in T_qQ \mapsto q \ .$$

Recall that if Q is a C^k manifold, then TQ is a C^{k-1} manifold. Because in local coordinates, π_{TQ} simply maps $(q_\kappa, \dot{q}_\kappa)$ to q_κ, we see that π_{TQ} is in fact a C^{k-1} submersion.

Now, the requirement that $v : Q \to TQ$ maps $q \in Q$ to an element of T_qQ, i.e. the requirement that v be a vector field, can quite elegantly be formulated as

$$\pi_{TQ} \circ v = \mathrm{id}_Q \ .$$

Remark 3.9 (The differential) The cotangent bundle T^*Q also has a bundle projection π_{T^*Q}, that sends $p \in (T_qQ)^*$ to q. A C^1 mapping $\alpha : Q \to T^*Q$ with the property that $\pi_{T^*Q} \circ \alpha = \mathrm{id}_Q$ is called a *one-form* on Q. It assigns to every $q \in Q$ a linear form $\alpha(q)$ on T_qQ.

When $f : Q \to \mathbb{R}$ is a real-valued C^1 function on Q, then its tangent map T_qf at q sends T_qQ to $T_{f(q)}\mathbb{R} \cong \mathbb{R}$. We usually denote it by

$$df(q) := T_qf \in (T_qQ)^* \ .$$

The map $df : Q \to T^*Q, q \mapsto df(q)$ is an example of a one-form on Q.

Remark 3.10 (Notation) If $v : Q \to TQ$ is a continuous vector field and $f : Q \to \mathbb{R}$ a C^1 function, then $v(q) \in T_qQ$ and $df(q) : T_qQ \to \mathbb{R}$, that is $df(q) \in (T_qQ)^*$. Thus one can define the function $v \cdot f : Q \to \mathbb{R}$ by

$$(v \cdot f)(q) = df(q) \cdot v(q) \ .$$

$v \cdot f$ is called the *derivative of f in the direction of v*. In local coordinates we have

$$(v \cdot f)(q) = \sum_{j=1}^n v_j(q) \frac{\partial f(q)}{\partial q_j} \ ,$$

where $v_j : U \subset \mathbb{R}^n \to \mathbb{R}$ are continuous functions on a coordinate patch. This shows that if v is C^k and f is C^{k+1}, then $v \cdot f$ is C^k. We observe that $v\cdot : f \mapsto v \cdot f$ is a first order differential operator - a first order differential operator is also called a *derivation*. In fact, the allocation $v \mapsto v\cdot$ allows us to interpret v itself as a first order differential operator. Hence we may choose to write

$$v = \sum_{j=1}^n v_j \frac{\partial}{\partial q_j} \ .$$

This is very common notation for a vector field in local coordinates.

3.8 Integral curves

If $\gamma : I \to Q$ is a C^1 curve in Q, then $\frac{d\gamma(t)}{dt} \in T_{\gamma(t)}Q$. We say that γ is an integral curve of the vector field v on Q if

$$\frac{d\gamma(t)}{dt} = v(\gamma(t))$$

for all $t \in I$. The existence and uniqueness theory for ordinary differential equations guarantees that if v is C^1, then for every initial condition $q_0 \in Q$, there exists a unique maximal integral curve $\gamma : I \to Q$ of v for which $\gamma(0) = q_0$.

Recall that the existence and uniqueness theory for ordinary differential equations moreover says that for every $t \in \mathbb{R}$ there exists an open subset $D_t \subset Q$ and a C^k mapping

$$\Phi^t : D_t \to Q$$

with the following properties

- $D_0 = Q$ and $\Phi^0 = \mathrm{id}_Q$.
- Φ^t is a diffeomorphism onto its image.
- $D_t \cap \Phi^s(D_s) = D_s \cap \Phi^t(D_t) = D_{t+s}$ and on this set,

  $$\Phi^t \circ \Phi^s = \Phi^{t+s} \ .$$

  This is called the group property.
- $\frac{d\Phi}{dt} = v \circ \Phi$, that is $\frac{d}{dt}\Phi^t(q) = v(\Phi^t(q))$ for all $q \in D_t$. That is, the curve $t \mapsto \Phi^t(q)$ is the unique maximal integral curve of v with initial condition q.

The map Φ^t is called the time-t flow of the vector field v. It is also, suggestively, denoted by

$$\Phi^t = e^{tv} \ ,$$

because of the third property. I will use this notation a lot.

3.9 Conjugation

Let Q be a C^k manifold, $v : Q \to TQ$ a C^{k-1} vector field and $\Phi : Q \to Q$ a C^k diffeomorphism of Q. Because $T\Phi$ maps $T_{\Phi^{-1}(q)}Q$ to T_qQ, the composition

$$T\Phi \circ v \circ \Phi^{-1} : Q \to TQ$$

is a C^{k-1} vector field on Q. This vector field is denoted Φ_*v and called the *pushforward* of v:

$$\Phi_*v(q) := (T_{\Phi^{-1}(q)}\Phi) \cdot v(\Phi^{-1}(q)) \in T_qQ \ .$$

Note that the identity $(\Psi \circ \Phi)^{-1} = \Phi^{-1} \circ \Psi^{-1}$, the chain rule $T(\Psi \circ \Phi) = T\Psi \circ T\Phi$ and the definition $\Phi_*v = T\Phi \circ v \circ \Phi^{-1}$ together imply that

$$(\Psi \circ \Phi)_*v = \Psi_*(\Phi_*v)$$

Proposition 3.11 *If $\gamma : I \to Q$ is an integral curve of v, then $\Phi \circ \gamma$ is an integral curve of Φ_*v.*

Proof: Let γ be an integral curve of v, that is $\gamma'(t) = v(\gamma(t))$ for all $t \in I$. Then, by the definition of the tangent map, $(\Phi \circ \gamma)'(t) = T_{\gamma(t)}\Phi \cdot \gamma'(t) = T_{\gamma(t)}\Phi \cdot v(\gamma(t)) = \Phi_* v(\Phi(\gamma(t))) = \Phi_* v((\Phi \circ \gamma)(t))$. $\square$

When v and w are vector fields on Q and $\Phi_* v = w$ for some diffeomorphism Φ, then we say that Φ *conjugates* v to w and that v and w are conjugate vector fields.

3.10 The Lie bracket

The collection of all C^k vector fields on the C^{k+1} manifold Q is denoted

$$\mathcal{X}^k(Q) := \{v : Q \to TQ \mid \pi_{TQ} \circ v = \mathrm{id}_Q,\ v \text{ is } C^k\} \ .$$

Clearly, vector fields can be added and multiplied by a scalar, because the tangent spaces T_qQ are linear spaces, so that $\mathcal{X}^k(Q)$ is a linear space. Apart from this, it turns out we can define an operation on it. This goes as follows.

When $u, v : Q \to TQ$ are two C^k vector fields on Q, then for small enough t, the flow e^{tu} is a diffeomorphism of an open neighborhood of every point $q \in Q$. We can use it to push forward the vector field v, to obtain the new vector field $(e^{tu})_* v$. Now we can define the *Lie bracket* of u and v as

$$[u, v] := \left.\frac{d}{dt}\right|_{t=0} (e^{tu})_* v \ .$$

It turns out that, if u and v are given in in local coordinates as

$$u = \sum_{j=1}^{n} u_j \frac{\partial}{\partial q_j} \text{ and } v = \sum_{j=1}^{n} v_j \frac{\partial}{\partial q_j} \ ,$$

then their Lie bracket is given by

$$[u, v] = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} \frac{\partial u_j}{\partial q_k} v_k - \frac{\partial v_j}{\partial q_k} u_k \right) \frac{\partial}{\partial q_j} \ .$$

From this expression, one observes that if u and v are C^k, then $[u, v]$ is C^{k-1}. Furthermore, it is immediately clear that the Lie bracket is antisymmetric:

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

A priori this is not so obvious from the definition. Moreover, a little computation shows that if u, v and w are C^2 vector fields, then we have the identity

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

This identity is called the *Jacobi identity.*

Definition 3.12 *A linear space V on which a* bracket operation $[\cdot,\cdot] : V \times V \to V$ *is defined for which*

$$[u,v] = -[v,u] \text{ and } [u,[v,w]] + [v,[w,u]] + [w,[u,v]] = 0 \text{ for all } u,v,w \in V \ ,$$

is called a Lie algebra*.*

We just proved that $\mathcal{X}^\infty(Q)$ is a Lie algebra.

3.11 Excursion: Second order vector fields

Recall that a mechanical system is described by a system of differential equations of second order. This means that in local coordinates $(q_\kappa, \dot{q}_\kappa) \in U_\kappa \times \mathbb{R}^n$ for TQ induced by local coordinates $\kappa : Q_\kappa \subset Q \to U_\kappa \subset \mathbb{R}^n$, the equations of motion have the form

$$\frac{dq_\kappa}{dt} = \dot{q}_\kappa \ , \ \frac{d\dot{q}_\kappa}{dt} = a_\kappa(q_\kappa, \dot{q}_\kappa) \ ,$$

for some function $a_\kappa : U_\kappa \times \mathbb{R}^n \to \mathbb{R}^n$.

In the coordinate invariant formulation, these differential equations define a vector field on TQ (not on Q), that is a map

$$v : TQ \to T(TQ) \text{ with the property that } \pi_{T(TQ)} \circ v = \mathrm{id}_{TQ} \ .$$

The property that v be a *second order vector field* is expressed by an additional requirement, that we shall investigate now:

First of all, we remark that the bundle projection of TQ maps TQ to Q and that it has a tangent map $T(\pi_{TQ}) : T(TQ) \to TQ$. On the other hand, $T(TQ)$ has its own bundle projection $\pi_{T(TQ)} : T(TQ) \to TQ$. In general, $\pi_{T(TQ)} \neq T(\pi_{TQ})$.

If q_κ are local coordinates for Q, then they induce local coordinates $(q_\kappa, \dot{q}_\kappa)$ for TQ, which in turn induce local coordinates $(q_\kappa, \dot{q}_\kappa, \delta q_\kappa, \delta\dot{q}_\kappa)$ for $T(TQ)$. In these local coordinates, $\pi_{T(TQ)}$ maps $(q_\kappa, \dot{q}_\kappa, \delta q_\kappa, \delta\dot{q}_\kappa) \mapsto (q_\kappa, \dot{q}_\kappa)$. At the same time, π_{TQ} sends $(q_\kappa, \dot{q}_\kappa) \mapsto q_\kappa$, and hence $T(\pi_{TQ})$ maps $(q_\kappa, \dot{q}_\kappa, \delta q_\kappa, \delta\dot{q}_\kappa)$ to $(q_\kappa, \delta q_\kappa)$. In local coordinates, a vector field $v : TQ \to T(TQ)$ sends $(q_\kappa, \dot{q}_\kappa)$ to $(q_\kappa, \dot{q}_\kappa, v_{q_\kappa}(q_\kappa, \dot{q}_\kappa), v_{\dot{q}_\kappa}(q_\kappa, \dot{q}_\kappa))$. The requirement that $v_{q_\kappa}(q_\kappa, \dot{q}_\kappa) = \dot{q}_\kappa$ therefore just means that

$$\pi_{T(TQ)} \circ v = T(\pi_{TQ}) \circ v \ . \tag{3.3}$$

Formula (3.3) is the coordinate invariant way of saying that v is a second order vector field. Of course, this is just a complicated way of saying something quite simple.

3.12 Excursion: Fiber bundles

A fiber bundle is a generalization of the Cartesian product of two manifolds. They are studied extensively by differential geometers. For the interested reader, we will give the relevant definitions here and discuss a few examples. And we will see that the tangent bundle and cotangent bundle are special kinds of fiber bundles, namely vector bundles.

Definition 3.13 *Let X, Y and Z be C^k manifolds of dimensions p, q and r respectively, with $p = q + r$. We say that X is a Z-*fiber bundle *over Y if there is a C^k surjective mapping*

$$\pi_X : X \to Y$$

with the property that for all $y \in Y$ there is a nonempty open neighborhood $Y_y \subset Y$ of y and a diffeomorphism $\phi_y : \pi_X^{-1}(Y_y) \to Y_y \times Z$, such that $\phi_y(x) \in \{\tilde{y}\} \times Z$ if $\pi_X(x) = \tilde{y}$.

The mapping π_X is called the *bundle projection* of the fiber bundle X, Y is called the *base manifold* of the fiber bundle, and Z is called the *fiber manifold*. The maps ϕ_y are called *local trivialisations* of X. Definition 3.13 implies that for every $y \in Y$, the *fiber* $\pi_X^{-1}(\{y\}) \subset X$ is diffeomorphic to Z. What's more, this fiber has an open neighborhood that is diffeomorphic to the Cartesian product of an open subset of Y and the manifold Z. One could say that a fiber bundle X consists of many copies of the fiber manifold Z, parameterized by elements of the base manifold Y.

An example of a fiber bundle is the Cartesian product manifold $X = Y \times Z$, for which $\pi_X(y, z) = y$, $Y_y = Y$ for all y and $\phi_y = \mathrm{id}_X$ for all y. We say that $X = Y \times Z$ is a *trivial* fiber bundle.

Not all fiber bundles are Cartesian products though. For instance if we take the 2-plane $\mathbb{R} \times \mathbb{R}$ and identify the points $(x, y) \equiv (x + 1, -y)$, then we obtain a manifold that is an $\mathbb{R}$-bundle over the circle $S^1 := \mathbb{R}/\mathbb{Z}$, which is not homeomorphic to $S^1 \times \mathbb{R}$. This manifold is called the *Möbius strip.* A slightly less known, but equally interesting example is the Hopf fibration:

Remark 3.14 (The Hopf fibration) Let

$$S^n := \{x \in \mathbb{R}^{n+1} \mid ||x|| = 1 \}$$

denote the n-dimensional unit sphere. H. Hopf constructed an explicit polynomial surjective mapping

$$\pi_{S^3} : S^3 \to S^2 \ ,$$

that satisfies definition 3.13 with $Z = S^1$. In other words, he showed that S^3 is an S^1-fiber bundle over S^2. The mapping π_{S^3} is known as the *Hopf fibration.* On the other hand, one can prove that S^3 is not homeomorphic to the Cartesian product $S^2 \times S^1$, for instance because S^3 is simply connected (every closed curve in S^3 can be continuously contracted to a point) and $S^2 \times S^1$ is not. △

Suppose that $y_1, y_2 \in Y$ and that $Y_{y_1} \cap Y_{y_2} \neq \emptyset$. Then the map $\phi_{y_1} \circ \phi_{y_2}^{-1}$ is a diffeomorphism of $(Y_{y_1} \cap Y_{y_2}) \times Z$ that, for each $y \in Y_{y_1} \cap Y_{y_2}$, sends $\{y\} \times Z$ to itself.

Definition 3.15 *Let X be a Z-fiber bundle over Y. We say that X is a* vector bundle *over Y if* **i)** *the fiber manifold Z is a vector space and* **ii)** *for each $y_1, y_2 \in Y$ and $y \in Y_{y_1} \cap Y_{y_2}$, the map $\phi_{y_1} \circ \phi_{y_2}^{-1}\big|_{\{y\} \times Z}$ is a linear isomorphism.*

One can now check quite easily that both the tangent bundle $X = TQ$ and the cotangent bundle T^*X of an n-dimensional C^k manifold are examples of a C^{k-1} vector bundle over $Y = Q$. In this case, the fibers are the linear spaces T_qQ and $(T_qQ)^*$ respectively, that is $Z = \mathbb{R}^n$, and the bundle projections π_{TQ} and π_{T^*Q} are as defined in section 3.7.

Some more terminology: If X is a fiber bundle over Y, then a smooth mapping $s : Y \to X$ with the property that $\pi_X \circ s = \mathrm{id}_Y$ is called a *section* of the fiber bundle X. A section assigns to every $y \in Y$ exactly one element in the fiber $\pi_X^{-1}(\{y\})$ over y. With this terminology, a vector field v on Q is precisely a section of TQ and a one-form becomes a section of T^*Q.

A *principal fiber bundle* is one for which the fiber manifold Z is actually a Lie group. We will study Lie groups in more detail in Chapter 6. Physicists call a section of a principal fiber bundle a *gauge*. Principal fiber bundles arise in several branches of physics, for instance in string theory.

3.13 Exercises

Exercise 3.1 (The circle as a submanifold of 2-space) *Consider the circle*

$$S = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \ \} \ .$$

- *Show that for every $(x, y) \in S$ there are an open neighborhood $U \subset \mathbb{R}^2$ of (x, y), an open interval $I \subset \mathbb{R}$ and a C^∞ function $f : I \to \mathbb{R}$ such that*

  $$S \cap U = \{(x, f(x)) \mid x \in I \ \} \text{ or } S \cap U = \{(f(y), y) \mid y \in I \ \} \ .$$

  Give U, I and f explicitly.

- *Show that for every $(x, y) \in S$ there are an open neighborhood $U \subset \mathbb{R}^2$ of (x, y), an open interval $I \subset \mathbb{R}$ and a C^∞ embedding $h : I \to \mathbb{R}^2$ such that $S \cap U = h(I)$. Give U, I and h explicitly.*

- *Show that for every $(x, y) \in S$ there are an open neighborhood $U \subset \mathbb{R}^2$ of (x, y) and a C^∞ submersion $g : U \to \mathbb{R}$ such that $S \cap U = g^{-1}(\{0_1\})$. Give U and g explicitly.*

- *Show that for every $(x, y) \in S$ there are an open neighborhood $U \subset \mathbb{R}^2$ of (x, y), an open neighborhood V of 0_2 in $\mathbb{R}^2$ and a C^∞ diffeomorphism $\Phi : U \to V$ such that $\Phi(S \cap U) = V \cap (\mathbb{R} \times \{0_1\})$. Give U, V and Φ explicitly.*

Exercise 3.2 (Some elementary topology) *Recall that a topological space X is called* compact *if every covering of X by open sets has a finite subcovering.*

Let X, Y be topological spaces and $h : X \to Y$ a continuous map. Show that if X is compact, then so is $h(X) \subset Y$. In particular, show that if X is compact and $Y \subset \mathbb{R}^n$ is a nonempty open subset, then there does not exist a homeomorphism $h : X \to Y$.

Remark 3.16 If $h : X \to Y$ is continuous and $h(X)$ is compact, then this does not imply that X is compact. A map $h : X \to Y$ with the special property that $h^{-1}(Z) \subset X$ is compact whenever $Z \subset Y$ is compact, is called *proper*.

Exercise 3.3 *Recall from exercise 3.1 that the circle*

$$S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \} \subset \mathbb{R}^2$$

is a one-dimensional C^∞ manifold. Prove that S^1 is compact. Use exercise 3.2 to show that S^1 is not homeomorphic to an open subset of $\mathbb{R}$.

4 Dynamics on Riemannian manifolds

Riemannian manifolds were introduced by Riemann in his Habilitationsthesis. His discovery was that by defining on every tangent space of a manifold an inner product, one can study concepts like curvature and shortest paths. In this chapter we will focus on the latter, the so-called geodesics.

4.1 Riemannian manifolds

We start with the notion of a *pseudo-inner product*:

Definition 4.1 *Let V be a finite-dimensional linear space. A pseudo-inner product on V is a mapping*

$$\beta : V \times V \to \mathbb{R}$$

with the following properties:

1. *Symmetry: $\beta(v,w) = \beta(w,v)$ for all $v,w \in V$.*
2. *Bilinearity: $\beta(v_1 + sv_2, w) = \beta(v_1, w) + s\beta(v_2, w)$ for all $v_1, v_2, w \in V$ and $s \in \mathbb{R}$.*
3. *Nondegeneracy: if $\beta(v,w) = 0$ for all $w \in V$, then $v = 0$.*

If in addition, we require

4. *Positivity: $\beta(v,v) \geq 0$,*

then β is called an inner product.

With this definition, we can define

Definition 4.2 *A C^k (pseudo-)Riemannian manifold is a C^k manifold Q for which a (pseudo-)inner product $\beta_q : T_qQ \times T_qQ \to \mathbb{R}$ is defined on each tangent space T_qQ, in such a way that this inner product depends in a C^k way on $q \in Q$.*

The family β_q of (pseudo-)inner products on T_qQ is called a *(pseudo-)Riemannian metric* on Q.

Recall that a coordinate chart $\kappa : q \mapsto q_\kappa \in U_\kappa \subset \mathbb{R}^n$ for Q induces a coordinate chart $D\kappa : \dot{q} \in T_qQ \mapsto (q_\kappa, \dot{q}_\kappa) \in U_\kappa \times \mathbb{R}^n$ for TQ. And, in order not to make the notation too heavy, let us suppress the subscript κ. In induced coordinates, the (pseudo-)inner product β_q on T_qQ takes the form

$$(\dot{q}^1, \dot{q}^2) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto \sum_{i,j=1}^{n} \beta_{ij}(q) \dot{q}^1_i \dot{q}^2_j \ .$$

The matrix $\beta(q)$ with coefficients $\beta_{ij}(q)$ is symmetric, nondegenerate and, if we require positivity, positive definite. Note also that it depends explicity on the chart κ. The requirement that $\beta(q)$ depends C^k on $q \in Q$ in Definition 4.2, just means that the functions $q \mapsto \beta_{ij}(q)$ on U are C^k.

Remark 4.3 (Quadratic forms) One can observe that a pseudo-inner product $\beta : V \times V \to \mathbb{R}$ on a linear space V defines a nondegenerate quadratic form $S_\beta : V \to \mathbb{R}$ by setting

$$S_\beta(v) := \frac{1}{2}\beta(v,v) \ .$$

On the other hand, given a nondegenerate quadratic form S on V, the unique pseudo-inner product β on V for which $S_\beta = S$, is given by

$$\beta_S(v,w) := \frac{1}{2}\left(S(v+w) - S(v) - S(w)\right) \ .$$

Hence, giving a nondegenerate quadratic form on a linear space is equivalent to giving a pseudo-inner product on that space. And: giving a pseudo-Riemannian metric β on a manifold Q is equivalent to giving a function $S : TQ \to \mathbb{R}$ of positions and velocities, for which for every $q \in Q$, the restriction $S|_{T_qQ}$ to T_qQ is a nondegenerate quadratic form. The requirement that the metric β is C^k therefore is equivalent to the requirement that $S_\beta : TQ \to \mathbb{R}$ is a C^k function.

This shows that the definition of a Riemannian metric given in Remark 2.3 is equivalent to definition 4.2. By "pseudo-Riemannian metric" we will sometimes mean β and sometimes also the corresponding kinetic energy $S : TQ \to \mathbb{R}$.

Remark 4.4 (The index of a metric) Let $\{e_1, \ldots, e_n\}$ be a basis for the n-dimensional linear space V. With respect to this basis, a symmetric bilinear form β on V has the form

$$\beta\left(\sum_{j=1}^n v_j^1 e_j, \sum_{k=1}^n v_k^2 e_k\right) = \sum_{j,k=1}^n \beta_{jk} v_j^1 v_k^2 \ ,$$

for certain coefficients $\beta_{jk} := \beta(e_j, e_k)$ with the property that $\beta_{jk} = \beta_{kj}$.

We define the *index* of β as the number of negative eigenvalues of β. This index is defined independent of the basis that we chose for V.

If β is a pseudo-Riemannian metric on a manifold Q, then each bilinear form β_q on T_qQ has its own index. This index is locally constant and hence constant on the connected components of Q. Therefore we can speak of *the* index of a pseudo-Riemannian metric on a connected manifold. A pseudo-Riemannian metric is a Riemannian metric if and only if its index is equal to zero.

4.2 Geodesics

Let β be a Riemannian metric (not a pseudo-Riemannian metric) on the manifold Q. If $\dot q \in T_qQ$, then we say that its *length* $||\dot q||_q$ is given by

$$||\dot q||_q := \sqrt{\beta_q(\dot q, \dot q)} \ .$$

Now if $\gamma : [a, b] \to Q$ is a smooth curve in Q, then $\frac{d\gamma(t)}{dt} \in T_{\gamma(t)}Q$, which allows us to define the length of the curve γ as

$$l(\gamma) := \int_a^b \left|\left|\frac{d\gamma(t)}{dt}\right|\right|_{\gamma(t)} dt \ .$$

The following proposition shows that this definition is independent of the parametrization of γ:

Proposition 4.5 *Let $\psi : [c, d] \to [a, b]$ be an orientation preserving diffeomorphism, that is: $\psi(c) = a, \psi(d) = b$ and $\frac{d\psi}{ds} > 0$. Then $l(\gamma \circ \psi) = l(\gamma)$.*

Proof:

$$l(\gamma \circ \psi) = \int_c^d \left|\left|\frac{d}{ds}(\gamma \circ \psi)(s)\right|\right|_{\gamma(\psi(s))} ds = \int_c^d \left|\left|\frac{d\gamma(t)}{dt}\right|\right|_{\gamma(t)}\Bigg|_{t=\psi(s)} \cdot \frac{d\psi(s)}{ds} ds = l(\gamma) \ ,$$

where the second equality follows from the chain rule and the bilinearity of $\beta_{\gamma(t)}$ and the last equality follows from a substitution of variables $t = \psi(s)$. $\square$

If $\frac{d\gamma(t)}{dt} \neq 0$ for all t, the ambiguity in the parameterization can be removed by requiring that γ be parameterized by *arclength*. We say that a curve $\gamma : [a, b] \to Q$ is parameterized by arclength if

$$l(\gamma|_{[a,t]}) = t - a \ .$$

This is of course equivalent to $||\frac{d\gamma(t)}{dt}||_{\gamma(t)} = 1$.

Note that $l(\gamma)$ is the action integral along γ of the Lagrangian $L : TQ \to \mathbb{R}$ defined by $L(q, \dot{q}) = \sqrt{\beta_q(\dot{q}, \dot{q})}$. We conclude that if $\gamma : [a, b] \to Q$ is a C^2 curve with $\gamma(a) = q_0, \gamma(b) = q_1$ is the shortest C^2 curve from q_0 to q_1, then γ must satisfy the Euler-Lagrange equation $[L]^\gamma = 0$.

Unfortunately though, the Lagrangian L is a bit nasty. First of all, it is not differentiable at the points $(q, \dot{q}) \in TQ$ for which $\dot{q} = 0$, because of the square root. A more serious problem is that L is degenerate: because $L(q, \lambda\dot{q}) = \lambda L(q, \dot{q})$ for all $\lambda > 0$, the derivative $\frac{\partial^2 L(q,\dot{q})}{\partial\dot{q}\partial\dot{q}}$ has $\dot{q}$ in its kernel. This implies that the Euler-Lagrange equations $[L]^\gamma = 0$ for L do not give rise to an explicit system of second order differential equations. Of course, this is closely related to the fact that any reparameterization of a shortest curve is also a shortest curve.

The solutions to $[L]^\gamma = 0$ become unique if we require that they are parameterized by arclength. Indeed, if γ is parameterized by arclength, then $L = 1$ along γ. Now recall the definition of the kinetic energy $T = \frac{1}{2}L^2$, i.e. $T(q, \dot{q}) = \frac{1}{2}\beta_q(\dot{q}, \dot{q})$ and observe that $dT = LdL$. This implies that $[L]^\gamma = [T]^\gamma$ if γ is parameterized by arclength.

T is of course a nondegenerate Lagrangian. Moreover, as T is a constant of motion for the equations $[T]^\gamma = 0$, the solution curves of $[T]^\gamma = 0$ for which $T = \frac{1}{2}$ are automatically solutions of $[L]^\gamma = 0$ parameterized by arclength. This inspires the following definition:

Definition 4.6 *A geodesic in a pseudo-Riemannian manifold* (Q, β) *is a solution to the Euler-Lagrange equations*

$$[S]^q = 0 \ ,$$

where the kinetic energy $S : TQ \to \mathbb{R}$ *is defined by* $S : \dot{q} \in T_qQ \mapsto \frac{1}{2}\beta_q(\dot{q}, \dot{q})$.

Note that in the definition of a geodesic we did not require the metric to be a Riemannian one. In a Riemannian manifold, the geodesics for which $S = \frac{1}{2}$ are those parameterized by arclength. Because S has the interpretation of kinetic energy, we also say that the geodesics describe the motion of a free particle in Q, where "free" refers to the absence of external forces.

Remark 4.7 (Unit tangent bundle) The flow of the Euler-Lagrange equations $[T]^q = 0$ on TQ is called the *geodesic flow*. Because S is a constant of motion for the geodesic flow, it leaves the so-called *unit tangent bundle*

$$(TQ)_1 := \{(q, \dot{q}) \in TQ \ |S(q, \dot{q}) = \frac{1}{2} \ \}$$

invariant. If β is a Riemannian metric, the intersection of $(TQ)_1$ with the tangent space T_qQ is diffeomorphic to a $n-1$-dimensional sphere. If moreover Q is compact, this implies that $(TQ)_1$ is a compact manifold, and hence the geodesic flow restricted to the unit tangent bundle is complete (i.e. solutions exist for all time).

One can prove that if the so-called *sectional curvatures* of a compact Riemannian manifold are negative, then the geodesic flow on $(TQ)_1$ is *mixing*. Loosely speaking, this means that the geodesic flow mixes or "stirs" the elements of $(TQ)_1$ very well.

Remark 4.8 (Einstein's general relativity) We say that a pseudo-Riemannian metric β on a manifold Q is a *Lorentzian metric* if its index is equal to one, and Q is then called a Lorentzian manifold. This is the setting of Einstein's theory of general relativity, in which Q is a 4-dimensional manifold called "space-time" and on which the Lorentzian metric describes a gravitational field.

We say that the curve $t \mapsto q(t), I \to Q$ in the Lorentzian manifold Q is *space-like* if $\beta(q(t))(dq(t)/dt, dq(t)/dt) > 0$ for all t, *time-like* if $\beta(q(t))(dq(t)/dt, dq(t)/dt) < 0$ for all t and *light-like* if $\beta(q(t))(dq(t)/dt, dq(t)/dt) = 0$ for all t.

Einstein's theory now says that a free massive relativistic particle is described by a time-like geodesic in Q for the Lorentzian metric, while light follows the light-like geodesics in space-time. No particle can travel on a space-like curves, as such a particle would move faster than light.

4.3 The geodesic equations

In local coordinates, the equations of motion for geodesics are found by writing out the Euler-Lagrange equations $[S]^q_i = 0$ for the kinetic energy

$$S(q, \dot{q}) = \frac{1}{2}\beta(q)(\dot{q}, \dot{q}) = \frac{1}{2}\sum_{i,j=1}^{n} \beta_{ij}(q)\dot{q}_i\dot{q}_j \ .$$

From the general formula (2.8) we see that these equations read $\frac{dq_l}{dt} = \dot{q}_l$ and

$$\sum_{m=1}^{n} \beta_{lm}(q) \frac{d\dot{q}_m}{dt} + \frac{1}{2} \sum_{j,k=1}^{n} \left(\partial_j \beta_{kl}(q) + \partial_k \beta_{jl}(q) - \partial_l \beta_{jk}(q) \right) \dot{q}_j \dot{q}_k \ . \tag{4.1}$$

If we now denote by

$$\beta^{lm}(q) = \left(\beta(q)^{-1} \right)_{lm}$$

the (l, m)-th element of the inverse of the matrix $\beta(q)$, then multiplying (4.1) by $\beta^{il}(q)$ and summing over l, we obtain the explicit formulas

$$\frac{dq_i}{dt} = \dot{q}_i \ , \ \frac{d\dot{q}_i}{dt} = - \sum_{j,k=1}^{n} \Gamma^i_{jk}(q) \dot{q}_j \dot{q}_k \ , \tag{4.2}$$

where

$$\Gamma^i_{jk}(q) := \frac{1}{2} \sum_{l=1}^{n} \beta^{il}(q) \left(\partial_j \beta_{kl}(q) + \partial_k \beta_{jl}(q) - \partial_l \beta_{jk}(q) \right) \ . \tag{4.3}$$

are called the *Christoffel symbols* of the metric.

The differential equations (4.2) have the following special property: if the curve $t \mapsto \gamma(t)$ solves (4.2) and $a \in \mathbb{R}$ is a constant, then the curve $t \mapsto \delta(t)$ defined by $\delta(t) := \gamma(at)$ also solves (4.2). It is straightforward to check this. A differential equation with this property is called a *spray*.

4.4 Isometries

A diffeomorphism $\Phi : Q \to Q$ of a pseudo-Riemannian manifold Q with pseudo-Riemannian metric β is called an *isometry* if

$$\beta_{\Phi(q)}(T_q\Phi \cdot \dot{q}_1, T_q\Phi \cdot \dot{q}_2) = \beta_q(\dot{q}_1, \dot{q}_2) \text{ for all } q \in Q \text{ and } \dot{q}_1, \dot{q}_2 \ .$$

It is easy to check that this implies that such a Φ preserves length: if $\gamma : [a, b] \to Q$ is a C^1 curve in Q, then $l(\Phi \circ \gamma) = l(\gamma)$.

Clearly, saying that Φ is an isometry is equivalent to saying that $T\Phi : TQ \to TQ$ preserves the kinetic energy function $S_\beta : TQ \to \mathbb{R}$, that is

$$S_\beta \circ T\Phi = S_\beta \ .$$

This in turn implies that if the curve γ in Q is a geodesic, then also $\Phi \circ \gamma$ is a geodesic. This follows for example from applying the following general theorem to $L = S_\beta$.

Theorem 4.9 (Symmetric Lagrangians) *Let Q be a C^2 manifold, $L : TQ \to \mathbb{R}$ a C^1 Lagrangian function and $\Phi : Q \to Q$ a C^2 diffeomorphism. Note that this implies that $T\Phi : TQ \to TQ$ is a C^1 diffeomorphism.*

Assume that $t \mapsto \gamma(t), I \to Q$ is a solution of the Euler-Lagrange equations $[L]^\gamma = 0$ and that L is invariant under $T\Phi$, that is

$$L \circ T\Phi = L \ .$$

Then $\Phi \circ \gamma$ solves the Euler-Lagrange equations as well, that is $[L]^{\Phi\circ\gamma} = 0$.

Proof: Recall the definition of the linear form

$$[L]^\gamma(t) : T_{\gamma(t)}Q \to \mathbb{R} \ ,$$

defined in local coordinates by formula (2.3). By Theorem 2.1, we have the identity

$$[L \circ T\Phi]^\gamma(t) = [L]^{\Phi\circ\gamma}(t) \circ T_{\gamma(t)}\Phi$$

for linear forms on $T_{\gamma(t)}Q$. This proves the theorem. $\square$

Theorem 4.9 expresses that if the Lagrangian L is symmetric, then this is reflected by a symmetry in the set of solutions of the Euler-Lagrange equations for L. In particular, if Φ is an isometry, then it sends geodesics to geodesics. An isometry is an example of a *symmetry*. Symmetries will be the topic of Chapter 5.

4.5 Excursion: the Jacobi metric

We will show in this section that solution curves of *natural* mechanical systems can be viewed as the geodesics of a special metric.

Let $L = S - V : TQ \to \mathbb{R}$ be a natural Lagrangian, which means that $S|_{T_qQ}$ is a nondegenerate quadratic form and V is constant on each T_qQ. In local coordinates $(q, \dot{q})$ for TQ this just means that $S(q, \dot{q}) = \frac{1}{2}\sum_{j,k=1}^n \beta_{jk}(q)\dot{q}_j\dot{q}_k$ and $V = V(q)$. In the same local coordinates, the Euler-Lagrange equations for such L read

$$\frac{d^2q_i}{dt^2} = -\sum_{j,k=1}^n \Gamma^i_{jk}(q)\frac{dq_j}{dt}\frac{dq_k}{dt} - \sum_{l=1}^n \beta^{il}(q)\frac{\partial V(q)}{\partial q_l} \ , \tag{4.4}$$

with Γ^i_{jk} as in (4.3). We saw before that the total energy $E = T + V : TQ \to \mathbb{R}$ is a constant of motion for equations (4.4). We now have the following theorem that characterizes the solution curves of equations (4.4) in terms of the geodesics of a special metric:

Theorem 4.10 (Jacobi-Maupertuis principle) *Let $S : TQ \to \mathbb{R}$ be a smooth pseudo-Riemannian metric and let $V : Q \to \mathbb{R}$ be a smooth potential energy function. Let $t \mapsto q(t), I \to Q$ be a curve in Q such that $E\left(q(t), \frac{dq(t)}{dt}\right) = e \in \mathbb{R}$ and $V(q(t)) \neq e$ for all t. Then the map $t \mapsto s(t), I \to \mathbb{R}$ defined by*

$$s(t) = 2\int_0^t e - V(q(\tau))d\tau \ .$$

is a diffeomorphism onto its image J. We denote its inverse by $s \mapsto t(s), J \to I$.

Moreover, the curve $t \mapsto q(t)$ in Q is a solution to the Euler-Lagrange equation $[S - V]^q = 0$, if and only if the curve $s \mapsto q(t(s)), J \to Q$ is a geodesic of the "Jacobi metric"

$$\tilde{S} = (e - V)S \ .$$

Proof: Because $\frac{ds(t)}{dt} = e - V(q(t)) \neq 0$, the inverse function theorem guarantees that $t \mapsto s(t)$ is a diffeomorphism onto its image. Let us denote the reparametrized curve by $s \mapsto \underline{q}(s) := q(t(s))$, or equivalently, $\underline{q}(s(t)) = q(t)$.

We work in local coordinates on Q now. Differentiation of the identity $q(t) = \underline{q}(s(t))$ with respect to t twice leads to the identities $\frac{dq_i}{dt} = 2(e - V(\underline{q}))\frac{d\underline{q}_i}{ds}$ and $\frac{d^2 q_i}{dt^2} = 4(e - V(\underline{q}))^2 \frac{d^2 \underline{q}_i}{ds^2} - 4(e - V(\underline{q}))\frac{d\underline{q}_i}{ds} \sum_{l=1}^n \frac{\partial V(q)}{\partial q_l}\Big|_{q=\underline{q}} \cdot \frac{d\underline{q}_l}{ds}$. We conclude that the curve $t \mapsto q(t)$ solves equations (4.4) if and only if $s \mapsto \underline{q}(s)$ satisfies

$$\frac{d^2 \underline{q}_i}{ds^2} = -\sum_{j,k=1}^n \Gamma^i_{jk} \frac{d\underline{q}_j}{ds}\frac{d\underline{q}_k}{ds} + \frac{1}{e - V}\frac{d\underline{q}_i}{ds}\sum_{l=1}^n \frac{\partial V}{\partial q_l}\frac{d\underline{q}_l}{ds} - \frac{1}{4\left(e - V\right)^2}\sum_{l=1}^n \beta^{il}\frac{\partial V}{\partial q_l} \ .$$

Using that $e - V = \frac{1}{2}\sum_{j,k=1}^n \beta_{jk}\frac{dq_j}{dt}\frac{dq_k}{dt} = 2\left(e - V\right)^2 \sum_{j,k=1}^n \beta_{jk}\frac{d\underline{q}_j}{ds}\frac{d\underline{q}_k}{ds}$ along solutions, we then finally find that

$$\frac{d^2 \underline{q}_i}{ds^2} = -\sum_{j,k=1}^n \tilde{\Gamma}^i_{jk}\frac{d\underline{q}_j}{ds}\frac{d\underline{q}_k}{ds} \ ,$$

in which

$$\tilde{\Gamma}^i_{jk} = \Gamma^i_{jk} - \frac{1}{2(e - V)}\left(\delta_{ik}\frac{\partial V}{\partial q_j} + \delta_{ij}\frac{\partial V}{\partial q_k} - \beta_{jk}\sum_{l=1}^n \beta^{il}\frac{\partial V}{\partial q_l}\right) \ . \tag{4.5}$$

Incidentally, the $\tilde{\Gamma}^i_{jk}$ are also exactly the Cristoffel symbols of the metric

$$\tilde{S}(q, \dot{q}) = (e - V(q))S(q, \dot{q})$$

defined on the subset of $q \in Q$ for which $V(q) \neq e$. This last claim is easy to verify by a short computation. $\square$

When $E = S + V = e$, then the condition that $V \neq e$ is equivalent to the condition that $S \neq 0$. Hence the Jacobi-Maupertuis principle holds for curves that never have zero kinetic energy. When S defines a Riemannian metric, then $S(q, \dot{q}) \neq 0$ if and only if $S(q, \dot{q}) > 0$ if and only if $\dot{q} \neq 0$ and the condition that $e \neq V$ implies that $e - V > 0$. Hence the Jacobi metric $(e - V)S$ then is automatically a Riemannian metric.

4.6 Gradient vector fields

A pseudo-Riemannian metric on a manifold Q allows us to define the geodesic vector field on TQ, as was shown in the previous sections. Now we will show how one can use it to construct vector fields on Q itself.

First of all, we remark that a pseudo-inner product $\beta : V \times V \to \mathbb{R}$ on a linear space V allows us to define a linear isomorphism from V to its dual

$$V^* := \{\xi : V \to \mathbb{R} \mid \xi \text{ is a linear map } \} ,$$

namely the map

$$\tilde{\beta} : v \mapsto \beta(v, \cdot) = \langle w \mapsto \beta(v, w)\rangle, V \to V^* .$$

Indeed, the nondegeneracy of β implies that this map is injective and hence bijective.

If (Q, β) is a pseudo-Riemannian manifold and $f : Q \to \mathbb{R}$ a C^1 function on Q, then the derivative $df(q)$ of f at q is an element of $(T_qQ)^*$. Now we can apply the inverse of $\tilde{\beta}_q$ to $df(q)$ to obtain a vector in T_qQ. The result is called the *gradient vector field of f* with respect to the metric β, and denoted $\text{grad}_\beta f$. It is implicitly defined by the relation

$$\beta_q((\text{grad}_\beta f)(q), v) := df(q) \cdot v \text{ for all } v \in T_qQ .$$

In local coordinates,

$$\text{grad}_\beta f = \sum_{i=1}^{n} \left(\sum_{l=1}^{n} \beta^{il}(q) \frac{\partial f}{\partial q_l} \right) \frac{\partial}{\partial q_i}$$

This shows that $\text{grad}_\beta f$ is a C^k vector field if f is C^{k+1} and β is C^k.

The *gradient flow* of a C^1 function $f : Q \to \mathbb{R}$ is defined as the flow of minus the gradient vector field of f. The crucial remark is that if Q is a Riemannian manifold (not a pseudo-Riemannian manifold!), then every gradient flow on it possesses a Lyapunov function: let $t \mapsto q(t)$ be an integral curve of $-\text{grad}_\beta f$, i.e. assume that $q(t)$ satisfies

$$\frac{dq(t)}{dt} = -\text{grad}_\beta f(q(t)) .$$

Then it is easy to compute that

$$\frac{d}{dt} f(q(t)) = df(q(t)) \cdot \frac{dq(t)}{dt} = \beta_{q(t)}(\text{grad}_\beta f(q(t)), \frac{dq(t)}{dt}) = -||\text{grad}_\beta f(q(t))||^2_{q(t)} \leq 0 .$$

In other words, the value of f always decreases along the solution curves of minus the gradient vector field of f, the only exception being at the points where $\text{grad}_\beta f = 0$, i.e. the equilibria of the gradient vector field. Due to the nondegeneracy of β, these points coincide with the stationary points of f, i.e. the points where $df = 0$.

Remark 4.11 (Infinite-dimensional gradient flows) Infinite dimensional gradient vector fields arise in many important, and often quite surprising, branches of mathematics and physics. Examples include pattern formation and reaction-diffusion equations, Morse-theory and various branches of topological dynamics.

An important and somewhat easily understood example is the curve-shortening flow that is used to study the geodesics on a Riemannian manifold (Q, β). The Riemannian manifold on which the curve-shortening flow takes place is an infinite-dimensional "manifold of smooth curves γ in Q", endowed with an appropriate metric. On this infinite-dimensional manifold, the length functional $\gamma \mapsto l(\gamma)$ produces a gradient flow that evolves an initial curve γ_0 into a family of curves γ_s with $l(\gamma_{s_2}) < l(\gamma_{s_1})$ if $s_2 > s_1$. The hope then is that γ_s converges to a geodesic as $s \to \infty$. It turns out that this is more or less what happens.

Of course, the study of infinite-dimensional gradient flows is much more involved than the study of finite-dimensional ones.

4.7 Exercises

Exercise 4.1 (Geodesics in submanifolds) *Let $U \subset \mathbb{R}^n$ be an open subset. U can be considered a Riemannian manifold with the Euclidean inner product*

$$\langle \dot{q}^1, \dot{q}^2 \rangle = \sum_{j=1}^{n} \dot{q}^1_j \dot{q}^2_j \text{ for } \dot{q}^i \in \mathbb{R}^n \ .$$

Suppose that $Q \subset U$ is a C^2 submanifold of $\mathbb{R}^n$ of dimension $d < n$. Then Q carries the metric β that it inherits from the surrounding space U, given by

$$\beta_q(\dot{q}^1, \dot{q}^2) := \langle \dot{q}^1, \dot{q}^2 \rangle \ , \text{ for } \dot{q}^i \in T_q Q \subset \mathbb{R}^n \text{ and } q \in Q.$$

Define $S : U \times \mathbb{R}^n \to \mathbb{R}$ by $S(q, \dot{q}) := \frac{1}{2}\langle \dot{q}, \dot{q} \rangle$.

- *Use the ideas of Section 2.7 to prove that $\gamma : I \to Q$ is a geodesic for the inherited metric if and only if for all $t \in I$,*

$$[S]^\gamma(t)|_{T_{\gamma(t)}Q} = 0$$

- *Prove that $[S]^\gamma(t) = \gamma''(t)$.*
- *Show that γ is a geodesic of the inherited metric if and only if for all $t \in I$, $\gamma''(t) \in \mathbb{R}^n$ is perpendicular to $T_{\gamma(t)}Q$.*

Exercise 4.2 (Geodesics on the sphere) *Denote by S^n the n-dimensional sphere:*

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

Prove that for every $x \in S^n$ and every nonzero $\dot{x} \in \mathbb{R}^n$ with $\langle x, \dot{x} \rangle = 0$, the curve

$$\gamma_{x,\dot{x}}(t) := \cos(||\dot{x}||t)x + \frac{\sin(||\dot{x}||t)}{||\dot{x}||}\dot{x}$$

is a geodesic with $\gamma_{x,\dot{x}}(0) = x, \gamma'_{x,\dot{x}}(0) = \dot{x}$.

Exercise 4.3 *[The hyperbolic half plane] I copied this exercise from the lecture notes on Classical Mechanics of J.J. Duistermaat.*

Let $H = \{z \in \mathbb{C} \mid \mathrm{Im}\ z > 0\}$ be the complex upper half plane, which can be viewed as an open subset of $\mathbb{R}^2$ by identifying $x + iy \in \mathbb{C}$ with $(x, y) \in \mathbb{R}^2$. Moreover, let $SL(2, \mathbb{R})$ be the group of 2×2 matrices A with real coefficients and $\det A = 1$. *For every $A \in SL(2, \mathbb{R})$, we define the fractional linear transformation $\Phi_A : \mathbb{C} \to \mathbb{C}$ by*

$$\Phi_A(z) := \frac{az + b}{cz + d} \text{ if } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} .$$

Prove the following results:

- *Φ_A is complex differentiable with derivative $\Phi'_A(z) = (cz + d)^{-2}$ and maps $\Phi_A(H) \subset H$. Moreover, $\Phi_I = \mathrm{id}_{\mathbb{C}}$ and $\Phi_A \circ \Phi_B = \Phi_{AB}$. Φ_A is a diffeomorphism of H with $(\Phi_A)^{-1} = \Phi_{A^{-1}}$.*
- *For every $z \in H$ and nonzero $v \in \mathbb{C}$ there is exactly one $A \in SL(2, \mathbb{R})/\{\pm 1\}$ with $\Phi_A(z) = i$ and $\Phi'(z)v$ a positive multiple of i.*
- *Let β be a Riemannian structure on H with the property that every fractional linear transformation Φ_A is an isometry of β. Then β is uniquely defined by β_i by*

  $$\beta_z(v, v) = \beta_i(\Phi'_A(z)v, \Phi'_A(z)v) \ ,$$

  for every $A \in SL(2, \mathbb{R})$ with $\Phi_A(z) = i$.
- *This defines a Riemannian structure on H for which every fractional linear transformation is an isometry if and only if*

  $$\beta_i(v, v) = \beta_i(\Phi'_A(i)v, \Phi'_A(i)v)$$

  for all $A \in SL(2, \mathbb{R})$ with $\Phi_A(i) = i$. Equivalently, there is a $c > 0$ such that $\beta_i(v, v) = c|v|^2$. Choose $c = 1$. In this way, we get

  $$\beta_z(v, v) = (\mathrm{Im}\ z)^{-2}|v|^2 \ , \ z \in H, v \in \mathbb{C} \ .$$
- *The reflection $S : x + iy \mapsto -x + iy$ is an isometry of this metric. If γ is a geodesic with $\gamma(0) = i$ and $\gamma'(0) = i$, then $\delta := S \circ \gamma$ is also a geodesic with $\delta(0) = i$ and $\delta'(0) = i$. We have that $\delta = \gamma$, i.e. $\gamma(t) = iy(t)$ for some positive real valued function $y(t)$. The condition that the geodesic is parameterized by arclength leads to the conclusion that $y'/y = 1$, so that $\gamma(t) = ie^t$.*
- *Every geodesic parametrized by arclength is of the form $\Phi_A \circ \gamma$ with γ as above. If $c \neq 0$ and $d \neq 0$ respectively, then*

  $$\lim_{t\to\infty} \delta(t) = \frac{a}{c} \text{ and } \lim_{t\to-\infty} \delta(t) = \frac{b}{d} \ .$$

  If $c \neq 0$ and $d \neq 0$, then the orbit of δ is a half circle with its center on the real axis. If $c = 0$ or $d = 0$, then the orbit of δ is a vertical half line.

Remark 4.12 H is called the *hyperbolic half plane.* The *hyperbolic metric* β makes it a surface of negative "curvature". The hyperbolic half plane is a standard example of a non-Euclidean geometry.

Exercise 4.4 (Einstein's special relativity) *In Einstein's theory of special relativity, one introduces the so-called Lorentzian space-time. Let us consider the case of one space-dimension, i.e. space-time is* $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$*. Let us denote its elements by* (t, x)*, that have the interpretation of the time- and space-coordinates respectively.*

Let $c > 0$ *be a real constant, with the interpretaton of the speed of light, and let* $\alpha > 0$ *be another constant, to be determined later. For* $(t, x) \in \mathbb{R}^2$*, let us now define on the tangent space* $T_{(t,x)}\mathbb{R}^2 \cong \mathbb{R}^2$*, the bilinear form*

$$\beta_{(t,x)} : ((\dot{t}^1, \dot{x}^1), (\dot{t}^2, \dot{x}^2)) \mapsto \alpha(-c^2 \dot{t}^1 \dot{t}^2 + \dot{x}^1 \dot{x}^2) \ .$$

- *Show that* β *defines a pseudo-Riemannian metric of index* 1*. It is called the* Lorentz metric*.*

- *Write down the geodesic equations that are defined by this metric. Show that the solution curves are straight lines. What distinguishes the time-like, light-like and space-like geodesics?*

- *Denote by* $S : T(\mathbb{R}^2) \to \mathbb{R}$ *the kinetic energy:* $S(t, x, \dot{t}, \dot{x}) = \frac{1}{2}\alpha(-c^2\dot{t}^2 + \dot{x}^2)$*. Let us parameterize a time-like geodesic by time, i.e. consider the time-like geodesic* $s \mapsto (s, sv)$*. Our experience tells us that in the "classical limit" where* $|v|$ *is small, the kinetic energy* S *should of course increase like* $\frac{1}{2}mv^2$*. Show that this implies that* $\alpha = m$ *and that* $S = -\frac{1}{2}mc^2\,(1 - (v/c)^2)$*.*

- *For simplicity, assume now that the speed of light is* $c = 1$*. A linear map* $L : \mathbb{R}^2 \to \mathbb{R}^2$ *is called a* Lorentz transformation *if it is an isometry of the Lorentz metric and the collection of all Lorentz transformation is called the* Lorentz group*. Show that a linear map* $L : \mathbb{R}^2 \to \mathbb{R}^2$ *is a Lorentz transformation if and only if its matrix has the form*

$$\text{mat } L = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

  with $A^2 - C^2 = 1$*,* $D^2 - B^2 = 1$ *and* $AB - CD = 0$*. Show that this is true if and only if there is a* $\phi \in \mathbb{R}$ *such that* $A = \pm\cosh\phi, B = \pm\sinh\phi, C = \pm\sinh\phi$ *and* $D = \pm\cosh\phi$*, while at the same time sign*$(AB) =$ *sign*(CD)*.*

- *Assume again that* $c = 1$*. Show that a Lorentz transformation sends the light-cone* $|t| = |x|$ *to itself. Similarly, show that a Lorentz transformation sends the hyperbolas* $t^2 - x^2 = E \neq 0$ *to themselves.*

Exercise 4.5 *Let* $S(q, \dot{q}) = \frac{1}{2}\sum_{j,k=1}^{n} \beta_{jk}(q)\dot{q}_j\dot{q}_k$*. Prove that the Cristoffel symbols of the metric* $\tilde{S}(q, \dot{q}) = (e - V(q))S(q, \dot{q})$ *are given by (4.5)*

Exercise 4.6 *We say that two elements $v, w \in V$ of an inner product space V are perpendicular if $\langle v, w \rangle = 0$.*

Let $f : Q \to \mathbb{R}$ be a C^1 function on a C^1 Riemannian manifold Q with Riemannian metric β. Let $q \in Q$, $f(q) = 0$ and assume that f is a submersion at q.

Prove that $(\mathrm{grad}_\beta f)(q) \in T_q Q$ is perpendicular to the tangent space $T_q f^{-1}(\{0\}) \subset T_q Q$ with respect to the inner product β_q on $T_q Q$.

5 Dynamical systems with symmetry

Many mechanical systems and differential equations in applications are symmetric, for instance under reflections, rotations, translations, etc. Symmetry is closely related to group theory. In this chapter, we will introduce Lie groups, and we will investigate their importance for and influence on dynamical systems.

Although this chapter will be more about mathematics than about mechanics, I hope it will make clear the relevance of Lie groups in mechanics.

5.1 Symmetry

Definition 5.1 *Let Q be a C^k manifold, $v : Q \to TQ$ a C^{k-1} vector field and $\Phi : Q \to Q$ a C^k diffeomorphism. We say that Φ is a* symmetry *of the vector field v if $\Phi_* v = v$.*

In other words, a symmetry conjugates a vector field to itself. Symmetries have the following important dynamical property:

Proposition 5.2 *Let v be a C^1 vector field on Q. If $\Phi : Q \to Q$ is a C^k symmetry of v and $\gamma : I \to Q$ is an integral curve of v, then also $\Phi \circ \gamma$ is an integral curve of v.*

Proof: This simply follows because Φ maps integral curves of v to integral curves of $\Phi_* v$. □

A crucial observation is that the collection of all the symmetries of a vector field form a *group.*

Theorem 5.3 *The collection of C^k symmetries of a C^{k-1} vector field v on a C^k manifold Q form a group, with multiplication defined by composition of maps. This group is called the symmetry group of v.*

Proof: First of all, id_Q fulfills the role of identity element.

Secondly, if Φ is a C^k diffeomorphism, then so is Φ^{-1}. The chain rule implies that $(\Phi^{-1})_*(\Phi_* v) = (\Phi^{-1} \circ \Phi)_* v = v$, so that if $\Phi_* v = v$, then $(\Phi^{-1})_* v = v$, that is if Φ is a C^k symmetry then Φ^{-1} is as well.

Similarly, if Φ and Ψ are C^k diffeomorphisms, then so is the composition $\Psi \circ \Phi$, and the identity $(\Psi \circ \Phi)_* v = \Psi_* \Phi_* v$ implies that if Ψ and Φ are both symmetries, then so is their composition. □

Remark 5.4 (Groups) In case you forgot, let me refresh your memory. A group is a set G in which a multiplication

$$(g_1, g_2) \mapsto g_1 g_2 \ , \ \ G \times G \to G$$

is defined, for which the following is true:

1. Multiplication is associative, that is $(g_1 g_2) g_3 = g_1 (g_2 g_3)$ for all $g_1, g_2, g_3 \in G$.

2. There is a unique *identity* element, denoted e (from the german word *Einheit*), which has the property that $eg = ge = g$ for all $g \in G$.

3. For every $g \in G$ there is a unique element $g^{-1} \in G$, called the inverse of g, for which $gg^{-1} = g^{-1}g = e$.

A group G is called *Abelian* or *commutative* if $g_1 g_2 = g_2 g_1$ for all $g_1, g_2 \in G$. Certainly not all groups are Abelian.

Simple examples of finite Abelian groups are $\mathbb{Z}/n\mathbb{Z}$ with addition modulo n and $(\mathbb{Z}/p\mathbb{Z})^* = \mathbb{Z}/p\mathbb{Z} - \{0\}$ (p prime) with multiplication. Examples of finite non-Abelian groups are the *polyhedral groups*, the symmetry groups of the platonic solids. An example of a countably infinite Abelian group is $\mathbb{Z}$ with addition.

5.2 Lie groups

Let us forget about mechanics for a while and think about groups and symmetry. In fact, in this section we shall be interested in groups that are at the same time differentiable manifolds. Such groups are called Lie groups:

Definition 5.5 *A C^k* Lie group *is a group G that is at the same time a C^k differentiable manifold such that*

1. *The inversion $G \to G$, $g \mapsto g^{-1}$ is a C^k map.*

2. *The multiplication $G \times G \to G$, $(g, h) \mapsto gh$ is a C^k map.*

If the group G is finite or countable, then we usually think of it as a C^∞ Lie group of dimension zero. This is done by giving it the topology in which every subset is an open set and defining for every $g \in G$ the chart $\kappa_g : \{g\} \to \mathbb{R}^0, g \mapsto 0_0$.

Example 5.6 (Simple Lie groups) A very simple example of a Lie group of dimension one is $\mathbb{R}\backslash\{0\}$ with the multiplication $(a, b) \mapsto ab$ that it inherits from $\mathbb{R}$. Being an open subset of $\mathbb{R}$, this is clearly a C^∞ manifold of dimension 1. Moreover, the multiplication $(a, b) \mapsto ab, \mathbb{R}\backslash\{0\} \times \mathbb{R}\backslash\{0\} \to \mathbb{R}\backslash\{0\}$ and the inversion $a \mapsto \frac{1}{a}, \mathbb{R}\backslash\{0\} \to \mathbb{R}\backslash\{0\}$ are obviously C^∞ maps. This shows that $\mathbb{R}\backslash\{0\}$ is a C^∞ Lie group of dimension one.

In the same way, $\mathbb{C}\backslash\{0\}$ is a C^∞ Abelian Lie group of dimension two. And the nonzero quaternions $\mathbb{H}\backslash\{0\}$ form an example of a non-Abelian C^∞ Lie group of dimension 4. $\triangle$

The most important examples of Lie groups are the matrix Lie groups, that is the subgroups of the *general linear group*

$$GL(n, \mathbb{R}) = \{A \in \mathbb{R}^{n\times n} \mid \det A \neq 0\} .$$

Being an open subset of $\mathbb{R}^{n\times n} \cong \mathbb{R}^{n^2}$, $GL(n, \mathbb{R})$ is an n^2-dimensional C^∞ manifold. Since matrix multiplication is a polynomial map and, by Cramer's rule, matrix inversion a rational map, $GL(n, \mathbb{R})$ is indeed a C^∞ Lie group. As we all know, the multiplication in

$GL(n.\mathbb{R})$ is noncommutative.

A subgroup $H \subset G$ of a Lie group G that is at the same time a submanifold of G, is of course a Lie group itself. Examples of Lie subgroups of $GL(n, \mathbb{R})$ are the *special linear group*

$$SL(n, \mathbb{R}) = \{A \in GL(n, \mathbb{R}) \mid \det A = 1\} \ ,$$

the orthogonal group

$$O(n, \mathbb{R}) = \{A \in GL(n, \mathbb{R}) \mid AA^* = \mathrm{id}\}$$

and the *special orthogonal group*

$$SO(n, \mathbb{R}) = \{A \in GL(n, \mathbb{R}) \mid AA^* = \mathrm{id} \ , \ \det A = 1\} \ .$$

Here A^* denotes the matrix transpose of A.

5.3 Lie theory

In this section I will present some of the theory of abstract Lie groups. We will see that, as manifolds, Lie groups have several quite special properties.

We start by defining, for every $h \in G$, the following two mappings from G to G:

$$L_h : G \to G, g \mapsto hg \ ,$$

$$R_h : G \to G, g \mapsto gh \ .$$

L_h is called 'left-translation by h' or 'left multiplication by h' and R_h is called 'right-translation by h' or 'right-multiplication by h'. Being the restriction of the multiplication map $G \times G \to G$ to $\{h\} \times G$ and $G \times \{h\}$ respectively, these maps are smooth. Furthermore, $(L_h)^{-1} = L_{h^{-1}}$ and $(R_h)^{-1} = R_{h^{-1}}$, which proves that L_h and R_h are diffeomorphisms of G.

Being a differentiable manifold, G has a tangent bundle TG consisting of the tangent spaces T_gG ($g \in G$). The tangent space at the identity element e is called the *Lie-algebra* of G, denoted

$$\mathfrak{g} := T_eG \ .$$

We remark that left-multiplication by h sends e to h: $L_h : e \mapsto h$. Hence, $T_eL_h : \mathfrak{g} \to T_hG$. Differentiation of $L_{h^{-1}} \circ L_h = L_h \circ L_{h^{-1}} = \mathrm{id_G}$ at e and h respectively shows that

$$(T_eL_h)^{-1} = T_hL_{h^{-1}} \ ,$$

i.e. $T_eL_h : \mathfrak{g} \to T_hG$ is an isomorphism. This proves the following interesting topological fact about the vector bundle TG:

Proposition 5.7 *Let G be a C^k Lie group. Then the map*

$$l : TG \to G \times \mathfrak{g} \ , \ \dot{g} \in T_gG \mapsto (g, T_gL_{g^{-1}} \cdot \dot{g})$$

is a global C^{k-1} diffeomorphism. Thus, TG is a trivial vector bundle over G.

The diffeomorphism l is called the 'left-trivialisation of TG'. One of the consequences of the above proposition that is not hard to prove, is that on every C^1 Lie group it is possible to define a nonvanishing volume form, that is every C^1 Lie group is *orientable*. In particular, no Lie group is diffeomorphic to the Möbius strip!

In the same way as we defined the left-trivialisation, one can define the right-trivialisation, which is a global diffeomorphism as well:

$$r : TG \to G \times \mathfrak{g} \ , \ \dot{g} \in T_gG \mapsto (g, T_gR_{g^{-1}} \cdot \dot{g}) \ .$$

5.4 Left invariant vector fields

The next step is to consider vector fields on the Lie group G. In fact, given a vector $X \in \mathfrak{g}$, we can define a vector field $v^l_X : G \to TG$ on G by

$$v^l_X(g) := T_eL_g \cdot X \in T_gG \ .$$

Proposition 5.8 *The vector field v^l_X is* left-invariant*, that is for any $h \in G$, we have*

$$(L_h)_* v^l_X = v^l_X \ .$$

Proof: $((L_h)_*v^l_X)(g) = T_{(L_h)^{-1}(g)}L_h \cdot v^l_X((L_h)^{-1}(g)) = T_{h^{-1}g}L_h \cdot v^l_X(h^{-1}g) = T_{h^{-1}g}L_h \cdot T_eL_{h^{-1}g} \cdot X = T_eL_g \cdot X = v^l_X(g)$. □

Also, a left-invariant vector field is uniquely determined by its value $v(e) \in \mathfrak{g}$ at the identity, because left-invariance implies that $v(g) = ((L_g)_*v)(g) = T_eL_g \cdot v(e)$. Thus, we have the following correspondence:

Proposition 5.9 *Let G be a C^{k+1} Lie group. Denote by $\mathcal{X}^k_l(G) \subset \mathcal{X}^k(G)$ the vector space of left-invariant C^k vector fields on G. Then the linear map*

$$X \mapsto v^l_X \ , \ \mathfrak{g} \to \mathcal{X}^k_l(G)$$

is a bijection.

Some people actually define the Lie algebra of G as the space of left-invariant vector fields on G. Finally, the above correspondence allows us to define a bracket in $\mathfrak{g}$:

Definition 5.10 *Let G be a C^k Lie group and let $X, Y \in \mathfrak{g}$. Then we define the Lie bracket $[X, Y] \in \mathfrak{g}$ as*

$$[X, Y] := -[v^l_X, v^l_Y](e) \text{ for all } X, Y \in \mathfrak{g} \ .$$

Here the right hand side is minus the Lie bracket of the vector fields v^l_X and v^l_Y on G, evaluated at the identity. The minus sign is just due to convention. This definition is designed in such a way that the following proposition holds:

Proposition 5.11 *The map $X \mapsto v^l_X$ from $\mathfrak{g}$ to $\mathcal{X}^k(G)$ is a Lie-algebra anti-homomorphism, i.e.*

$$[v^l_X, v^l_Y] = -v^l_{[X,Y]} \ .$$

Here, the bracket on the left hand side denotes the Lie bracket of $\mathcal{X}^k(G)$ and the bracket on the right hand side is the Lie bracket of $\mathfrak{g}$ defined above.

Proof: By definition, both the vector field $v^l_{[X,Y]}$ and the Lie bracket $-[v^l_X, v^l_Y]$ take the value $[X, Y]$ in e. Because the Lie bracket of two left-invariant vector fields is again left invariant, both vector fields are left invariant. Because the value at e determines a left invariant vector field uniquely, these vector fields are equal. $\square$

The proposition immediately implies that the Lie bracket on $\mathfrak{g}$ inherits the properties of the Lie bracket for vector fields:

Proposition 5.12 *The above Lie bracket is anti-symmetric, that is*

$$[X, Y] = -[Y, X] \ ,$$

and satisfies the Jacobi identity

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0 \ .$$

Proof: By definition and the anti-symmetry of the Lie bracket of vector fields, $[X, Y] = -[v^l_X . v^l_Y](e) = [v^l_Y, v^l_X](e) = -[Y, X]$. Moreover,

$$[X, [Y, Z]] = [X, -[v^l_Y, v^l_Z](e)] = [X, v^l_{[Y,Z]}(e)] = -[v^l_X, v^l_{[Y,Z]}](e) = v^l_{[X,[Y,Z]]}(e) \ ,$$

and similarly for the terms $[Y, [Z, X]]$ and $[Z, [X, Y]]$, so that

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = v^l_{[X,[Y,Z]]+[Y,[Z,X]]+[Z,[X,Y]]}(e) = 0 \ ,$$

because of the Jacobi identity for vector fields. $\square$

This proves that the Lie bracket on $\mathfrak{g}$ is really a Lie bracket and that with this Lie bracket, the Lie algebra $\mathfrak{g}$ has really become a Lie algebra.

Remark 5.13 (Right invariance) A vector field v on G is called 'right-invariant' if $(R_h)_* v = v$ for all $h \in G$, and there is also a bijection between the Lie algebra $\mathfrak{g}$ of G and the space of right-invariant vector fields on G. To add to the confusion, some authors define the Lie algebra of G as the space of right-invariant vector fields on G.

5.5 Group actions

Before discussing some important examples, let us first give the general definition:

Definition 5.14 *Let G be a C^k Lie group and Q a C^k manifold. An* action *of G on Q is a C^k map*

$$a : G \times Q \to Q$$

that satisfies

$$a(e, q) = q \text{ for all } q \in Q \ , \tag{5.1}$$

$$a(g, a(h, q)) = a(gh, q) \text{ for all } g, h \in G \text{ and } q \in Q \ . \tag{5.2}$$

Despite the name, group actions have nothing to do with action integrals.

If $a : G \times Q \to Q$ is a C^k action of a C^k Lie group G on a C^k manifold Q, then let us denote, for $g \in G$, the C^k map

$$a_g : q \mapsto a(g, q) \ , \ Q \to Q \ .$$

The identities (5.1) and (5.2) just mean that $a_e = \mathrm{id}_Q$ and $a_g \circ a_h = a_{gh}$. Now observe that this implies that $a_g \circ a_{g^{-1}} = a_{g^{-1}} \circ a_g = \mathrm{id}_Q$, that is each a_g is a diffeomorphism. The action thus defines a map

$$g \mapsto a_g \ , \ G \to Diff^k(Q) \ ,$$

where $Diff^k(Q)$ denotes the collection of C^k diffeomorphisms of Q. Of course, $Diff^k(Q)$ is a group under composition of diffeomorphisms. The identity $a_g \circ a_h = a_{gh}$ just means that the above map is a group homomorphism!

Remark 5.15 (The diffeomorphism group) Let Q be a C^k manifold, with $k \geq 1$. Under certain conditions, for instance that Q be compact, it can be shown that the diffeomorphism group $Diff^k(Q)$ is actually an infinite dimensional Lie group - an infinite dimensional smooth manifold for which the group operations are differentiable maps. Unfortunately, it goes beyond this course to show how this works. As we will see in the next chapter, the diffeomorphism group plays an important role in fluid dynamics. △

Example 5.16 (Left and right multiplication) Let G be a C^k Lie group. Then it is easy to check that the C^k mappings

$$L : (h, g) \mapsto hg \text{ and } R : (h, g) \mapsto gh^{-1} \ , G \times G \to G$$

define actions of G on itself! L is called the action by left multiplication and R is called the action by right multiplication. △

Example 5.17 (The adjoint action and the adjoint representation) Another example from Lie group theory is the *adjoint action* of a C^k Lie group on itself, that is defined by

$$\mathbf{Ad} : G \times G \to G \ , \ (h, g) \mapsto hgh^{-1} \ .$$

The map

$$\mathbf{Ad}_h := \ g \mapsto hgh^{-1} \ , \ G \to G$$

is called *conjugation by* h. It is easy to check that $\mathbf{Ad}$ defines an action of G on itself, that is that it is a C^k map from $G \times G$ to G, that $\mathbf{Ad}_e = \mathrm{id}_G$ and that $\mathbf{Ad}_{h_1} \circ \mathbf{Ad}_{h_2} = \mathbf{Ad}_{h_1 h_2}$.

The next thing to observe is that, for every $h \in G$, $\mathbf{Ad}_h : e \mapsto e$ keeps the identity element fixed. Therefore, its tangent map is a linear map that sends $\mathfrak{g}$ to $\mathfrak{g}$:

$$\mathbf{ad}_h := T_e \mathbf{Ad}_h : \ \mathfrak{g} \to \mathfrak{g} \ .$$

Differentiation of $\mathbf{Ad}_h \circ \mathbf{Ad}_{h^{-1}} = \mathbf{Ad}_{h^{-1}} \circ \mathbf{Ad}_h = \mathrm{id}_G$ at e reveals that $\mathbf{ad}_h$ is invertible, its inverse being $\mathbf{ad}_{h^{-1}}$. In other words, $\mathbf{ad}_h$ is an automorphism of $\mathfrak{g}$. Furthermore, differentiating $\mathbf{Ad}_{h_1} \circ \mathbf{Ad}_{h_2} = \mathbf{Ad}_{h_1 h_2}$ at e leads to the conclusion that

$$\mathbf{ad}_{h_1} \circ \mathbf{ad}_{h_2} = \mathbf{ad}_{h_1 h_2} \ ,$$

This shows that

$$\mathbf{ad} : G \times \mathfrak{g} \to \mathfrak{g} \ , (g, X) \mapsto \mathbf{ad}_g(X)$$

defines an action of G on $\mathfrak{g}$ *by means of linear mappings.* An action of a Lie group G on a linear space V by means of linear mappings is called a *representation* of G in V, so that $\mathbf{ad}$ is a representation of G in its own Lie algebra $\mathfrak{g}$. It is called the *adjoint representation* of G.

The importance of the adjoint representation lies in the fact that it measures at the infinitesimal level the noncommutativity of the group G. Indeed, if G is an Abelian group, then $\mathbf{Ad}_h = \mathrm{id}_G$ for all $h \in G$, and hence $\mathbf{ad}_h = \mathrm{id}_{\mathfrak{g}}$ for all $h \in G$. This shows that whenever G is commutative, $\mathbf{ad}$ is the *trivial representation* of G in $\mathfrak{g}$. At the same time, if $\mathbf{ad} : G \to GL(\mathfrak{g})$ is not the trivial representation, then this means that G is noncommutative. △

5.6 Symmetry actions

We give an action a special name if it acts by means of symmetries of a vector field:

Definition 5.18 *Let $a : G \times Q \to Q$ be an action of G on Q and $v : Q \to TQ$ a vector field on Q. We say that a is a* symmetry action *for v if*

$$(a_g)_* v = v \ \ \textit{for all } g \in G \ .$$

If a is a symmetry action for v then we also say that v is G-equivariant. Note that probably it would have been more correct to call v "a-equivariant", but the above terminology is more or less standard.

A special role is played by the fixed point set of a group action:

Definition 5.19 *Let G be a C^k Lie group that acts on a C^k manifold Q by means of C^k diffeomorphisms. Then the fixed point set of G is defined as the collection of points in Q that are left fixed by this action:*

$$\text{Fix } G := \{q \in Q \mid a(g,q) = q \ \forall \ g \in G\} \ .$$

Again, it probably would have been better to speak of "Fix a", but we will stick to the convention.

The importance of this definition is expressed in the following important theorem:

Theorem 5.20 (The fixed point set of a symmetry group is flow-invariant) *Let $k \geq 1$, Q a C^k manifold and $v : Q \to TQ$ a C^{k-1} vector field on Q. Suppose that v is G-equivariant. Let $\gamma : I \to Q$ be an integral curve of v defined on an interval $I \subset \mathbb{R}$.*

If $\gamma(t_0) \in \text{Fix } G$ for some $t_0 \in I$, then $\gamma(t) \in \text{Fix } G$ for all $t \in I$.

Proof: Let $\gamma : I \to Q$ be an integral curve of v and $g \in G$ arbitrary. Then by Proposition 5.2, the curve $a_g \circ \gamma$ is an integral curve of v as well. Moreover, $(a_g \circ \gamma)(t_0) = \gamma(t_0)$ because $\gamma(t_0) \in \text{Fix } G$. But integral curves with the same initial condition are unique, and we conclude that $a_g \circ \gamma = \gamma$, i.e. $\gamma(t)$ is fixed by a_g. $\square$

5.7 Excursion: Time reversal symmetry

Recall Newton's equations of motion for a mechanical system subject to a force that depends only on the positions:

$$m_i \frac{d^2 q_i}{dt^2} = \phi_i(q) \ , \ i = 1, \ldots, n \ . \tag{5.3}$$

Here, q is element of an open subset U of $\mathbb{R}^n$. Let $t_1 < t_2$ and assume that $t \mapsto q(t) \in U$ is a C^2 solution curve of equation (5.3) defined for $t \in (t_1, t_2)$.

Then we can create a new curve in U by time reversal. That is, we define a new time $\tau := -t \in (-t_2, -t_1)$ and set

$$\tilde{q}(\tau) := q(-\tau) \in U \ .$$

Note that the image of the curve $\tau \mapsto \tilde{q}(\tau)$ is the same as that of the curve $t \mapsto q(t)$, but it is 'traversed' in the opposite direction.

Using equations (5.3) and the chain rule, it is easy to compute that the new curve satisfies

$$m_i \frac{d^2 \tilde{q}_i(\tau)}{d\tau^2} = m_i \frac{d^2}{d\tau^2} q_i(-\tau) = (-1)^2 m_i \left. \frac{d^2 q_i(t)}{dt^2} \right|_{t=-\tau} = \phi_i(q(t))|_{t=-\tau} = \phi_i(\tilde{q}(\tau)) \ , \ i = 1, \ldots, n \ .$$

This shows that, if $t \mapsto q(t)$ solves the equations (5.3), then $\tau \mapsto \tilde{q}(\tau) = q(-\tau)$ is a solution as well. Physically this means the following: if one makes a movie of $q(t)$ as it moves in U, then from watching this movie, one would not be able to decide if it was being played forward or backward! One says that Newton's equations (5.3) are *time reversible* or simply *reversible.*

Let us try to understand this reversibility geometrically. Recall that Newton's equations define a differential equation on TQ, which in local coordinates takes the form

$$\frac{dq_i}{dt} = \dot{q}_i \ , \ m_i \frac{d\dot{q}_i}{dt} = \phi_i(q) \ , \ i = 1, \ldots, n \ .$$

Solutions to Newton's equations are integral curves in TQ of the vector field $v : TQ \to T(TQ)$, which in local coordinates maps $U \times \mathbb{R}^n$ to $U \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ by

$$v(q, \dot{q}) = (q, \dot{q}, v_q(q, \dot{q}), v_{\dot{q}}(q, \dot{q})) \ , \tag{5.4}$$

where $v_q(q, \dot{q}) = \dot{q}$ and $v_{\dot{q}}(q, \dot{q}) = (\frac{1}{m_1}\phi_1(q), \ldots, \frac{1}{m_n}\phi_n(q))$. Now we can explain the reversibility of Newton's equations by defining the *velocity reversal* map

$$R : TQ \to TQ \ , \ \dot{q} \in T_qQ \to -\dot{q} \in T_qQ \ . \tag{5.5}$$

In coordinates it takes the form

$$R : U \times \mathbb{R}^n \to U \times \mathbb{R}^n, (q, \dot{q}) \mapsto (q, -\dot{q}) \ .$$

In local coordinates, the tangent map $TR : T(TQ) \to T(TQ)$ sends $(q, \dot{q}, \delta q, \delta \dot{q}) \mapsto (q, -\dot{q}, \delta q, -\delta \dot{q})$, so that it is not hard to check that

$$TR \circ v \circ R^{-1} : (q, \dot{q}) \mapsto (q, \dot{q}, -v_q(q, \dot{q}), -v_{\dot{q}}(q, \dot{q})) \ .$$

This proves that the vector field $R_* v$ on TQ is equal to minus the vector field v. Apparently, R acts a bit like a symmetry of v. In fact, R is called a *time reversal symmetry* of the vector field v:

Definition 5.21 *Let Q be a C^k manifold and $v : Q \to TQ$ be a C^{k-1} vector field on Q. A C^k diffeomorphism $R : Q \to Q$ is called a C^k* reversing symmetry *or* time reversal symmetry *if*

$$R_* v = -v \ .$$

To avoid confusion, we remark that in the case of Newton's equations, the vector field is defined on TQ, so in the above definition, Q should be replaced by TQ and TQ by $T(TQ)$.

The reversibility of Newton's equations can geometrically be understood from the following proposition, which shows that, just like symmetries, reversing symmetries map integral curves to integral curves. But only if we also reverse time:

Proposition 5.22 *Let Q be a C^k manifold, $v : T \to TQ$ a C^{k-1} vector field and $R : Q \to Q$ a C^k reversing symmetry of v. Assume that $\gamma : (t_1, t_2) \to Q$ is an integral curve of the vector field v, with $t_i \in \mathbb{R} \cup \{\pm\infty\}$ and $t_1 < t_2$.*

Then $R \circ \gamma \circ -\mathrm{id} : t \mapsto R(\gamma(-t))$, $(-t_2, -t_1) \to Q$ is also an integral curve of v.

Proof: Let γ be an integral curve of v, that is $\gamma'(t) = v(\gamma(t))$ for all $t \in (t_1, t_2)$. Then, by the chain rule, $(R \circ \gamma \circ -\mathrm{id})'(t) = T_{\gamma(-t)}R \cdot \gamma'(-t) \cdot (-1) = -T_{\gamma(-t)}R \cdot v(\gamma(-t)) = -T_{R^{-1}((R\circ\gamma)(-t))}R \cdot v(R^{-1}((R \circ \gamma)(-t))) = -(R_*v)((R \circ \gamma)(-t))) = v((R \circ \gamma \circ -\mathrm{id})(t))$. $\square$

Although Proposition 5.22 shows that there are similarities between symmetry and reversing symmetry, there are also large differences. For instance, the fixed point set of a time reversal symmetry in general is not flow-invariant. On the other hand, there is the following result, due to Birkhoff:

Theorem 5.23 (Birkhoff's reversing symmetry theorem) *Suppose that $R : Q \to Q$ is a reversing symmetry of the vector field $v : Q \to TQ$. Suppose that $R^2 = \mathrm{id}_Q$ and that $\gamma : I \to Q$ is a maximal integral curve of v with the property that it intersects Fix R at least twice. Then $I = \mathbb{R}$ and γ is periodic.*

Proof: Suppose that $\gamma(t_1), \gamma(t_2) \in \mathrm{Fix}\ R$ with $t_i \in I$ and $t_1 < t_2$. If $\gamma(t_1) = \gamma(t_2)$, then $I = \mathbb{R}$ and the solution is periodic. If not, define $\hat{\gamma} : [t_1, 2t_2 - t_1] \to Q$ by

$$\hat{\gamma}(t) = \begin{cases} \gamma(t) & \text{when } t \in [t_1, t_2)\ , \\ R(\gamma(2t_2 - t)) & \text{when } t \in [t_2, 2t_2 - t_1]\ . \end{cases}$$

Then $\lim_{t \downarrow t_2} \hat{\gamma}(t) = \lim_{t \uparrow t_2} \hat{\gamma}(t) = \hat{\gamma}(t_i)$, that is $\hat{\gamma}$ is continuous. By construction, it is also an integral curve of v. Moreover, $\hat{\gamma}(2t_2 - t_1) = \hat{\gamma}(t_1)$, that is $\hat{\gamma}$ is periodic. $\square$

A diffeomorphism $R : Q \to Q$ with the property that $R^2 = \mathrm{id}_Q$, is called an *involution*. The velocity reversal $R : TQ \to TQ$ defined in (5.5) is an example of an involution.

Note that if R is a time reversal symmetry of a vector field v, then $(R^2)_*v = (R_*)^2 v = -(-v) = v$. In other words, R^2 is a symmetry of v. Thus, time reversal symmetries do not form a group. But the collection of the symmetries *and* time reversal symmetries of v do form a group, which is called the *reversing symmetry group* of v. This name is a little misleading: the reversing symmetry group contains symmetries as well as reversing symmetries.

5.8 Exercises

Exercise 5.1 *Prove that the torus*

$$\mathbb{R}^n/\mathbb{Z}^n = \{x + \mathbb{Z}^n\ | x \in \mathbb{R}^n\}\ ,$$

with addition modulo $\mathbb{Z}^n$, is an Abelian C^∞ Lie group of dimension n.

Exercise 5.2 (Matrix Lie groups) *The collection of all invertible $n \times n$ matrices is called the n-th general linear group:*

$$GL(n, \mathbb{R}) := \{A \in \mathbb{R}^{n \times n}\ |\ \det A \neq 0\}\ .$$

Prove that:

- *$GL(n,\mathbb{R})$ is an open subset of $\mathbb{R}^{n\times n}$ and hence a manifold of dimension n^2.*
- *$GL(n,\mathbb{R})$ is a group under matrix multiplication.*
- *$GL(n,\mathbb{R})$ is a C^∞ Lie group, that is the multiplication and inversion are C^∞ maps.*
- *For $A \in GL(n,\mathbb{R})$, the tangent space $T_A GL(n,\mathbb{R})$ is isomorphic to $\mathbb{R}^{n\times n}$. Remark: The Lie algebra $T_I GL(n,\mathbb{R})$ is denoted $\mathfrak{gl}(n,\mathbb{R})$.*
- *For $A \in GL(n,\mathbb{R})$, left translation over A is the map $L_A : B \mapsto AB$.*
- *The tangent map $T_I L_A$ sends $E \in \mathfrak{gl}(n,\mathbb{R})$ to $AE \in T_A GL(n,\mathbb{R})$.*
- *For $E \in \mathfrak{gl}(n,\mathbb{R})$, the unique left-invariant vector field v_E^l on $GL(n,\mathbb{R})$ which takes the value $v_E^l(I) = E$, is given by $v_E^l(A) = AE$.*
- *The curve $t \mapsto A(t)$ in $GL(n,\mathbb{R})$ is an integral curve of v_E^l, if and only if $\frac{dA(t)}{dt} = A(t)\cdot E$.*
- *The integral curves $t \mapsto A(t)$ of the left invariant vector field v_E^l are given by*
$$A(t) = A(0)\cdot e^{tE} \ .$$
- *Let $E_1, E_2 \in \mathfrak{gl}(n,\mathbb{R})$. The Lie bracket $[E_1,E_2] := -[v_{E_1}^l, v_{E_2}^l](I)$ is given by the commutator*
$$[E_1,E_2] = E_1\cdot E_2 - E_2\cdot E_1 \ .$$

Exercise 5.3 (Identical particles) *Let $n \in \mathbb{N}$. We consider the group of permutations of n elements, that is the group*

$$S_n := \{\sigma : \{1,\ldots,n\} \to \{1,\ldots,n\} \mid \sigma \text{ is bijective } \} \ .$$

- *Show that the mapping*
$$\pi : S_n \times \mathbb{R}^n \to \mathbb{R}^n \ , \ (\sigma,(q_1,\ldots,q_n)) \mapsto (q_{\sigma(1)},\ldots,q_{\sigma(n)})$$
defines an action of S_n on $\mathbb{R}^n$ by means of linear automorphisms. We call it the representation of coordinate permutations.
- *Show that for all $\sigma \in S_n$, the tangent map $T(\pi_\sigma) : T\mathbb{R}^n \to T\mathbb{R}^n$ is given by*
$$T(\pi_\sigma) : (q,\dot{q}) \mapsto (\pi_\sigma\cdot q, \pi_\sigma\cdot\dot{q}) \ .$$
- *Show that the mapping*
$$\Pi : S_n \times T\mathbb{R}^n \to T\mathbb{R}^n \ , \ (\sigma,(q,\dot{q})) \mapsto T(\pi_\sigma)\cdot(q,\dot{q})$$
defines a representation of S_n in $T\mathbb{R}^n$.

- *Assume that $V : \mathbb{R}^n \to \mathbb{R}$ is an S_n-invariant potential energy, that is*
$$V(q_{\sigma(1)}, \dots, q_{\sigma(n)}) = V(q_1, \dots, q_n)$$
for all $\sigma \in S_n$. Prove that the equations of motion
$$\frac{dq_i}{dt} = \dot{q}_i \ , \ m_i \frac{d\dot{q}_i}{dt} = -\frac{\partial V(q)}{\partial q_i} \ .$$
are Π-equivariant, if and only if $m_i = m_j$ for all $1 \le i, j \le n$.

- *Assume that $m_i = m_j$ for all $1 \le i, j \le n$. Let $t \mapsto q(t), I \to \mathbb{R}^n$ be a solution of the equations of motion and let $1 \le k < l \le n$. Show that if $q_k(t_0) = q_l(t_0)$ and $\dot{q}_k(t_0) = \dot{q}_l(t_0)$, then $q_k(t) = q_l(t)$ for all $t \in I$.*

Exercise 5.4 (The Lie algebra $\mathfrak{so}(3, \mathbb{R})$) *The 3-dimensional special orthogonal group is defined as*
$$SO(3, \mathbb{R}) := \{A \in \mathbb{R}^{3\times3} \mid A \cdot A^* = I \ , \ \det A = 1\},$$
where A^* denotes the transpose of A. You may assume without proof that $SO(3, \mathbb{R})$ is a 3-dimensional Lie subgroup of the 9-dimensional Lie group $GL(3, \mathbb{R})$.

- *Show that the Lie algebra $\mathfrak{so}(3, \mathbb{R}) := T_I SO(3, \mathbb{R})$ consists of the skew-symmetric matrices:*
$$\mathfrak{so}(3, \mathbb{R}) := \{E \in \mathbb{R}^{3\times3} \mid E + E^* = 0 \ \} \ .$$

- *Show that the mapping*
$$\sigma : \begin{pmatrix} 0 & e_1 & e_2 \\ -e_1 & 0 & e_3 \\ -e_2 & -e_3 & 0 \end{pmatrix} \mapsto \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} \ , \ \mathfrak{so}(3, \mathbb{R}) \to \mathbb{R}^3$$
is a linear isomorphism.

- *Prove that*
$$\sigma([E_1, E_2]) = \sigma(E_2) \times \sigma(E_1) \ ,$$
where $[E_1, E_2] := E_1 \cdot E_2 - E_2 \cdot E_1$ is the matrix commutator and $a \times b \in \mathbb{R}^3$ denotes the usual cross product of 3-vectors, that is
$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{pmatrix} \ .$$

Remark: We say that σ is a Lie algebra anti-homomorphism from $\mathfrak{so}(3, \mathbb{R})$ with the matrix commutator to $\mathbb{R}^3$ with the cross product.

Exercise 5.5 *Let G be a C^1 Lie group and $X \in \mathfrak{g}$. Recall the definition of the unique left-invariant vector field $v^l_X : G \to TG$ that takes the value X at the identity and the left trivialisation $l : TG \to G \times \mathfrak{g}$. Prove that the composition $l \circ v^l_X : G \to G \times \mathfrak{g}$ sends*
$$l \circ v^l_X : g \mapsto (g, X) \ .$$

6 Mechanics on Lie groups

In some mechanical systems the configuration is naturally determined by an element of a Lie group G. Such mechanical systems are described by a differential equation on the tangent bundle TG of the group. The most famous example is the motion of a rigid body, which can be viewed as the geodesic motion on $SO(3,\mathbb{R})$ with respect to a left-invariant metric. Quite remarkably, various partial differential equations that describe fluid flows arise in exactly the same way. This chapter is devoted to these examples. It turns out that these examples also have a lot of symmetry, so that a lot of the theory of the previous chapters will come together now.

6.1 The rigid body

Sometimes, a Lie group is the natural configuration space of a mechanical system. An example is the motion of a free rigid body. The configuration of a rigid body is completely determined by a rotation matrix that describes the orientation of the body with respect to a reference configuration. That is, the configuration manifold is

$$SO(3,\mathbb{R}) = \{A : \mathbb{R}^3 \to \mathbb{R}^3 \mid A^*A = I \ , \det A = 1\} \ ,$$

the Lie group of all rotation matrices of three-dimensional space.

So how do we determine the equations of motion for the rigid body in the absence of external forces -such as gravity? In this case, the Lagrangian of the rigid body is the kinetic energy and this kinetic energy depends only on the velocity of the body. We remark that the velocity $\dot{A}$ of the rigid body with position $A \in SO(3,\mathbb{R})$ is an element of

$$T_A SO(3,\mathbb{R}) = \{\dot{A} \in \mathbb{R}^{3\times 3} \mid A^*\dot{A} + \dot{A}^*A = 0 \ \} \ .$$

To determine how much kinetic energy is stored in this velocity we need to specify a quadratic form S_A on $T_A SO(3,\mathbb{R})$ or, equivalently, an inner product β_A. Let us start by choosing some inner product β_I on the Lie algebra

$$T_I SO(3,\mathbb{R}) = \mathfrak{so}(3,\mathbb{R}) = \{X \in \mathbb{R}^{3\times 3} \mid X^* + X = 0\} \ .$$

We do not worry about the exact form of β_I now, as it will depend on the exact shape and properties of the rigid body.

Now let $t \mapsto A(t)$ be a curve in $SO(3,\mathbb{R})$ with $A(0) = I$ and $A'(0) = X \in \mathfrak{so}(3,\mathbb{R})$, then the instantaneous kinetic energy of the rigid body at $t = 0$ is $\frac{1}{2}\beta_I(X, X)$. But what if $A(0) = A \neq I$, and therefore $A'(0) \in T_{A(0)}SO(3,\mathbb{R})$? Then we apply the left multiplication $L_{A(0)^{-1}} : SO(3,\mathbb{R}) \to SO(3,\mathbb{R})$ to obtain a curve with $A(0)^{-1}A(t)|_{t=0} = I$ and $(A(0)^{-1}A(t))'|_{t=0} = A(0)^{-1}A'(0) \in \mathfrak{so}(3,\mathbb{R})$. The point is that this corresponds to choosing a new "identity" in $SO(3,\mathbb{R})$, i.e. the motions $A(t)$ and $A(0)^{-1}A(t)$ should have the same kinetic energy at $t = 0$. In other words, the kinetic energy is given by

$$T(A, \dot{A}) = \frac{1}{2}\beta_A(\dot{A}, \dot{A}) := \frac{1}{2}\beta_I(A^{-1}\dot{A}, A^{-1}\dot{A}).$$

This corresponds to choosing a metric β on $SO(3,\mathbb{R})$ that depends on β_I by

$$\beta_A(\dot{A}_1, \dot{A}_2) := \beta_I(A^{-1}\dot{A}_1, A^{-1}\dot{A}_2) \ .$$

It is easy to check that, by construction, this metric is *left-invariant*, that is it satisfies

$$(L_B)_*\beta = \beta \ , \tag{6.1}$$

for every $B \in SO(3,\mathbb{R})$. Here, the pushforward of the metric β under the left-multiplication by B, $(L_B)_*\beta$, is defined as

$$((L_B)_*\beta)_A(\dot{A}_1, \dot{A}_2) := \beta_{B^{-1}A}(T_A L_{B^{-1}} \cdot \dot{A}_1, T_A L_{B^{-1}} \cdot \dot{A}_2) \ .$$

Also, it is not hard to check that equation (6.1) holds if and only if $L_B : SO(3,\mathbb{R}) \to SO(3,\mathbb{R})$ is an isometry of β. In other words, a metric β is left-invariant if and only if for every $B \in SO(3,\mathbb{R})$, the left-multiplication L_B is an isometry of β.

The equations of motion of the rigid body are the equations for geodesics on $SO(3,\mathbb{R})$ with respect to this left invariant metric β.

6.2 Euler-Poincaré reduction

The rigid body motion is a special example of a variational principle on a Lie group with a left-invariant Lagrangian.

In general, let G be a Lie group and assume that on $\mathfrak{g}$ some smooth function $L_e : \mathfrak{g} \to \mathbb{R}$ is defined. This L_e extends to a Lagrangian L on TG by setting, for $\dot{g} \in T_gG$,

$$L(\dot{g}) := L_e(T_g L_{g^{-1}} \cdot \dot{g}) \ .$$

By construction this L is left invariant, that is $L \circ TL_h = L$ for all $h \in G$, and it is the unique left invariant Lagrangian that equals L_e on $\mathfrak{g}$.

For the rigid body, $G = SO(3,\mathbb{R})$ and $L_e(X) = \frac{1}{2}\beta_e(X,X)$. In the discussion that follows, L does not have to be a (pseudo-)Riemannian metric though.

Recall that a curve $t \mapsto \gamma(t), [a,b] \to G$ satisfies the Euler-Lagrange equations $[L]^\gamma(t) = 0$ if and only if it is stationary for the action integral

$$A(\gamma) = \int_a^b L\left(\gamma'(t)\right) dt \ ,$$

with respect to C^2 variations with fixed endpoints. Also, I would like to remark that Theorem (4.9) applies. But we can say more:

Theorem 6.1 (Euler-Poincaré reduction) *The following are equivalent:*

- *$\gamma : [a,b] \to G$ satisfies the Euler-Lagrange equations for the left-invariant Lagrangian L.*

- *The curve* $\lambda(t) = T_{\gamma(t)}L_{\gamma(t)^{-1}} \cdot \gamma'(t), [a,b] \to \mathfrak{g}$ *is stationary for the 'reduced' action-integral*
$$a(\lambda) = \int_a^b L_e(\lambda(t))dt$$
with respect to all variations of $t \mapsto \lambda(t)$ *of the form*
$$(t,\varepsilon) \mapsto \lambda(t) + \varepsilon\left(\frac{d\delta(t)}{dt} + [\lambda(t), \delta(t)]\right) ,$$
with zero endpoints $\delta(a) = \delta(b) = 0$.

Proof: We start by remarking that the left-invariance of the Lagrangian implies that
$$\int_a^b L(\gamma'(t))dt = \int_a^b L_e(\lambda(t))dt .$$
Assume the left hand side is stationary for variations with fixed endpoints. Thus, let $(t,\varepsilon) \mapsto \tilde{\gamma}(t,\varepsilon), [a,b] \times (-\varepsilon_0, \varepsilon_0) \to G$ be a variation of $t \mapsto \gamma(t) = \tilde{\gamma}(t,0)$ with fixed endpoints, i.e. $\tilde{\gamma}(a,\varepsilon) = \tilde{\gamma}(a,0)$ and $\tilde{\gamma}(b,\varepsilon) = \tilde{\gamma}(b,0)$ for all ε. Then $\tilde{\lambda} : [a,b] \times (-\varepsilon_0, \varepsilon_0) \to \mathfrak{g}$ defined by $\tilde{\lambda}(t,\varepsilon) := T_{\tilde{\gamma}(t,\varepsilon)}L_{\tilde{\gamma}(t,\varepsilon)^{-1}} \cdot \gamma'(t,\varepsilon)$ defines a variation of $t \mapsto \lambda(t) = \tilde{\lambda}(t,0)$. Define
$$\delta(t) := T_{\gamma(t)}L_{\gamma(t)^{-1}} \cdot \frac{\partial\tilde{\gamma}(t,0)}{\partial\varepsilon} \in \mathfrak{g} .$$
Then $\delta(a) = \delta(b) = 0$ because $\tilde{\gamma}$ has fixed endpoints. I claim that
$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \tilde{\lambda}(t,\varepsilon) = \delta'(t) + [\lambda(t), \delta(t)] . \tag{6.2}$$
To prove formula (6.2), let us pretend that G is a matrix Lie group. It turns out that formula (6.2) is also true in the case that G is a general abstract Lie group, where the bracket is the Lie bracket of $\mathfrak{g}$. The proof of this general result is a little cumbersome though, and I will skip it in these notes.

Anyway, when G is a matrix Lie group, then $\tilde{\lambda}(t,\varepsilon) = \tilde{\gamma}(t,\varepsilon)^{-1} \cdot \frac{\partial\gamma(t,\varepsilon)}{\partial t}$ and $\delta(t) = \tilde{\gamma}(t,0)^{-1} \cdot \frac{\partial\gamma(t,0)}{\partial\varepsilon}$. We hence find that
$$\frac{\partial\tilde{\lambda}(t,\varepsilon)}{\partial\varepsilon} = \frac{\partial}{\partial\varepsilon}\left(\tilde{\gamma}(t,\varepsilon)^{-1}\right) \cdot \frac{\partial\tilde{\gamma}(t,\varepsilon)}{\partial t} + \tilde{\gamma}(t,\varepsilon)^{-1}\frac{\partial^2\tilde{\gamma}(t,\varepsilon)}{\partial\varepsilon\partial t}$$
and
$$\delta'(t) = \frac{\partial}{\partial t}\left(\tilde{\gamma}(t,0)^{-1}\right) \cdot \frac{\partial\tilde{\gamma}(t,0)}{\partial\varepsilon} + \tilde{\gamma}(t,0)^{-1}\frac{\partial^2\tilde{\gamma}(t,0)}{\partial t\partial\varepsilon} .$$
Formula (6.2) now follows from consecutively evaluating these two identities in $\varepsilon = 0$, subtracting them and using the theorem for interchanging the order of differentiation. Moreover, one also needs to know that $\frac{\partial}{\partial t}\left(\gamma(t,\varepsilon)^{-1}\right) = -\gamma(t,\varepsilon)^{-1} \cdot \frac{\partial\gamma(t,\varepsilon)}{\partial t} \cdot \gamma(t,\varepsilon)^{-1}$ (and

similarly for the derivative with respect to ε) and that $[\lambda, \delta] = \lambda \cdot \delta - \delta \cdot \lambda$.

The left-invariance of L means that

$$A(\tilde{\gamma}(\cdot, \varepsilon)) = \int_a^b L(\tilde{\gamma}'(t, \varepsilon))dt = \int_a^b L_e(\tilde{\lambda}(t, \varepsilon))dt = a(\tilde{\lambda}(\cdot, \varepsilon)) \ .$$

Thus, if a is stationary at λ with respect to variations of the form $\delta' + [\lambda, \delta]$, then A is stationary with respect to variations with fixed endpoints.

Also, one can produce the variation $\lambda + \varepsilon\,(\delta' + [\lambda, \delta])$ of λ by choosing a variation $\tilde{\gamma}$ to γ for which $\frac{\partial \tilde{\gamma}(t,0)}{\partial \varepsilon} = T_{\gamma(t)} L_{\gamma(t)^{-1}} \cdot \delta(t)$, which implies that if γ is stationary for A, then so is λ for a with respect to the allowed variations. $\square$

The process of reducing the Euler-Lagrange equations on $TG \cong G \times \mathfrak{g}$ to the 'Euler-Poincaré equations' on $\mathfrak{g}$ is called Euler-Poincaré reduction, because $\mathfrak{g}$ has only half the dimension of TG. Once one has found a solution curve $t \mapsto \lambda(t) \in \mathfrak{g}$, one may try to *reconstruct* the solution curve γ in G. This is done by writing

$$\frac{dg(t)}{dt} = T_e L_{g(t)} \cdot \lambda(t) \ ,$$

which is a nonautonomous, first order differential equation on G. Given an initial condition $g(0) = g_0$, its solutions are unique.

6.3 The rigid body continued

According to the previous section the curve $t \mapsto A(t)$ is a geodesic in $SO(3, \mathbb{R})$ for the left-invariant metric induced by the inner product β_e on T_eG if and only if the curve $t \mapsto \Omega(t) := A^{-1}(t) \cdot \frac{dA(t)}{dt}$ in $\mathfrak{so}(3, \mathbb{R})$ is stationary for the action integral

$$\int_a^b \frac{1}{2} \beta_e(\Omega(t), \Omega(t))dt$$

with respect to all variations of $t \mapsto \Omega(t)$ of the form $t \mapsto \Omega(t) + \varepsilon\,(\Xi'(t) + [\Omega(t), \Xi(t)])$ with $\Xi(a) = \Xi(b) = 0$, i.e.

$$0 = \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \int_a^b \frac{1}{2} \beta_e(\Omega + \varepsilon(\Xi' + [\Omega, \Xi]), \Omega + \varepsilon(\Xi' + [\Omega, \Xi]))dt = \int_a^b \beta_e(\Omega, \Xi' + [\Omega, \Xi])dt$$

for all curves $t \mapsto \Xi(t)$ in $\mathfrak{so}(3, \mathbb{R})$ with fixed endpoints.

Let us make this discussion a bit more concrete by identifying $\mathfrak{so}(3, \mathbb{R})$ with $\mathbb{R}^3$ by means of the Lie-algebra anti-homomorphism of Exercise 5.4. In other words, let us write $\omega(t) = \sigma(\Omega(t))$ and $\xi(t) = \sigma(\Xi(t))$. Also, let us choose a specific inner product on $\mathfrak{so}(3, \mathbb{R})$. For instance if I is some symmetric positive definite 3×3 matrix, called the *inertia tensor*, then we can define

$$\beta_e(X, Y) := \langle I\sigma(X), \sigma(Y) \rangle = \langle Ix, y \rangle \ ,$$

if $x = \sigma(X)$ and $y = \sigma(Y)$. Then, using that $[\sigma^{-1}(\xi), \sigma^{-1}(\omega)] = \sigma^{-1}(\omega \times \xi)$, we find that for all curves $t \mapsto \xi(t) \in \mathbb{R}^3$ with $\xi(a) = \xi(b) = 0$, we have

$$0 = \int_a^b \langle I\omega(t), \frac{d\xi(t)}{dt} + \xi(t) \times \omega(t)\rangle = \int_a^b \langle -I\frac{d\omega(t)}{dt} + \omega(t) \times I\omega(t), \xi(t)\rangle dt \ .$$

Here we have performed a partial integration, using that $\xi(a) = \xi(b) = 0$, and the identity $\langle a, b \times c\rangle = \langle b, c \times a\rangle$. By the usual arguments, we have thus obtained the Euler equations for the rigid body:

$$I\frac{d\omega}{dt} = \omega \times I\omega \ . \tag{6.3}$$

6.4 Eulerian fluid equations

Some important fluid dynamical equations can be written in Euler-Poincaré form. The (Lie) group under consideration here is the group $Diff^{\infty}(X)$ of C^{∞} diffeomorphisms of a bounded open subset $X \subset \mathbb{R}^n$ with C^{∞} boundary ∂X. X has the interpretation of a fluid container. An element $x \in X$ has the interpretation of a fluid particle's reference position or 'fluid label'. Each diffeomorphism $\phi \in Diff^{\infty}(X)$ then represents a possible fluid configuration, where $\phi(x)$ has the interpretation of the position of the fluid element with fluid label x. The requirement that ϕ be a diffeomorphism is to prevent the fluid from developing shocks, particle collapse and other kinds of singularities.

We will say that $u : [a, b] \to Diff^{\infty}(X), t \mapsto u(t, \cdot)$ is a C^{∞} curve of diffeomorphisms of X if u is C^{∞} as a map from $[a, b] \times X$ to X. Such a curve describes a possible fluid motion, so that $t \mapsto u(t, x), [a, b] \to X$ describes the trajectory of the fluid element with label x. One often also requires that the fluid motions leave the volume-form $d_n x$ on X invariant, that is that the fluid is incompressible. In this case, the fluid motions are curves in $SDiff^{\infty}(X)$, the (Lie) group of volume preserving diffeomorphisms of X.

As usual, the Lagrangian fluid dynamical equations are obtained from Hamilton's principle: given a Lagrangian function $L = L(u, \dot{u})$, defined for diffeomorphisms $u : X \to X$ and velocities $\dot{u} := \frac{\partial u(t,\cdot)}{\partial t} : X \to \mathbb{R}^n$, it is postulated that $t \mapsto u(t)$ is a physical fluid flow if and only if

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \int_a^b L\left(\tilde{u}(t, \cdot, \varepsilon), \frac{\partial \tilde{u}(t, \cdot, \varepsilon)}{\partial t}\right) dt = 0$$

for every C^{∞} variation $\tilde{u} : [a, b] \times (-\varepsilon_0, \varepsilon_0) \to (S)Diff^{\infty}(X)$ of u (i.e. $\tilde{u}(t, x, 0) = u(t, x)$) with fixed endpoints $\tilde{u}(a, x, \varepsilon) = u(a, x)$, $\tilde{u}(b, x, \varepsilon) = u(b, x)$.

Usually $L(u, \dot{u})$ is the integral over X of some density depending on u, $\dot{u}$ and their x-derivatives, i.e.

$$L(u, \dot{u}) = \int_X l(u(x), Du(x), \ldots, D^k u(x); \dot{u}(x), D\dot{u}(x), \ldots, D^m \dot{u}(x)) \ d_n x \ ,$$

but in more exotic applications L can involve integrals over the boundary of X, for instance if surface tension is taken into account. The simplest possible Lagrangians do not depend on the $x-$derivatives of u and $\dot{u}$, i.e. $L(u,\dot{u}) = \int_X l(u,\dot{u})d_nx$. Such fluids are sometimes called nonviscous or 'ideal'. An example is $L(u,\dot{u}) = \int_X \frac{1}{2}\sum_{i=1}^n (\dot{u}_i(x))^2 d_nx$, the total kinetic energy of the fluid.

Let's note that an ideal Lagrangian has a remarkable symmetry: if $\phi : X \to X$ is any volume-preserving diffeomorphism, then

$$L(u\circ\phi^{-1},\dot{u}\circ\phi^{-1}) = \int_X l(u(\phi^{-1}(x)),\dot{u}(\phi^{-1}(x)))d_nx = \int_X l(u(\tilde{x}),\dot{u}(\tilde{x}))d_n\tilde{x} = L(u,\dot{u}) \ .$$

This simply follows from the substitution of variables $x = \phi(\tilde{x})$, using that $\det\left(\frac{\partial\phi(\tilde{x})}{\partial\tilde{x}}\right) = 1$ if ϕ is volume-preserving. This means that the ideal Lagrangian is right-invariant, i.e. invariant under the action $(u,\dot{u}) \mapsto (u\circ\phi^{-1},\dot{u}\circ\phi^{-1})$ of $SDiff^\infty(X)$ by right multiplication. This symmetry is called the *relabeling symmetry* of an ideal fluid and it expresses that fluid elements can be given another name or label according to any volume-preserving diffeomorphism ϕ^{-1} without changing the value of the Lagrangian. If u is itself volume-preserving, then $L(u,\dot{u}) = L(\mathrm{id},\dot{u}\circ u^{-1}) =: L_{\mathrm{id}}(\dot{u}\circ u^{-1})$.

We shall now sketch the derivation of the Euler-Poincaré equations for a right-invariant Lagrangian on $SDiff^\infty(X)$ from a variational principle, and in particular we shall derive the Euler equations for an ideal incompressible fluid.

Instead of deriving the Euler-Lagrange equations for u, we shall exploit the right-invariance of the Lagrangian L to derive Euler-Poincaré equations for the curve of Eulerian velocity fields $\lambda : [a,b]\times X \to \mathbb{R}^n$ defined as

$$\lambda = \text{“}\ \frac{\partial u}{\partial t}\circ u^{-1}\ \text{”} : (t,x) \mapsto \left.\frac{\partial u(t,\tilde{x})}{\partial t}\right|_{\tilde{x}=u(t,\cdot)^{-1}(x)} ,$$

or implicitly:

$$\lambda(t,u(t,x)) = \frac{\partial u(t,x)}{\partial t} \ .$$

$\lambda(t,x)$ simply has the interpretation of the velocity of the fluid element that is at position x at time t. Let us discuss some properties of λ. First of all, when $x\in\partial X$, then $u(t,x)\in\partial X$ for each t, whence $\lambda(t,u(t,x)) = \frac{\partial u(t,x)}{\partial t} \in T_{u(t,x)}\partial X$, so that we can conclude that λ is tangent to ∂X.

The second property follows from differentiating the identity $\lambda_j(t,u(t,x)) = \frac{\partial u_j(t,x)}{\partial t}$ with respect to x_k to obtain

$$\frac{\partial}{\partial t}\frac{\partial u_i(x,t)}{\partial x_k} = \sum_{l=1}^n \frac{\partial\lambda_i(t,u(x,t))}{\partial x_l}\frac{\partial u_l(t,x)}{\partial x_k} ,$$

i.e. $\frac{\partial}{\partial t}\frac{\partial u(x,t)}{\partial x} = \frac{\partial \lambda(t,u(x,t))}{\partial x}\frac{\partial u(t,x)}{\partial x}$. This and the fact that $\det \frac{\partial u(t,x)}{\partial x} = 1$ identically, leads to the conclusion that

$$0 = \frac{\partial}{\partial t}\det\left(\frac{\partial u(t,x)}{\partial x}\right) = \frac{d}{dh}\bigg|_{h=0}\det\left(\frac{\partial u(t+h,x)}{\partial x}\right) =$$
$$\frac{d}{dh}\bigg|_{h=0}\det\left(\frac{\partial u(t+h,x)}{\partial x}\left(\frac{\partial u(t,x)}{\partial x}\right)^{-1}\right)\det\left(\frac{\partial u(t,x)}{\partial x}\right) = \operatorname{tr}\left(\frac{\partial \lambda(t,u(t,x))}{\partial x}\right) ,$$

where we have used that $\frac{d}{d\varepsilon}\big|_{\varepsilon=0}\det(I+\varepsilon E) = \operatorname{tr}(E)$. We conclude that for each t, the vector field $x \mapsto \lambda(t,x)$ on X is divergence-free: $\operatorname{div}(\lambda) = \sum_{j=1}^n \frac{\partial \lambda_j}{\partial x_j} = 0$.

After these preparations, let $u : [a,b] \to SDiff^\infty(X)$ be a C^∞ curve of volume preserving diffeomorphisms and let $\tilde{u} : [a,b]\times(-\varepsilon_0,\varepsilon_0) \to SDiff^\infty(X)$ be a C^∞ variation of u with fixed endpoints, i.e. $\tilde{u}(t,x,0) = u(t,x)$, $\tilde{u}(a,x,\varepsilon) = u(a,x)$ and $\tilde{u}(b,x,\varepsilon) = u(b,x)$. Define $\tilde{\lambda}, \tilde{\delta} : [a,b]\times X\times(-\varepsilon_0,\varepsilon_0) \to \mathbb{R}^n$ implicitly by

$$\tilde{\lambda}(t,\tilde{u}(t,x,\varepsilon),\varepsilon) = \frac{\partial \tilde{u}(t,x,\varepsilon)}{\partial t} , \tag{6.4}$$
$$\tilde{\delta}(t,\tilde{u}(t,x,\varepsilon),\varepsilon) = \frac{\partial \tilde{u}(t,x,\varepsilon)}{\partial \varepsilon} , \tag{6.5}$$

and set $\delta(t,x) := \tilde{\delta}(t,x,0)$. Note that $\lambda(t,x) = \tilde{\lambda}(t,x,0)$ and that $\tilde{\delta}(a,x,\varepsilon) = \tilde{\delta}(b,x,\varepsilon) = 0$ because $\tilde{u}$ has fixed endpoints. Moreover, a similar argument as above shows that $x \mapsto \tilde{\lambda}(t,x,\varepsilon)$ and $x \mapsto \tilde{\delta}(t,x,\varepsilon)$ are tangent to ∂X and divergence-free.

Our main observation now is that differentiating (6.4) with respect to ε and (6.5) with respect to t, evaluating the resulting equations in $(t,x,\varepsilon) = (t,\tilde{u}(t,\cdot,0)^{-1}(\tilde{x}),0)$ and using that $\frac{\partial^2 \tilde{u}_i}{\partial t\partial\varepsilon} = \frac{\partial^2\tilde{u}_i}{\partial\varepsilon\partial t}$, we find (please perform this computation if you don't believe it!):

$$\frac{\partial \tilde{\lambda}_i(t,x,\varepsilon)}{\partial \varepsilon}\bigg|_{\varepsilon=0} = \frac{\partial \delta_i(t,x)}{\partial t} + \sum_{j=1}^n\left(\frac{\partial \delta_i(t,x)}{\partial x_j}\lambda_j(t,x) - \frac{\partial \lambda_i(t,x)}{\partial x_j}\delta_j(t,x)\right) .$$

This proves the "only if" part of:

Theorem 6.2 (Euler-Poincaré for incompressible ideal fluids) *Let $L = L(u,\dot{u}) = L_{\text{id}}(\dot{u}\circ u^{-1})$ be a right invariant Lagrangian function, defined for volume preserving diffeomorphisms $u : X \to X$ and vector fields $\dot{u} : X \to \mathbb{R}^n$. Then the following are equivalent:*

- *The curve $t \mapsto u(t)$ in $SDiff^\infty(X)$ is stationary for the action integral $\int_a^b L(u(t),\frac{\partial u(t)}{\partial t})dt$ with respect to volume-preserving variations with fixed endpoints.*
- *The curve of divergence-free vector fields $t \mapsto \lambda(t) := \frac{\partial u(t)}{\partial t}\circ u(t)^{-1}$ is stationary for the action integral $\int_a^b L_{\text{id}}(\lambda(t))dt$ with respect to variations $\tilde{\lambda}$ of λ of the form*

$$\tilde{\lambda} = \lambda + \varepsilon\left(\frac{\partial \delta}{\partial t} + \sum_{j=1}^n\left(\frac{\partial \delta}{\partial x_j}\lambda_j - \frac{\partial \lambda}{\partial x_j}\delta_j\right)\right) ,$$

such that δ is tangent to ∂X, divergence-free and $\delta(a,x) = \delta(b,x) = 0$.

I leave the "if" part of this theorem as a (rather difficult) exercise.

As an example, let us compute the Euler equations for an ideal fluid, that is the Euler-Poincaré equations for the kinetic Lagrangian

$$L(u,\dot u) = ||\dot u||^2_{L_2} := \int_X \frac{1}{2}\sum_{i=1}^n (\dot u_i(x))^2 d_n x = \int_X \frac{1}{2}\sum_{i=1}^n ((\dot u_i \circ u^{-1})(x))^2 d_n x \ .$$

According to Theorem 6.2, for all curves $(t,x)\mapsto \delta(t,x), [a,b]\times X \to \mathbb{R}^n$ with $\delta(t,\cdot)$ tangent to ∂X, $\operatorname{div}(\delta(t,\cdot)) = 0$ and $\delta(a,x) = \delta(b,x) = 0$, the curve $t\mapsto \lambda(t) := \frac{\partial u(t)}{\partial t}\circ u(t)^{-1}$ should then satisfy

$$\left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \int_a^b \int_X \frac{1}{2}\sum_{i=1}^n \left(\lambda_i(x) + \varepsilon\left[\frac{\partial \delta_i}{\partial t} + \sum_{j=1}^n\left(\frac{\partial\delta_i}{\partial x_j}\lambda_j - \frac{\partial \lambda_i}{\partial x_j}\delta_j\right)\right]\right)^2 dx = 0 \ .$$

Integration by parts yields that the left hand side of this expression is equal to

$$\int_a^b \int_X \sum_{i=1}^n \lambda_i \left(\frac{\partial \delta_i}{\partial t} + \sum_{j=1}^n\left(\frac{\partial\delta_i}{\partial x_j}\lambda_j - \frac{\partial \lambda_i}{\partial x_j}\delta_j\right)\right) dxdt = \tag{6.6}$$

$$\int_a^b \int_X \sum_{i=1}^n -\delta_i \left(\frac{\partial \lambda_i}{\partial t} + \sum_{j=1}^n \lambda_j \frac{\partial \lambda_i}{\partial x_j} + \lambda_i \operatorname{div}(\lambda)\right) + \left(\frac{1}{2}\sum_{j=1}^n \lambda_j^2\right)\operatorname{div}(\delta)\ dxdt =$$

$$\int_a^b \int_X \sum_{i=1}^n -\delta_i \left(\frac{\partial \lambda_i}{\partial t} + \sum_{j=1}^n \lambda_j \frac{\partial \lambda_i}{\partial x_j}\right) dxdt \ .$$

If $X\subset\mathbb{R}^n$ is bounded, then the space of vector fields on X tangent to ∂X is the L_2-orthogonal sum of the divergence free vector fields and the gradient vector fields. This is a consequence of Hodge-de Rham theory and unfortunately it would go too far to explain this in detail. Nevertheless, we have derived the Euler equations for an ideal incompressible fluid:

$$\frac{\partial \lambda_i}{\partial t} + \sum_{j=1}^n \lambda_j \frac{\partial \lambda_i}{\partial x_j} = -\frac{\partial p}{\partial x_i}\ ,\ \operatorname{div}(\lambda) = 0 \ .$$

The function p is implicitly defined by the condition $\operatorname{div}(\lambda) = 0$ and is called the *pressure* of the fluid.

The above derivation of Euler's equations for an incompressible fluid as the Euler-Poincaré equations for the geodesics on the special diffeomorphism group with a right-invariant metric, is due to Arnol'd.

6.5 Exercises

Exercise 6.1 (The rigid body dynamics) *Recall Euler's equations for the rigid body*

$$I\frac{d\omega}{dt} = \omega\times I\omega \ , \tag{6.7}$$

in which the inertia matrix I is symmetric and nondegenerate. Using the explicit Euler equations for the rigid body (6.3), show that the energy $E = \frac{1}{2}\langle I\omega, \omega\rangle$ is a constant of motion.

Can you give another argument why E is a constant of motion? Hint: View E as a function on $TSO(3,\mathbb{R})$.

Prove that the solutions of the rigid body equations lie on ellipsoids. Prove that they stay bounded and are defined for all time.

Show moreover that the function $J(\omega) = \langle I\omega, I\omega\rangle$ is conserved. Can you draw the joint level sets of E and J in $\mathbb{R}^3$?

Exercise 6.2 (The hyperbolic half plane again) *In this exercise we will again study the hyperbolic half plane $H = \{z \in \mathbb{C} \mid \mathrm{Im\ z} > 0\}$ and we will give it the structure of a noncommutative Lie group.*

- *For $x, y \in \mathbb{R}$, $y > 0$, let $\phi_{x,y} : \mathbb{R} \to \mathbb{R}$ be the affine mapping defined by*

  $$\phi_{x,y}(s) = x + ys \ .$$

  Show that $\phi_{x,y} \circ \phi_{\tilde{x},\tilde{y}} = \phi_{x+y\tilde{x},y\tilde{y}}$. In other words, the collection of affine mappings

  $$G = \{\phi_{x,y} \mid x, y \in \mathbb{R}, y > 0\}$$

  is closed under composition. Show that G is a group under composition. Show that the identity element of G is $\phi_{0,1}$ and that $\phi^{-1}_{x,y} = \phi_{-\frac{x}{y},\frac{1}{y}}$.

- *Prove that the mapping $x+iy \mapsto \phi_{x,y}$ from H to G is bijective. Prove that the unique group operation $* : H \times H \to H$ that makes this mapping a group isomorphism, is given by*

  $$(x + iy) * (\tilde{x} + i\tilde{y}) = (x + y\tilde{x}) + iy\tilde{y} \ .$$

  From now on, $*$ will be "the" multiplication on H and we will forget about the ordinary multiplication of complex numbers.

- *Prove that $*$ makes H into an analytic Lie group. Show that the identity element of H is i and that $(x + iy)^{-1} = \frac{-x}{y} + i\frac{1}{y}$.*

- *Write $z = x + iy$. Show that $L_{z^{-1}} : H \to H$ sends $\tilde{x} + i\tilde{y}$ to $\frac{\tilde{x}-x}{y} + i\frac{\tilde{y}}{y}$. Now show that $T_zL_{z^{-1}}$ maps $v = v_x + iv_y \in T_zH \cong \mathbb{C}$ to $\frac{1}{y}v := \frac{v_x}{y} + i\frac{v_y}{y} \in T_iH \cong \mathbb{C}$.*

- *On T_iH let us define an inner product by*

  $$\langle v, v\rangle := |v|^2 \text{ for } v \in T_iH \cong \mathbb{C} \ .$$

  Prove that the unique left-invariant metric β on H with the property that $\beta_i(v, v) = \langle v, v\rangle$ for all $v \in T_iH$, is given by the hyperbolic metric of Exercise 4.3, that is

  $$\beta_z(v, v) = \frac{1}{(\mathrm{Im}\ z)^2}|v|^2 \ .$$

- *For $\xi = \xi_x + i\xi_y \in T_iH \cong \mathbb{C}$, show that the unique left invariant vector field v_l^ξ on H is given by $v_l^\xi(z) = (\mathrm{Im}\ z)\xi$. Show that the Lie bracket in the Lie algebra T_iH is given by*
$$[\xi, \eta] := \eta_x\xi_y - \eta_y\xi_x \in \mathbb{R} \subset \mathbb{C} \ .$$

- *Let $t \mapsto z(t) \in H$ be a geodesic for the hyperbolic metric and write $z(t) = x(t) + iy(t)$. Show that $\lambda(t) := T_{z(t)}L_{z(t)^{-1}} \cdot \frac{dz(t)}{dt} = \frac{1}{y(t)}\left(\frac{dx(t)}{dt} + i\frac{dy(t)}{dt}\right)$ is a solution of the Euler-Poincaré equations*
$$\frac{d\lambda_x}{dt} = -\lambda_x\lambda_y \ , \ \frac{d\lambda_y}{dt} = \lambda_x^2 \ .$$
Show that the energy $\frac{1}{2}|\lambda|^2$ is a constant of motion for these equations. Draw the phase portrait.

Remark 6.3 One can repeat this procedure and define for $x \in \mathbb{R}^n$ and $y > 0$ the mappings $\phi_{x,y} : \mathbb{R}^n \to \mathbb{R}^n$, thus producing a "hyperbolic half-$(n+1)$-space".

Exercise 6.3 (Ideal compressible fluids) *Let $X \subset \mathbb{R}^n$ be a bounded open subset. For an "ideal compressible fluid" the Lagrangian function is the total kinetic energy*

$$L(u, \dot{u}) = \int_X \frac{1}{2}\sum_{i=1}^{n}(\dot{u}_i(x))^2 d_nx \ .$$

This Lagrangian is not right-invariant under the group $Diff^\infty(X)$. Why? For a C^∞ family of diffeomorphisms $(t, x) \mapsto u(t, x), [a, b] \times X \to X$ show that the following are equivalent:

- *u is stationary for the action integral $\int_a^b L\left(u(t), \frac{\partial u(t)}{\partial t}\right) dt$ with respect to variations with fixed endpoints.*
- *$(t, x) \mapsto u(t, x)$ solves the Euler-Lagrange equations $\frac{\partial^2 u_i(t,x)}{\partial t^2} = 0$ for all $i = 1, \ldots, n$.*
- *For each $x \in X$, the curve $t \mapsto u(t, x)$, $[a, b] \to X$ is a solution to the second order ordinary differential equation $\frac{\partial^2 u_i(t)}{\partial t^2} = 0$ $(i = 1, \ldots, n)$.*
- *For each $x \in X$, $t \mapsto u(t, x)$, $[a, b] \to X$ is a geodesic in X for the Euclidean metric.*
- *$u_i(t, x) = u_i(0, x) + \left.\frac{\partial u_i(t,x)}{\partial t}\right|_{t=0} \cdot t$ for all $i = 1, \ldots, n$.*
- *$\lambda := \frac{\partial u}{\partial t} \circ u^{-1}$ (i.e. $\lambda_i(t, u(t, x)) = \frac{\partial u_i(t,x)}{\partial t}$ for all $i = 1, \ldots, n$) satisfies the Euler equation for an ideal compressible fluid*
$$\frac{\partial \lambda_i}{\partial t} + \sum_{j=1}^{n}\lambda_j\frac{\partial \lambda_i}{\partial x_j} = 0 \text{ for all } i = 1, \ldots, n \ .$$

Remark 6.4 (Burgers' equation) For $n = 1$ the Euler equation for an ideal compressible fluid is called Burgers' equation: $\frac{\partial\lambda}{\partial t} + \lambda\frac{\partial\lambda}{\partial x} = 0$.

Exercise 6.4 (1-dimensional EPDiff) *Let $X = (\alpha, \beta) \subset \mathbb{R}$. We shall study the Euler-Poincaré equation on the group $Diff^\infty(X)$ of diffeomorphisms of X as follows. Let $l : \mathbb{R}^{k+1} \to \mathbb{R}$ be a C^∞ function. For a diffeomorphism $u : X \to X$ and vector field $\dot{u} : X \to \mathbb{R}$ with $\dot{u}(\alpha) = \dot{u}(\beta) = 0$, let*

$$L(u, \dot{u}) := \int_X l\left((\dot{u} \circ u^{-1})(x), \frac{d}{dx}(\dot{u} \circ u^{-1})(x), \ldots, \frac{d^k}{dx^k}(\dot{u} \circ u^{-1})(x)\right) dx \ .$$

Prove that

- *For any diffeomorphism $\phi : X \to X$, $L(u \circ \phi^{-1}, \dot{u} \circ \phi^{-1})) = L(u, \dot{u})$. This means that L is right-invariant under the action of $Diff^\infty(X)$.*
- *The curve of diffeomorphisms $(t, x) \mapsto u(t, x), [a, b] \times X \to X$ is stationary for the action $\int_a^b L\left(u(t), \frac{du(t)}{dt}\right) dt$ with respect to variations with fixed endpoints if and only if the curve of vector fields $\lambda : [a, b] \times X \to \mathbb{R}$ defined by $\lambda(t, u(t, x)) = \frac{du(t,x)}{dt}$ (i.e. $\lambda := \frac{du}{dt} \circ u^{-1}$) is stationary for the action $\int_X l(\lambda(x), \frac{d\lambda(x)}{dx}, \ldots, \frac{d^k\lambda(x)}{dx^k})dx$ with respect to variations of the form*

$$\lambda(t, x) + \varepsilon\left(\frac{\partial\delta(t, x)}{\partial t} + \frac{\partial\lambda(t, x)}{\partial x}\delta(t, x) - \frac{\partial\delta(t, x)}{\partial x}\lambda(t, x)\right) \ ,$$

  where $\delta : [a, b] \times X \to \mathbb{R}$ is an arbitrary C^∞ curve of vector fields on X with $\delta(t, \alpha) = \delta(t, \beta) = \delta(a, x) = \delta(b, x) = 0$.

Exercise 6.5 (Burgers' equation as EPDiff) *In exercise 6.4, choose*

$$L(u, \dot{u}) = ||\dot{u} \circ u^{-1}||_{L_2}^2 := \int_X \frac{1}{2}(\dot{u}(u^{-1}(x))^2 dx \ .$$

Show that the Euler-Poincaré equations for $\lambda := \frac{\partial u}{\partial t} \circ u^{-1}$ read

$$\frac{\partial\lambda}{\partial t} = 3\lambda\frac{\partial\lambda}{\partial x} \ .$$

Exercise 6.6 (The Camassa-Holm equation) *In exercise 6.4, choose*

$$L(u, \dot{u}) = ||\dot{u} \circ u^{-1}||_{H^1}^2 = \int_X \frac{1}{2}\left((\dot{u} \circ u^{-1})(x)\right)^2 + \frac{1}{2}\left(\frac{d}{dx}(\dot{u} \circ u^{-1})(x)\right)^2 \ dx \ .$$

Show that this Lagrangian gives rise to the Euler-Poincaré equation

$$\left(1 - \frac{\partial^2}{\partial x^2}\right)\frac{\partial\lambda}{\partial t} = 3\lambda\frac{\partial\lambda}{\partial x} - 2\frac{\partial\lambda}{\partial x}\frac{\partial^2\lambda}{\partial x^2} - \lambda\frac{\partial^3\lambda}{\partial x^3} \ .$$

This equation is called the Camassa-Holm equation. It is famous because it is integrable and because it admits a special type of peaked soliton solutions, called 'peakons'.

7 Hamiltonian systems

In this section, we will show that the Euler-Lagrange equations (2.8) for a nondegenerate Lagrangian are equivalent to the famous equations of Hamilton.

7.1 The Legendre transform

Let $L : TQ \to \mathbb{R}$ be such a nondegenerate Lagrangian. We define the "momentum" variables

$$p_j = \frac{\partial L(q, \dot{q})}{\partial \dot{q}_j} \in \mathbb{R} \ .$$

Again under the assumption that $\frac{\partial^2 L(q,\dot{q})}{\partial \dot{q}_j \partial \dot{q}_k}$ is invertible, the transformation

$$(q, \dot{q}) \mapsto (q, p) = (q, \frac{\partial L(q, \dot{q})}{\partial \dot{q}})$$

is a local diffeomorphism. Hence we may write $p_j = p_j(q, \dot{q})$ and $\dot{q}_j = \dot{q}_j(q, p)$. If we now also express the constant of motion (2.9) in terms of the new variables

$$H(q, p) := h(q, \dot{q}(q, p)) = \sum_{j=1}^{n} p_j \dot{q}_j(q, p) - L(q, \dot{q}(q, p)) \ ,$$

then we observe that

$$\frac{\partial H(q, p)}{\partial p_j} = \dot{q}_j(q, p) + \sum_{k=1}^{n} p_k \frac{\partial \dot{q}_k(q, p)}{\partial p_j} - \sum_{k=1}^{n} \frac{\partial L(q, \dot{q}(q, p))}{\partial \dot{q}_k} \frac{\partial \dot{q}_k(q, p)}{\partial p_j} = \dot{q}_j(q, p) \ ,$$

because of the definition of p_k. A similar computation leads to the conclusion that

$$\frac{\partial H(q, p)}{\partial q_j} = -\frac{\partial L(q, \dot{q}(q, p))}{\partial q_j} \ .$$

This makes the Euler-Lagrange equations $\frac{dq_j}{dt} = \dot{q}_j, \frac{d}{dt}\frac{\partial L(q,\dot{q})}{\partial \dot{q}_j} = \frac{\partial L(q,\dot{q})}{\partial q_j}$ for the curve $t \mapsto (q(t), \dot{q}(t))$ in TQ equivalent to the equations

$$\frac{dq_j}{dt} = \frac{\partial H(q, p)}{\partial p_j} \ , \ \frac{dp_j}{dt} = -\frac{\partial H(q, p)}{\partial q_j} \tag{7.1}$$

for the curve $t \mapsto (q(t), p(t))$. Equations (7.1) are called Hamilton's equations of motion for the *Hamiltonian function* H. The transformation of the Lagrangian L into the Hamiltonian H is traditionally called the "Legendre transformation".

The momentum $p(q, \dot{q}) := \frac{\partial L(q,\dot{q})}{\partial \dot{q}}$ does not have the interpretation of an element of $T_q Q$. In fact, $p(q, \dot{q})$ is the total derivative of the function $\dot{q} \mapsto L(q, \dot{q})$ at the point $\dot{q}$. Sometimes it is also called the "fiber derivative" of L as the differentiation is only in the direction of the fiber $T_qQ \subset TQ$. The derivative acts on $v \in T_qQ$ by $p(q,\dot q)(v) = \frac{d}{d\varepsilon}\big|_{\varepsilon=0} L(q,\dot q + \varepsilon v)$, so $p(q,\dot q)$ is actually a linear map from T_qQ to $\mathbb{R}$, i.e. $p(q,\dot q) \in (T_qQ)^*$. Recall that the manifold of all q's and p's is called the cotangent bundle T^*Q of Q: $T^*Q := \cup_{q\in Q}(T_qQ)^*$.

For an arbitrary C^1 function $H : T^*Q \to \mathbb{R}$ of positions and momenta, Hamilton's equations of motion define a vector field on T^*Q, namely

$$X_H := \sum_{j=1}^n \frac{\partial H}{\partial p_j}\frac{\partial}{\partial q_j} - \frac{\partial H}{\partial q_j}\frac{\partial}{\partial p_j} \ . \tag{7.2}$$

X_H is called the "Hamiltonian vector field" of H.

7.2 The tautological one-form and the canonical two-form

Let us start by introducing a one-form on the cotangent bundle T^*Q of a manifold Q. Recall the canonical projection $\pi_{T^*Q} : T^*Q \to Q$, which maps $p \in (T_qQ)^*$ to q. Its tangent map $T_{(q,p)}(\pi_{T^*Q})$ thus sends $T_{(q,p)}(T^*Q)$ to T_qQ. But p is a linear form on T_qQ, hence this allows us to define a one-form τ on T^*Q, as the composition

$$\tau_{(q,p)} := p \circ T_{(q,p)}(\pi_{T^*Q}) \ .$$

In local coordinates, π_{T^*Q} sends (q,p) to q and hence $T_{(q,p)}(\pi_{T^*Q}) : (q,p,\delta q,\delta p) \mapsto (q,\delta q)$, so $\tau_{(q,p)} : (\delta q,\delta p) \mapsto \sum_{j=1}^n p_j\delta q_j$, that is

$$\tau = \sum_{j=1}^n p_j dq_j \ .$$

The one-form τ has the following remarkable property:

Proposition 7.1 *View a one-form α on Q as a map $\alpha : Q \to T^*Q$. The one-form τ on T^*Q defined above is the unique one-form on T^*Q with the property that*

$$\alpha^*\tau = \alpha \tag{7.3}$$

for all one-forms α on Q.

Proof Let us prove this statement by a computation in local coordinates, writing $(q,\delta q)$ for an element $\dot q$ of T_qQ, $\alpha_q = \sum_{j=1}^n \alpha_j(q)dq_j$ for an arbitrary one-form on Q and $\sigma_{(q,p)} = \sum_{j=1}^n \sigma_j^q(q,p)dq_j + \sigma_j^p(q,p)dp_j$ for an arbitrary one-form on T^*Q. Using that $\alpha^*\sigma = \sigma\circ T\alpha$, the equality $\alpha^*\sigma(\dot q) = \alpha(\dot q)$ becomes

$$\sum_{j=1}^n \left(\sigma_j^q(q,\alpha(q)) + \sum_{k=1}^n \sigma_k^p(q,\alpha(q))\frac{\partial\alpha_k(q)}{\partial q_j}\right)\delta q_j = \sum_{j=1}^n \alpha_j(q)\delta q_j \ .$$

This yields equality for all $\delta q \in \mathbb{R}^n$ and all $\alpha : Q \to T^*Q$ if and only if $\sigma_j^q(q,p) = p_j$ and $\sigma_j^p = 0$, that is if and only if $\sigma = \tau$. $\square$

Because of the last property, τ is called the "tautological one-form" of T^*Q.

The exterior derivative

$$\omega = -d\tau = \sum_{j=1}^{n} dq_j \wedge dp_j$$

is called the "canonical two-form" of T^*Q. Recall that this two-form assigns to two tangent vectors $(\delta q^1, \delta p^1), (\delta q^2, \delta p^2) \in T_{(q,p)}(T^*Q)$ the number

$$\omega_{(q,p)}((\delta q^1, \delta p^1), (\delta q^2, \delta p^2)) = \sum_{j=1}^{n} (\delta q^1)_j (\delta p^2)_j - (\delta q^2)_j (\delta p^1)_j =$$

$$(\delta q_1^1, \ldots, \delta q_n^1; \delta p_1^1, \ldots, \delta p_n^1) \begin{pmatrix} 0 & \mathrm{Id} \\ -\mathrm{Id} & 0 \end{pmatrix} \begin{pmatrix} \delta q_1^2 \\ \ldots \\ \delta q_n^2 \\ \delta p_1^2 \\ \ldots \\ \delta p_n^2 \end{pmatrix} .$$

The matrix

$$J := \begin{pmatrix} 0 & \mathrm{Id} \\ -\mathrm{Id} & 0 \end{pmatrix}$$

is sometimes called the *standard symplectic matrix.*

Proposition 7.2 *The Hamiltonian vector field X_H in formula (7.2) is uniquely determined by the relation*

$$i_{X_H}\omega = dH \tag{7.4}$$

The way equation (7.4) should be read is as follows: X_H is vector field on T^*Q, and ω a two-form on T^*Q, so that $i_{X_H}\omega := \omega(X_H, \cdot)$ is a one-form on T^*Q. As H is a function on T^*Q, dH is also a one form on T^*Q. These two one-forms are apparently equal.

Proof (of Proposition (7.2): Applied to a tangent vector $(\delta q, \delta p) \in T_{(q,p)}(T^*Q)$, both the left and the right hand side of (7.4) give the value

$$\sum_{j=1}^{n} \frac{\partial H}{\partial q_j}\delta q_j + \frac{\partial H}{\partial p_j}\delta p_j \ .$$

$\square$

In fact, it is not hard to see that X_H is uniquely determined by the relation (7.4). This follows because for every $(q,p) \in T^*Q$, the mapping $w \mapsto i_w\omega$ defines an isomorphism between $T_{(q,p)}(T^*Q)$ to $T^*_{(q,p)}(T^*Q)$. This is true because the matrix J is invertible. A two-form with this property is called *nondegenerate*.

7.3 Symplectic Geometry

Let us now generalize the setting of a cotangent bundle:

Definition: A symplectic manifold is a differentiable manifold M on which a two-form ω is defined that satisfies

1. ω is closed, i.e. $d\omega = 0$ and
2. ω is nondegenerate, i.e. $w \mapsto i_w\omega : T_mM \to (T_mM)^*$ is invertible for every $m \in M$.

As above, for every smooth function $H : M \to \mathbb{R}$ the Hamiltonian vector field X_H on M can be defined by the relation

$$i_{X_H}\omega = dH \ .$$

With the above definitions, T^*Q is an example of a symplectic manifold. On the other hand, the so-called *Darboux lemma* says that every symplectic manifold locally looks like a cotangent bundle, that is near every point $m \in M$ there exists a coordinate system (q,p) for M such that $\omega = \sum_{j=1}^n dq_j \wedge dp_j$. In particular, this implies that M must be even-dimensional: $\dim M = 2n$.

Remark 7.3 (deRham cohomology) Let M be an arbitrary smooth manifold. Recall that for an arbitrary k-form α on M we can define the $k+1$-form $d\alpha$ on M. α is called closed if $d\alpha = 0$ and α is called exact if $\alpha = d\beta$ for some $k-1$-form β. We have that $d(d\beta) = 0$ for all forms β, so that if $\alpha = d\beta$, then $d\alpha = 0$, that is every exact form is automatically closed.

The inverse is true only locally, that is if α is a $k-$form with $d\alpha = 0$, then the so-called *Poincaré lemma* says that every point $m \in M$ possesses an open neighborhood U_m together with a $k-1$-form β on U_m such that $\alpha|_{U_m} = d\beta$. In other words, the Poincaré lemma says that every closed form is locally exact.

As we saw above, the canonical symplectic form $\sum_{j=1}^n dq_j \wedge dp_j$ on T^*Q is exact, and hence closed. On the other hand, in Darboux coordinates on an arbitrary symplectic manifold (M,ω), we have $\omega = \sum_{j=1}^n dq_j \wedge dp_j$, so ω is exact at least on every Darboux coordinate patch.

Globally the structure of a symplectic manifold may differ from that of a cotangent bundle though, if only because an arbitrary closed two-form on a manifold need not be exact. Let us try to make this statement more precise by making the following definitions:

Denote by $Z^k(M)$ the space of all closed k-forms on M and by $B^k(M)$ the space of exact $k-$forms on M. Clearly, $B^k(M) \subset Z^k(M)$. Then one can define the *k-th deRham cohomology group of M* as

$$H^k_{dR}(M) = Z^k(M)/B^k(M) \ .$$

Recall that the elements of $H^k_{dR}(M)$ are the equivalence classes of elements of $Z^k(M)$ whose difference is in $B^k(M)$, that is for $\alpha \in Z^k(M)$,

$$[\alpha] = \{\tilde{\alpha} \in Z^k(M) \mid \tilde{\alpha} - \alpha \in B^k(M) \ \} \in H^k_{dR}(M) \ .$$

The properties of $H^k_{dR}(M)$ depend strongly on the topological properties of M. For instance, it is well-known that $H^1_{dR}(M) = 0$ if and only if M is simply connected. Also, one can show that if M is compact, $H^k_{dR}(M) \cong H_k(M)$ where $H_k(M)$ is the so-called *singular homology* of M, defined purely in terms of the possible triangulations of M. Singular homology is usually treated in a course on algebraic topology.

With all this terminology we can say that the symplectic form ω on M is exact if and only if the cohomology class $[\omega] \in H^2_{dR}(M)$ of ω is equal to zero.

Remark 7.4 (Symplectic topology) The existence of a symplectic form on a manifold M puts rather strong topological restrictions on M: not every manifold admits a nondegenerate closed 2-form. What's more, the $2k$-forms $\omega^k := \omega \wedge \ldots \wedge \omega$ (k times) also are closed and nondegenerate, and in particular, every symplectic manifold is *orientable*, i.e. it admits a nowhere-zero volume-form, namely ω^n.

For instance, it is a known fact from topology that the only $2n$ dimensional sphere that admits a nondegenerate $2k$-form for every $1 \leq k \leq n$ is the 2-sphere. Hence we may conclude that the only sphere that can carry a symplectic structure, is the 2-sphere.

7.4 The dynamics of Hamiltonian systems

A Hamiltonian vector field X_H has a number of special properties: first of all,

Proposition 7.5 *H is a constant of motion for the flow of X_H.*

Proof: This follows as $dH \cdot X_H = i_{X_H} dH = i_{X_H} i_{X_H} \omega = \omega(X_H, X_H) = 0$ because of antisymmetry of ω that holds for every k-form. $\square$

Definition: Let $\Psi : M \to M$ be a diffeomorphism on M, v a vector field on M and α a k-form on M. The pullback $\Psi^* v$ of v under Ψ is the vector field on M defined by

$$\Psi^* v(m) = (T_m \Psi)^{-1} v(\Psi(m))$$

and the pullback $\Psi^* \alpha$ of α under Ψ is the k-form on M defined by

$$(\Psi^* \alpha)_m(v_1, \ldots, v_k) = \alpha_{\Psi(m)}(T_m \Psi \cdot v_1, \ldots, T_m \Psi \cdot v_k)$$

If we let Ψ be the time-t flow (for small t) of a vector field v on M, then it is possible to define the k-form $\mathcal{L}_v\alpha$ on M by

$$(\mathcal{L}_v\alpha)_m(v_1,\ldots,v_n) := \lim_{t\to 0}\frac{1}{t}\left((e^{tv})^*\alpha-\alpha\right)(v_1,\ldots,v_n),\ \text{i.e. } \mathcal{L}_v\alpha = \frac{d}{dt}\bigg|_{t=0}(e^{tv})^*\alpha$$

By Cartan's magic formula we have for every k-form α that

$$\mathcal{L}_v\alpha = di_v\alpha + i_v d\alpha$$

Proposition 7.6 *The flow of a Hamiltonian system leaves the symplectic form invariant, i.e.* $(e^{tX_H})^*\omega = \omega$.

Proof: $\mathcal{L}_{X_H}\omega = i_{X_H}d\omega + di_{X_H}\omega = 0 + ddH = 0.$ □

Remark 7.7 (Locally Hamiltonian vector fields) In general, a smooth vector field v satisfies $\mathcal{L}_v\omega = 0$, if and only if $(e^{tv})^*\omega = \omega$. As $d\omega = 0$, Cartan's magic formula gives that $\mathcal{L}_v\omega = di_v\omega$. Hence the Poincaré lemma guarantees that if $\mathcal{L}_v\omega = 0$, then $i_v\omega = dh$ locally, and hence v is called *locally Hamiltonian* if $\mathcal{L}_v\omega = 0$. Whether or not v is globally Hamiltonian, i.e. $i_v\omega = dH$ for some globally defined function $H : M \to \mathbb{R}$, depends on whether $i_v\omega$ is globally exact, that is whether the cohomology class $[i_v\omega] \in H^1_{dR}(M)$ is equal to zero. This is automatically the case if M is simply connected, because $H^1_{dR}(M) = 0$ if M is simply connected.

A diffeomorphism $\Psi : M \to M$ with the property that $\Psi^*\omega = \omega$ is called *symplectic*. Clearly, the flow of any (locally) Hamiltonian vector field consists of symplectic transformations.

One important consequence of the above proposition is that a Hamiltonian flow preserves the volume form $\omega^n = \omega\wedge\ldots\wedge\omega$:

$$(e^{tX_H})^*(\omega^n) = ((e^{tX_H})^*\omega)^n = \omega^n\ .$$

Another particularly convenient property of symplectic diffeomorphisms is that they transform Hamiltonian vector fields into Hamiltonian vector fields:

Proposition 7.8 *Let M be a symplectic manifold with symplectic form ω, $H : M \to \mathbb{R}$ a Hamiltonian function and $\Psi : M \to M$ a symplectic diffeomorphism, i.e. $\Psi^*\omega = \omega$. Then*

$$\Psi^* X_H = X_{\Psi^* H}$$

in which $\Psi^*H := H\circ\Psi : M \to \mathbb{R}$.

Proof: For $v \in T_mM$ we have $(i_{\Psi^*X_H}\omega)_m(v) = (i_{\Psi^*X_H}\Psi^*\omega)_m(v) = (\Psi^*\omega)_m(\Psi^*X_H(m), v) = \omega_{\Psi(m)}(X_H(\Psi(m)), T_m\Psi\cdot v) = dH(\Psi(m))\cdot T_m\Psi\cdot v = d(\Psi^*H)(m)\cdot v.$ □

One more definition is that of the Poisson bracket of two Hamiltonian functions. Let

$F, G : M \to \mathbb{R}$ be two Hamiltonian functions, then we define a third Hamiltonian function $\{F, G\} : M \to \mathbb{R}$ by setting

$$\{F, G\} = \omega(X_F, X_G) = dF \cdot X_G = -dG \cdot X_F$$

In Darboux coordinates (q, p), one computes that

$$\{F, G\} = \sum_{j=1}^{n} \frac{\partial F}{\partial q_j}\frac{\partial G}{\partial p_j} - \frac{\partial G}{\partial q_j}\frac{\partial F}{\partial p_j}$$

This then leads to the conclusion that the space of C^∞ functions on M is a Lie-algebra. In particular, there is the so-called Jacobi-identity

$$\{F, \{G, H\}\} + \{G, \{H, F\}\} + \{H, \{F, G\}\}$$

At the same time we can for two vector fields v and w on an arbitrary m-dimensional manifold X define the vector field $[v, w]$, called the Lie-bracket of v and w, by

$$[v, w] := \left.\frac{d}{dt}\right|_{t=0} (e^{tv})^* w$$

The well-known expression for $[v, w]$ in local coordinates

$$[v, w] = \sum_{j=1}^{m} \left(\sum_{k=1}^{m} \frac{\partial w_j}{\partial x_k} v_k - \frac{\partial v_j}{\partial x_k} w_k \right) \frac{\partial}{\partial x_j}$$

leads to the conclusion that

$$[X_F, X_G] = -X_{\{F,G\}}$$

This means that the map $F \mapsto X_F$ is a so-called Lie-algebra anti-homomorphism.

Definition: Two functions F and G on M are said to be *in involution* if $\{F, G\} = 0$.

First of all the involution property for F and G implies, because $\{F, G\} = dF \cdot X_G = -dG \cdot X_F$, that F is a constant of motion for the vector field X_G and G is a constant of motion for the vector field X_F. In other words, the flow of X_G leaves the level sets of the function F invariant and vice versa. At the same time the involution property for F and G implies that $[X_F, X_G] = 0$, i.e. the vector fields X_F and X_G -and therefore there flows !- commute.

7.5 Excursion: a Hamiltonian wave equation

Some partial differential equations have a remarkable Hamiltonian formulation. We will discuss in this section a simple example of a nonlinear wave equation on the circle $S^1 := \mathbb{R}/\mathbb{Z}$, as follows. Given a C^1 function $F : \mathbb{R} \to \mathbb{R}$, we look for a C^2 function $u : [a,b] \times S^1 \to \mathbb{R}$ of time $t \in [a,b]$ and position $x \in S^1$ that satisfies

$$u_{tt} = u_{xx} + F'(u) \; . \tag{7.5}$$

Remark 7.9 (The Lagrangian formulation) Equation (7.5) has a Lagrangian formulation, as we discussed before: setting $\dot{u} = u_t$ as usual, one defines the kinetic and potential energy of u at time t as

$$T(\dot{u}(t,\cdot)) := \int_{S^1} \frac{1}{2} \left(\dot{u}(t,x)\right)^2 \, dx \text{ and } V(u(t,\cdot)) := \int_{S^1} \frac{1}{2} \left(u_x(t,x)\right)^2 - F(u(t,x)) \, dx \; ,$$

and observes that u is stationary for the action integral

$$A(u) := \int_a^b T(u_t(t,\cdot)) - V(u(t,\cdot)) \, dt$$

with respect to smooth variations with fixed endpoints, if and only if u satisfies (7.5).

The Hamiltonian formulation of equation (7.5) comes about as follows. Denote by $C^k(S^1)$ the space of C^k functions from S^1 to $\mathbb{R}$, and define the "Hamiltonian function"

$$\mathcal{H} : C^2(S^1) \times C^0(S^1) \to \mathbb{R} \; , \; \mathcal{H}(u,v) := \int_{S^1} \frac{1}{2} \left(v(x)\right)^2 + \frac{1}{2} \left(u_x(x)\right)^2 - F(u(x)) \, dx \; .$$

The derivative of $\mathcal{H}$ at the point $(u,v) \in C^2(S^1) \times C^0(S^1)$ is the linear form

$$d\mathcal{H}(u,v) \in \left(C^2(S^1) \times C^0(S^1)\right)^* \; ,$$

defined by its action on $(u_2, v_2) \in C^2(S^1) \times C^0(S^1)$:

$$d\mathcal{H}(u,v) \cdot (u_2,v_2) := \left.\frac{d}{d\varepsilon}\right|_{\varepsilon=0} \mathcal{H}((u,v) + \varepsilon(u_2,v_2)) =$$
$$\int_{S^1} v(x)v_2(x) - (F'(u(x)) + u_{xx}(x))u_2(x) \, dx \; .$$

The last equality follows from a partial integration.

At the same time, for $(u_1,v_1), (u_2,v_2) \in C^0(S^1) \times C^0(S^1)$, we can define

$$\Omega((u_1,v_1),(u_2,v_2)) := \int_{S^1} u_1(x)v_2(x) - u_2(x)v_1(x) \, dx \; .$$

Clearly, Ω is an anti-symmetric bilinear form on $C^0(S^1) \times C^0(S^1)$, that is

$$\Omega((u_1,v_1) + s(\tilde{u}_1,\tilde{v}_1),(u_2,v_2)) = \Omega((u_1,v_1),(u_2,v_2)) + s\Omega((\tilde{u}_1,\tilde{v}_1),(u_2,v_2)) \text{ and}$$
$$\Omega((u_1,v_1),(u_2,v_2)) = -\Omega((u_2,v_2),(u_1,v_1)) \; .$$

Also, Ω is "weakly nondegenerate" in the following sense: if $\Omega((u_1, v_1), (u_2, v_2)) = 0$ for all $(u_2, v_2) \in C^0(S^1) \times C^0(S^1)$, then by continuity this implies that $(u_1, v_1) = 0$. In other words, the linear map

$$\Omega^{\#} : (u_1, v_1) \mapsto \Omega((u_1, v_1), \cdot) : C^0(S^1) \times C^0(S^1) \to (C^0(S^1) \times C^0(S^1))^*$$

has trivial kernel and hence is injective. A weakly nondegenerate anti-symmetric bilinear form on a linear space is called a *weak symplectic form*.

Finally, it is easy to see that

$$\Omega((v, u_{xx} + F'(u)), (u_2, v_2)) = d\mathcal{H}(u, v) \cdot (u_2, v_2) \ .$$

In other words, the equations

$$u_t = v \ , \ v_t = u_{xx} + F'(u)$$

can be viewed as Hamilton's equations of motion on the infinite-dimensional space $C^2(S^1) \times C^0(S_1)$ equipped with the weak symplectic form Ω, and for the Hamiltonian function $\mathcal{H}$.

The weak nondegeneracy of Ω implies that these equations are indeed unique, so that it is allowed to speak of "the" Hamiltonian equations of $\mathcal{H}$. On the other hand, if $\alpha(u, v) : C^0(S^1) \times C^0(S^1) \to \mathbb{R}$ is just any linear map, the equation $\Omega(X(u, v), \cdot) = \alpha(u, v)$ may not always have a solution $X(u, v) \in C^0(S^1) \times C^0(S_1)$, because Ω is not necessarily strongly symplectic.

7.6 Exercises

Exercise 7.1 *Prove that the Poisson bracket satisfies the Jacobi-identity.*

Exercise 7.2 *Prove that the map $F \mapsto X_F$ is a Lie-algebra anti-homomorphism, i.e. that $[X_F, X_G] = -X_{\{F,G\}}$.*

Exercise 7.3 *Consider the 2-sphere $S^2 := \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$ with its standard volume-form as a symplectic manifold. Compute the Hamiltonian vector field generated by the Hamiltonian function $H(x, y, z) = z$. Draw its phase portrait.*

Exercise 7.4 *Consider the harmonic oscillator $\dot{x} = y, \dot{y} = -x$, for $(x, y) \in \mathbb{R}^2$. With its volume form $dx \wedge dy$, $\mathbb{R}^2$ is a symplectic manifold. Prove that the harmonic oscillator is a Hamiltonian system. What is the Hamiltonian function?*

Next, let us make the coordinate change $\Psi : (x, y) \mapsto (I, \phi)$, where

$$I = \frac{1}{2}(x^2 + y^2) \text{ and } \phi = \arg(x, y) \ .$$

Here, $\arg(x, y)$ denotes the angle that (x, y) makes with the positive x-axis. What are the natural domain and range of this coordinate change? Compute the symplectic form in the new coordinates, i.e. compute $\Psi_*(dx \wedge dy)$. What is $\Psi_* H$, i.e. the Hamiltonian in the new coordinates? Finally, compute the Hamiltonian vector field in the new coordinates.

Further reading

Most of the material in these lecture notes can be found in the existing literature in one form or the other, although most of the time with much less detail. For further reading on the subject, I would like to refer you to the following texts.

The bible of Geometric Mechanics, which treats classical mechanics in a differential geometric framework, is

- Abraham, R., Marsden, J.E., *Foundations of Mechanics*. The Benjamin/Cummings Publ. Co., Reading, Mass., 1987.

Easier to read, and with a focus on computations-by-hand is

- Arnol'd, V.I., *Mathematical Methods of Classical Mechanics*, Graduate Texts in Mathematics 60, Springer-Verlag, 1978.

A nice introduction to the application of geometry and topology in fluid mechanics can be found in

- Arnol'd, V.I. and Khesin, B.A., *Topological Methods in Hydrodynamics*, Graduate Texts in Mathematics 60, Springer-Verlag, 1978.

Several examples of classical mechanical systems, treated in a slightly formal way, arise in the textbook

- Cushman, R.H. and Bates, L.M., *Global aspects of Classical Mechanical Systems*, Birkhäuser Verlag, 1997.

An already standard, advanced, but nice book on Lie groups is:

- Duistermaat, J.J. and Kolk, J.A.C., *Lie groups*, Springer-Verlag, 2000.

Finally, introductory texts about the restricted three-body problem are found in

- Meyer, K.R., *Periodic Solutions of the N-Body-Problem*, Lecture Notes in Mathematics, Springer-Verlag, 1999.

- Meyer, K.R., Hall, G.R., *Introduction to Hamiltonian dynamical systems and the N-body problem*, Applied Math. Sciences 90, Springer-Verlag, 1992.