Mathematical Physics

Bergfinnur Durhuus and Jan Philip Solovej

Preface

These are notes for the course Mathematical Physics at the university of Copenhagen for students in their second or third year of study. The course consists of two parts the first of which is on Classical Mechanics corresponding to the first three chapters, while the second part is on Quantum Mechanics corresponding to the remaining three chapters. The two parts can be read independently.

The main focus is on theoretical developments rather than elaborating on concrete physical systems, which the students are supposed to encounter in regular physics courses. Nevertheless, some examples are included for the purpose of illustration. The prerequisites for the course are standard courses on calculus and linear algebra. We have tried to make the notes essentially self-contained, but presumably some parts, such as the more technical parts of Chapter 2 and the more abstract parts of Chapter 4, may be conceived as difficult without some prior acquaintance with concepts like series expansions, function spaces and normed spaces. The reader may prefer to skip those parts in a first reading and then return for a more thorough study later.

October 2014
Bergfinnur Durhuus Jan Philip Solovej

Contents

Chapter 1

Newtonian Mechanics

In these notes, classical mechanics will be studied as a mathematical model for the description of physical systems consisting of a certain (generally finite) number of particles with fixed masses subject to mutual interactions and possibly to external forces as well.

From the 17th century onwards, beginning with Newton's foundational work [10], such "Newtonian" models have provided an accurate description of a vast variety of mechanical systems, e. g. the solar system, and continue to be of fundamental importance in physics, even though the advent of quantum mechanics and special relativity in the early 20th century has revealed that Newtonian mechanics should be viewed as an approximation valid only in a certain regime, nowadays called *the classical regime*. Moreover, classical mechanics continues to be a significant source of motivation and inspiration for mathematical research in subjects such as ordinary differential equations, calculus of variations, and differential geometry.

1.1 Classical space-time

We shall begin – rather abstractly – by introducing the notion of space-time in classical mechanics. A central paradigm of the classical framework is that *time is absolute,* which reflects the universality of simultaneous events in classical space-time (in striking contrast to special relativity where the simultaneity of events is relative).

The basic ingredient of the mathematical model of classical space-time we are about to describe is a set, $\mathcal{M}$, called the *universe*, and whose elements are referred to as *events*. It is assumed that the events can be labeled by four real numbers, i. e., there exists a bijective map $\underline{x} : \mathcal{M} \to \mathbb{R}^4$, called a *coordinate system* on $\mathcal{M}$. Composing one coordinate system $\underline{x}$ with a bijective map $\psi : \mathbb{R}^4 \to \mathbb{R}^4$ we obtain a second coordinate system $\underline{y} = \psi \circ \underline{x}$. Thus coordinate systems are far from being unique. We call ψ the *coordinate transformation* from $\underline{x}$-coordinates to $\underline{y}$-coordinates.

Next, we postulate the existence of two functions $T : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ and $d : D \to \mathbb{R}$, where

$$D = \{(p, q) \in \mathcal{M} \times \mathcal{M} \mid T(p, q) = 0\}.$$

The function T is called the *time-difference function* and $T(p,q)$ has the physical interpretation of the lapse of time between the event p and the event q as measured by standard clocks. Two events p and q are said to be *simultaneous*, if $T(p,q) = 0$ holds. Thus D is the set of pairs of simultaneous events, and the physical interpretation of $d(p,q)$ is the *(spatial) distance* between the simultaneous events p and q. Note that the distance between non-simultaneous events has no physical meaning and is not defined.

In view of the physical interpretation of the functions T and d, these must be subject to further constraints, which bring in the Euclidean structure of three-dimensional space. This is achieved by postulating the existence of a coordinate system $\underline{x} = (x_1, x_2, x_3, t)$ such that

$$d(p,q) = \|x(p) - x(q)\|\,, \quad (p,q) \in D\,, \tag{1.1}$$

and

$$T(p,q) = t(q) - t(p)\,, \quad p,q \in \mathcal{M}\,, \tag{1.2}$$

where we use the notation $x = (x_1, x_2, x_3)$ and $\|\cdot\|$ denotes the *Euclidean norm* in $\mathbb{R}^3$ defined by

$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}\,.$$

A coordinate system $\underline{x}$ fulfilling (1.1) and (1.2) will be called a *Galilean coordinate system*. Likewise, we call $x_1(p)$, $x_2(p)$, $x_3(p)$ the *space-coordinates* and $t(p)$ the *time-coordinate* of the event $p \in \mathcal{M}$.

The relation (1.1) expresses the Euclidean nature of the distance d. A few remarks on the Euclidean norm and its properties will be useful for the following discussion. We define the *scalar product* (also called *inner product*) of two vectors $x, y \in \mathbb{R}^3$, written as $\langle x, y\rangle$ or simply as $x \cdot y$, by

$$x \cdot y = \langle x, y\rangle = x_1y_1 + x_2y_2 + x_3y_3\,.$$

Note that

$$\|x\|^2 = \langle x, x\rangle > 0 \quad \text{for } x \neq \underline{0} = (0,0,0)\,,$$

and it is clear from the definition that $\langle x, y\rangle$ is a linear real-valued function of $x \in \mathbb{R}^3$, for fixed y, as well as of y for fixed x. Moreover, we have $\langle x, y\rangle = \langle y, x\rangle$. We say, that $\langle \cdot, \cdot\rangle$ is a symmetric and positive definite *bilinear form* on $\mathbb{R}^3$.

Given two non-zero vectors $x, y \in \mathbb{R}^3$, we define the *angle* $\theta \in [0, \pi]$ between x and y through the formula

$$\cos\theta = \frac{x \cdot y}{\|x\|\|y\|}\,. \tag{1.3}$$

Note that the definition (1.3) presupposes that the right-hand side belongs to the interval $[-1, 1]$, which is a consequence of the so-called *Cauchy-Schwarz inequality*; see Exercise 1.1 d). Furthermore, we say that x and y are *orthogonal*, written as $x \perp y$, if $x \cdot y = 0$.

A mapping $S : \mathbb{R}^3 \to \mathbb{R}^3$ is called an *isometry* if it preserves the Euclidean distance, that is if

$$\|S(x) - S(y)\| = \|x - y\|\,, \quad x, y \in \mathbb{R}^3\,.$$

It can be shown (see Exercise 1.2) that any isometry $S : \mathbb{R}^3 \to \mathbb{R}^3$ is of the form

$$S(x) = a + S_0(x)\,,$$

where $a \in \mathbb{R}^3$ is a fixed vector and S_0 is a linear isometry, also called an *orthogonal transformation*. Moreover, any such orthogonal transformation is either a rotation around an axis through $\underline{0}$ or it is a rotation around an axis through $\underline{0}$ composed with a reflection in a plane through $\underline{0}$ (see Exercise 1.4). A matrix that represents an orthogonal transformation with respect to the standard basis for $\mathbb{R}^3$ is called an *orthogonal matrix*, and the set of orthogonal matrices is denoted by $O(3)$. They are discussed further in Exercises 1.3 and 1.4.

The following result gives a characterization of coordinate transformations between Galilean coordinate systems.

Proposition 1.1. *Let $\underline{x}$ and $\underline{y}$ be two Galilean coordinate systems. Then there exists a constant $t_0 \in \mathbb{R}$, a function $\varphi : \mathbb{R} \to \mathbb{R}^3$ and a function $A : \mathbb{R} \to O(3)$, such that $\underline{y} = \psi \circ \underline{x}$, where*

$$\psi(x,t) = (\varphi(t) + A(t)x,\, t - t_0)\,, \quad x \in \mathbb{R}^3\,,\ t \in \mathbb{R}\,. \tag{1.4}$$

Remark. We see from equation (1.4) that the time coordinates with respect to $\underline{x}$ and $\underline{y}$ are related by a constant shift. This property is commonly expressed by saying that *"time is absolute"* in classical physics, since time is uniquely determined once a zero-point has been chosen. Consequently, we shall in the following frequently restrict attention to systems with a common time coordinate.

Proof of Proposition 1.1. Choose a fixed event $q_0 \in \mathcal{M}$ and write $\underline{x} = (x,t)$ and $\underline{y} = (y,s)$. Then, according to (1.2), we have

$$t(p) - t(q_0) = s(p) - s(q_0)\,, \quad p \in \mathcal{M}\,,$$

and hence

$$s(p) = t(p) - t_0\,, \quad p \in \mathcal{M}\,, \tag{1.5}$$

where $t_0 = t(q_0) - s(q_0)$.

Next, let $p, q \in \mathcal{M}$ be two events such that $t(p) = t(q)$. By (1.1), this means that p and q are simultaneous events. Therefore, equation (1.1) implies

$$\|y(p) - y(q)\| = \|x(p) - x(q)\| \quad \text{if } t(p) = t(q), \tag{1.6}$$

since $\underline{x} = (x,t)$ and $\underline{y} = (y,s)$ are both Galilean coordinate systems. Further, by the fact that $\underline{x} : \mathcal{M} \to \mathbb{R}^4$ is bijective, there exists a unique $p_\tau \in \mathcal{M}$ for each $\tau \in \mathbb{R}$ such that $x(p_\tau) = (0,0,0,\tau)$ holds. Equivalently, this means that $p_\tau = \underline{x}^{-1}(0,0,0,\tau)$.

Defining

$$\varphi(\tau) = y(p_\tau) \quad \text{and} \quad S_\tau(v) = y(\underline{x}^{-1}(v,\tau)) - \varphi(\tau), \quad \text{for } \tau \in \mathbb{R} \text{ and } v \in \mathbb{R}^3, \tag{1.7}$$

we claim that eq. (1.6) implies

$$\|S_\tau(v) - S_\tau(w)\| = \|v - w\| \quad \text{for } v, w \in \mathbb{R}^3. \tag{1.8}$$

Indeed, setting $p = \underline{x}^{-1}(v,\tau)$ and $q = \underline{x}^{-1}(w,\tau)$, it follows that $t(p) = t(q) = \tau$, and thus the events p and q are simultaneous and (1.6) applies. It is clear that $x(p) = v$, $x(q) = w$, and $S_\tau(v) - S_\tau(w) = y(p) - y(q)$. Hence (1.8) holds true.

By (1.8), we see that $S_\tau : \mathbb{R}^3 \to \mathbb{R}^3$ is an isometry for any $\tau \in \mathbb{R}$ fixed. In fact, we conclude from Exercise 1.3 that $S_\tau(v)$ must be a linear isometry, since $S_\tau(\underline{0}) = y(\underline{x}^{-1}(\underline{0}, \tau)) - \varphi(\tau) = y(p_\tau) - y(p_\tau) = \underline{0}$. Hence, it follows from the remark about isometries above (see also Exercise 1.4) that

$$S_\tau(v) = A(\tau)v, \quad v \in \mathbb{R}^3\,,$$

where $A(\tau) \in O(3)$ is an orthogonal 3×3-matrix depending on τ. Inserting this into the definition (1.7) of S_τ and recalling (1.5) we get the expression (1.4) for $\psi = \underline{y} \circ \underline{x}^{-1}$. □

The spatial content of eq. (1.4) can be described geometrically as follows. As already noted in the proof, the function φ describes the motion of the origin of x-coordinates with respect to the coordinate system $\underline{y}$. More precisely, setting $x = 0$ in (1.4), we see that $\varphi(t)$ is the y-coordinate set of the origin of x-coordinates at $\underline{x}$-time t. Similarly, denoting by $\varepsilon_1 = (1, 0, 0)$, $\varepsilon_2 = (0, 1, 0)$, $\varepsilon_3 = (0, 0, 1)$ the standard basis vectors of the vector space $\mathbb{R}^3$ and choosing two simultaneous events p, q at $\underline{x}$-time t such that $x(p) - x(q) = \varepsilon_i$ we see that

$$y(p) - y(q) = A(t)\varepsilon_i = A_i(t)\,,$$

where $A_i(t)$ denotes the i'th column of $A(t)$. In other words, the columns of $A(t)$ are the y-coordinates of the standard basis vectors in x-coordinate space at $\underline{x}$-time t. As previously mentioned (see also Exercise 1.4) the orthogonal matrix $A(t)$ represents a rotation around an axis through the origin, possibly composed with a reflection in a plane. We call $A(t)$ *orientation preserving* if the reflection is absent, and ψ is by definition orientation preserving if this is the case for all t. This allows us to characterize an orientation preserving Galilean coordinate transformation as being the composition of a time-dependent rotation around a (time-dependent) axis, represented by $A(t)$, composed with a time-dependent spatial translation, represented by φ.

1.2 Galilean principle of relativity and inertial systems

By a *mechanical system* we mean a finite collection of point-particles, called $i = 1, \ldots, N$, each of which is identified by an event p_i at any instant of time. More precisely, given a Galilean coordinate system $\underline{x} = (x, t)$, a mechanical system consists of N trajectories $x_1(t), \ldots, x_N(t)$, $t \in \mathbb{R}$, in $\mathbb{R}^3$, whose form is determined by fundamental laws of motion, and the essential goal of classical mechanics is to determine those laws and their consequences. We shall restrict our attention to ordinary differential equations as candidates for the fundamental laws of mechanics; in other words, the trajectories are determined as solutions to a set of equations containing the space coordinates $x_i(t)$ and a number of their derivatives w. r. t. time t. When discussing fundamental laws and principles, the particle systems we have in mind are ideal in the sense that they serve as models for the whole universe consisting of N particles. In practice, finite subsystems of the observed universe that can be considered essentially isolated from the rest, often called *closed systems*, do to a very good approximation obey the fundamental laws and, indeed, all experiments and observations used to verify those laws are performed on such systems. Thus, the solar system can to a good approximation be regarded as an isolated system on time scales that are short compared to its rotation time around the center of the galaxy, approximately 10^8 years, with the sun and planets being treated as point particles.

It is a well known experimental fact that the basic laws governing the behavior of mechanical systems are not invariant under arbitrary Galilean coordinate transformations. A falling body behaves differently in a coordinate system at rest at the surface of the earth and in a freely falling system. In fact, the Galilean transformations are sufficiently general that there are no interesting differential equations preserving their form under all such transformations.

The Galilean principle of relativity, described in 1632 by Galilei in his treatise "Dialogue Concerning the Two Chief World Systems"[7], is an invariance principle restricting the possible forms of the basic equations of motion for a mechanical system under a restricted set of coordinate transformations. We formulate this principle as follows.

Galilean principle of relativity. *There exist particular Galilean coordinate systems, called inertial systems, in which the equations of motion for mechanical systems assume identical form. Any coordinate system in uniform motion relative to an inertial system is likewise an inertial system.*

We shall discuss the invariance issue in more detail in the next section. The second part of the statement can, in view of the results of the preceding section, be stated as follows: If $\underline{x}$ is an inertial system of coordinates, and $\psi : \mathbb{R}^4 \to \mathbb{R}^4$ is given by

$$\psi(x,t) = (a + vt + Ax,\, t + s)\,, \quad x \in \mathbb{R}^3\,, t \in \mathbb{R}\,, \tag{1.9}$$

then $\underline{y} = \psi \circ \underline{x}$ is also an inertial coordinate system. Here $\varphi(t) = a + vt$ describes the motion of the origin of x-coordinates on a straight line through a with constant velocity v in y-coordinate space, and the matrix A represents a constant rotation of the x-coordinate axes around a fixed axis through the origin of x-coordinates, possibly composed with reflection in a fixed plane. The parameter s is a (fixed) time translation. We call ψ as given by (1.9) an *inertial coordinate transformation* and denote it by $\psi[a, s, v, A]$.

It is important to note, that the set of inertial coordinate transformations forms a *group* with respect to composition of mappings.

Definition 1.2. *A* **group** *is a set G with a distinguished element $e \in G$, called the* identity element*, and two maps*

$$G \times G \ni (g,h) \mapsto gh \in G, \qquad G \ni g \mapsto g^{-1} \in G\,, \tag{1.10}$$

called multiplication *and* inversion*, respectively, such that*

$$g_1(g_2g_3) = (g_1g_2)g_3, \qquad ge = eg = g, \qquad gg^{-1} = g^{-1}g = e$$

for all $g, g_1, g_2, g_3 \in G$.

The set $O(3)$ of orthogonal matrices is a group with the standard matrix multiplication and matrix inversion (see Exercise 1.3).

Thus, saying that the inertial coordinate transformations form a group means that the composition of two such transformations is an inertial coordinate transformation, that the inverse of an inertial coordinate transformation is also an inertial coordinate transformation and that the identity mapping of $\mathbb{R}^4$ is an inertial coordinate transformation. The last

property is obvious, since the identity mapping is obtained when $a = s = v = 0$ and $A = I$, the 3×3 identity matrix. The other two properties follow from the formula

$$\psi[a_1, s_1, v_1, A_1] \circ \psi[a_2, s_2, v_2, A_2] = \psi[a_1 + v_1 s_2 + A_1 a_2,\ s_1 + s_2,\ v_1 + A_1 v_2,\ A_1 A_2]\,, \quad (1.11)$$

which is obtained by straightforward computation (see Exercise 1.5). In particular, one gets

$$\psi[a, s, v, A]^{-1} = \psi[-A^{-1}(a - vs),\ -s,\ -A^{-1}v,\ A^{-1}]\,, \quad (1.12)$$

and every inertial coordinate transformation can be decomposed into four basic types of transformations:

time translation	$T_s = \psi[\underline{0}, s, \underline{0}, I]$
space translation	$S_a = \psi[a, 0, \underline{0}, I]$
boost	$B_v = \psi[\underline{0}, 0, v, I]$
$O(3)$-transformation	$R_A = \psi[\underline{0}, 0, \underline{0}, A]$

according to

$$\psi[a, s, v, A] = T_s \circ B_v \circ S_a \circ R_A. \quad (1.13)$$

The group of inertial coordinate transformations is often called the *Galilei group*.

Remark 1.3. The Galilean principle of relativity was generalized by A. Einstein in 1905 beyond mechanical systems to include electromagnetism. Since its content is otherwise the same, except that the notion of space-time and the definition of an inertial coordinate system are different in special relativity, we shall in the following refer to the Galilean principle of relativity simply as the principle of relativity.

1.3 Newton's laws of motion

Having set the stage of classical space-time, we now proceed to formulate the physical laws in terms of equations. More specifically, we will see that these so-called Newtonian equations of motion are ordinary differential equations subject to restrictions imposed by the Galilean principle of relativity.

An ordinary differential equation of order $n \geqslant 1$, describing the motion of a point $x(t)$ in $\mathbb{R}^k$ as a function of the variable t, is an equation of the form

$$\frac{d^n x}{dt^n} = f(x, \frac{dx}{dt}, \ldots, \frac{d^{n-1}x}{dt^{n-1}}, t)\,, \quad (1.14)$$

where f is a function with values in $\mathbb{R}^k$ defined on a subset $\Omega \subseteq \mathbb{R}^{kn} \times \mathbb{R}$. We will mostly assume that Ω is the whole space $\mathbb{R}^{kn} \times \mathbb{R}$ and that f is a C^∞-function, i.e. all its partial derivatives exist and are continuous at all points of Ω. A solution to (1.14) is a function $x : I \to \mathbb{R}^k$, where I is an interval, such that $(x(t), x^{(1)}(t), \ldots, x^{(n-1)}(t), t) \in \Omega$ for $t \in I$, and such that

$$x^{(n)}(t) = f(x(t), x^{(1)}(t), \ldots, x^{(n-1)}(t), t) \quad t \in I\,.$$

As usual, the i'th derivative $x^{(i)}$ of the vector function $x = (x_1, \ldots, x_k)$ is defined by differentiating the coordinates:

$$x^{(i)}(t) = (x_1^{(i)}(t), \ldots, x_k^{(i)}(t)) \,.$$

The variable t will in the following be chosen to be the time coordinate in some inertial system $\underline{x}$. For a system consisting of a single (point-)particle, we have $k = 3$ and $x(t)$ is the spatial coordinate set of the particle at time t. For a system consisting of N particles, we collect their coordinate sets $x_1(t), \ldots, x_N(t)$ into one set $x(t) = (x_1(t), \ldots, x_N(t)) \in \mathbb{R}^{3N}$, such that $k = 3N$ in this case.

Let us first consider the case of a first order equation for a single particle

$$\frac{dx}{dt} = f(x,t) \,.$$

Under an inertial coordinate transformation with $A = I$ and $s = 0$, we have $y = x + a + vt$ and hence $\frac{dy}{dt} = \frac{dx}{dt} + v$. The equation can therefore be rewritten in the form

$$\frac{dy}{dt} = f(y - a - vt, t) + v \,.$$

Thus, the two equations have identical form if and only if the function f fulfills

$$f(x,t) = f(x - a - vt, t) + v$$

for arbitrary $x, a, v \in \mathbb{R}^3$ and $t \in \mathbb{R}$. Since the left hand side is independent of a it follows that f is independent of x. Hence $f(x,t) = f(t)$ fulfills $f(t) = f(t) - v$, which clearly is impossible. We conclude that an equation of first order is incompatible with the principle of relativity.

Next, consider an equation of second order for a single particle

$$\frac{d^2x}{dt^2} = f(x, \frac{dx}{dt}, t) \,.$$

Under an inertial coordinate transformation with $A = I$ and $s = 0$ we have $\frac{d^2x}{dt^2} = \frac{d^2y}{dt^2}$ and the equation can be rewritten as

$$\frac{d^2y}{dt^2} = f(y - a - vt, \frac{dy}{dt} - v, t) \,.$$

The two equations have the same form if and only if f satisfies

$$f(x,z,t) = f(x - a - vt, z - v, t)$$

for all $x, z, a, v \in \mathbb{R}^3$ and $t \in \mathbb{R}$. As above this implies that $f(x,z,t) = f(t)$ is independent of x, z. Next, applying an inertial coordinate transformation with $a = s = v = 0$, we have $y = Ax$ and hence $\frac{dy}{dt} = A\frac{dx}{dt}$ and $\frac{d^2y}{dt^2} = A\frac{d^2x}{dt^2}$, since A is constant. Therefore, we can rewrite the equation as

$$\frac{d^2y}{dt^2} = Af(t) \,.$$

Invariance of the equation requires

$$f(t) = Af(t)\,,$$

which implies $f(t) = \underline{0}$, since $\underline{0}$ is the only vector invariant under arbitrary rotations. We conclude that the only second order equation for a single particle that is compatible with the principle of relativity is

$$\frac{d^2x}{dt^2} = \underline{0}\,.$$

The solutions to this equation are of the form

$$x(t) = a + vt\,,$$

where $a, v \in \mathbb{R}^3$ are constants. This means that the possible motions of the particle are along straight lines with constant velocity v, which we recognize as the contents of **Newton's first law** of motion. We have shown that it is a consequence of the principle of relativity and the assumption that the equation of motion is a second order differential equation.

We promote this latter assumption to a general principle for systems with an arbitrary number of particles, i .e. we postulate:

Newton's second law. *In an inertial system the equation of motion for a system of N particles is a second order differential equation for the coordinates of the particles as functions of time.*

This means that the equation of motion is of the form

$$\ddot{x}_i = f_i(x, \dot{x}, t) \qquad i = 1, \ldots, N\,, \tag{1.15}$$

where $x = (x_1, \ldots, x_N)$ and we have introduced the notation $\dot{x}$ and $\ddot{x}$ for the first and second time derivatives of x, respectively. As in the case of a single particle, the principle of relativity implies constraints on the possible forms of the right hand side of (1.15). One finds (see Exercise 1.6) that, as a consequence of invariance under translations and boosts, f can depend only on the coordinate differences $x_i - x_j$ and the velocity differences $\dot{x}_i - \dot{x}_j$, $1 \leqslant i < j \leqslant N$. Furthermore, f_i is independent of time t because of invariance under time shifts T_s. We see that invariance of (1.15) under $O(3)$-transformations implies (see Exercise 1.7) that the f_i's must obey the identity

$$f_i(Ax_1, \ldots, Ax_N, A\dot{x}_1, \ldots, A\dot{x}_N) = Af_i(x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_N)\,. \tag{1.16}$$

This requirement is, however, far from determining the form of the f_i's uniquely (see Exercise 1.7).

A prime example of a mechanical system respecting the Galilean principle of relativity is a system of N particles interacting according to the **universal law of gravitation** which is the statement that (1.15) holds with

$$f_i(x) = \sum_{j \neq i} \frac{k_j(x_j - x_i)}{\|x_j - x_i\|^3}\,, \tag{1.17}$$

where $k_1, \ldots, k_N$ are constants characteristic of the individual particles, i. e., k_i does not depend on which other particles the i'th one is brought into interaction with and it does not depend on the chosen inertial coordinate system. By definition $k_i = Gm_i$, where

$$G = 6.670 \cdot 10^{-14} \mathrm{gm}^3/\mathrm{s}^2$$

is *Newton's constant*, and m_i is called the **mass** of particle i. Note that f_i is independent of the particle velocities $\dot{x}_i$ and clearly fulfills (1.16). Note also that f_i is singular if other particle positions coincide with that of particle i. By measuring the accelerations $\ddot{x}_i$ for a suitable set of particle configurations one can determine the masses m_i from (1.15) and (1.17) and once this has been done, equation (1.15) provides in principle a complete description of the motion of the particles, as discussed in more detail below, apart from the lack of a prescription how to deal with the possible occurrence of singularities when particles coalesce.

The problem of solving (1.15) with f_i given by (1.17) is often referred to as the **N-body problem**. As is well known, the 2-body problem is exactly solvable. The standard way of doing this is first to observe that the center of mass

$$x_c = \frac{m_1}{m_1 + m_2} x_1 + \frac{m_2}{m_1 + m_2} x_2$$

has vanishing acceleration, i.e. the total momentum is conserved. Hence we can choose the inertial system of coordinates such that x_c is a rest at the origin which effectively reduces the system of equations (1.15) to a one-particle problem that can be solved. Notice, however, that although the one-particle problem mentioned also has the form of Newtons equation (1.15) for $N = 1$ it is not invariant under inertial coordinate transformations, since it is valid only in the particular inertial system chosen.

For $N \geqslant 3$ the complete solution to the N-body problem is a long standing unsolved problem. If one of the masses, say m_1, is much larger that the other ones, such as is the case for the solar system, one can obtain an approximate solution by disregarding all terms on the right-hand side of (1.15) not containing m_1. This implies that the acceleration of particle 1 vanishes and the acceleration of particle i for $i \geqslant 2$ only depends on x_1 and x_i. It follows that, in the inertial system of coordinates where $x_1 = 0$, the equations (1.15) reduce to $N - 1$ independent one-particle equations as above. This is the approximation used to obtain, e. g., the elliptic orbits of the planets around the sun.

Hence we see that, although the fundamental equations of mechanics are assumed to be invariant under inertial coordinate transformations, the relevant equations describing reductions or subsystems, or approximations to those equations, need not be invariant under inertial coordinate transformations.

Multiplying the i'th equation (1.15) by the mass m_i and defining the **force** F_i acting on particle i by $F_i = m_i f_i$ we obtain the more familiar form of Newton's second law

$$\boxed{m_i \ddot{x}_i = F_i\,, \quad i = 1, \ldots, N\,.} \tag{1.18}$$

In particular, the gravitational force acting on a particle i in a system of N particles as above is the sum

$$F_i^{grav} = \sum_{j \neq i} F_{ij}^{grav} \tag{1.19}$$

of individual contributions

$$F_{ij}^{grav} = -Gm_i m_j \frac{x_i - x_j}{\|x_i - x_j\|^3}\,. \tag{1.20}$$

from the other particles $j \neq i$.

Remark 1.4. Let us mention at this point that a more general and historically more correct version of Newton's second law is obtained by replacing m_i on the left-hand side of (1.15) by the so-called *inertial mass* $\bar{m}_i$. In this way, two characteristic constants are associated with each individual particle, the inertial mass $\bar{m}_i$ and the *gravitational mass* m_i. We leave it to the reader to contemplate that by observing the motion of two particles with masses $m_1, \bar{m}_1$ and $m_2, \bar{m}_2$, respectively, obeying the general version of Newton's second law and subject to the gravitational interaction force given by (1.20) one can measure

$$\frac{\bar{m}_1}{\bar{m}_2} \quad \text{and} \quad G\frac{m_1}{\bar{m}_1}m_2\,.$$

By varying $\bar{m}_1$ one can hence determine the dependence of $\mu = \frac{m_1}{\bar{m}_1}$ on $\bar{m}_1$. It has been found experimentally that the value of the ratio μ is a universal constant, which can be set to 1 with the proper definition of Newton's constant. The fact that μ is a universal constant is called the **equivalence principle**. It's validity has been implicitly assumed in the formulation (1.18) of Newton's second law and is particularly important as a basic ingredient in the general theory of relativity. The origin of the equivalence principle dates back to the investigations by Galilei around year 1600 of the motion of falling bodies and has since then gained ever stronger experimental support. In particular, the famous Eötvös experiment in 1889 and later refinements have shown that μ equals 1 with an accuracy of order 10^{-10} for various substances. Furthermore, it constitutes the basis of Einstein's general theory of relativity form 1915.

Besides gravity the most important fundamental forces acting in the classical regime are the electromagnetic forces. As mentioned earlier the complete laws of electromagnetism in the form of Maxwell's equations do not satisfy the principle of relativity in the form stated above. However, for particle velocities small compared to the velocity of light and for small particle accelerations such that radiation effects can be ignored the motion of N charged particles is to a good approximation determined by (1.18) with

$$F_i = F_i^{grav} + F_i^{em}\,,$$

where the electromagnetic force F_i^{em} on particle i is given by

$$F_i^{em} = \sum_{j\neq i} F_{ij}^{em}\,, \tag{1.21}$$

with

$$F_{ij}^{em} = Ke_ie_j\frac{x_i - x_j}{\|x_i - x_j\|^3} + \frac{K}{c^2}e_ie_j\dot{x}_i \times \frac{\dot{x}_j \times (x_i - x_j)}{\|x_i - x_j\|^3}\,. \tag{1.22}$$

Here the first term on the right-hand side is the **Coulomb force** and

$$K = 8.99 \cdot 10^{12}\mathrm{gm}^3/\mathrm{A}^2\mathrm{s}^4$$

is called *Coulomb's constant*, while

$$c = 2.998 \cdot 10^8 \text{m/s}$$

is the velocity of light, and the constant e_i, characteristic of particle i, is called its **charge**. The second term on the right-hand side of (1.22) is the magnetic contribution to F_i^{em} arising from the motion of the other particles $j \neq i$. We remark that the mass of a charged particle can be determined from Newton's law by letting it interact with neutral particles, i. e., particles with charge 0, where-after the charge can be determined from its interaction with other charges particles. The reader may easily verify that the Coulomb force satisfies the transformation property (1.16) and is invariant under boosts and translations. Also the magnetic term in (1.22) satisfies (1.16) but obviously is not invariant under boosts. In fact, this term can be viewed as a relativistic correction to the Coulomb interaction and can under certain circumstances be ignored, but in other situations, for instance if particle i is located in the vicinity of a steady current in a conducting wire, then F_i^{em} may be the dominant force acting on particle i.

Comparing the Coulomb force with the gravitational force between, say, an electron and a proton one finds that the ratio is of order 10^{37}. In this sense the Coulomb attraction is much stronger than the gravitational attraction. However, the fact that the charge can assume both positive and negative values, contrary to the mass, means that screening effects may occur as witnessed by our every day experience with macroscopic bodies made up of a large number of negatively charged electrons and an approximately equal number of oppositely charged protons. For such approximately neutral subsystems the effective interaction can well be dominated by the gravitational force, such as is the case for the planets and the sun. Even for small neutral systems of molecular size the effective electromagnetic force falls off faster than quadratically with distance. On the other hand, the electromagnetic forces between constituents of matter also give rise to effective macroscopic forces such as elastic forces.

To sum up, the Newtonian description of the motion of a (system of) particle(s) is based on the assumption that a force represented by a vector $F_i(x_i, t) \in \mathbb{R}^3$ acts on each particle i with coordinates (x_i, t) and that Newton's second law (1.18) holds in an inertial system of coordinates. In principle the force F_i is derivable from fundamental interactions satisfying the principle of relativity but in practice reductions or approximations violating this principle may occur. Although we have here expressed F_i explicitly only as a function of the coordinates of the particle i on which the force acts, it may depend on other variables, including the particle velocities, as seen above. On the other hand, F_i does not depend on higher derivatives of the coordinates since, otherwise, the equation would not be of second order. Thus, in general, to determine the motion of a system of particles it is not possible to solve the equation (1.18) for each individual particle. Instead, one must solve simultaneously the whole system of coupled differential equations for the particles. This is discussed further in Section 1.5 below.

1.4 Conservative force fields

In this section we shall discuss in some detail time independent forces and the concepts of *work* and *energy*.

A *vector field* is a function $F : \mathcal{O} \to \mathbb{R}^k$, where $\mathcal{O} \subseteq \mathbb{R}^k$ is an open set and $k \in \mathbb{N}$ is fixed. If F represents a force acting on a (system of) particle(s) we call it a **force field**. For a single particle in physical space we have $k = 3$, whereas for N interacting particles in 3-dimensional space we naturally have $k = 3N$, as explained at the beginning of the previous section. Other examples of vector fields in physics are the electric and magnetic fields.

Definition 1.5. *A* **curve** *in $\mathbb{R}^k$ is a continuous function $\gamma : I \to \mathbb{R}^k$ defined on an interval $I \subseteq \mathbb{R}$. We say that γ is piecewise C^1 if it has a continuous derivative except at finitely many points $t_1 < t_2 < \cdots < t_n$, where the left and right derivatives are required to exist, but may be different. Equivalently, γ is continuously differentiable in each sub-interval $[t_i, t_{i+1}]$, for each $i = 0, 1, \ldots, n$, with the convention $t_0 = a$, $t_{n+1} = b$.*

Thinking of a curve γ as representing the trajectory of a (system of) particles we also call γ a **motion** *and the derivative $\dot\gamma(t)$ is called the* **velocity** *at time t, whenever it exists.*

We say that a curve $\widetilde\gamma : J \to \mathbb{R}^k$ is a **reparametrization** *of the curve $\gamma : I \to \mathbb{R}^k$ if there exists a strictly monotone continuous function $\varphi : J \to I$, with $\varphi(J) = I$ such that $\widetilde\gamma = \gamma \circ \varphi$. If φ is increasing then γ and $\tilde\gamma$ are said to have the same* **orientation**, *and otherwise they have opposite orientation.*

Definition 1.6. *Let $F : \mathcal{O} \to \mathbb{R}^k$ be a continuous force field and $\gamma : [a, b] \to \mathcal{O}$ a piecewise C^1-curve. We define the* **line integral**, *or* **work integral**, *of F along γ to be*

$$W_\gamma(F) = \int_a^b F(\gamma(t)) \cdot \dot\gamma(t) dt \,,$$

where $\cdot$ denotes the standard scalar product on $\mathbb{R}^k$. In case γ is not C^1, the integral over $[a, b]$ is (as usual) defined as the sum of integrals over the sub-intervals on which γ is C^1.

An important property of the work integral is that it is unchanged under orientation preserving reparametrizations as a consequence of the following result.

Lemma 1.7. *If γ and φ are as in Definition 1.5, with γ and φ piecewise C^1 and φ increasing, then*

$$W_\gamma(F) = W_{\gamma\circ\varphi}(F).$$

Proof. By suitably splitting the interval I into sub-intervals we may assume φ and γ to be C^1. Then the result follows by using integration by substitution with $t = \varphi(s)$:

$$W_{\gamma\circ\phi}(F) = \int_c^d F(\gamma(\varphi(s)) \cdot \dot\gamma(\varphi(s))\varphi'(s) ds = \int_a^b F(\gamma(t)) \cdot \dot\gamma(t) dt = W_\gamma(F) \,.$$

□

A frequently used notation reflecting the reparametrization invariance of $W_\gamma(F)$ is

$$W_\gamma(F) = \int_\gamma F \cdot \mathbf{dr} \,.$$

Note also that the work integral changes sign under a change of orientation of γ (see Exercise 1.9).

For a single particle of mass m moving along a curve γ the kinetic energy at time t is defined as

$$E_{kin}(t) = \frac{1}{2} m \|\dot{\gamma}(t)\|^2 .$$

The following result shows that for a motion fulfilling Newton's second law the work integral of the force equals the change of the particle's kinetic energy.

Theorem 1.8. *For a particle of mass m moving along a C^2-curve $\gamma : I \to \mathcal{O}$ subject to Newton's second law* (1.18) *under the influence of a continuous force field $F : \mathcal{O} \to \mathbb{R}^k$, the identity*

$$E_{kin}(t_2) - E_{kin}(t_1) = \int_{t_1}^{t_2} F(\gamma(t)) \cdot \dot{\gamma}(t) dt$$

holds for all $t_1, t_2 \in I$.

Proof. By differentiation and using Newton's second law we arrive at

$$\frac{d}{dt} \left(\frac{1}{2} m \|\dot{\gamma}(t)\|^2 \right) = m \dot{\gamma}(t) \cdot \ddot{\gamma}(t) = F(\gamma(t)) \cdot \dot{\gamma}(t) .$$

Integrating both sides and using the fundamental theorem of calculus then gives the result. □

As seen in Lemma 1.7 $W_\gamma(F)$ is invariant under increasing reparametrizations of γ. We now consider a very important class of force fields for which the work integral depends only on the end points of the motion. An equivalent way of saying this is that the work integral along any closed curve, i.e. a curve with coinciding endpoints, is zero (see Exercise 1.10).

Definition 1.9. *A continuous force field $F : \mathcal{O} \to \mathbb{R}^k$ is said to be* **conservative** *if*

$$W_\gamma(F) = 0.$$

for all piecewise C^1-curves $\gamma : [a, b] \to \mathcal{O}$ with $\gamma(a) = \gamma(b)$.

As we shall see below the gravitational force field (1.20) is conservative on the set where it is defined (i. e., when no two particles are on top of each other). Examples of force fields which are not conservative can be found in Exercise 1.11.

The next result gives a characterization of conservative force fields in terms of associated potentials as follows.

Theorem 1.10. *A continuous force field $F : \mathcal{O} \to \mathbb{R}^k$ is conservative if and only if there exists a C^1-function $V : \mathcal{O} \to \mathbb{R}$, called a* **potential** *or* **potential energy** *of F, such that $F = -\nabla V$.*

Proof. If there exists a function V as stated in the theorem we have for all piecewise C^1-curves $\gamma : [a, b] \to \mathcal{O}$ that

$$W_\gamma(F) = -\int_a^b \nabla V(\gamma(t)) \cdot \dot{\gamma}(t) dt = -\int_a^b \frac{d}{dt} V(\gamma(t)) dt = V(\gamma(a)) - V(\gamma(b))$$

which vanishes if $\gamma(a) = \gamma(b)$. Hence F is conservative.

On the other hand, if F is conservative we shall now construct a function V such that $-\nabla V = F$. By a simple argument we can assume $\mathcal{O}$ is *pathwise connected*, i. e., for all points $x, x_0 \in \mathcal{O}$ there exists a piecewise C^1-curve $\gamma : [a, b] \to \mathcal{O}$ such that $\gamma(a) = x_0$ and $\gamma(b) = x$. Let x_0 be fixed and define

$$V(x) = -\int_a^b F(\gamma(t)) \cdot \dot{\gamma}(t) dt \,,$$

where $\gamma : [a, b] \to \mathcal{O}$ is any such piecewise C^1-curve. Since F is conservative the value of $V(x)$ above does not depend on the choice of curve γ from x_0 to x (see Exercise 1.10).

In order to differentiate V at $x = (x_1, \ldots, x_k)$ we use that $\mathcal{O}$ is open and choose a ball $B \subseteq \mathcal{O}$ centered at x. Varying the first coordinate we consider another point $x' = (x_1', x_2, \ldots, x_k)$ in the ball B, and connect x to x' by the curve $\mu : t \to x + (t, 0, \ldots, 0)$, along the first axis, where t is between 0 and $x_1' - x_1$. In the definition of $V(x')$ we choose the curve connecting x_0 to x' as the curve γ connecting x_0 to x followed by μ from x to x'. This gives

$$V(x') - V(x) \;\; = \;\; -\int_{x_1}^{x_1'} F_1(t, x_2, \ldots, x_k) dt \,,$$

where F_1 is the first coordinate of F. It now follows from the fundamental theorem of calculus that

$$\frac{V(x') - V(x)}{x_1' - x_1} \to -F_1(x)$$

as $x_1' \to x_1$. Applying the same argument for the other coordinates we conclude that $\nabla V(x) = -F(x)$. Since F is continuous it follows that V is C^1. □

The above result allows us to formulate the theorem of **conservation of energy** for a single particle acted on by a conservative force.

Theorem 1.11. *For a particle of mass m moving along a C^2-curve $\gamma : I \to \mathcal{O}$ according to Newton's second law under the influence of a conservative force field $F = -\nabla V$ on $\mathcal{O} \subseteq \mathbb{R}^3$, the total energy*

$$E(t) = \frac{1}{2} m \|\dot{\gamma}(t)\|^2 + V(\gamma(t))$$

is conserved in the sense that it is a constant function of t.

Proof. By differentiation we have from Newton's second law

$$E'(t) = m\dot{\gamma}(t) \cdot \ddot{\gamma}(t) + \nabla V(\gamma(t)) \cdot \dot{\gamma}(t) = (m\ddot{\gamma}(t) - F(\gamma(t))) \cdot \dot{\gamma}(t) = 0.$$

□

A force field $F : \mathcal{O} \to \mathbb{R}^k$, where $\mathcal{O} = \mathbb{R}^k$ or $\mathcal{O} = \mathbb{R}^k \setminus \{0\}$, is *rotationally invariant* if it fulfills

$$F(Ax) = AF(x) \,, \quad x \in \mathcal{O} \,,$$

for all special orthogonal $k \times k$-matrices A, i. e., A is invertible, $A^{-1} = A^t$ and $\det A = 1$. If $k = 3$ this is equivalent to stating that F is invariant under the transformation (1.16)

with $N = 1$ for all special orthogonal 3×3-matrices A. It is easy to show that rotational invariance implies that F can be written in the form

$$F(x) = f(\|x\|^2)x\,, \quad x \in \mathcal{O}\,, \tag{1.23}$$

where f is a function defined on $[0, \infty[$ or on $]0, \infty[$. Examples of such forces are the gravitational attraction of a point particle and the Coulomb force of a charged point particle at the origin both of which have the form

$$F(x) = \frac{k}{\|x\|^3}\,x\,, \quad x \neq 0\,. \tag{1.24}$$

They have the property of being conservative as a consequence of the following result.

Theorem 1.12. *All rotationally invariant force fields are conservative*

Proof. Write F in the form (1.23). For a piecewise C^1-curve $\gamma : [a, b] \to \mathcal{O}$ we then have

$$\begin{aligned} W_\gamma(F) &= \int_a^b f(\|\gamma(t)\|^2)\gamma(t) \cdot \dot{\gamma}(t)dt \\ &= \frac{1}{2}\int_a^b f(\|\gamma(t)\|^2)\frac{d\|\gamma(t)\|^2}{dt}dt = \frac{1}{2}\int_{\|\gamma(a)\|^2}^{\|\gamma(b)\|^2} f(s)ds, \end{aligned}$$

which explicitly depends only on the endpoints $\gamma(a)$ and $\gamma(b)$. □

If we compare with how the potential V was defined in the proof of Theorem 1.10 we see from the above proof that the potential of a spherically symmetric force field may be chosen spherically symmetric, i.e. to depend only on $\|x\|$. In particular, the potential of the force (1.24) can be chosen to be

$$V(x) = -\frac{k}{\|x\|}\,.$$

Next, consider N particles moving in $\mathbb{R}^3$ subject to a force field $F : \mathcal{O} \to \mathbb{R}^{3N}$, where $\mathcal{O} = \mathbb{R}^{3N}$ or $\mathcal{O} = \mathbb{R}^{3N}_{\neq} \equiv \{x \mid x_i \neq x_j \text{ for all } i \neq j\}$. Writing $F = (F_1, \ldots, F_N)$, where $F_i : \mathcal{O} \to \mathbb{R}^3$, this means that $F_i(x)$ is the force acting on particle i if the particle positions are given by $x = (x_1, \ldots, x_N)$. We say that the forces F_i are conservative if F is conservative as a force field on $\mathcal{O} \subseteq \mathbb{R}^{3N}$. By Theorems 1.10 and 1.12 this means that there is a C^1-function V on $\mathcal{O}$ such that

$$F_i = -\nabla_{x_i} V(x_1, \ldots, x_N)\,.$$

We note the following immediate generalization of Theorem 1.11.

Theorem 1.13. *For N particles with masses $m_1, \ldots, m_N$, moving along C^2-curves $(\gamma_1(t), \ldots, \gamma_N(t))$ according to Newton's second law under the influence of conservative forces $F_i = -\nabla_i V$, the total energy*

$$\boxed{E = \sum_i \frac{1}{2} m_i \dot{\gamma}_i(t)^2 + V(\gamma_1(t), \ldots, \gamma_N(t))}$$

is conserved.

Proof. The statement follows from

$$E'(t) = \sum_i m_i \dot{\gamma}_i(t) \cdot \ddot{\gamma}_i(t) + \sum_i \nabla_{x_i} V(\gamma_1(t), \ldots, \gamma_N(t)) \cdot \dot{\gamma}_i(t) = 0.$$

□

The gravitational force between N point-particles with masses $m_1, \ldots, m_N$ located at points $x_1, \ldots, x_N$ in $\mathbb{R}^3$ is conservative on $\mathcal{O} = \mathbb{R}^{3N}_{\neq}$. In fact, the potential

$$V(x_1, \ldots, x_N) = -\sum_{1 \leqslant i < j \leqslant N} G \frac{m_i m_j}{\|x_i - x_j\|}, \tag{1.25}$$

fulfills

$$\nabla_{x_i} V(x_1, \ldots, x_N) = \sum_{j \neq i} G m_i m_j \frac{x_i - x_j}{\|x_i - x_j\|^3}. \tag{1.26}$$

We end this subsection by giving a characterization of translation invariant conservative forces.

Theorem 1.14. *The potential of a translation invariant conservative force field F acting on a system of N particles in $\mathbb{R}^3$ can be written as*

$$V(x_1, \ldots, x_N) = W(x_2 - x_1, \ldots, x_N - x_1) + F_0 \cdot (x_1 + \ldots + x_N), \tag{1.27}$$

where $W(y_2, \ldots, y_N)$ is a function of $3(N-1)$ variables and $F_0 \in \mathbb{R}^3$ is a constant vector.

Note that the last term in (1.27) is the potential of a constant force F_0 acting on each particle. One refers to the first term on the right-hand side, which depends on the relative positions of the particles only, as the *interior potential* and to F_0 as an *exterior constant force*. The fact that the interior interaction is written with x_1 as subtraction point is not important, see Exercise 1.13.

Proof. Translation invariance means that $F(x_1, \ldots, x_N) = F(x_1 - a, \ldots, x_N - a)$, that is

$$\nabla_{x_i} V(x_1, \ldots, x_N) = \nabla_{x_i} V(x_1 - a, \ldots, x_N - a)$$

for all $a \in \mathbb{R}^k$ and all $i = 1, \ldots, N$. This implies that

$$V(x_1, \ldots, x_N) = V(x_1 - a, \ldots, x_N - a) + g(a),$$

for all $a \in \mathbb{R}^3$ and some function $g : \mathbb{R}^3 \to \mathbb{R}$ with $g(0) = 0$. Differentiating this equation w. r. t. a gives by the chain rule

$$\nabla g(a) = \sum_{i=1}^{N} \nabla_{x_i} V(x_1 - a, \ldots, x_N - a) = \sum_{i=1}^{N} \nabla_{x_i} V(x_1, \ldots, x_N).$$

In other words, both sides of this equation are constant. Hence $g(a) = b \cdot a + c$ for some $b \in \mathbb{R}^3$ and $c \in \mathbb{R}$. Since $g(0) = 0$ we have $c = 0$. Hence $g(a) = b \cdot a$.

Choosing $a = x_1$ gives

$$\begin{aligned} V(x_1, \ldots, x_N) &= V(0, x_2 - x_1, \ldots, x_N - x_1) + b \cdot x_1 \\ &= V(0, x_2 - x_1, \ldots, x_N - x_1) - \frac{1}{N} b \cdot ((x_2 - x_1) + \ldots + (x_N - x_1)) \\ &\quad + \frac{1}{N} b \cdot (x_1 + \ldots + x_N) \\ &= W(x_2 - x_1, \ldots, x_N - x_1) + F_0 \cdot (x_1 + \ldots + x_N), \end{aligned}$$

where

$$W(y_2, \ldots, y_N) = V(0, y_2, \ldots, y_N) - \frac{1}{N} b \cdot (y_2 + \ldots + y_N)$$

and $F_0 = \frac{1}{N} b$. □

In the particular case $N = 2$, this result implies that the interior potential of two interacting particles subject to translation invariant conservative forces is of the form $W(x_2 - x_1)$. Hence the force with which particle 1 acts on particle 2 is $F_{12} = -\nabla W(x_2 - x_1)$ which is equal in size but opposite to $F_{21} = \nabla W(x_2 - x_1)$, the force with which particle 2 acts on particle 1. This is the *action-reaction principle* or **Newton's third law**. If we also assume that the force field is invariant under rotations in $\mathbb{R}^3$ we get the *strong action-reaction principle* that $F_{12} = -F_{21}$ is directed along the line joining the particles (see Exercise 1.7).

1.5 Configuration space, phase space and state space

That the equation of motion for a particle is of second order is intimately connected to the empirical fact that one can in general specify arbitrarily the location and the velocity of a particle at a given instant of time, and once this has been done the motion of the particle is uniquely determined by the acting forces, and similarly for systems of particles. The following theorem on the existence and uniqueness of solutions to a second order differential equation is crucial in this connection.

Theorem 1.15. *Let $f : \Omega \to \mathbb{R}^k$ be a C^∞-function defined on an open subset Ω of $\mathbb{R}^{2k+1}$ and let $x_0, \dot{x}_0 \in \mathbb{R}^k$ and $t_0 \in \mathbb{R}$ be given such that $(x_0, \dot{x}_0, t_0) \in \Omega$. Then there exists a unique solution $x : I \to \mathbb{R}^k$ to the equation*

$$\ddot{x} = f(x, \dot{x}, t) \tag{1.28}$$

defined on a maximal open interval I such that $t_0 \in I$ and the initial conditions $x(t_0) = x_0$ and $\dot{x}(t_0) = \dot{x}_0$ hold.

We shall not prove this theorem here. A proof can be found in [3] §7. More detailed results can be found in, e. g., [4]. The regularity assumption made for f is unnecessarily strong, but, on the other hand, some regularity condition stronger than continuity is needed (see Exercise 1.18).

In order to apply the theorem to the equations of motion for a system of N particles,

$$m_i \ddot{x}_i = F_i\,, \quad i = 1, \ldots, N\,, \tag{1.29}$$

we set as before $k = 3N$, $x = (x_1, \dots, x_N)$, and $f = (m_1^{-1}F_1, \dots, m_N^{-1}F_N)$ in terms of which the equations take the form (1.28). Thus, assuming that each F_i is a C^∞-function of the positions and velocities of the particles and of time, we conclude that there exists a unique maximal solution with prescribed initial values $(x_1(t_0), \dots, x_N(t_0))$ and $(\dot{x}_1(t_0), \dots, \dot{x}_N(t_0))$.

A result analogous to Theorem 1.15 is valid for equations of arbitrary order $n \geqslant 1$, in which case the values of x and all derivatives up to order $n-1$ at some instant t_0 need to be specified in order to ensure existence and uniqueness of a solution. In particular, for a first order equation the initial condition consists in specifying the value of x itself only, at a given time. We shall make use of this instance of the theorem in Chapter 3 when discussing Hamilton's equations. For a third order equation one would need to specify location and velocity as well as the acceleration at a given instant of time. Since this would be in conflict with experience such an equation is not a candidate equation of motion.

Note that the theorem does not in general provide solutions defined on the full time axis, called globally defined solutions, but only ensure existence on a subinterval. For linear equations, that is when each $F_i : \mathbb{R}^{2k+1} \to \mathbb{R}^{2k}$ is a linear function of x and $\dot{x}$, for fixed time t, all solutions are in fact globally defined, but for non-linear F_i solutions may grow arbitrarily large in finite time or develop other types of singularities (see Exercises 1.15 and 1.17). For the sake of simplicity, we will assume in the following that all solutions are globally defined.

It is convenient to describe the solutions to (1.29) as motions in the *phase space* $\mathbb{R}^{6N}$ of the system. Here we think of phase space as the space of initial data, and a motion in phase space is a function $z : I \to \mathbb{R}^{6N}$, where I is an interval (see Definition 1.5). Theorem 1.15 asserts that specifying a point $z_0 = (x_0, \dot{x}_0)$ in phase space and a time t_0 there is a unique curve $z : t \to (x(t), \dot{x}(t))$ determined by (1.29) which passes through the given point at time t_0, i.e. such that $z(t_0) = z_0$. Assuming that all solutions are globally defined we obtain in this way for each fixed t a mapping $\Psi_{t,t_0} : \mathbb{R}^{6N} \to \mathbb{R}^{6N}$ defined by

$$\Psi_{t,t_0}(z_0) = z(t) \quad \text{for } z(t_0) = z_0 \in \mathbb{R}^{6N}\,.$$

Since this holds for all $t_0, t \in \mathbb{R}$ it follows from Theorem 1.15 (see Exercise 1.16) that $\Psi_{t_0,t}$ is bijective and fulfills

$$\Psi_{t_2,t_1} \circ \Psi_{t_1,t_0} = \Psi_{t_2,t_0} \quad \text{and} \quad (\Psi_{t_1,t_0})^{-1} = \Psi_{t_0,t_1} \tag{1.30}$$

for arbitrary $t_0, t_1, t_2 \in \mathbb{R}$. The family of maps Ψ_{t,t_0}, $t, t_0 \in \mathbb{R}$, is called the *phase flow* of the equations (1.29).

It is also useful to describe the solutions to (1.29) as curves in *configuration space* $\mathbb{R}^{3N}$, considered as the space of initial configurations in space of the system. Since the initial positions alone do not determine the temporal development of the system there is no analogue of the phase flow in configuration space. The dimension of configuration space is half the dimension of phase space and is called the *number of degrees of freedom of the system*.

Finally, let us mention that a point in phase space of a system of N particles is often called a *state* of the system. Thus, given a state of the system at a given instant of time, its state is determined at all times, provided the acting forces are known. This notion of state is, however, too restrictive for some purposes. Thus, for typical thermodynamic systems,

the number of particles is sufficiently large that their individual positions and velocities cannot be determined in practice. A more convenient description of such a macroscopic system is obtained by using a probability distribution on phase space. More precisely, a *thermodynamic state* of the system at a given time t_0 is given by a non-negative function p on phase space, fulfilling

$$\int_{\mathbb{R}^{6N}} p(x,v)d^{3N}xd^{3N}v = 1\,.$$

The probability distribution p_t at time t is then given by

$$p_t(x,v) = p(\Psi_{t,t_0}^{-1}(x,v)) = p(\Psi_{t_0,t}(x,v))\,.$$

Note that $\Psi_{t_0,t}$ is the "backwards flow" from t to t_0. The fact, that this is still a probability distribution is known as *Liouville's Theorem*, which will be proven in Chapter 3. *Thermodynamic equilibrium states* are defined as states that are invariant under the phase flow in the sense that $p_t = p$ for all t.

As one can imagine, a general thermodynamic equilibrium state can be very complicated as it depends on a large number of variables – the $6N$ variables labelling states of the *microscopic* particles. In thermodynamics one is mostly concerned with states that depend on a small number of *macroscopic physical variables*, e. g. volume and temperature, besides the phase space variables.

As an example we consider N particles moving freely in a box $[0,L]^3 \subseteq \mathbb{R}^3$ subject to the assumption that when a particle hits the boundary of the box the component of its velocity orthogonal to the boundary is reflected. The phase flow then preserves the lengths of the individual velocities $v_1,\ldots,v_N$. The so-called *Maxwell-Boltzmann distribution* provides an example of a thermodynamic equilibrium state for such a system. It depends on the volume $V = L^3$ and temperature $T > 0$ of the system and is given by (see Exercise 1.19)

$$p(x_1,\ldots,x_N,v_1,\ldots,v_N) = L^{-3N}\prod_{i=1}^{N}\left[\left(\frac{m_i}{2\pi k_{\mathrm{B}}T}\right)^{3/2}\exp\left(-\frac{\frac{1}{2}m_iv_i^2}{k_{\mathrm{B}}T}\right)\right], \tag{1.31}$$

where the constant $k_{\mathrm{B}} = 1.38\times10^{-23}$ Joule/Kelvin is *Boltzmann's constant*. Since p depends only on the lengths of the velocities which are unchanged during the time evolution, it is a thermodynamic equilibrium state (see Exercise 1.19).

In this state the average kinetic energy, which equals the internal energy U of the ideal gas, is (see Exercise 1.20)

$$U = \sum_{i=1}^{N}\int\frac{1}{2}m_iv_i^2p(x_1,\ldots,x_N,v_1,\ldots,v_N)d^{3N}xd^{3N}v = \frac{3}{2}Nk_{\mathrm{B}}T. \tag{1.32}$$

In Exercise 1.21 it is shown how to arrive at the *ideal gas equation*

$$PV = Nk_{\mathrm{B}}T\,, \tag{1.33}$$

where P denotes the pressure.

1.6 Digression: Exact and closed differentials

In this section, we give a very brief introduction to the subject of *differential forms*, which provides an elegant machinery engulfing line integrals, surface integrals and alike in a unified and coordinate-independent manner. In particular, all the classical integral theorems in vector calculus, such as the theorems of Gauss, Gauss–Green, and Stokes, can be formulated in terms of differential forms as one strikingly simple formula:

$$\int_{\partial A} \alpha = \int_A d\alpha \,.$$

Here $d\alpha$ denotes the so-called exterior derivative of a differential form α, and A is some "reasonable" object which can be integrated over (e. g. a smooth curve) and ∂A is the "boundary" of A (e. g., if A is a smooth curve, then ∂A consists of its endpoints). For a basic introduction to the theory of differential forms, we refer to V.I. Arnol'd: Mathematical Methods of Classical Mechanics, chapter 7.

Of course, the beauty of the theory of differential forms does not come without a price to pay: an introduction requires a great deal of linear algebra and analysis. We shall here limit ourselves to a very short digression into so-called *differential 1-forms* or simply *1-forms* on $\mathbb{R}^k$, partly motivated by the characterization of conservative force fields in Theorem 1.10, and to a brief discussion of their application in thermodynamics.

In previous sections we have discussed functions and vector fields on the configuration space modeled by (a subset of) $\mathbb{R}^k$, the coordinates of which have a common interpretation and are measured in the same units. In particular, it makes good sense in this case to think of the gradient of a function as a vector field in space, since the gradient at a point makes sense as a vector independently of the orthogonal coordinate system chosen to calculate it (see Exercise 1.14). It can, indeed, be characterized as pointing in the direction in which the function has the fastest increase and its size is the derivative of the function in this direction.

In physics, however, we frequently encounter functions f of several variables, that represent fundamentally different quantities. In thermodynamics, for instance, one may consider the pressure $P(U, V)$ of a system as a function of the internal energy U and the volume V of the system. In this case, it is not too meaningful to think of the gradient of P as a vector. In fact, the two components $\frac{\partial P}{\partial U}$ and $\frac{\partial P}{\partial V}$ have different units, hence the length of such a vector or the angle between such vectors does not have any physical meaning. Instead, it is more appropriate to introduce the notion of *differential*

$$dP = \frac{\partial P}{\partial U} dU + \frac{\partial P}{\partial V} dV \,,$$

where dP, dU, dV are often interpreted as infinitesimal quantities.

We now give a rigorous definition of the differential of a function which attributes a precise meaning to the expression above.

Definition 1.16. *If $f : \mathcal{O} \to \mathbb{R}$ is a C^1-function on an open set $\mathcal{O} \subseteq \mathbb{R}^k$ we define the* **differential** *of f at the point $x \in \mathcal{O}$ to be the linear map $df_x : \mathbb{R}^k \to \mathbb{R}$ given by*

$$df_x(X) = \sum_{i=1}^{k} \frac{\partial f}{\partial x_i}(x) X_i, \quad X = (X_1, \ldots, X_k) \in \mathbb{R}^k \,.$$

As usual, we identify the coordinates x_i, $i = 1, \ldots, k$, with the coordinate functions $x = (x_1, \ldots, x_k) \mapsto x_i$. The differential of x_i is then given by

$$dx_i(X) = X_i.$$

Note that dx_i is the same map at all points. Hence we may write

$$df_x = \frac{\partial f}{\partial x_1}(x)dx_1 + \ldots \frac{\partial f}{\partial x_k}(x)dx_k,$$

or simply

$$df = \frac{\partial f}{\partial x_1}dx_1 + \ldots \frac{\partial f}{\partial x_k}dx_k,$$

where we have omitted explicit reference to the point at which the differential of f is calculated. Thus df is a differential 1-form according to the following definition.

Definition 1.17. *A* **(differential) 1-form** ω *on an open subset of* $\mathcal{O} \subseteq \mathbb{R}^k$ *is a map from* $\mathcal{O}$ *to the linear functions from* $\mathbb{R}^k$ *to* $\mathbb{R}$ *of the form*[1]

$$\omega_x = \sum_{i=1}^{k} \omega_i(x)dx_i,$$

where the functions $\omega_i : \mathcal{O} \to \mathbb{R}$, $i = 1, \ldots, k$, *are continuous. Thus, for a vector* $X \in \mathbb{R}^k$ *we have*

$$\omega_x(X) = \sum_{i=1}^{k} \omega_i(x)X_i$$

If ω *is the differential of a function, that is if there is a* C^1*-function* $f : \mathcal{O} \to \mathbb{R}$ *such that* $\omega = df$*, we say that* ω *is an* **exact** 1*-form.*

For any vector field $F : \mathcal{O} \to \mathbb{R}^k$ there is a corresponding 1-form ω_F given by $\omega_F = F_1dx_1 + \ldots + F_kdx_k$ or, equivalently,

$$\omega_F(X) = F \cdot X\,, \quad X \in \mathbb{R}^k\,.$$

For vector fields we introduced the notion of line integral in Definition 1.6. The corresponding notion for 1-forms is defined as follows.

Definition 1.18. *If* $\omega = \sum_{i=1}^{k} \omega_i dx_i$ *is a 1-form on an open set* $\mathcal{O} \subseteq \mathbb{R}^k$ *and* $\gamma : I \to \mathcal{O}$ *is a piecewise* C^1*-curve we define the* **line integral** *of* ω *along* γ *by*

$$\int_\gamma \omega = \int_I \omega_{\gamma(t)}(\dot\gamma(t))dt = \sum_{i=1}^{k} \int_I \omega_i(\gamma(t))\dot\gamma_i(t)dt.$$

Evidently,

$$\int_\gamma \omega_F = \int_\gamma F \cdot \mathbf{dr}\,,$$

for any vector field F, and consequently Theorem 1.10 for vector fields may be translated into the following statement about 1-forms.

[1] Any map from $\mathcal{O}$ to the linear functions from $\mathbb{R}^k$ to $\mathbb{R}$ can be written in this form. The extra requirement is that the coordinate functions ω_i are continuous.

Theorem 1.19. *A 1-form ω defined on an open set $\mathcal{O} \subseteq \mathbb{R}^k$ is exact if and only if $\int_\gamma \omega = 0$ for any closed piecewise C^1-curve.*

For a C^2-function $f : \mathcal{O} \to \mathbb{R}^k$ it is well-known that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Thus, a 1-form $\omega = \sum_{i=1}^k \omega_i dx_i$ such that the functions ω_i are C^1, in which case we say that ω is (of class) C^1, can only be exact if

$$\frac{\partial \omega_i}{\partial x_j} = \frac{\partial \omega_j}{\partial x_i}, \tag{1.34}$$

for all $i, j = 1, \ldots, k$.

Definition 1.20. *A 1-form ω of class C^1 is said to be* **closed** *if it satisfies the relation (1.34) for all $i, j = 1, \ldots, k$.*

Hence, we have seen that exact differential forms are necessarily closed:

$$\boxed{\text{If } \omega \text{ is } C^1 \text{ and exact, then } \omega \text{ is closed}}$$

However, as can be seen by simple examples (see Exercise 1.23), the converse statement is not true in general. But it turns out that this depends on the shape of the domain $\mathcal{O}$ on which the differential is defined. A very useful sufficient condition ensuring that the notions of exactness and closedness of 1-forms on $\mathcal{O}$ coincide l sufficient condition ensuring that the notions of exactness and closedness of 1-forms on $\mathcal{O}$ coincide is that $\mathcal{O}$ is *simply connected*, which means that any closed curve in $\mathcal{O}$ can be continuously deformed to a point entirely within $\mathcal{O}$. The following theorem is a precise statement of this result. A proof in the case when $\mathcal{O}$ is a star-shaped set can be found in Exercise 1.24.

Theorem 1.21. *If the open set $\mathcal{O} \subseteq \mathbb{R}^k$ is simply connected then any 1-form ω on $\mathcal{O}$ of class C^1 is exact if and only if it is closed.*

Remark 1.22. The set $\mathbb{R}^2 \backslash \{0\}$ is not simply connected. Exercise 1.23 provides an example of a 1-form which is closed but not exact on this set.

Example 1.23 (Magnetic fields in two dimensions)**.** We consider a charged particle moving in an open subset $\mathcal{O}$ of the two-dimensional plane $\mathbb{R}^2$ under the influence of a magnetic field perpendicular to the plane of motion. Hence, the magnetic field is described by a function $B : \mathcal{O} \to \mathbb{R}$ which equals the third component of the three-component magnetic field vector.

A 1-form $\alpha = \alpha_1 dx + \alpha_2 dy$ on $\mathcal{O}$ is called a vector potential of B if $B = \frac{\partial \alpha_2}{\partial x} - \frac{\partial \alpha_1}{\partial y}$. Hence, if α is closed the corresponding magnetic field vanishes. Strangely, this does not necessarily mean that the vector potential has no effect. Indeed, if the differential is not exact, it may have an observable effect on the motion of a quantum mechanical particle. A particular instance is that of a charged particle moving in the set $\mathbb{R}^2 \setminus \{0\}$ under the influence of the vector potential given by the closed 1-form in Exercise 1.23. It is then observed that the motion of the particle is different from that of a free particle. This effect is called the Bohm-Aharonov effect.

Definition 1.24. *If ω is a 1-form on an open set $\mathcal{O} \subseteq \mathbb{R}^k$, we say that a continuous function $\tau : \mathcal{O} \to \mathbb{R}$ with τ not identically equal to zero, is an* **integrating factor** *for ω if the differential $\tau\omega$ is exact.*

We quote without proof the following theorem on the existence of integrating factors in the case of two variables.

Theorem 1.25. *Any 1-form ω on $\mathcal{O} \subseteq \mathbb{R}^2$ locally has an integrating factor, i. e., for any $x \in \mathcal{O}$ there exists a disc $D \subseteq \mathcal{O}$ centered at x, such that ω has an integrating factor on D.*

Example 1.26 (Integrating factors, temperature and entropy)**.** As a last example we discuss a thermodynamical system described by two macroscopic variables U and V, the internal energy and the volume, respectively. The pressure of the system, i. e., the force per area exerted by the system on the boundary of its domain, is then a function $P(U, V)$.

The 1-form

$$\omega = dU + P(U, V)dV$$

plays an important role in the thermodynamic description of such a system. The significance of this 1-form can heuristically be described as follows. If the system receives an infinitesimal energy in the form of heat this will be either stored as an infinitesimal internal energy dU or used for the system to do work, i. e., to change its volume infinitesimally. The work done by the system in changing its volume by dV is PdV (why?). Thus the total heat received by the system equals $dU + PdV$.

The 1-form ω is not exact, but it has an integrating factor, which is the reciprocal temperature $T(U, V)^{-1}$ (this is indeed a way to introduce the temperature). Thus there exists a function $S(U, V)$ such that

$$dS = T(U, V)^{-1}(dU + PdV)\,.$$

The function S, which is determined in this way up to an overall additive constant, is called the *entropy* of the system. The fact that the heat differential has an integrating factor follows, at least locally, from Theorem 1.25, since the system depends only on two variables U and V.

Examples of systems that depend on more than two variables can be obtained by bringing several, say M, systems as above in thermal contact, meaning that they can transfer their internal energies. Such a combined system is described by the total internal energy U and the volumes $V_1, V_2, \ldots, V_M$ of the individual subsystems. The heat differential is in this case

$$\omega = dU + P_1 dV_1 + \ldots + P_M dV_M\,,$$

where $P_1, \ldots, P_M$ are the pressures of the subsystems and are functions of $U, V_1, \ldots, V_M$. It is an important consequence of the so-called *Second Law of Thermodynamics* that ω has the common temperature of the combined system as integrating factor, that is we have again

$$TdS = dU + P_1 dV_1 + \ldots + P_M dV_M\,,$$

where the entropy S and temperature T are functions of $U, V_1, \ldots, V_M$.

Exercises

Exercise 1.1. Prove the following statements relating to the scalar product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^3$.

a) If the vectors $e_1, e_2, e_3 \in \mathbb{R}^3$ fulfill

$$\langle e_i, e_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} ,$$

then (e_1, e_2, e_3) is a basis for $\mathbb{R}^3$. It is called an *orthonormal basis* for $\mathbb{R}^3$.

b) Let (e_1, e_2, e_3) be an orthonormal basis for $\mathbb{R}^3$ and let $v \in \mathbb{R}^3$. Use the linearity properties of the scalar product to show that the coordinate set (x_1, x_2, x_3) of v with respect to the basis (e_1, e_2, e_3) is given by

$$x_i = \langle v, e_i \rangle , \quad i = 1, 2, 3 .$$

c) Use the linearity properties of the scalar product to verify the identities

$$\begin{aligned} \langle v, w \rangle &= \frac{1}{2}\|v\|^2 + \frac{1}{2}\|w\|^2 - \frac{1}{2}\|v - w\|^2 \\ \langle v, w \rangle &= \frac{1}{4}\|v + w\|^2 - \frac{1}{4}\|v - w\|^2 \end{aligned}$$

d) Derive the *Cauchy-Schwarz inequality*

$$|\langle v, w \rangle| \leqslant \|v\|\|w\| , \quad v, w \in \mathbb{R}^3 .$$

Hint. One possible way to proceed is to consider the polynomial $f(x) = \|v + xw\|^2$ and use that, since it is nonnegative, its discriminant must be less than or equal to zero.

Exercise 1.2. Let $S : \mathbb{R}^3 \to \mathbb{R}^3$ be an isometry. Prove that

$$S(v) = a + S_0(v) ,$$

where $a \in \mathbb{R}^3$ is a fixed vector and $S_0 : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear isometry, in the following two steps.

a) Set

$$a = S(\underline{0})$$

and define S_0 by

$$S_0(v) = S(v) - a , \quad v \in \mathbb{R}^3 .$$

Verify that $S_0(\underline{0}) = \underline{0}$ and that S_0 is an isometry, and use the first identity in Exercise 1.1c) to show

$$\langle S_0(v), S_0(w) \rangle = \langle v, w \rangle , \quad v, w \in \mathbb{R}^3 .$$

b) Show, using a), that

$$\|S_0(\lambda v) - \lambda S_0(v)\|^2 = 0$$
$$\|S_0(v+w) - S_0(v) - S_0(w)\|^2 = 0\,,$$

and conclude that S_0 is linear.

Exercise 1.3. Assume $S_0 : \mathbb{R}^3 \to \mathbb{R}^3$ is an orthogonal transformation and let A be the 3×3-matrix representing S_0 with respect to the standard basis $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ for $\mathbb{R}^3$, that is

$$S_0(v) = Av\,, \quad v \in \mathbb{R}^3\,.$$

a) Show that the columns of A constitute an orthonormal basis for $\mathbb{R}^3$ and that this is equivalent to

$$A^t A = I\,,$$

where A^t denotes the transpose of A (defined by substituting the columns of A by the corresponding rows), and I is the 3×3 identity matrix.

Conclude that A is invertible and that

$$A^{-1} = A^t\,. \tag{1.35}$$

Matrices fulfilling (1.35) are called orthogonal matrices and the set of orthogonal 3×3-matrices is denoted by $O(3)$.

b) Show that A is an orthogonal matrix if and only if the corresponding orthogonal transformation S_0 maps an orthonormal basis to an orthonormal basis.

c) Show that I is an orthogonal matrix and that, if A, B are orthogonal matrices, then AB and A^{-1} are also orthogonal matrices.

Exercise 1.4. In this exercise A denotes an orthogonal 3×3-matrix and S_0 the corresponding linear map, $S_0 v = Av\,,\ v \in \mathbb{R}^3$.

a) Let $\lambda \in \mathbb{R}$. Show that the equation

$$Av = \lambda v$$

has a solution $v \neq \underline{0}$ if and only if

$$\det(A - \lambda I) = 0\,. \tag{1.36}$$

Show that the left hand side of (1.36) is a third order polynomial in λ and that it has at least one root $\lambda_1 \in \mathbb{R}$ and conclude from $\|Av\| = \|v\|$ that

$$\lambda_1 = \pm 1.$$

b) Explain that there exists a vector $v_1 \in \mathbb{R}^3$ such that

$$Av_1 = \lambda_1 v_1 \quad \text{and} \quad \|v_1\| = 1\,,$$

and that, given such a vector v_1, there exist vectors $v_2, v_3 \in \mathbb{R}^3$ such that (v_1, v_2, v_3) is an orthonormal basis for $\mathbb{R}^3$.

Show that the matrix representing S_0 with respect to the basis (v_1, v_2, v_3) is of the form

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{pmatrix},$$

where the columns $(0, a, c)$ and $(0, b, d)$ are orthogonal and of norm 1. Show that it can be written in one of the following two forms:

$$B_1 = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, \qquad B_2 = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & \sin\theta & -\cos\theta \end{pmatrix},$$

for some $\theta \in [0, 2\pi[$.

c) i) If $\lambda_1 = 1$, show that B_1 represents a rotation through an angle θ around the v_1-axis, i. e., show that vectors parallel to v_1 are left invariant by S_0 and that if v is orthogonal to v_1 then $S_0(v)$ is also orthogonal to v_1 with same norm and forming an angle θ with v.

ii) If $\lambda_1 = 1$, show that B_2 represents a rotation through an angle θ around the v_1-axis composed with reflection in the (v_1, v_2)-plane.

iii) If $\lambda_1 = -1$, show that B_1 represents a rotation through an angle θ around the v_1-axis composed with reflection in the (v_2, v_3)-plane.

iv) If $\lambda_1 = -1$, show that B_2 represents a rotation through an angle π around the axis obtained by rotating the v_2-axis through an angle $\theta/2$ in the (v_2, v_3)-plane.

Exercise 1.5. Verify formulas (1.11), (1.12) and (1.13).

Exercise 1.6. Show that, for a system of N particles, the principle of relativity implies that the force acting on a particle can depend only on the spatial coordinate differences and the velocity differences in an inertial system of coordinates, and that it transforms according to equation (1.16) under an inertial coordinate transformation $\psi[0, 0, 0, A]$.

Exercise 1.7. a) Derive the expression (1.23) for a rotationally invariant force field in $\mathbb{R}^3$.

b) Show, for the case of two interacting particles, that if the forces F_i, $i = 1, 2$, are of the form

$$F_1 = g_1(\|x_1 - x_2\|)(x_2 - x_1), \qquad F_2 = g_2(\|x_1 - x_2\|)(x_1 - x_2),$$

where g_1 and g_2 are C^∞-functions on the positive real axis, then the equations of motion (1.15) are invariant under inertial coordinate transformations.

c) Show, that if the forces are assumed to be independent of velocities then invariance of the equations of motion under inertial coordinate transformations implies that the forces must be of the form given above (for $x_1 \neq x_2$).

Exercise 1.8. a) Show that the force field $F_1(x, y) = (x^2 + y, y^2 + x)$ is conservative on $\mathbb{R}^2$.

b) Find the work integral of F_1 along a motion going in a straight line from $(0, 0)$ to $(1, 0)$ and then in a straight line from $(1, 0)$ to $(1, 1)$.

c) Find the work integral for the force field $F_2(x, y) = (y^2, x^2)$ along the same motion as above.

d) Is the force field F_2 above conservative?

Exercise 1.9. Let $F : \Omega \to \mathbb{R}^k$ be a continuous vector field and $\gamma : [a, b] \to \Omega$ a piecewise C^1 curve and $\varphi : [c, d] \to [a, b]$ a strictly monotone piecewise C^1-function with $\varphi(c) = b$ and $\varphi(d) = a$. Let $\tilde{\gamma} = \gamma \circ \varphi$.

Show that

$$\int_\gamma F \cdot \mathbf{dr} \;\; = \;\; -\int_{\tilde{\gamma}} F \cdot \mathbf{dr}$$

Exercise 1.10. Show that a force field is conservative if and only if the work integral of the field along any piecewise C^1-curve depends only on the endpoints of the motion.

Exercise 1.11. Show that the vector fields

$$F(x, y) = (-y, x)\,, \qquad (x, y) \in \mathbb{R}^2\,,$$

and

$$G(x, y) = \left(\frac{-y}{\sqrt{x^2 + y^2}}, \frac{x}{\sqrt{x^2 + y^2}} \right)\,, \qquad (x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}\,,$$

are not conservative.

Hint. Evaluate the integrals of F and G along a suitable closed curve.

Exercise 1.12. Prove the relation (1.26) for the gravitational force.

Exercise 1.13. Given a function $W(y_2, \ldots, y_N)$ of $N - 1$ variables in $\mathbb{R}^k$ show that there exists a function W' of $N - 1$ variables in $\mathbb{R}^k$ such that

$$W(x_2 - x_1, x_3 - x_1, \ldots, x_N - x_1) = W'(x_1 - x_2, x_3 - x_2 \ldots, x_N - x_2).$$

Exercise 1.14. a) Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a C^1-function. For any 3×3 matrix let $g_A(x) = f(Ax)$. Show that $\nabla g_A(x) = A^t \nabla f(Ax)$.

b) We want to investigate how the gradient changes under a coordinate transformation given by an invertible 3×3 matrix A. Thus introduce new coordinates $y = Ax$. In the new coordinates a C^1-function $f : \mathbb{R}^3 \to \mathbb{R}$ becomes $g(y) = f(A^{-1}y)$. Show, using the result of the previous question, that if and only if the matrix A represents an orthogonal transformation (see Exercise 1.3) does the gradient transform like a vector, i. e. $\nabla g(y) = A \nabla f(A^{-1}y)$, for all functions f.

Exercise 1.15. Find all maximal solutions to the first order differential equation

$$\frac{dx}{dt} = 1 + x^2\,, \qquad (x,t) \in \mathbb{R}^2\,,$$

and verify that they are not globally defined.

Exercise 1.16. Prove the relations (1.30) for the phase flow.

Exercise 1.17. Consider a particle in one dimension with potential energy $U(x) = -x^4$, i. e., its equation of motion is

$$m\ddot{x} = -U'(x) = 4x^3\,.$$

a) Show that the energy

$$E = \frac{1}{2}m\dot{x}^2 - x^4$$

is conserved.

b) Use a) to show that for a solution $x(t)$, $t \in [t_1, t_2]$, the time $t_2 - t_1$ it takes to move from a point x_1 to a point x_2 is given by

$$t_2 - t_1 = \int_{x_1}^{x_2} \frac{dx}{\sqrt{\frac{2}{m}(E + x^4)}}\,,$$

assuming that $x_1 < x_2$ and that $\dot{x} \neq 0$ on $[t_1, t_2]$.

c) Use b) to show that, for $E > 0$, the particle escapes to infinity in finite time, i. e., the solution is not globally defined.

Exercise 1.18. Consider the first order differential equation

$$\frac{dx}{dt} = f(x,t)\,,$$

where the continuous function $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$f(x,t) = \begin{cases} \sqrt{x} & \text{if } x \geqslant 0 \\ 0 & \text{if } x < 0\,. \end{cases}$$

Find two different solutions x_1 and x_2 such that $x_1(0) = x_2(0) = 0$.

Exercise 1.19. Use the fact that $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$ to show that the Maxwell-Boltzmann distribution (1.31) is a probability distribution, i. e., that

$$\int p(x_1, \ldots, x_N, v_1, \ldots, v_N) d^{3N}x d^{3N}v = 1$$

Exercise 1.20. Use the fact that $\frac{d}{d\alpha}\int_{-\infty}^{\infty} e^{-\alpha x^2} dx = -\int_{-\infty}^{\infty} x^2 e^{-\alpha x^2} dx$ (can you prove this?) and $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$ to show the identity (1.32). Show that the corresponding relation in k space dimensions would be $U = \frac{k}{2} N k_{\mathrm{B}} T$. This is often referred to as the *equipartition of energy*, i. e., that each degree of freedom contributes an equal amount $\frac{1}{2} N k_{\mathrm{B}} T$ to the total energy. If the particles have internal degrees of freedom they will also contribute to the total energy.

Exercise 1.21 (Ideal gas equation)**.** We study again the system of N particles in the box $[0, L]^3$ given by the Maxwell-Boltzmann distribution (1.31). When a particle with velocity $v = (v^1, v^2, v^3)$, $v^1 \leqslant 0$, hits the wall $x^1 = 0$ its velocity changes to $(-v^1, v^2, v^3)$. If its mass is m then its *momentum* has therefore changed by the amount $-2mv^1$ in the first coordinate direction. The lost momentum, i. e. $2mv^1$, is said to be transferred to the wall. The absolute value of the expected total transfer of momentum per unit time and unit area is by definition *the pressure* P of the system.

Argue that a particle whose velocity component in the x^1-direction is v_i^1 will hit the wall $x^1 = 0$ approximately $\frac{|v_i^1|t}{2L}$ times during a time interval of length t, for t large, and hence that the momentum transfer to the wall per unit time by particle i is $-m_i(v_i^1)^2/L$. Use the result of the previous exercise to show that the expected transfer of momentum to the wall per unit time and unit area by all particles equals $-L^{-3}Nk_{\mathrm{B}}T$. Hence we have found that $P = L^{-3}Nk_{\mathrm{B}}T$ which is the ideal gas equation (1.33).

Exercise 1.22 (Entropy of an ideal gas)**.** Consider an ideal gas of N particles. We have seen that the pressure $P(U, V)$ and temperature $T(U, V)$ are given as functions of the energy U and volume V from the relations

$$VP(U, V) = Nk_{\mathrm{B}}T(U, V), \quad U = \frac{3}{2}Nk_{\mathrm{B}}T(U, V).$$

Show that $T(U, V)^{-1}$ is an integrating factor for the heat differential

$$dU + P(U, V)dV$$

and that the entropy must be given by

$$S(U, V) = Nk_{\mathrm{B}} \ln \Big(C(V/N)(U/N)^{3/2} \Big)$$

for some constant C independent of U and V. Show that C is independent of N if we assume that S, V, and U are proportional to N.

Exercise 1.23. Consider the differential

$$\alpha = -\frac{y}{x^2 + y^2}dx + \frac{x}{x^2 + y^2}dy$$

defined on the set $\mathbb{R}^2 \setminus \{0\}$.

a) Show that α is a closed differential.

b) Show that if $\gamma : [0, 2\pi] \to \mathbb{R}^2$ is the motion $\gamma(t) = (\cos(t), \sin(t))$ then $\int_\gamma \alpha \neq 0$ and conclude that α is not exact.

Exercise 1.24. The purpose of this exercise is to prove Theorem 1.21 under the additional assumption that the set $\mathcal{O} \subseteq \mathbb{R}^k$ is *star-shaped.* By this, we mean that there exists a point $x_0 \in \mathcal{O}$ such that for all $x \in \mathcal{O}$ the line segment connecting x_0 and x lies entirely in $\mathcal{O}$.

Under this assumption, we can assume without loss of generality that $x_0 = 0$ holds. Then define the function $V : \mathcal{O} \to \mathbb{R}$ by

$$V(x) = \sum_{j=1}^{k} \int_0^1 \omega_j(tx)x_j \, dt \, .$$

Briefly explain why V is well-defined. Next, verify that V has the desired property, i. e.

$$dV = \omega \quad \text{on } \mathcal{O}\,,$$

which means that

$$\frac{\partial V}{\partial x_i}(x) = \omega_i(x) \quad \text{for } x \in \mathcal{O} \text{ and } i = 1, \ldots, k\,.$$

To prove this, show (by using that ω is closed) the identity

$$\frac{\partial V}{\partial x_i} = \int_0^1 t \frac{d}{dt} \omega_i(tx)\, dt + \int_0^1 \omega_i(tx)\, dt\,.$$

Finally, integrate by parts to conclude the proof, also using that $\frac{d}{dt} g(tx) = x \cdot \nabla g(tx)$ for any differentiable function g of k variables.

Chapter 2

Lagrangian Mechanics

The main results of this chapter will be the formulation of *Lagrangian mechanics* and *Hamilton's principle of least action* which provide an interpretation of Newton's equation of motion as a variational equation under quite general circumstances. The value of this reformulation of classical mechanics can hardly be overestimated, both as a tool for solving concrete physics problems and as a mathematical framework for classical mechanics.

2.1 Calculus of variations

The *calculus of variations* is concerned with *functionals* and their extremal points or, more generally, their stationary points. In fact, many fundamental equations in physics and geometry arise from a so-called *variational principle,* and the calculus of variations defines a rigorous framework for deriving such equations and studying their properties.

By a functional Φ we simply mean a mapping $\Phi : X \to \mathbb{R}$, where X is some given function space. Thus, a functional is just a real-valued function depending on functions. For example, the assignment

$$\Phi(\gamma) = \int_0^1 \|\dot{\gamma}(t)\|\, dt$$

defines a functional on the space of continuously differentiable functions γ from $[0,1]$ to $\mathbb{R}^k$ or, in the notation of Definition 1.5 of Chapter 1, the space of C^1 curves defined on $[0,1]$. In geometric terms, the functional $\Phi(\gamma)$ given above is the *length* of the C^1 curve γ. We shall therefore in the following denote this functional by $\mathcal{L}$ instead of Φ (see Definition 2.2 below). Applying the calculus of variations to $\mathcal{L}$ the classical statement that the shortest curve connecting to given points in $\mathbb{R}^k$ is a straight line will be demonstrated.

Having the application to classical mechanics in mind, we shall restrict ourselves henceforth to functionals defined on spaces of curves in $\mathbb{R}^k$. For an open subset $\mathcal{U}$ of $\mathbb{R}^k$ the space of C^1-curves $\gamma : [a,b] \to \mathcal{U}$ on the interval $[a,b]$ will be denoted by $C^1([a,b],\mathcal{U})$. For notational simplicity we shall mostly consider the case $\mathcal{U} = \mathbb{R}^k$ although the central results, such as Theorems 2.10 and 2.17, are valid for general open sets $\mathcal{U}$ with obvious modifications. A similar statement holds true if $C^1([a,b],\mathcal{U})$ is replaced by the larger space of piecewise C^1-curves in $\mathcal{U}$.

The space $C^1([a,b],\mathbb{R}^k)$ is a clearly a vector space with standard pointwise addition of functions and multiplication by constants. If $\xi, \eta \in \mathbb{R}^k$ are two given points, we shall also

be considering the subset

$$C^1_{\xi,\eta}([a,b],\mathbb{R}^k) = \left\{\gamma \in C^1([a,b],\mathbb{R}^k) \;\middle|\; \gamma(a)=\xi,\ \gamma(b)=\eta\right\}.$$

In particular, the space $C^1_{0,0}([a,b],\mathbb{R}^k)$ consists of curves that start and end at 0. This is a linear subspace of $C^1([a,b];\mathbb{R}^k)$.

We define the *distance* between two curves γ,γ' in $C^1([a,b],\mathbb{R}^k)$ to be $\|\gamma-\gamma'\|_1$, where the *norm* $\|\gamma\|_1$ is given by

$$\|\gamma\|_1 = \max\left\{\|\gamma(t)\|,\ \|\dot\gamma(t)\| : t\in[a,b]\right\}.$$

The following basic properties of the norm are easily verified:

i) $\|\lambda\gamma\|_1 = |\lambda|\|\gamma\|_1$ for all $\lambda\in\mathbb{R}$.

ii) $\|\gamma+\gamma'\|_1 \leqslant \|\gamma\|_1 + \|\gamma'\|_1$ (triangle inequality).

iii) $\|\gamma\|_1 = 0$ if and only if $\gamma(t)=0$ for all t.

Having introduced a distance on the space $C^1([a,b];\mathbb{R}^k)$, we can also define what *continuity* of a functional means.

Definition 2.1. *A* **functional** *is a real-valued function $\Phi : D_\Phi \to \mathbb{R}$ defined on a subset $D_\Phi \subseteq C^1([a,b],\mathbb{R}^k)$. We say that Φ is* **continuous** *at $\gamma\in D_\Phi$ if the following holds: for every $\varepsilon>0$ there exists a $\delta>0$ such that*

$$|\Phi(\gamma)-\Phi(\gamma')|<\varepsilon \quad \textit{for all } \gamma'\in D_\Phi \textit{ fulfilling } \|\gamma-\gamma'\|_1<\delta.$$

Example 2.2. The following two classical examples of functionals on $C^1([a,b];\mathbb{R}^k)$ will be discussed in more detail later in this chapter:

a) The *length* functional

$$\mathcal{L}(\gamma) = \int_a^b \|\dot\gamma(t)\|dt = \int_a^b \left(\dot\gamma_1(t)^2+\cdots+\dot\gamma_k(t)^2\right)^{1/2} dt. \tag{2.1}$$

b) The *free action* (in geometry often referred to as the *energy*)

$$\mathcal{S}_0(\gamma) = \frac{1}{2}\int_a^b \|\dot\gamma(t)\|^2 dt = \frac{1}{2}\int_a^b (\dot\gamma_1(t)^2+\cdots+\dot\gamma_k(t)^2)dt. \tag{2.2}$$

Both of these functionals are continuous. In fact, we find that

$$|\mathcal{L}(\gamma)-\mathcal{L}(\gamma')| = \left|\int_a^b \left(\|\dot\gamma(t)\|-\|\dot\gamma'(t)\|\right)dt\right| \leqslant \int_a^b \|\dot\gamma(t)-\dot\gamma'(t)\|dt \leqslant (b-a)\|\gamma-\gamma'\|_1,$$

where the inequality

$$\left|\|\dot\gamma(t)\|-\|\dot\gamma'(t)\|\right| \leqslant \|\dot\gamma(t)-\dot\gamma'(t)\|,$$

which follows from the triangle inequality above, has been used. It is now clear that, given $\varepsilon>0$, we can choose $\delta=\varepsilon/(b-a)$ in order to satisfy the continuity requirement for $\mathcal{L}$. The continuity of $\mathcal{S}_0$ is left as an exercise for the reader (see Exercise 2.1).

Example 2.3. The functional Φ on $C^1([a,b],\mathbb{R}^k)$ defined by

$$\Phi(\gamma) = \begin{cases} 1 & \text{if } \|\gamma\|_1 > 1 \\ 0 & \text{if } \|\gamma\|_1 \leqslant 1 \end{cases}$$

is not continuous at all curves γ. We leave the (simple) proof of this fact to the reader.

For C^1 functions $f : \mathcal{O} \to \mathbb{R}$ of finitely many variables we defined in Section 1.6 the differential $df_x : \mathbb{R}^k \to \mathbb{R}$ which for each $x \in \mathcal{O}$ is a linear map. We now generalize this notion to functionals.

Definition 2.4. *A functional Φ, defined on $D_\Phi = C^1_{\xi,\eta}([a,b],\mathbb{R}^k)$ or $D_\Phi = C^1([a,b],\mathbb{R}^k)$, is called* **differentiable** *at $\gamma \in D_\Phi$, if there exists a continuous and linear functional $d\Phi_\gamma : C^1_{0,0}([a,b],\mathbb{R}^k) \to \mathbb{R}$ such that*

$$\Phi(\gamma + h) = \Phi(\gamma) + d\Phi_\gamma(h) + r(h),$$

for all $h \in C^1_{0,0}([a,b],\mathbb{R}^k)$, where the remainder term $r(h)$ satisfies

$$\lim_{\|h\|_1 \to 0} \frac{r(h)}{\|h\|_1} = 0\,.$$

More explicitly, this last condition means that for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|r(h)| \leq \varepsilon \|h\|_1$$

if $\|h\|_1 \leq \delta$. We call $d\Phi_\gamma$ the **differential** *of Φ at γ.*

Although it may seen quite technical at first reading this definition is entirely analogous to the definition of differentiability for functions of several variables. It simply says that there is a linear approximation to the functional Φ close to γ with an error $r(h)$ which goes to zero faster than linearly as $\|h\|_1 \to 0$. In many examples, the error is explicitly of quadratic order, that is $|r(h)| \leqslant c\|h\|_1^2$ for some constant c. If Φ is differentiable at all curves in its domain of definition we call Φ differentiable.

We list two important facts relating to differentiability in the following proposition whose proof is deferred to Exercises 2.2 and 2.3.

Proposition 2.5. *Let Φ be a functional defined on $C^1([a,b];\mathbb{R}^k)$ or $C^1_{\xi,\eta}([a,b];\mathbb{R}^k)$.*

a) The differential $d\Phi_\gamma$ is unique, if it exists.

b) If Φ is differentiable at γ, then Φ is continuous at γ.

The next definition and the subsequent result have close analogues for functions of finitely many variables.

Definition 2.6. *If Φ is a differentiable functional defined on $D_\Phi = C^1([a,b],\mathbb{R}^k)$ or $D_\Phi = C^1_{\xi,\eta}([a,b],\mathbb{R}^k)$, we call $\gamma \in D_\Phi$ a* **stationary point** *of Φ if the differential $d\Phi_\gamma$ vanishes on $C^1_{0,0}([a,b],\mathbb{R}^k)$, that is*

$$d\Phi_\gamma(h) = 0$$

for all $h \in C^1_{0,0}([a,b],\mathbb{R}^k)$.

Theorem 2.7. *Let Φ be a differentiable functional defined on $D_\Phi = C^1([a,b],\mathbb{R}^k)$ or $D_\Phi = C^1_{\xi,\eta}([a,b],\mathbb{R}^k)$. If $\gamma \in D_\Phi$ is an* **extremal point** *for Φ, i. e., a point where Φ attains either a maximal or a minimal value, then γ is a stationary point for Φ.*

Proof. Without loss of generality assume that Φ attains a local maximum at γ (otherwise, replace Φ by $-\Phi$), and consider the function

$$f(s) = \Phi(\gamma + sh)\,, \quad s \in \mathbb{R}\,,$$

for an arbitrary fixed $h \in C^1_{0,0}([a,b],\mathbb{R}^k)$.

Using that Φ is differentiable and $d\Phi_\gamma(sh) = sd\Phi_\gamma(h)$, we have

$$|f(s) - f(0) - sd\Phi_\gamma(h)| \leqslant |r(sh)|\,,$$

where $\frac{r(sh)}{\|sh\|_1} \to 0$ as $s \to 0$. It follows that the function $f(s)$ is differentiable at $s = 0$ and that $f'(0) = d\Phi_\gamma(h)$. But since Φ has a local maximum at γ, it is clear that $f(s)$ attains a local maximum at $s = 0$, and we conclude that $f'(0) = 0$ by usual single-variable calculus. Therefore, we have $d\Phi_\gamma(h) = f'(0) = 0$. Since this holds for arbitrary $h \in C^1_{0,0}([a,b];\mathbb{R}^k)$ we have shown that $d\Phi_\gamma = 0$, and hence γ is a stationary point for Φ. □

2.2 Euler-Lagrange equations

The free action $\mathcal{S}_0$ defined in (2.2) is an example of a differentiable functional. This fact is quite easy to verify directly using Definition 2.4. However, it is also a consequence of the following general result concerning differentiability of a class of functionals that includes both the length functional and the free action.

Definition 2.8. *Let $L : \mathbb{R}^k \times \mathbb{R}^k \times [a,b] \to \mathbb{R}$ be a continuous function. We define the corresponding functional $\Phi^L : C^1([a,b],\mathbb{R}^k) \to \mathbb{R}$ by*

$$\Phi^L(\gamma) = \int_a^b L(\gamma(t), \dot\gamma(t), t)dt. \tag{2.3}$$

The integral above makes sense since the integrand $t \mapsto L(\gamma(t), \dot\gamma(t), t)$ is continuous. Functionals of the form (2.3) are called **Lagrange functionals** *and the function L is called a* **Lagrange function**, *or* **Lagrangian**, *for the functional Φ^L.*

Theorem 2.9. *If $L(q,u,t)$ is a C^2 function of its variables $(q,u,t) \in \mathbb{R}^k \times \mathbb{R}^k \times [a,b]$, then the functional Φ^L given by (2.3) is differentiable at all $\gamma \in C^1([a,b],\mathbb{R}^k)$, and for $\gamma \in C^2([a,b],\mathbb{R}^k)$ its differential is given by*

$$d\Phi^L_\gamma(h) = \int_a^b \sum_{i=1}^k h_i(t)\left[\frac{\partial L}{\partial q_i}(\gamma(t),\dot\gamma(t),t) - \frac{d}{dt}\left(\frac{\partial L}{\partial u_i}(\gamma(t),\dot\gamma(t),t)\right)\right]dt \tag{2.4}$$

for $h \in C^1_{0,0}([a,b],\mathbb{R}^k)$.

Sketch of proof. We give here the main idea of the proof and its main steps. The complete proof with technical details can be found in the appendix at the end of this chapter.

For $\gamma \in C^2([a,b];\mathbb{R}^k)$ and $h \in C^1_{0,0}([a,b];\mathbb{R}^k)$, one finds by using Taylor's theorem for $L(q,u,t)$ (see the appendix for details) that

$$\begin{aligned}\Phi^L(\gamma+h) - \Phi^L(\gamma) &= \int_a^b \Big[L(\gamma+h,\dot\gamma+\dot h,t) - L(\gamma,\dot\gamma,t)\Big]\, dt \\ &= \int_a^b \sum_{i=1}^k \left[\frac{\partial L}{\partial q_i}h_i + \frac{\partial L}{\partial u_i}\dot h_i\right] dt + O(\|h\|_1^2) = F(h) + R\,,\end{aligned}$$

where

$$F(h) = \int_a^b \sum_{i=1}^k \left[\frac{\partial L}{\partial q_i}h_i + \frac{\partial L}{\partial u_i}\dot h_i\right] dt \quad \text{and} \quad R = O(\|h\|_1^2).$$

Clearly, $F(h)$ is a linear functional, and one also verifies (see Exercise 2.5) that $F(h)$ is continuous. Hence $\Phi(\gamma)$ is differentiable and we have $d\Phi_\gamma(h) = F(h)$.

Finally, to bring $F(h)$ into the desired form, we integrate by parts to obtain

$$\int_a^b \sum_{i=1}^k \frac{\partial L}{\partial u_i}\dot h_i\, dt = -\int_a^b \sum_{i=1}^k h_i \frac{d}{dt}\left(\frac{\partial L}{\partial u_i}\right) dt + \left(\sum_{i=1}^k h_i \frac{\partial L}{\partial u_i}\right)\Big|_{t=a}^{t=b},$$

where the last term vanishes due to $h(a) = h(b) = 0$. This completes our sketch of the proof of Theorem 2.9. □

The next result is central and introduces the so-called Euler-Lagrange equations as an equivalent characterization of stationary points of functionals $\Phi^L(\gamma)$ of the form (2.3).

Theorem 2.10. *Let $L : \mathbb{R}^k \times \mathbb{R}^k \times [a,b]$ is a C^2 function and define the functional $\Phi^L(\gamma)$ by (2.3). Then $\gamma \in C^2([a,b],\mathbb{R}^k)$ is a stationary point for Φ^L if and only if γ satisfies the system of differential equations*

$$\boxed{\frac{d}{dt}\left(\frac{\partial L}{\partial u_i}(\gamma(t),\dot\gamma(t),t)\right) - \frac{\partial L}{\partial q_i}(\gamma(t),\dot\gamma(t),t) = 0} \tag{2.5}$$

where $i = 1,\dots,k$. These equations are called the **Euler-Lagrange equations** *for the functional Φ^L.*

Proof. It is clear from the expression (2.4) for the functional $d\Phi^L$ that if γ satisfies the Euler-Lagrange equations then $d\Phi^L_\gamma(h) = 0$ for all $h \in C^1_{0,0}([a,b],\mathbb{R}^k)$.

Conversely, if the curve γ is a stationary point for Φ^L we have that $d\Phi^L_\gamma(h_1,0,\cdots,0) = 0$ for all functions $h_1 \in C^1_{0,0}([a,b],\mathbb{R})$, that is

$$\int_a^b h_1(t)\left[\frac{\partial L}{\partial q_1}(\gamma(t),\dot\gamma(t),t) - \frac{d}{dt}\left(\frac{\partial L}{\partial u_1}(\gamma(t),\dot\gamma(t),t)\right)\right] dt = 0.$$

From Lemma 2.11 below we conclude that

$$\frac{\partial L}{\partial q_1}(\gamma(t),\dot\gamma(t),t) - \frac{d}{dt}\left(\frac{\partial L}{\partial u_1}(\gamma(t),\dot\gamma(t),t)\right) = 0$$

for all $t \in [a,b]$. This shows that eq. (2.5) holds when $i = 1$. Similarly it follows for $i = 2,3,\dots,k$. □

Lemma 2.11 (Fundamental Lemma of the Calculus of Variations)**.** *Let* $f : [a, b] \to \mathbb{R}$ *be continuous and suppose that*

$$\int_a^b f(t)h(t)dt = 0 \quad \textit{for all } h \in C^1_{0,0}([a, b], \mathbb{R}).$$

Then $f(t) = 0$ *for all* $t \in [a, b]$.

Proof. It suffices to show that $f(t) = 0$ for all t in the open interval (a, b). By continuity of $f(t)$ this implies $f(a) = f(b) = 0$ as well.

To show that $f(t) = 0$ for all $t \in (a, b)$, we argue by contradiction as follows. Suppose there exists $t_0 \in (a, b)$ such that $f(t_0) \neq 0$. Without loss of generality, we can assume that $c := f(t_0) > 0$ holds. (Otherwise, we replace f by $-f$). By continuity, there is some $\delta > 0$ such that $f(t) \geqslant c/2$ for $t \in (t_0 - \delta, t_0 + \delta)$. Now, we can choose (see Exercise 2.6) a function $h \in C^1_{0,0}([a, b], \mathbb{R})$ with the following properties:

$$h(t) > 0 \text{ for } t \in (t_0 - \delta, t_0 + \delta) \quad \text{and} \quad h(t) = 0 \text{ if } t \notin (t_0 - \delta, t_0 + \delta).$$

Therefore, we can deduce

$$\int_a^b f(t)h(t)\,dt = \int_{t_0-\delta}^{t_0+\delta} f(t)h(t)\,dt \geqslant c/2 \int_{t_0-\delta}^{t_0+\delta} h(t)\,dt > 0,$$

which gives the desired contradiction to $\int_a^b f(t)h(t)\,dt = 0$. □

We end this subsection by studying in more detail the extremal properties of the length functional $\mathcal{L}$ and the free action $\mathcal{S}_0$. Given points $\xi, \eta \in \mathbb{R}^k$ we want to determine the curves joining ξ and η that minimize the length or the free action. Of course, we expect that the curves that minimize the length are straight line segments. The same turns out to be the case for the free action, but there is a subtle difference, as we will see, due to the fact that $\mathcal{L}$ is reparametrization invariant whereas $\mathcal{S}_0$ is not.

The free action is the functional corresponding to the Lagrangian $L(q, u, t) = \frac{1}{2}\|u\|^2$. This is clearly a C^2-function and we can use Theorems 2.7 and 2.10 to conclude that C^2-curves that minimize the free action must satisfy the Euler-Lagrange equations. Since L in this case is independent of q we get immediately

$$\frac{d}{dt}\frac{\partial L}{\partial u_i}(\dot\gamma(t)) = \ddot\gamma_i(t) = 0\,.$$

Integrating twice we get $\gamma(t) = u_0 t + q_0$, where $q_0, u_0 \in \mathbb{R}^k$ are integration constants that are to be determined such that $\gamma(a) = \xi$ and $\gamma(b) = \eta$. This gives the unique solution

$$\gamma_0(t) = \xi + \frac{\eta - \xi}{b - a}(t - a)\,, \tag{2.6}$$

representing the line segment from ξ to η.

The length functional $\mathcal{L}$ has the corresponding Lagrangian $L(q, u, t) = \|u\|$. This is not a C^2-function at $u = 0$. If we restrict attention away from this point, and only consider *regular curves* with non-vanishing velocity, we can still use the method above although we did not exactly treat that case. Hence, a regular curve $\gamma \in C^2_{\xi,\eta}([a, b], \mathbb{R}^k)$ that minimizes

the length is a stationary point for $\mathcal{L}$ and satisfies the Euler-Lagrange equations. Using that $\frac{\partial\|u\|}{\partial u_i} = \frac{u_i}{\|u\|}$, for $u \neq 0$, and that $\|u\|$ is independent of q, these equations take the form

$$\frac{d}{dt}\frac{\partial L}{\partial u_i}(\dot\gamma(t)) = \frac{d}{dt}(\dot\gamma_i(t)\|\dot\gamma(t)\|^{-1}) = 0, i = 1, \dots, k$$

This means that the direction $\dot\gamma(t)\|\dot\gamma(t)\|^{-1}$ of the velocity vector $\dot\gamma(t)$ is a constant unit vector $e \in \mathbb{R}^k$, that is

$$\dot\gamma(t) = f(t)e\,,$$

where f is a positive C^1-function on $[a, b]$. By integration this gives

$$\gamma(t) = \xi + F(t)e\,,$$

where $F(t) = \int_a^t f(s)ds$ is a strictly increasing C^2 function vanishing at a, and we have used $\gamma(a) = \xi$. Using also $\gamma(b) = \eta$ we get $e = F(b)^{-1}(\eta - \xi)$. This shows that the solutions found are exactly the increasing C^2-reparametrisations of the line segment γ_0 given by (2.6), which corresponds to the particular choice $F(t) = \frac{t-a}{b-a}$.

We emphasize that we have not yet proven that the curves found are really minimizers for the respective functionals, but only that they are the only possible stationary points. The additional arguments required to prove this are provided in Exercises 2.9 and 2.10.

2.3 Change of variables

In this section we discuss briefly how functionals and their differentials change under coordinate transformations.

Definition 2.12. *A C^1 function $\psi = (\psi_1, \dots, \psi_k) : \mathcal{O} \to \mathbb{R}^k$ defined on an open set $\mathcal{O} \subseteq \mathbb{R}^k$ is called a* **coordinate transformation** *if it is an injective map and*

$$\det D\psi(\widetilde q) \neq 0 \tag{2.7}$$

for all $\widetilde q \in \mathcal{O}$, where $D\psi(\widetilde q)$ is the Jacobian matrix of ψ,

$$D\psi(\widetilde q) = \begin{pmatrix} \frac{\partial\psi_1}{\partial\widetilde q_1}(\widetilde q) & \cdots & \frac{\partial\psi_1}{\partial\widetilde q_k}(\widetilde q) \\ \vdots & \ddots & \vdots \\ \frac{\partial\psi_k}{\partial\widetilde q_1}(\widetilde q) & \cdots & \frac{\partial\psi_k}{\partial\widetilde q_k}(\widetilde q) \end{pmatrix}.$$

The requirement (2.7) ensures by the inverse mapping theorem that the image $\mathcal{U} = \psi(\mathcal{O})$ is an open subset of $\mathbb{R}^k$ and that the inverse mapping $\psi^{-1} : \mathcal{U} \to \mathcal{O}$ is a C^1-function, i.e. ψ^{-1} is also a coordinate transformation. If $q = \psi(\widetilde q)$, we think of $\widetilde q \in \mathcal{O}$ as new coordinates being related to old coordinates $q \in \mathcal{U}$ by the function ψ.

Example 2.13. a) For the Galilean coordinate transformations introduced in Chapter 1 we have $\mathcal{O} = \mathcal{U} = \mathbb{R}^4$ and one easily sees that

$$D\psi(x, t) = A(t)\,,$$

and hence $\det D\psi(x, t) = \pm 1$.

b) The well known change from Cartesian coordinates (x_1, x_2) to polar coordinates (r, θ) in the plane is given by

$$\psi(r, \theta) = (r\cos\theta, r\sin\theta)$$

and $\mathcal{O} = \mathbb{R}_+ \times (-\pi, \pi)$, while $\mathcal{U} = \mathbb{R}^2 \setminus \{(x_1, 0) \mid x_1 \leqslant 0\}$. In this case,

$$D\psi(r, \theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \tag{2.8}$$

with $\det D\psi(r, \theta) = r$.

Given a functional $\Phi : C^1_{\xi,\eta}([a, b], \mathcal{U}) \to \mathbb{R}$ and a coordinate transformation $\psi : \mathcal{O} \to \mathcal{U}$ as above the transformed functional $\widetilde{\Phi}$ is defined on $C^1_{\widetilde{\xi},\widetilde{\eta}}([a, b], \mathcal{O})$, where $\widetilde{\xi} = \psi^{-1}(\xi)$, $\widetilde{\eta} = \psi^{-1}(\eta)$, by

$$\widetilde{\Phi}(\widetilde{\gamma}) = \Phi(\psi \circ \widetilde{\gamma})\,. \tag{2.9}$$

In particular, for a Lagrange functional Φ_L we get

$$\widetilde{\Phi}^L(\widetilde{\gamma}) = \int_a^b L(\psi(\widetilde{\gamma}(t)), D\psi(\widetilde{\gamma}(t))\dot{\widetilde{\gamma}}(t), t)dt = \int_a^b \widetilde{L}(\widetilde{\gamma}(t), \dot{\widetilde{\gamma}}(t), t)dt = \Phi^{\widetilde{L}}(\widetilde{\gamma})\,, \tag{2.10}$$

where the chain rule

$$\frac{d}{dt}\psi(\widetilde{\gamma}(t)) = D\psi(\widetilde{\gamma}(t))\dot{\widetilde{\gamma}}(t)$$

has been used and we have defined the transformed Lagrangian $\widetilde{L}$ by

$$\widetilde{L}(\widetilde{q}, \widetilde{u}, t) = L(\psi(\widetilde{q}), D\psi(\widetilde{q})\widetilde{u}, t)\,. \tag{2.11}$$

Our goal is to demonstrate that a curve $\widetilde{\gamma}$ in $\mathcal{O}$ is a stationary point of $\tilde{\Phi}$ if and only if the transformed curve $\gamma = \psi \circ \widetilde{\gamma}$ is a stationary point of Φ. For this we need the following result.

Proposition 2.14. *Let the functionals $\widetilde{\Phi}$ and Φ be related by a C^3 coordinate transformation ψ as in (2.9). If Φ is differentiable at $\gamma \in C^1([a, b], \mathcal{U})$ then $\widetilde{\Phi}$ is differentiable at $\widetilde{\gamma} = \psi^{-1} \circ \gamma$ and*

$$d\widetilde{\Phi}_{\widetilde{\gamma}}(h) = d\Phi_\gamma((D\psi \circ \widetilde{\gamma})h) \tag{2.12}$$

for $h \in C^1_{0,0}([a, b], \mathbb{R}^k)$.

Proof. Since $\mathcal{O}$ is open and $\widetilde{\gamma}$ lies in $\mathcal{O}$ we can choose $\epsilon > 0$ small enough such that $\widetilde{\gamma} + h$ lies in $\mathcal{O}$ if $\|h\|_1 \leqslant \epsilon$. For such an h we set $\mu(t) = \psi(\gamma(t) + h(t))$. By Taylor expanding the function $f(s) = \psi(\widetilde{\gamma}(t) + sh(t))$ to second order around $s = 0$ for fixed t and setting $s = 1$ we get that

$$\begin{aligned} \psi_i(\widetilde{\gamma}(t) + h(t)) &= \psi_i(\widetilde{\gamma}(t)) + \sum_{j=1}^k \frac{\partial\psi_i}{\partial\widetilde{q}_j}(\widetilde{\gamma}(t))h_j(t) \\ &+ \frac{1}{2}\sum_{l,m=1}^k \frac{\partial^2\psi_i}{\partial\widetilde{q}_j\partial\widetilde{q}_l}(\widetilde{\gamma}(t) + s_ih(t))h_l(t)h_m(t)\,, \end{aligned}$$

for some $s_i \in [0,1]$ depending on t. Since ψ is C^2 the second derivatives on the right-hand side of this equation are bounded by a constant independent of t for $\|h\|_1 \leqslant \epsilon$ (see Appendix 2.6 at the end of this chapter for a similar argument). Using the above equation for $i = 1, \ldots, k$ it follows that

$$\|\mu(t) - \gamma(t) - D\psi(\widetilde{\gamma}(t))h(t)\| \leqslant M\|h(t))\|^2\,,$$

for all $t \in [a,b]$, where M is a constant and we have used that $\gamma(t) = \psi(\widetilde{\gamma}(t))$. Similarly, considering the function $g(s) = \frac{d}{dt}\psi(\widetilde{(t)} + sh(t))$ for fixed t and using that ψ is C^3, we get

$$\|\dot{\mu}(t) - \dot{\gamma}(t) - \frac{d}{dt}D\psi(\widetilde{\gamma}(t))h(t)\| \leqslant M(\|h(t)\| + \|\dot{h}(t)\|)^2$$

for all $t \in [a,b]$, if $\|h\|_1 \leqslant \epsilon$ and provided M is chosen large enough and ϵ small enough. The last two inequalities obviously imply

$$\|\mu - \gamma - (D\psi \circ \widetilde{\gamma})h\|_1 \leqslant 4M\|h\|_1^2 \quad \text{for } \|h\|_1 \leqslant \epsilon\,. \tag{2.13}$$

Recalling the definition of differentiability we have

$$\begin{aligned}
\widetilde{\Phi}(\widetilde{\gamma} + h) - \widetilde{(}\gamma) &= \Phi(\mu) - \Phi(\gamma) \\
&= d\Phi_\gamma(\mu - \gamma) + r(\mu - \gamma) \\
&= d\Phi_\gamma((D\psi \circ \widetilde{\gamma})h + r'(h)) + r(D\psi \circ \widetilde{\gamma})h + r'(h)) \\
&= d\Phi_\gamma((D\psi \circ \widetilde{\gamma})h) + d\Phi_\gamma(r'(h)) + r(D\psi \circ \widetilde{\gamma})h + r'(h))\,, \qquad (2.14)
\end{aligned}$$

where $r(h)$ fulfills the requirements of Definition 2.4 and $|r'(h)|$ is of second order in $\|h\|_1$ by (2.13). Using, finally, that $d\Phi_\gamma$ is continuous it follows easily that the last two terms in (2.14) tend to zero faster than linearly as $|h\|_1 \to 0$. Since $d\Phi_\gamma((D\psi \circ \widetilde{\gamma})h)$ is linear and continuous in h this completes the proof that $\widetilde{\Phi}$ is differentiable at $\widetilde{\gamma}$ if Φ is differentiable at γ, and that (2.12) holds.

The converse statement follows from this by using that ψ^{-1} is a coordinate transformation. □

As a direct consequence of Proposition 2.14 we get the desired result on stationary points.

Corollary 2.15. *If $\Phi, \tilde{\Phi}$ and ψ are given as in Proposition 2.14 then $\widetilde{\gamma} \in C^1_{\widetilde{\xi},\widetilde{\eta}}([a,b],\mathcal{O})$ is a stationary point for $\widetilde{\Phi}$ if and only if $\gamma = \psi \circ \widetilde{\gamma} \in C^1_{\xi,\eta}([a,b],\mathcal{U})$ is a stationary point for Φ, where $\xi = \psi(\widetilde{\xi})$, $\eta = \psi(\widetilde{\eta})$.*

Proof. Since ψ^{-1} is a coordinate transformation it suffices to show that if $\gamma \in C^1_{\xi,\eta}([a,b],\mathcal{U})$ is a stationary point of Φ then $\widetilde{\gamma}$ is a stationary point of $\widetilde{\Phi}$, that is $d\widetilde{\Phi}_{\widetilde{\gamma}} = 0$ if $d\Phi_\gamma = 0$. But this is an immediate consequence of (2.12). □

In the particular case, where $\Phi = \Phi^L$ is given by a Lagrangian $L(q,u,t)$ this implies that the Euler-Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial L}{\partial u_i}(\gamma(t),\dot{\gamma}(t),t)\right) - \frac{\partial L}{\partial q_i}(\gamma(t)),\dot{\gamma}(t)),t) = 0, \quad i = 1,\ldots,k, \tag{2.15}$$

hold for γ if and only if the Euler-Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial \widetilde{L}}{\partial \widetilde{u}_i}(\widetilde{\gamma}(t),\dot{\widetilde{\gamma}}(t),t)\right) - \frac{\partial \widetilde{L}}{\partial \widetilde{q}_i}(\widetilde{\gamma}(t),\dot{\widetilde{\gamma}}(t),t) = 0\,, \quad i = 1,\dots,k, \tag{2.16}$$

hold for $\widetilde{\gamma} = \psi^{-1} \circ \gamma$, where $\widetilde{L}(\widetilde{q},\widetilde{u},t)$ is given by (2.11).

Remark 2.16. It is sometimes relevant to also consider time dependent coordinate changes $q = \psi(\widetilde{q},t)$, where for each fixed $t \in \mathbb{R}$ the map $\psi_t : \widetilde{q}) \to \psi(\widetilde{q},t)$ is defined on a fixed open set $\mathcal{O} \subseteq \mathbb{R}^k$ and satisfies the properties above, and ψ is assumed to be a C^1 mapping of all variables including t. We then denote by $D\psi$ the Jacobian matrix w. r. t. $\widetilde{q}$ and by $\dot{\psi}$ the partial time derivative. Given a functional Φ defined on $C^1_{\xi,\eta}([a,b],\mathcal{U})$ as above, such that $\psi_t(\mathcal{O}) \subseteq \mathcal{U}$ for all t, we define the transformed functional $\widetilde{\Phi}$ on $C^1_{\widetilde{\xi},\widetilde{\eta}}([a,b],\mathcal{O})$ by

$$\widetilde{\Phi}(\widetilde{\gamma}) = \Phi(\psi \circ \widetilde{\gamma})\,,$$

where $\psi \circ \widetilde{\gamma}$ (by a slight abuse of notation) here means the motion $t \mapsto \psi(\gamma(t),t) = \psi_t(\gamma(t))$, $t \in [a,b]$, and $\widetilde{\xi} = \psi_a^{-1}(\xi)$, $\widetilde{\eta} = \psi_b^{-1}(\eta)$.

As before we find that if $\Phi = \Phi^L$ is a Lagrangian functional then $\widetilde{\Phi}(\widetilde{\gamma}) = \Phi^{\widetilde{L}}(\widetilde{\gamma})$ where the transformed Lagrangian $\widetilde{L}$ is now given by

$$\widetilde{L}(\widetilde{q},\widetilde{u},t) = L(\psi(\widetilde{q},t), D\psi(\widetilde{q},t)\widetilde{u} + \dot{\psi}(\widetilde{q},t),t). \tag{2.17}$$

Proposition 2.14 also holds for such time dependent coordinate transformations with essentially the same proof. As a consequence, equation (2.15) also implies equation (2.16) in this case with $\widetilde{L}$ given by (2.17).

2.4 Hamilton's principle

We now consider a system of N particles with masses $m_1, m_2, \dots, m_N$ in an inertial system of coordinates. As in Chapter 1 we use the notation $x = (x_1,\dots,x_N)$ for the particle coordinates. Given a conservative force field F on $\mathcal{U} \subseteq \mathbb{R}^{3N}$ with corresponding potential $V : \mathcal{U} \to \mathbb{R}$, such that $F_i = -\nabla V$, $i = 1,\dots,N$, the Lagrangian $L : \mathcal{U} \times \mathbb{R}^{3N} \to \mathbb{R}$ of the particle system is defined by

$$L(x,v) = \sum_{i=1}^{N} \frac{1}{2} m_i \|v_i\|^2 - V(x)\,. \tag{2.18}$$

The corresponding functional Φ^L is called the **action functional** of the system; it will be denoted by $\mathcal{S}$ and

$$\mathcal{S}(\gamma) = \int_{t_1}^{t_2} \left(\sum_{1}^{N} \frac{1}{2} m_i \|\dot{\gamma}_i(t)\|^2 - V(\gamma(t)\right) dt \tag{2.19}$$

is called the **action** of the motion $\gamma = (\gamma_1,\dots,\gamma_N)$. Notice that the free action $\mathcal{S}$ introduced in (2.2) is indeed the action of a particle of unit mass moving freely without any force acting.

The following main result of the present chapter states that Newton's equations for the particle system in question are obtained as the Euler-Lagrange equations for the Lagrangian (2.18).

Theorem 2.17 (Hamilton's principle). *For a system of N particles with Lagrangian (2.18) where V is a C^2-function on $\mathcal{U} \subseteq \mathbb{R}^{3N}$, it holds that a C^2 curve γ is a stationary point for the action functional $\mathcal{S}$ on the set $C^1_{\xi,\eta}([t_1,t_2],\mathcal{U})$, if and only if γ is a solution to Newton's equations of motion* (1.18) *subject to the conditions $\gamma(t_1)=\xi$, $\gamma(t_2)=\eta$.*

Proof. We see that L given by (2.18) is a C^2 function and hence γ is a stationary point of $\mathcal{S}$ if and only if it satisfies the corresponding Euler-Lagrange equations. Using that $\frac{\partial L}{\partial v_i} = m_i v_i$ these equations take the form

$$m_i \ddot{\gamma}_i(t) + \nabla_i V(\gamma(t)) = 0, \qquad i = 1, \dots, N,$$

which are exactly Newton's equations. □

It follows from Hamilton's principle and Theorem 2.7 that if a C^2 motion γ minimizes (or maximizes) the action $\mathcal{S}$ on the set $C^1_{\xi,\eta}([t_1,t_2],\mathcal{U})$ then it is a solution of Newton's equations of motion. For this reason Hamilton's principle is often referred to as the **principle of least action**. As seen from the example treated in Exercise 2.15 the solutions to the equations of motion are not always extremal points of the action.

Given L as above let $\psi : \mathcal{O}_0 \to \mathbb{R}^3$ be a coordinate transformation of three variables and denote the new coordinates of particle i by q_i such that $x_i = \psi(q_i)$, $i = 1, \dots, N$. Then ψ gives rise to a coordinate transformation ψ^N of $3N$ variables defined by $\psi^N(q_1, \dots, q_N) = (\psi(q_1), \dots, \psi(q_N))$ on the set $\mathcal{O} = \{(q_1, \dots, q_N) \in {\mathcal{O}_0}^N \mid (\psi(q_1), \dots, \psi(q_N)) \in \mathcal{U}\}$. According to (2.11) the transformed Lagrangian with respect to ψ^N is given by

$$\widetilde{L}(q_1, \dots, q_N, u_1, \dots, u_N) = \sum_{i=1}^{N} \frac{1}{2} m_i \|D\psi(q_i)u_i\|^2 - V(\psi(q_1), \dots, \psi(q_N)) \,. \tag{2.20}$$

In case ψ is time dependent we have instead

$$\widetilde{L}(q_1, \dots, q_N, u_1, \dots, u_N, t) = \sum_{i=1}^{N} \frac{1}{2} m_i \|D\psi(q_i,t)u_i + \dot{\psi}(q_i,t)\|^2 - V\left(\psi(q_1,t), \dots, \psi(q_N,t)\right) . \tag{2.21}$$

We also call $\widetilde{L}$ the Lagrangian of the particle system in q-coordinates.

Now combining Hamilton's principle with the equivalence of (2.15) and (2.16) for this particular coordinate transformation we obtain the following result stating that Newton's equations can be written in Euler-Lagrange form with respect to any local system of coordinates, inertial or not.

Corollary 2.18. *For a system of N particles with Lagrangian (2.18) where V is a C^2 function on $\mathcal{U} \subseteq \mathbb{R}^{3N}$ and a coordinate transformation $x = \psi(q)$ defined on $\mathcal{O}_0 \subseteq R^3$ we have that Newton's equations of motion expressed in q-coordinates take the form*

$$\frac{d}{dt}\left(\nabla_{u_i} \widetilde{L}(\widetilde{\gamma}(t), \dot{\widetilde{\gamma}}(t))\right) - \nabla_{q_i} \widetilde{L}(\widetilde{\gamma}(t), \dot{\widetilde{\gamma}}(t)) = 0 \,, \quad i = 1, \dots, N \,, \tag{2.22}$$

in the sense that $\widetilde{\gamma} = (\widetilde{\gamma}_1, \dots, \widetilde{\gamma}_N)$ is a solution of this system of equations if and only if $\gamma = (\psi \circ \widetilde{\gamma}_1, \dots, \psi \circ \widetilde{\gamma}_N)$ is a solution of Newton's equations, and where $\widetilde{L}$ is defined by (2.20).

For a time dependent coordinate transformation $x = \psi(q,t)$ the same statement holds with $\widetilde{L}$ defined by (2.21) and $\gamma(t) = (\psi_t(\widetilde{\gamma}_1(t)), \dots, \psi_t(\widetilde{\gamma}_N(t)))$.

Example 2.19. The Lagrangian of the two-dimensional harmonic oscillator is given by

$$L(x_1, x_2, v_1, v_2) = \frac{1}{2}m(v_1^2 + v_2^2) - \frac{1}{2}k(x_1^2 + x_2^2)$$

where m, k are positive constants. The Euler-Lagrange equations are

$$m\ddot{x}_i(t) + kx_i(t) = 0\,, \quad i = 1, 2\,,$$

where we use the notation $(x_1(t), x_2(t))$ instead of $\gamma(t)$ for the particle trajectory. The solutions to this equation are of the form

$$x_i(t) = A_i \cos(\omega t + \varphi_i)\,,$$

where the amplitudes A_1, A_2 are non-negative constants, $\omega = \sqrt{\frac{k}{m}}$ is the angular frequency of the oscillator, and φ_1, φ_2 are initial phases.

Transforming to polar coordinates (r, θ) and using (2.8) the Lagrangian takes the form

$$\widetilde{L}(r, \theta, u_1, u_2) = \frac{1}{2}m(u_1^2 + r^2u_2^2) - \frac{1}{2}kr^2\,.$$

The equation of motion for r becomes

$$m\ddot{r}(t) - mr(t)\theta(t) + kr(t) = 0\,.$$

Since $\widetilde{L}$ is independent of θ the corresponding equation of motion becomes a conservation law,

$$\frac{d}{dt}mr^2\dot{\theta} = 0\,,$$

stating that the angular momentum $mr^2\dot{\theta}$ is a constant of the motion.

Definition 2.20. *For a particle system with Lagrangian $\widetilde{L}(q, u)$ in q-coordinates the* **generalized momentum** *of particle i or the* **conjugate variable** *to q_i is defined by*

$$p_i = \nabla_{u_i} L(q, u)\,,$$

and a variable q_i is called **cyclic** *if the Lagrangian is independent of q_i.*

Note that in an inertial system of coordinates it follows from (2.18) that the generalized momentum equals the standard momentum

$$p_i = m_i v_i\,,$$

and if q_i is a cyclic variable then the generalized momentum p_i is a constant of the motion, that is

$$\frac{dp_i}{dt} = \frac{d}{dt}\left(\nabla_{u_i} L(\gamma(t), \dot{\gamma}(t))\right) = 0\,,$$

if γ solves the equations of motion (2.22). We shall discuss conjugate variables further in the next chapter.

We point out that while Newton's equations in Chapter 1 have been treated as an *initial value problem*, where the equations of motion are supplemented by initial values of

the particle positions and velocities, the version obtained in Theorem 2.17 is a so-called *boundary value problem*, where the positions ξ and η are specified at at two different times t_1 and t_2. In particular, we emphasize that whereas Theorem 1.15 in Section 1.5 ensures existence and uniqueness of a solution to the initial value problem there is no such result available for boundary value problems. Indeed, it may happen in this case that no solution exists or there may be several solutions. This is illustrated in Exercise 2.15.

Remark 2.21. In view of Hamilton's principle one possible strategy for showing the existence of solutions to Newton's equation with given boundary values would be to try to show that there is a minimizer or maximizer for the action. It is well known that on Euclidean space $\mathbb{R}^k$ a continuous function has extremal points on a closed and bounded set. This property can be traced back to the fact that any such set is *sequentially compact*, i. e., that any sequence of points in the set has a convergent subsequence. The corresponding result is not true for closed and bounded subsets of $C^1_{x,y}([a,b],\mathbb{R}^k)$ (see Exercise 2.19). Nevertheless, in many cases it is still possible to conclude that functionals have minimizers or maximizers by enlarging the set $C^1_{x,y}([a,b],\mathbb{R}^k)$ to a space on which sequential compactness holds for suitable subsets and in this way conclude that there is an extremal point which is a solution to the Euler-Lagrange equations. This approach is beyond the scope of these notes and in general requires discussing the theory of generalized functions, also called distributions.

2.5 Constrained motion

In this section we provide a formulation of Hamilton's principle for constrained motion of an N-particle system, that is each particle trajectory is constrained to be on a subset M of $\mathbb{R}^3$. Besides the assumption that the motion takes place in a conservative force field it is also assumed that the constraining force on each particle is orthogonal to the subset M, which is frequently expressed by saying that the constraint is *holonomic*. Newton's equations for constrained motion of this kind turn out to be rather complicated but, as we shall see, Hamilton's principle for the motion can be adapted in a straight-forward way by simply allowing only motions in M in the variation of the action.

First, let us restrict attention to the case where M is the graph of a C^2 function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^2$ is an open set. Thus

$$M = \left\{(x_1, x_2, x_3) \,\middle|\, (x_1, x_2) \in A,\ x_3 = f(x_1, x_2)\right\}.$$

Henceforth, we set $x' = (x_1, x_2)$ such that $x = (x', x_3)$.

Except for the assumption that the constraining force is perpendicular to the graph this force is not explicitly given, but will be derived from the requirement that the motion is constrained to M. As mentioned we assume that, additionally, the motion is influenced by an external conservative force field $F = -\nabla V$, where $V : \mathbb{R}^{3N} \to \mathbb{R}$ is C^2. (Actually, V only needs to be defined in some open set containing M^N.)

Let us consider first a single particle of mass m in which case the motion is a C^2 function $\gamma = (\gamma^1, \gamma^2, \gamma^3) : I \to M$ from an interval I to the graph M, and hence can be written as $\gamma(t) = (\gamma'(t), f(\gamma'(t))$ where $\gamma' = (\gamma^1, \gamma^2) : I \to \mathbb{R}^2$ is the motion of the first pair of coordinates of γ. Similarly we write the force as $F = (F', F^3)$.

Applying the chain rule, the acceleration in the third coordinate direction is given by

$$\begin{aligned}\ddot{\gamma}^3(t) &= \sum_{i,j=1}^{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(\gamma'(t))\dot{\gamma}^i(t)\dot{\gamma}^j(t) + \sum_{i=1}^{2} \frac{\partial f}{\partial x_i}(\gamma'(t))\ddot{\gamma}^i(t) \\ &= \dot{\gamma}'(t) \cdot D^2 f\, \dot{\gamma}'(t) + \nabla f \cdot \ddot{\gamma}'(t)\,, \end{aligned} \tag{2.23}$$

where we have introduced the *Hesse matrix* for f

$$D^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix}$$

which, together with ∇f, is evaluated at $\gamma'(t)$ in the second line of (2.23).

The total force on the particle is $F_{\text{tot}} = F + F_{\text{c}}$, where F_{c} is the constraining force, such that Newton's second law takes the form

$$m\ddot{\gamma}(t) = F + F_{\text{c}}.$$

Splitting both sides of the equation into components parallel to and perpendicular to M, the parallel part provides, as we shall see, the information needed to calculate γ' and the perpendicular part may then be used to calculate the constraining force F_{c} which, however, we shall not need in the following. From multivariable calculus we know that the vector $(-\nabla f(x'), 1) \in \mathbb{R}^3$ is perpendicular to the graph of f at the point $(x', f(x'))$. Normalizing this vector we get that

$$n = (1 + \|\nabla f\|^2)^{-1/2}(-\nabla f, 1)$$

is a unit vector orthogonal to M at each of its points. Subtracting the normal components of both sides of Newton's equation we get

$$m\ddot{\gamma}(t) - m(\ddot{\gamma}(t) \cdot n) = F - (F \cdot n)n\,,$$

where we have used that F_{c} is perpendicular to the graph such that $F_{\text{c}} - (F_{\text{c}} \cdot n)n = 0$.

It turns out that, in order to determine γ', we need only consider the first two coordinates of this vector equation, that is

$$m\ddot{\gamma}' - m(\ddot{\gamma} \cdot n)n' = F' - (F \cdot n)n', \tag{2.24}$$

where

$$n' = -(1 + \|\nabla f\|^2)^{-1/2}\nabla f\,.$$

It is not difficult to see that the equation for $\gamma^3(t)$ can be derived from this equation, but this is not important for the present argument and is left as an exercise for the reader. Using (2.23) we have

$$\begin{aligned}\ddot{\gamma}(t) \cdot n &= (1 + \|\nabla f\|^2)^{-1/2}(\ddot{\gamma}^3(t) - \ddot{\gamma}'(t) \cdot \nabla f) \\ &= (1 + \|\nabla f\|^2)^{-1/2}\, \dot{\gamma}'(t)\, \cdot D^2 f\, \dot{\gamma}'(t)\,, \end{aligned}$$

and inserting this into (2.24) gives

$$m\ddot{\gamma}'(t) + m\frac{\dot{\gamma}'(t)\, \cdot D^2 f\, \dot{\gamma}'(t)}{1 + \|\nabla f\|^2}\nabla f = F'(\gamma(t)) + \frac{F^3(\gamma(t)) - F'(\gamma(t)) \cdot \nabla f}{1 + \|\nabla f\|^2}\nabla f\,. \tag{2.25}$$

This equation implies, in particular, that the vector $m\ddot{\gamma}'(t) - F'(\gamma(t))$ is parallel to $\nabla f(\gamma'(t))$, which means that $m\ddot{\gamma}'(t)$ and $F'(\gamma(t))$ have identical components orthogonal to $\nabla f(\gamma'(t))$. Hence, at points $\gamma'(t)$ where $\nabla f \neq 0$ we have

$$m\ddot{\gamma}'(t) - m\frac{\ddot{\gamma}'(t)\cdot\nabla f}{\|\nabla f\|^2}\nabla f = F'(\gamma(t)) - \frac{F'(\gamma(t))\cdot\nabla f}{\|\nabla f\|^2}\nabla f\,. \tag{2.26}$$

Multiplying equation (2.25) by $1+\|\nabla f\|^2$ and equation (2.26) by $\|\nabla f\|^2$ and subtracting the two equations we arrive at

$$m\ddot{\gamma}'(t) + m\left(\ddot{\gamma}'(t)\cdot\nabla f + \dot{\gamma}'(t)\,\cdot D^2 f\dot{\gamma}'(t)\right)\nabla f = F'(\gamma(t)) + F^3(\gamma(t))\nabla f\,. \tag{2.27}$$

Note that this equation may be written as $\ddot{\gamma}'(t) = h(\gamma'(t),\dot{\gamma}'(t))$ for some C^1 vector valued function h. Thus, according to Theorem 1.14 there is a unique local solution γ' near any time t_0 when initial conditions $\gamma'(t_0)$ and $\dot{\gamma}'(t_0)$ are specified. This holds also more generally for an N-particle system in which case we obtain N equations of the form (2.27) with m,γ',F replaced by m_i,γ_i',F_i for $i=1,\dots,N$. Our goal is now to show that this rather complicated system of equations of motion follows from a simple Hamilton's principle.

Theorem 2.22 (Hamilton's principle for constrained motion)**.**
Newton's equations for the motion $\gamma=(\gamma_1,\dots,\gamma_N):[a,b]\to M^N$ for N particles constrained to lie on the graph M of a C^2-function $f:\mathcal{O}\to\mathbb{R}$, where $\mathcal{O}\subseteq\mathbb{R}^2$ is open, such that $\gamma_i(t)=(\gamma_i'(t),f(\gamma_i'(t)))$ coincide with the Euler-Lagrange equations for the action

$$\mathcal{S}^{\rm c}(\gamma') = \int_a^b \left(\frac{1}{2}\sum_{i=1}^N m_i\dot{\gamma}(t)^2 - V(\gamma(t))\right)dt,$$

defined on $C^1([a,b],\mathcal{O})$. In other words, the constrained motion satisfies Hamilton's principle for the action functional for unconstrained motion restricted to the set of motions in M^N.

Proof. We again restrict attention to a single particle of mass $m=m_1$, since the general case follows by repeating the argument for each particle separately with the other particle motions fixed. Thus we need to derive the Euler-Lagrange equations for the action given with $N=1$ and showing that they coincide with (2.27).

Since $\dot{\gamma}(t) = (\dot{\gamma}'(t),\nabla f(\gamma'(t))\cdot\dot{\gamma}'(t))$ we may rewrite the action as

$$\mathcal{S}^{\rm c}(\gamma') = \int_a^b L(\gamma'(t),\dot{\gamma}'(t))dt,$$

where

$$L(q',v') = \frac{1}{2}m\|v'\|^2 + \frac{1}{2}m(\nabla f(q')\cdot v')^2 - V(q',f(q'))\,.$$

The Euler-Lagrange equations then have the form

$$\begin{aligned}&\frac{d}{dt}\left(m\dot{\gamma}'(t) + m(\nabla f\cdot\dot{\gamma}'(t))\nabla f\right)\\ &\quad= \; m(\nabla f\cdot\dot{\gamma}'(t))\,D^2 f\,\dot{\gamma}'(t) - \nabla_{x'}V(\gamma'(t),f(\gamma'(t))) - (\partial_{x_3}V)(\gamma'(t),f(\gamma'(t)))\nabla f\,.\end{aligned}$$

Recalling that $-\nabla_{x'}V = F'$ and $-\partial_{x_3}V = F^3$, it is straightforward to see that this equation is equivalent to (2.27). □

Example 2.23. Consider a pendulum consisting of a massless rod of length $\ell > 0$ with one end fixed at the point $(0, 0, \ell)$ and with a point particle of mass $m > 0$ at the other end. Under the influence of the constant gravitational force $F = (0, 0, -mg)$ the particle then performs a constrained motion on the sphere with radius ℓ and center $(0, 0, \ell)$, that is

$$x_1^2 + x_2^2 + (x_3 - \ell)^2) = \ell^2 \,.$$

Solving this equation for small oscillations around the equilibrium position $x_1 = x_2 = 0$ gives $x_3 = f(x_1, x_2)$, where

$$f(x_1, x_2) = \ell - \sqrt{\ell^2 - x_1^2 - x_2^2}\,, \quad \text{that is} \quad f(x') = \ell - \sqrt{\ell^2 - \|x'\|^2}\,.$$

Using the potential $V(x) = mg(x_3 - \ell)$ the unconstrained action of the motion $\gamma(t) = (\gamma^1(t), \gamma^2(t), \gamma^3(t))$ is

$$\mathcal{S}(\gamma) = \int_a^b \left(\frac{1}{2} m(\|\dot\gamma\|^2) - mg(\gamma^3(t) - \ell)\right) dt\,.$$

Inserting $\gamma(t) = (\gamma'(t), f(\gamma'(t)))$ gives the constrained action

$$\mathcal{S}^c(\gamma') = \int_a^b \left(\frac{1}{2} m \left(\|\dot\gamma'(t)\|^2 + \frac{(\gamma'(t)\cdot\dot\gamma'(t))^2}{\ell^2 - \|\gamma'(t)\|^2}\right) + mg\sqrt{\ell^2 - \|\gamma'(t)\|^2}\right) dt$$

with corresponding Lagrangian

$$L(x', v') = \frac{1}{2} m \left(\|v'\|^2 + \frac{(x'\cdot v')^2}{\ell^2 - \|x'\|^2}\right) + mg\sqrt{\ell^2 - \|x'\|^2}\,.$$

After some reduction ones finds that the Euler-Lagrange equations then take the form

$$\ddot\gamma^i(t) + \frac{\ddot\gamma'(t)\cdot\gamma'(t))}{\ell^2 - \|\gamma'(t)\|^2}\gamma^i(t) = -\left(\frac{\|\dot\gamma'(t)\|^2}{\ell^2 - \|\gamma'(t)\|^2} + \frac{(\gamma'\cdot\dot\gamma'(t))^2}{(\ell^2 - \|\gamma'(t)\|^2)^2}\right)\gamma^i(t) - \frac{g\gamma^i(t)}{\sqrt{\ell^2 - \|\gamma'(t)\|^2}}\,,$$

for $i = 1, 2$. Clearly this is a quite complicated non-linear system of two coupled equations. Notice, however, that the constant motion $\gamma'(t) = (0, 0)$ is a solution and, more generally, setting $\gamma^1(t) = \ell \sin\theta(t)$ and $\gamma^2(t) = 0$ one obtains for $i = 1$ the familiar form of the equation of motion for the pendulum in the $(x_1, x_3)-$plane

$$\ddot\theta(t) = -\frac{g}{\ell}\sin\theta(t)\,.$$

Remark 2.24. Hamilton's principle can be generalised in various ways. In particular, one may consider motion in $\mathbb{R}^k$ for arbitrary integer $k \geqslant 2$. If M is again chosen to be the graph of a C^2 function of $k-1$ variables and the constraining force is assumed to be orthogonal to M, the proof of Hamilton's principle given above can be generalized in a straight-forward manner.

One can also consider more general subsets M than graphs of functions, but still require that the constraining force is orthogonal the M. A particularly important class of sets $M \subseteq \mathbb{R}^k$ to consider are those of the form

$$M = \{\phi(x') \mid x' \in \mathcal{O}\}\,,$$

where $\phi : \mathcal{O} \to \mathbb{R}^k$ is a C^2 function defined on an open set $\mathcal{O} \subseteq \mathbb{R}^r$ and whose Jacobi matrix $D\phi(x')$ has rank r, i. e., its columns are linearly independent, at all points x'. In this case the constrained action $\mathcal{S}^{\mathrm{c}} : C^1([a,b],\mathcal{O}) \to \mathbb{R}$ for a single particle, which is the unconstrained action restricted to the set of constrained motions, is given by

$$\mathcal{S}^{\mathrm{c}}(\gamma') = \int_a^b \left(\frac{1}{2}m\dot{\gamma}(t)^2 - V(\gamma(t))\right) dt,$$

where $\gamma(t) = \phi(\gamma'(t))$. Hamilton's principle then says that Newton's equations for the constrained motion are equivalent to the Euler-Lagrange equations for this restricted action. We shall not prove this here since it requires geometric considerations beyond the scope of these notes.

From a physical point of view one would expect that the constrained motion on M only depends on M and not on the choice of function ϕ as long as $\phi(\mathcal{O}) = M$. In the language of Hamilton's principle this follows from the discussion of coordinate transformations in Section 2.3. Indeed, choosing a C^3 coordinate transformation $\psi : \widetilde{\mathcal{O}} \to \mathcal{O}$ and setting $\widetilde{\phi} = \phi \circ \psi$ we have $M = \phi(\mathcal{O}) = \widetilde{\phi}(\widetilde{\mathcal{O}})$, and the constrained action corresponding to $\widetilde{\phi}$ is obtained as the transformed constrained action for ϕ. Hence, the results of Section 2.3 tell us that the two sets of Euler-Lagrange equations are equivalent.

It is even possible to generalize Hamilton's principle to time dependent constrained motion, in which case $\phi : \mathcal{O} \times I \to \mathbb{R}^k$ is a C^3 function of a time variable t in addition to x'. Setting $\phi_t(x') = \phi(x', t)$, the set $M_t = \phi_t(\mathcal{O})$ to which the motion is constrained is time dependent but the same principle as stated above still holds with $\gamma(t) = \phi(\gamma'(t), t)$.

2.6 Appendix

In this appendix we provide the full details of the proof of Theorem 2.9.

Notice first that the functional $h \mapsto d\Phi^L_\gamma(h)$ defined in (2.4) is linear and continuous on $C^1_{0,0}([a,b],\mathbb{R}^k)$ (see Exercise 2.5). To prove differentiability of Φ^L it then suffices to show that

$$\left|\Phi^L(\gamma + h) - \Phi^L(\gamma) - d\Phi^L_\gamma(h)\right| \leqslant C\|h\|_1^2, \tag{2.28}$$

for all $h \in C^1_{0,0}([a,b],\mathbb{R}^k)$ with $\|h\|_1 \leqslant 1$ and some constant $C > 0$ depending only on L and γ, since this implies that $\lim_{\|h\|_1 \to 0} \frac{r(h)}{\|h\|_1} = 0$, using the notation in Definition 2.4.

For given h, let the continuous function f of two variables be given by

$$f(s,t) = L(\gamma(t) + sh(t), \dot{\gamma}(t) + s\dot{h}(t), t).$$

Then

$$\Phi^L(\gamma + h) - \Phi^L(\gamma) = \int_a^b (f(1,t) - f(0,t))dt \tag{2.29}$$

Moreover, by the chain rule

$$\partial_s f(s,t) = \sum_{i=1}^k \left(h_i(t)\frac{\partial L}{\partial q_i} + \dot{h}_i(t)\frac{\partial L}{\partial v_i}\right)(\gamma(t) + sh(t), \dot{\gamma}(t) + s\dot{h}(t), t).$$

Using that $h(a) = h(b) = 0$, an integration by parts on the right-hand side gives

$$\begin{aligned}\int_a^b \partial_s f(0,t)dt &= \int_a^b \sum_{i=1}^k \left(h_i(t) \frac{\partial L}{\partial q_i}(\gamma(t), \dot\gamma(t), t) + \dot h_i(t) \frac{\partial L}{\partial v_i}(\gamma(t), \dot\gamma(t), t) \right) \\ &= \int_a^b \sum_{i=1}^k \left(h_i(t) \left(\frac{\partial L}{\partial q_i}(\gamma(t), \dot\gamma(t), t) - \frac{d}{dt}\frac{\partial L}{\partial v_i}(\gamma(t), \dot\gamma(t), t) \right) \right) \\ &= d\Phi_\gamma^L(h). \end{aligned} \tag{2.30}$$

Combining (2.29) and (2.30) it is seen that in order to obtain (2.28) it is sufficient to show that

$$\left| \int_a^b \{ f(1,t) - f(0,t) - \partial_s f(0,t) \} dt \right| \leqslant C \|h\|_1^2, \tag{2.31}$$

if $\|h\|_1 \leqslant 1$.

Using the remainder estimate in Taylor's formula the inequality

$$|f(1,t) - f(0,t) - \partial_s f(0,t)| \leqslant \frac{1}{2} M \tag{2.32}$$

will hold if there exists a number M such that $|\partial_s^2 f(s,t)| \leqslant M$ for $s \in [0,1]$ and $t \in [a,b]$. By the chain rule, again, we have

$$\partial_s^2 f(s,t) = \sum_{i,j=1}^k \left(h_i(t) h_j(t) \frac{\partial^2 L}{\partial q_i \partial q_j} + \dot h_i(t) \dot h_j(t) \frac{\partial^2 L}{\partial v_i \partial v_j} + 2 h_i(t) \dot h_j(t) \frac{\partial^2 L}{\partial q_i \partial v_j} \right),$$

where the 2nd order partial derivatives above should be evaluated at the point $(\gamma(t) + sh(t), \dot\gamma(t) + s\dot h(t), t)$. If $\|h\|_1 \leqslant 1$ and $\|\gamma\|_1 \leqslant K$ this point belongs to the closed and bounded set

$$B_{K+1} \times B_{K+1} \times [a,b],$$

where B_r denotes the closed ball of radius r centered at the origin in $\mathbb{R}^k$. Since it is assumed that all second order derivatives of L are continuous it follows that they are bounded by some number $K_L > 0$ on this set. We hence have

$$|\partial_s^2 f(s,t)| \leqslant 4k^2 K_L \|h\|_1^2,$$

for all h with $\|h\|_1 \leqslant 1$. Consequently, (2.32) holds with $M = 4k^2 K_L \|h\|_1^2$.

In turn, from (2.32) it follows that (2.31) holds with $C = 2(b-a)k^2 K_L$ for all h such that $\|h\|_1 \leqslant 1$. This completes the proof of Theorem 2.9.

Exercises

Exercise 2.1. Show that the free action $\mathcal{S}_0$ given by (2.2) is a continuous functional on $C^1([a,b], \mathbb{R}^k)$.

Exercise 2.2. Show that the differential of a functional is uniquely determined if it exists.

Exercise 2.3. Show that a differentiable functional defined on $C^1_{x,y}([a,b],\mathbb{R}^k)$ is continuous.

Exercise 2.4. Let $f : [a,b] \to \mathbb{R}^k$ be a continuous function and define

$$\Phi(\gamma) = \int_a^b f(t) \cdot \gamma(t) dt\,,$$

where $\gamma : [a,b] \to \mathbb{R}^k$ is a curve. Show that Φ is differentiable and equals its differential. Determine the stationary points of Φ.

Exercise 2.5. Show that if $\gamma \in C^1([a,b],\mathbb{R}^k)$ then the functional $h \mapsto d\Phi^L_\gamma(h)$ defined on $C^1_{0,0}([a,b],\mathbb{R}^k)$ by (2.4), for a C^2-function L, is linear and satisfies

$$|d\Phi^L_\gamma(h)| \leqslant C\|h\|_1,$$

for some constant C depending only on γ and L. Prove that $d\Phi^L_\gamma$ is continuous. (See also Exercise 2.4.)

Exercise 2.6. Construct a function h with the properties used in the proof of Lemma 2.11.

Exercise 2.7. Let the functional Φ^L be given by a Lagrange function, which does not depend explicitly on t.

Show that, if γ is a solution to the corresponding Euler-Lagrange equations, then

$$E(t) = \dot\gamma \cdot \frac{\partial L}{\partial \dot\gamma}(\gamma(t),\dot\gamma(t)) - L(\gamma(t),\dot\gamma(t))$$

is a constant function of $t \in [a,b]$. We say that E is a conserved quantity.

Exercise 2.8. Let the functional Φ be defined by

$$\Phi(\gamma) = \int_a^b \frac{\sqrt{1+\dot\gamma(t)^2}}{\gamma(t)} dt\,,$$

where $\gamma : [a,b] \to \mathbb{R}_+$ is a positive C^1 function.

a) Show that the corresponding Euler-Lagrange equation is

$$\gamma\ddot\gamma + \dot\gamma^2 + 1 = 0\,.$$

b) To solve this equation, use first the result of Exercise 2.7 to show that any solution γ fulfills

$$\gamma^2 + (\gamma\dot\gamma)^2 = \textit{constant}\,.$$

Then proceed to find all solutions and verify that for given $x, y > 0$ there is exactly one solution γ such that $\gamma(a) = x$ and $\gamma(b) = y$.

Hint. It may be useful to rewrite the conservation equation in terms of $y = \gamma^2$. The correct form of the solutions is $\gamma(t) = \sqrt{-t^2 + Bt + C}$, where B, C are constants.

Exercise 2.9. Let $\gamma : [a, b] \to \mathbb{R}^k$ be C^1. Consider the functions $g(t) = \|\gamma(t) - \gamma(a)\|$ and

$$f(t) = \int_a^t \|\dot{\gamma}(s)\| ds.$$

Show that if $\gamma(t) \neq \gamma(a)$ then $f'(t) \geqslant g'(t)$. (*Hint.* Use the Cauchy-Schwarz inequality.) Conclude that a straight line minimizes the length functional.

Exercise 2.10. Show that the motion (2.6) minimizes the free action $\mathcal{S}_0$.
Hint. It may be useful to compute the value of $\mathcal{S}_0$ for the difference between an arbitrary curve and the solution (2.6).

Exercise 2.11. Show that if $L : \mathbb{R}^{2k} \times [a, b] \to \mathbb{R}$ is a C^2 function then the functional Φ^L is differentiable at all $\gamma \in C^1([a, b], \mathbb{R}^k)$ with a functional given by the expression in the first line of equation (2.30).

Exercise 2.12. Show that the length functional in polar coordinates in the plane has the form

$$\widetilde{\mathcal{L}}(\widetilde{\gamma}) = \int_a^b \sqrt{\dot{r}(t)^2 + r(t)^2 \dot{\theta}(t)^2} dt\,,$$

where $\widetilde{\gamma}(t) = (r(t), \theta(t))$, for $t \in [a, b]$.

Write also the free action $\mathcal{S}_0$ in polar coordinates and find the corresponding Euler-Lagrange equations.

Exercise 2.13. Let the time dependent coordinate transformation of 2 variables be given by

$$\begin{aligned} x_1 &= \cos(\omega t)\widetilde{x}_1 - \sin(\omega t)\widetilde{x}_2 \\ y_1 &= \sin(\omega t)\widetilde{x}_1 + \cos(\omega t)\widetilde{x}_2\,, \end{aligned}$$

corresponding to an orthogonal coordinate system rotating with constant angular velocity ω relative to an inertial system around their common origin.

Consider a free particle in 2-dimensions, i. e., its motion is determined by the free action $\mathcal{S}_0$.

a) Show that the transformed equations of motion can be written as the Euler-Lagrange equations for a functional with a Lagrangian function which is independent of time.

b) Write down the equations of motion and interpret the different terms.

Exercise 2.14. Generalize the proof of Lemma 2.14 to time dependent coordinate transformations.

Exercise 2.15. In this exercise we study the 1-dimensional motion of a particle of mass $m > 0$ moving in a harmonic oscillator potential $V : \mathbb{R} \to \mathbb{R}$ given by $V(q) = \frac{1}{2}kq^2$, where $k > 0$ is the spring constant. Thus the Lagrange function for this system is

$$L(q, v) = \frac{1}{2}mv^2 - \frac{1}{2}kq^2\,,$$

and we denote the corresponding action functional by $\mathcal{S}$.

a) Determine all solutions $\gamma(t)$ to the equation of motion satisfying $\gamma(0) = 0$.

b) Show that all these solutions satisfy $\gamma(\pi\sqrt{m/k}) = 0$.

c) Conclude that the equation of motion with the boundary values $\gamma(0) = 0$ and $\gamma(\pi\sqrt{m/k}) = 0$ has infinitely many solutions. Prove that all these solutions have vanishing action $\mathcal{S}(\gamma) = 0$.

d) Conclude that the equation of motion with the boundary values $\gamma(0) = 0$ and $\gamma(\pi\sqrt{m/k}) = x$ has no solution for all $x \in \mathbb{R} \setminus \{0\}$.

e) Show that for all $T > 0$ and $T \notin \left\{ n\pi\sqrt{m/k} \;\middle|\; n \in \mathbb{N} \right\}$ the equation of motion with boundary values $\gamma(0) = 0$ and $\gamma(T) = 0$ has exactly one solution and that this solution also satisfies $\mathcal{S}(\gamma) = 0$.

f) Prove that for all $T > 0$ we can find a C^1 function $\gamma : [0, T] \to \mathbb{R}$ with $\gamma(0) = 0$ and $\gamma(T) = 0$ such that the corresponding action $\mathcal{S}(\gamma) < 0$, if the spring constant $k > 0$ is sufficiently large.

g) Show that for all $T > 0$ and all $k > 0$ we can find a C^1 function $\gamma : [0, T] \to \mathbb{R}$ with $\gamma(0) = 0$ and $\gamma(T) = 0$ such that the corresponding action $\mathcal{S}(\gamma) > 0$.

h) Conclude that there are values of $T > 0$ and $k > 0$ such that the solutions to the equation of motion with boundary values $\gamma(0) = 0$ and $\gamma(T) = 0$ is neither a minimizer nor a maximizer for the action $\mathcal{S}$ on the set $C^1_{0,0}([0, T], \mathbb{R})$.

Exercise 2.16.

a) Consider the force field $F : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$F(x, y) = (2x + y, x + 2y).$$

Show that F is a conservative force field.

b) Show that the motion $(x(t), y(t)) = (e^{\sqrt{3}t}, e^{\sqrt{3}t})$ solves Newton's equations of motion corresponding to a particle of mass $m = 1$ moving in the above force field.

c) Write down the action corresponding to a 2-dimensional motion of a particle of mass $m = 1$ in the force field from (a) starting at time $t = 0$ and ending at time $t = 1$. Calculate the action for the motion in b).

Exercise 2.17. A particle of charge e moving with velocity v in $\mathbb{R}^3$ in a magnetic field $\mathbf{B} : \mathbb{R}^3 \to \mathbb{R}^3$ is subject to the *Lorentz* force $ev \times \mathbf{B}$ where $\times$ here refers to the vector cross product.

a) Consider motion $\gamma(t) = (\gamma_1(t), \gamma_2(t), 0)$ in the xy-plane with the magnetic field being $(0, 0, B)$, where $B : \mathbb{R}^2 \to \mathbb{R}$. Show that Newton's equations for a particle of mass m and charge e become

$$\begin{aligned} m\ddot{\gamma}_1(t) &= e\dot{\gamma}_2(t)B(\gamma_1(t), \gamma_2(t)) \\ m\ddot{\gamma}_2(t) &= -e\dot{\gamma}_1(t)B(\gamma_1(t), \gamma_2(t)). \end{aligned}$$

b) Show that if B is continuous then there exist vector fields $A : \mathbb{R}^2 \to \mathbb{R}^2$, $A(x, y) = (A_1(x, y), A_2(x, y))$ such that $B(x, y) = \partial_x A_2(x, y) - \partial_y A_1(x, y)$ (see Example 1.23). How many such vector fields can you find?

c) Show that Newton's equations follow from Hamilton's principle with Lagrange function

$$L(q, v) = \frac{1}{2}mv^2 + eA(q) \cdot v\,, \quad q, v \in \mathbb{R}^2\,.$$

Exercise 2.18. Write down the expression for the free action $\mathcal{S}_0(\gamma) = \frac{1}{2}\int_a^b \|\dot\gamma(t)\|^2 dt$ in spherical coordinates (r, θ, φ), given by

$$(x, y, z) = r(\cos\theta\cos\varphi, \cos\theta\sin\varphi, \sin\theta)\,,$$

for $r > 0, -\pi/2 \leqslant \theta \leqslant \pi/2, 0 \leqslant \varphi \leqslant 2\pi$. Determine the corresponding Euler-Lagrange equations and show that straight lines through the origin, with linear parametrization, are solutions.

Exercise 2.19. Consider the sequence of functions $g_n : [-1, 1] \to \mathbb{R}$, $n = 1, 2 \ldots$

$$g_n(x) = \begin{cases} 1, & -1 \leqslant x < -n^{-1} \\ -nx, & -n^{-1} \leqslant x \leqslant n^{-1} \\ -1, & n^{-1} < x \leqslant 1 \end{cases}$$

a) show that $g_n \in C([-1, 1])$, i.e., that g_n continuous for all n.

b) Let $f_n(x) = \int_{-1}^x g_n(y)dy$ for $n = 1, 2, \ldots$. Argue that $f_n \in C^1([-1, 1])$ and that $\|f_n\|_1 \leqslant 1$.

c) Let $f_\infty(x) = 1 - |x|$. Show that $f_\infty \notin C^1([-1, 1])$, but that it is piecewise C^1, and that

$$\max_{x\in[-1,1]} |f_n(x) - f_\infty(x)| \to 0, \qquad \text{as } n \to \infty.$$

d) A subset of $C^1([-1, 1])$ is said to be sequentially compact if any sequence (f_n) of functions in the set has a convergent subsequence (f_{n_k}), i.e., a subsequence such that $\|f_{n_k} - f\|_1 \to 0$ for some $f \in C^1([-1, 1])$. Prove, using the result above that the unit ball $\{f \in C^1([-1, 1]) \mid \|f\|_1 \leqslant 1\}$ in $C^1([-1, 1])$ is *not* sequentially compact.

Chapter 3

Hamiltonian Mechanics

The main goal in this chapter is to reformulate the Euler-Lagrange equations for a mechanical system, discussed in Chapter 2, as a system of first order differential equations in phase space, called **Hamilton's equations**. Such a reformulation has several advantages. On the one hand, it provides a natural basis for analyzing the phase flow of the system and its qualitative properties. As we shall see, it also turns out to be a useful setting for uncovering and investigating symmetries of the physical system and associated conserved quantities. It should also be remarked that there is a close analogy between the Hamiltonian formalism of classical mechanics and the so-called *canonical quantum mechanical formalism* whose deeper aspects we shall not have the possibility to discuss in these notes.

3.1 Convex functions

As a preparation for discussing Hamiltonian mechanics, we need first to introduce convex functions and establish some of their basic properties. In our applications to classical mechanics the convex functions under consideration are generally assumed to be differentiable. However, non-differentiable convex functions occur naturally in thermodynamics and statistical mechanics, e. g. in the description of phase transitions, and as such their properties are highly relevant to physics. To attain a sufficient degree of generality we shall therefore not assume differentiability except when explicitly stated.

Definition 3.1. *For $x, y \in \mathbb{R}^k$ and $0 < \alpha < 1$ we say that the point $\alpha x + (1-\alpha)y$ is a* convex combination *of x and y. This point is on the line segment between x and y and all points on this line segment can be written in this way for some $0 \leqslant \alpha \leqslant 1$.*

Definition 3.2. *A subset $\mathcal{C} \subseteq \mathbb{R}^k$ is said to be a* **convex set** *if, whenever $x, y \in \mathcal{C}$, then the whole line segment joining x and y is also in $\mathcal{C}$. This may be rephrased as saying that*

$$\alpha x + (1-\alpha)y \in \mathcal{C} \quad \textit{for all } 0 \leqslant \alpha \leqslant 1\,.$$

Evidently, the convex subsets of the real line $\mathbb{R}$ are exactly the intervals. A convex and a non-convex set in the plane $\mathbb{R}^2$ are illustrated in Fig. 3.1. Examples of convex subsets of $\mathbb{R}^k$ are, apart from $\mathbb{R}^k$ itself, the *open balls*

$$B(a;r) = \{x \in \mathbb{R}^k \mid \|x-a\| < r\}\,, \quad a \in \mathbb{R}^k,\ r > 0\,.$$

Indeed, if $x, y \in B(a;r)$ and $0 \leqslant \alpha \leqslant 1$, the triangle inequality yields

$$\|\alpha x+(1-\alpha)y-a\| = \|\alpha(x-a)+(1-\alpha)y\| \leqslant \alpha\|x-a\|+(1-\alpha)\|y-a\| < \alpha r+(1-\alpha)r = r\,,$$

which proves the claim. Similarly, the *closed balls*

$$B(a;r) = \{x \in \mathbb{R}^k \mid \|x-a\| \leqslant r\}\,, \quad a \in \mathbb{R}^k,\ r > 0\,.$$

are convex.

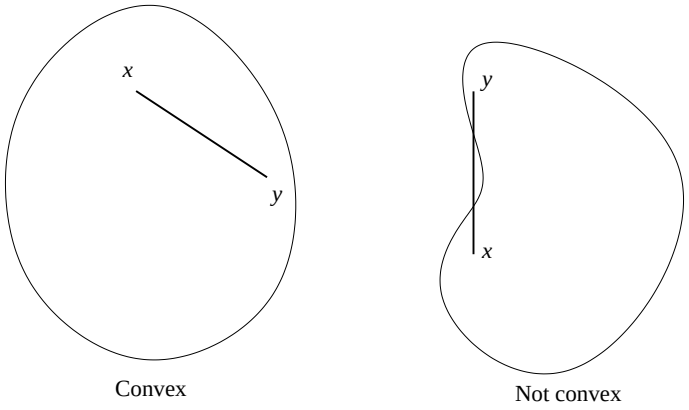

Figure 3.1: A convex and a non-convex set in $\mathbb{R}^2$

Definition 3.3. *A real-valued function f defined on a convex set $\mathcal{C} \subseteq \mathbb{R}^k$ is said to be a* **convex function** *if, for all $x, y \in \mathcal{C}$ and all $0 < \alpha < 1$, it holds that*

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y)\,. \tag{3.1}$$

The function is **strictly convex** *if*

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y) \tag{3.2}$$

for all $x, y \in \mathcal{C}$ with $x \neq y$ and $0 < \alpha < 1$.

We say that f is (strictly) **concave** *if $-f$ is (strictly) convex.*

Inequality (3.1), resp. (3.2), expresses the fact that the graph of the function f lies below, resp. strictly below, the (open) line segment joining the two points $(x, f(x))$ and $(y, f(y))$ on the graph of f as illustrated in Figure 3.2.

Example 3.4.
a) A function $f : \mathbb{R}^k \to \mathbb{R}$ of the form

$$f(x) = a \cdot x + b\,, \quad x \in \mathbb{R}^k\,,$$

where $a \in \mathbb{R}^k$ and $b \in \mathbb{R}$ is called *affine*. Such functions satisfy (3.1) with equality sign holding (verify this) and are hence both convex and concave, their graphs being straight lines.

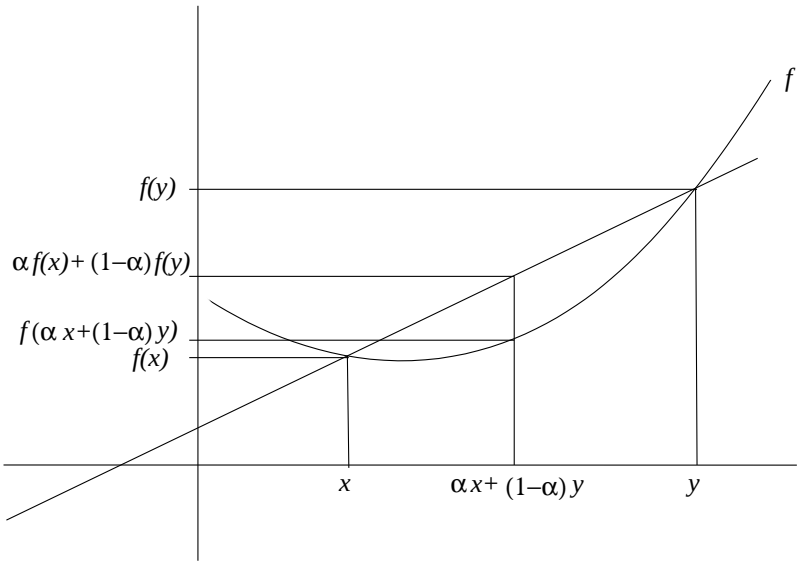

Figure 3.2: Graphical illustration of inequality (3.1)

b) The function

$$g(x) = k\|x\|,\ x \in \mathbb{R}^k,$$

is convex if $k > 0$, as a consequence of the triangle inequality (verify this). Note that in this case g is not differentiable at $x = \underline{0}$.

The following lemma is useful for constructing convex functions from other convex functions and will be applied repeatedly in the following.

Lemma 3.5.

a) Let $f : \mathcal{C} \to \mathbb{R}$ and $g : \mathcal{C} \to \mathbb{R}$ be two convex functions defined on the same convex set $\mathcal{C}$. Then $\lambda f + \mu g$ is a convex function on $\mathcal{C}$, if $\lambda, \mu \geqslant 0$.

b) Let $\mathcal{I}$ be an arbitrary set and for each $i \in \mathcal{I}$ let f_i be a convex function defined on a convex subset $\mathcal{C}_i$ of $\mathbb{R}^k$. Define the function f by

$$f(x) = \sup\{f_i(x) \mid i \in \mathcal{I}\}$$

for each x such that the sup *on the right hand side is finite, that is f is defined on the set*

$$\mathcal{C} = \{x \in \mathbb{R}^k \mid x \in \mathcal{C}_i \text{ for all } i \in \mathcal{I}, \text{ and } f_i(x) \leqslant M \text{ for all } i \in \mathcal{I} \text{ for some } M \in \mathbb{R}\}$$

Then f is a convex function.

Proof. The proof of a) is easy and is left for the reader. In order to prove b) we need to argue first that $\mathcal{C}$ is a convex set. So let $x, y \in \mathcal{C}$ and choose M_x and M_y in $\mathbb{R}$ such that $f_i(x) \leqslant M_x$ and $f_i(y) \leqslant M_y$ for all $i \in \mathcal{I}$. For $0 < \alpha < 1$ we then have by (3.1) that

$$f_i(\alpha x + (1-\alpha)y) \leqslant \alpha f_i(x) + (1-\alpha) f_i(y) \leqslant \alpha M_x + (1-\alpha) M_y$$

for all $i \in \mathcal{I}$, which shows that $\alpha x(1 - \alpha y \in \mathcal{C}$ and hence that $\mathcal{C}$ is convex. Choosing M_x and M_y above to be $\sup\{f_i(x) \mid i \in \mathcal{I}\}$ and $\sup\{f_i(y) \mid i \in \mathcal{I}\}$, respectively, and using the definition of f the previous inequality states that

$$f_i(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y)$$

for all $i \in \mathcal{I}$. By the definition of $\sup A$ for a set $A \subseteq \mathbb{R}$ as the smallest number larger than or equal to all numbers in A this implies $f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y)$. This proves that f is convex. □

The following characterization of convex functions of a single variable is particularly useful for analyzing regularity properties such as continuity and differentiability of convex functions in general.

Lemma 3.6. *A function $f : I \to \mathbb{R}$ of one variable defined on an interval $I \subseteq \mathbb{R}$ is convex if an only if*

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leqslant \frac{f(x_3) - f(x_1)}{x_3 - x_1} \leqslant \frac{f(x_3) - f(x_2)}{x_3 - x_2}\,. \tag{3.3}$$

holds for all $x_1, x_2, x_3 \in I$ such that $x_1 < x_2 < x_3$. Moreover, f is strictly convex if and only if (3.3) *holds with strict inequalities.*

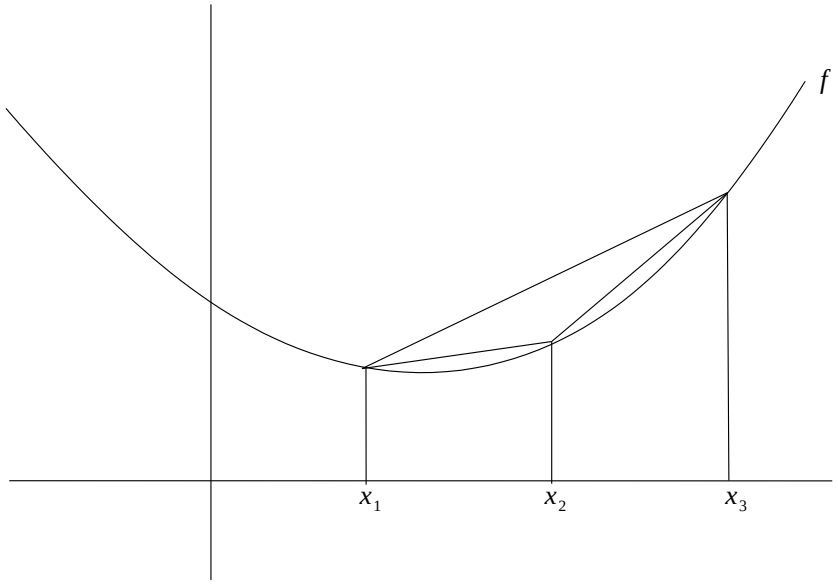

Figure 3.3: Graphical illustration of inequalities (3.3)

Proof. Let x_1, x_2, x_3 be as stated and consider the two points $(x_1, f(x_1))$ and $(x_3, f(x_3))$ on the graph of f and a third point (x_2, y) on the vertical line through $(x_2, 0)$ (see Figure 3.3). Clearly, (x_2, y) is below the line segment s_{13} connecting $(x_1, f(x_1))$ and $(x_3, f(x_3))$ if and only if the slope of s_{13} is larger than the slope of the line segment s_{12} connecting $(x_1, f(x_1))$ and (x_2, y) and smaller than that of the line segment s_{23} connecting (x_2, y) and $(x_3, f(x_3))$.

In particular, $(x_2, f(x_2))$ is below s_{13} if and only is these inequalities holds for $y = f(x_2)$. But this is exactly the contents of (3.3).

The statement for strictly convex functions is proven similarly. □

Note that if we define the function $m_f(x, y)$ of two variables $x, y \in I$ for $x \neq y$ as the slope of the line segment connecting $(x, f(x))$ and $(y, f(y))$, i. e.

$$m_f(x, y) := \frac{f(x) - f(y)}{x - y}, \qquad x, y \in I,\ x \neq y\,,$$

the first inequality of (3.3) shows that m_f is a monotone increasing function of y for $y > x$, if x is kept fixed, while the second shows that this also holds true for $y < x$. Since m_f is symmetric in x and y we conclude that m_f is a monotonic increasing function in each variable separately.

We next apply this result to derive the following central fact about convex functions of one variable.

Theorem 3.7. *Let $f : I \to \mathbb{R}$ be a convex function defined on an interval $I \subseteq \mathbb{R}$. Then the left and right derivatives*

$$f'_-(x_0) := \lim_{x \to x_0-} \frac{f(x) - f(x_0)}{x - x_0} \quad \textit{and} \quad f'_+(x_0) := \lim_{x \to x_0+} \frac{f(x) - f(x_0)}{x - x_0}$$

exist at every interior point x_0 in I. Moreover both $f_-(x)$ and $f_+(x)$ are monotone increasing functions on the interior of I and fulfill

$$f'_-(x) \leqslant f'_+(x) \qquad \textit{and} \quad f'_+(x) \leqslant f'_-(y) \ \textit{ if } x < y\,. \tag{3.4}$$

Proof. Let x_0 be an interior point of I. Then there exist numbers x_1 in I to the left of x_0 and x_3 to the right of x_0 and for such numbers the inequality (3.3) holds with x_2 replaced by x_0. In particular, we have

$$m_f(x_1, x_0) \leqslant m_f(x_3, x_0)\,.$$

Since the left hand side is an increasing function of x_1 it converges as $x_1 \to x_0-$ to a limit $f'_-(x_0$ fulfilling

$$f'_-(x_0) \leqslant m_f(x_3, x_0)\,.$$

Using this inequality we get similarly that the right hand side converges as $x_3 \to x_0+$ to a limit $f'_+(x_0)$ and that

$$f'_-(x_0) \leqslant f'_+(x_0)\,.$$

This proves existence of $f'_\pm(x_0)$ and the former of the stated inequalities.

For $x, y \in I$ we again consider (3.3) with x_1 replaced by x, x_3 replaced by y, and $x < x_2 < y$, such that

$$m_f(x, x_2) \leqslant m_f(x_2, y)$$

Here the left hand side decreases to $f'_+(x)$ as $x_2 \to x+$ and hence

$$f'_+(x) \leqslant m_f(x_2, y)\,.$$

Now, the right hand side of this inequality converges to $f'_-(y)$ as $x_2 \to y-$. Hence $f'_+(x) \leqslant f'_-(y)$ as claimed. Combining the two inequalities gives

$$f'_+(x) \leqslant f'_-(y) \leqslant f'_+(y)\,,$$

which shows that f'_+ is increasing. Similarly, we get that f_- is increasing. This finishes the proof. □

Note that (3.4) implies that if $f'_-(x) \neq f'_+(x)$, then f'_- and f'_+ are not continuous at x. Since a monotonic function has at most a countable number of discontinuities, it follows that any convex function of one variable is differentiable everywhere except possibly at a finite or countable number of points. We have already seen that non-differentiable convex functions exist. Below we prove that convex functions are at least continuous at interior points of their domain. Before doing we shall discuss differentiable convex functions in more detail.

Proposition 3.8. *Let $f : I \to \mathbb{R}$ be a differentiable function defined on an open interval I. Then f is convex if and only if f' is a monotonic increasing function on I. Moreover, f is strictly convex if and only if f' is strictly increasing.*

Proof. That f' is increasing on I if f is convex follows from Lemma 3.7. Assume next that f' is increasing and let $x_1 < x_2 < x_3$ be in I. By the mean value theorem there exist y_1, y_2 such that $x_1 < y_1 < x_2 < y_2 < x_3$ and

$$m_f(x_1, x_2) = f'(y_1) \quad \text{and} \quad m_f(x_2, x_3) = f'(y_2)\,.$$

Since $y_1 < y_2$ we conclude that $m_f(x_1, x_2) \leqslant m_f(x_2, x_3)$, which evidently implies that $(x_2, f(x_2))$ lies below the line segment connecting $(x_1, f(x_1))$ and $(x_2, f(x_2))$. This proves that f is convex.

The statement for strictly convex functions is proven similarly. □

Corollary 3.9. *Let $f : I \to \mathbb{R}$ be a twice differentiable function defined on an open interval I. Then f is convex if and only if $f''(x) \geqslant 0$ for all $x \in I$. Moreover, f is strictly convex if $f''(x) > 0$ for all $x \in I$.*

Proof. Since f' is differentiable it is well known that it is monotone increasing if and only if $f'' \geqslant 0$ on I. Furthermore, f' is strictly increasing if $f'' > 0$ on I. Hence the corollary follows from Proposition 3.8. □

Corollary 3.10. *Let $f : C \to \mathbb{R}$ be a two times differentiable function defined on an open convex set $C \subseteq \mathbb{R}^k$. Then f is convex if and only if*

$$\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x)\, a_i a_j \geqslant 0 \tag{3.5}$$

for all $x \in C$ and all $a = (a_1, a_2, \ldots, a_k) \in \mathbb{R}^k$. Moreover, if (3.5) holds with strict inequality for all $x \in C$ and all $a \neq \underline{0}$ in $\mathbb{R}^k$, then f is strictly convex.

Proof. Let $x \in C$ and let $a \in \mathbb{R}^k \setminus \{\underline{0}\}$ be given. Defining the affine function $\ell_{x,a}(t) = x + ta,\ t \in \mathbb{R}$, we have that the function $f \circ \ell_{x,a}$ is defined on an open interval $I_{x,a}$ containing 0 and is convex if f is convex (verify this!). Conversely, assume $f \circ \ell_{x,a}$ is convex for all choices of $x \in C$ and $a \in \mathbb{R}^k \setminus \{\underline{0}\}$. For $x_1, x_2 \in C$, choose $x = x_1$ and $a = x_2 - x_1$ such that $\ell_{x,a}(t) = x_1 + t(x_2 - x_1) = tx_2 + (1-t)x_1$ for $t \in [0,1]$. Using that $\ell_{x,a}$ is convex we get

$$\begin{aligned} f(\alpha x_1 + (1-\alpha)x_2) &= f(\ell_{x,a}(1-\alpha)) = f \circ \ell_{x,a}(\alpha \cdot 0 + (1-\alpha)\cdot 1)) \\ &\leqslant \alpha f \circ \ell_{x,a}(0) + (1-\alpha) f \circ \ell_{x,a}(1) = \alpha f(x_1) + (1-\alpha) f(x_2) \end{aligned}$$

for $\alpha \in [0,1]$. This shows that (3.1) holds and hence that f is convex. These considerations prove that establishing convexity of a function f of k variables defined on a convex set C is equivalent to showing convexity of all one-variable functions $f \circ \ell_{x,a}$, where $x \in C$ and $a \in \mathbb{R}^k$, $a \neq \underline{0}$. We briefly express this latter property by saying that f is convex on any line segment in C.

Clearly, $f \circ \ell_{x,a}$ is a twice differentiable function if f is, and applying the chain rule twice we get

$$(f \circ \ell_{x,a})''(t) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x + ta)\, a_i a_j \,. \tag{3.6}$$

Using Corollary 3.9 it follows by setting $t = 0$ in this identity that (3.5) holds if f is convex. Conversely, if (3.5) holds it follows from Corollary 3.9 and (3.6) that f is convex.

The statement concerning strict convexity follows similarly. □

Remark 3.11. A real $k \times k$-matrix $A = (a_{ij})$ is called *positive semi-definite* if

$$\langle x, Ax \rangle = \sum_{i,j=1}^{n} a_{ij}\, x_i x_j \geqslant 0 \,, \quad \text{for all } x = (x_1, \ldots, x_k) \in \mathbb{R}^k \,. \tag{3.7}$$

If strict inequality hold is (3.7) for all $x \neq \underline{0}$ then A is called *positive definite*.

For a twice differentiable function f of k variables the $k \times k$-matrix $D^2 f(x)$, whose ij-matrix element equals $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ is called the *Hesse matrix* of f at x (and was already introduced for functions of two variables in Section 2.5). Hence Corollary 3.10 can be expressed by saying that f is convex if and only if the Hesse matrix of f is positive semi-definite at all points of C, and a sufficient condition for f to be strongly convex is that $D^2 f(x)$ is positive definite for all $x \in C$.

In case f is a C^2 function we know that $D^2 f(x)$ is a symmetric matrix. In this case it can be shown that $D^2 f(x)$ is positive semi-definite, resp. positive definite, if and only if all its eigenvalues are non-negative, resp. positive (see Exercise 3.5).

Example 3.12. a) The function $f(x) = x^a$, defined for $x > 0$, is C^2 and $f''(x) = a(a-1)x^{a-2}$. Using Corollary 3.10 it follows that f is convex if and only if $a \geqslant 1$ or $a \leqslant 0$.

b) The quadratic function $f(x,y) = x^2 + 3xy + 3y^2$ on $\mathbb{R}^2$ has constant Hesse matrix

$$D^2 f = \begin{pmatrix} 2 & 3 \\ 3 & 6 \end{pmatrix} ,$$

which has eigenvalues $4 \pm \sqrt{13}$. Since these are positive f is strictly convex by Corollary 3.10 and the previous remark.

We now return to the general convex functions. First, recall that a *hyperplane* in $\mathbb{R}^k$ is a set of the form

$$H = \{x \in \mathbb{R}^k \mid N \cdot x = c\}\,, \tag{3.8}$$

where $c \in \mathbb{R}$ is a constant and $N \in \mathbb{R}^k$, $N \neq \underline{0}$, is called a *normal vector* to H, since N is orthogonal to the vector $x_1 - x_2$ for any $x_1, x_2 \in H$. This also shows that if we set $H_0 = \{x \in \mathbb{R}^k \mid N \cdot x = 0\}$, and x_0 is some point in H then H_0 is a $(k-1)$-dimensional subspace of $\mathbb{R}^k$ and

$$H = x_0 + H_0 := \{x_0 + x \mid x \in H_0\}\,.$$

In other words, a hyperplane is a translated $(k-1)$-dimensional subspace of $\mathbb{R}^k$.

In particular, the graph of an affine function $h(x) = \mu \cdot x + a$, $x \in \mathbb{R}^k$, where $\mu \in \mathbb{R}^k$ and $a \in \mathbb{R}$ are fixed, is a hyperplane in $\mathbb{R}^{k+1}$

$$H = \{(x, x_{k+1}) \mid x_{k+1} = \mu \cdot x + a\,,\ \ x \in \mathbb{R}^k\} \tag{3.9}$$

with normal vector $N = (\mu, -1)$.

The following theorem establishes the existence of hyperplanes touching the graph of a convex function f at any given point not on the boundary of its domain of definition and such that the graph of f has no points below the hyperplane.

Theorem 3.13. *Let $f : \mathcal{C} \to \mathbb{R}$ be a convex function defined on a convex set $\mathcal{C} \subseteq \mathbb{R}^k$. For each interior point $x_0 \in \mathcal{C}$, there exists at least one vector $\mu \in \mathbb{R}^k$ such that*

$$f(x) \geqslant f(x_0) + \mu \cdot (x - x_0). \tag{3.10}$$

The graph of the affine function $h(x) = f(x_0) + \mu \cdot (x - x_0)$ is called a **supporting hyperplane for f at x_0**.

Moreover, the function f has partial derivatives at x_0 if and only if the supporting hyperplane is unique. In this case $\mu = \nabla f(x_0)$.

Proof. We give the proof first in the case of one variable, i. e. if $k = 1$.

That x_0 is an interior point of $\mathcal{C}$ means that there is some open interval $I \subset \mathcal{C}$ such that $x_0 \in I$. From Theorems 3.6 and 3.7 follows that that the left and right derivatives $\mu_\pm$ of f at x_0 exist and that

$$\mu_+ = \inf\{\frac{f(x) - f(x_0)}{x - x_0} \mid x > x_0\} \quad \text{and} \quad \mu_- = \sup\{\frac{f(x) - f(x_0)}{x - x_0} \mid x < x_0\}\,.$$

Hence, we have

$$f(x) \geqslant f(x_0) + \mu(x - x_0)$$

for all x if and only if $\mu \in [\mu_-, \mu_+]$. This proves the theorem for one variable.

For a convex function of several variables, the problem is more complicated. A proof of the existence of a supporting hyperplane is contained in Exercise 3.9. The proof is based on utilizing the functions $f \circ \ell_{x_0,a}$ as in the proof of Theorem 3.10. In particular, if the f has partial derivatives the function $f \circ \ell_{x_0,\varepsilon_i}$, where ε_i denotes a unit vector along the i'th

coordinate axis, is a differentiable convex function of one variable, and we may apply the result just proven to conclude that there is a unique supporting line along each coordinate direction. Since the function is convex, it is not difficult to see that these lines span a unique supporting hyperplane. □

We note the following simple but important consequences concerning maxima and minima of convex functions.

Corollary 3.14. *The following holds for any convex function $f : \mathcal{C} \to \mathbb{R}$ of k variables.*
a) If f has a maximum at an interior point x_0 of $\mathcal{C}$ then f is constant on $\mathcal{C}$. In particular, any convex function assumes its maximal value, if it exists, on the boundary of its domain of definition.
b) Any local minimum of f is also a global minimum of f. If f is differentiable at an interior point x_0 of $\mathcal{C}$, then f has a minimum at x_0 if and only if $\nabla f(x_0) = \underline{0}$.

Remark 3.15. The same statements, with "maximum" and "minimum" interchanged, hold for concave functions.

Proof. Assume f has a maximum at an interior point x_0 and let μ be as in Theorem 3.13. Then $f(x_0) \geqslant f(x) \geqslant f(x_0)+\mu\cdot(x-x_0)$ for all $x \in \mathcal{C}$. It follows that $\mu\cdot(x-x_0) \leqslant 0$, $x \in \mathcal{C}$,, and since x_0 is an interior point of $\mathcal{C}$ this implies $\mu = \underline{0}$. Hence $f(x) = f(x_0)$, $x \in \mathcal{C}$, which proves a).

Assume f has a local minimum at $x_0 \in \mathcal{C}$ and let $x \in \mathcal{C}$ be arbitrary. Then the convex function $g(t) = f(tx + (1-t)x_0)$, $t \in [0,1]$ has a local minimum at $t = 0$. Hence,

$$h(t) = \frac{g(t) - g(0)}{t} \geqslant 0 \ \text{ for } t > 0 \text{ small enough}\,.$$

But h is an increasing function on $[0,1]$ by Lemma 3.6 and hence $0 \leqslant h(1) = g(1) - g(0) = f(x) - f(x_0)$, i. e. $f(x_0) \leqslant f(x)$. This proves the first statement in b).

If f is differentiable at an interior point x_0 of $\mathcal{C}$ and $\nabla f(x_0) = \underline{0}$, then Theorem 3.13 implies that $\mu = 0$ in (3.10) and hence f has a minimum at x_0. The necessity of the condition $\nabla f(x_0) = \underline{0}$ for x_0 to be a (local) minimmum of f is well known for any function f differentiable at x_0. This proves the second part of b). □

One further important consequence of Theorem 3.13 is the following result.

Theorem 3.16. *A convex function $f : \mathcal{C} \to \mathbb{R}$ defined on an open convex set $\mathcal{C} \subseteq \mathbb{R}^k$ is continuous.*

Remark 3.17. If $\mathcal{C}$ is not an open set, then f can be discontinuous at the boundary of $\mathcal{C}$, even for $k = 1$. See Exercise 3.3 for a simple example.

Proof. We prove this first for convex functions of one variable. Let $x_- < x_0 < x_+$ be in the domain of f. Then by Lemma 3.6 we have (see Figure 3.4)

$$a_- \leqslant \frac{f(x) - f(x_0)}{x - x_0} \leqslant a_+\,, \quad \text{for } x_- < x < x_+\,,$$

where

$$a_- = \frac{f(x_-) - f(x_0)}{x_- - x_0}, \quad a_+ = \frac{f(x_+) - f(x_0)}{x_+ - x_0}\,.$$

Thus

$$|f(x) - f(x_0)| \leqslant \max\{|a_-|, |a_+|\}|x - x_0|\,, \quad \text{for } x_- < x < x_+\,,$$

which proves the continuity of f at x_0.

Alternatively, one could prove the continuity of $f(x)$ using the fact that the left and right derivatives $f'_-(x)$ and $f'_+(x)$ exist. This alternative is left to the reader as an exercise.

Now assume $k > 1$ and let Q_λ be a k-dimensional cube of side length $\lambda > 0$ centered at x_0. We choose λ small enough such that $Q_\lambda \subseteq \mathcal{C}$, which is possible since $\mathcal{C}$ is open.

By Corollary 3.14 we conclude that the restriction of f to Q_λ must take its maximal value at one of the 2^k corner points of Q_λ. If we let λ approach 0 the corners of Q_λ will trace out straight lines approaching x_0. From the one-dimensional case we know that f is continuous when restricted to each of these straight lines. Hence the values of f at the 2^k corners will approach $f(x_0)$ as λ tends to 0. We conclude that

$$\lim_{\lambda \to 0} \max_{x \in Q_\lambda} f(x) = f(x_0).$$

On the other hand, we know that f has a supporting hyperplane at x_0, i. e., there exists $\mu_0 \in \mathbb{R}^k$ such that

$$f(x) \geqslant f(x_0) + \mu_0 \cdot (x - x_0).$$

It follows that $\lim_{\lambda \to 0} \min_{x \in Q_\lambda} f(x) = f(x_0)$ and consequently

$$\lim_{\lambda \to 0} \max_{x \in Q_\lambda} |f(x) - f(x_0)| = 0\,.$$

Hence we have shown that f is continuous at x_0. □

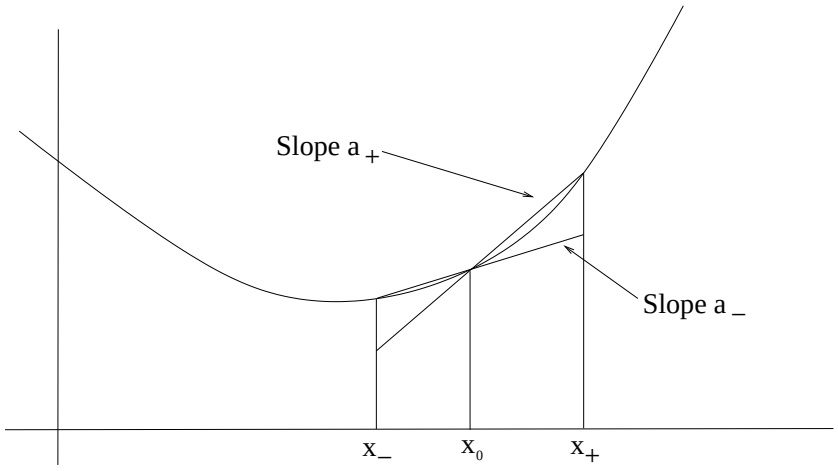

Figure 3.4: Continuity of convex function

3.2 Legendre transform

The Hamilton function of a mechanical system will be defined in Section 3.3 by applying a so-called Legendre transformation to the Lagrange function L w. r. t. the velocity variables. As discussed briefly in Section 3.5 this transformation also has important applications in thermodynamics. We here give a definition of the Legendre transform first introduced by W. Fenchel in 1949 and which does not assume the function to be neither convex nor differentiable. This form of the Legendre transform is also called the *Legendre-Fenchel transform.*

Definition 3.18. *Let $f : A \to \mathbb{R}$ be a function defined on a subset $A \subseteq \mathbb{R}^k$. The* **Legendre transform** *f^* of f is defined by*

$$f^*(p) = \sup\{x \cdot p - f(x) \mid x \in A\}$$

for all $p \in \mathbb{R}^k$ where this supremum on the right hand side is finite. We shall denote the domain of f^* by D_{f^*} and call p the conjugate variable *to x.*

Let us first note the following important fact.

Theorem 3.19. *The Legendre transform f^* of any function f of k variables is a convex function.*

Proof. This follows immediately from Lemma 3.5 b) by setting $\mathcal{I} = \mathcal{A}$ and $f_x(p) = p \cdot x - f(x)$, $p \in \mathbb{R}^k$, for each $x \in A$ and noting that f_x is an affine function, and hence also convex. □

It should be noted that D_{f^*} is non-empty if and only if f is bounded below by some affine function. The following result furnishes a key link between the Legendre transform and supporting hyperplanes.

Proposition 3.20. *Let $f : \mathcal{C} \to \mathbb{R}$ be a convex function of k variables and let $x_0 \in \mathcal{C}$. Then the graph of the function $h(x) = f(x_0) + p \cdot (x - x_0)$ is a supporting hyperplane for f at $x_0 \in \mathcal{C}$ if and only if $p \in D_{f^*}$ and $f(x_0) + f^*(p) = p \cdot x_0$. In particular, the conclusion holds for*

$$p = \nabla f(x_0)\,, \tag{3.11}$$

if x_0 is an interior point of $\mathcal{C}$ at which the partial derivatives of f exist.

Proof. By definition the graph of h is a supporting hyperplane for f if and only if

$$p \cdot (x - x_0) + f(x_0) \leqslant f(x), \quad x \in \mathcal{C}\,,$$

that is

$$p \cdot x - f(x) \leqslant p \cdot x_0 - f(x_0)\,, \quad x \in \mathcal{C}\,.$$

This is, however, equivalent to $p \in D_{f^*}$ and $f^*(p) = p \cdot x_0 - f(x_0)$. This proves the first claim. The second claim then follows from the last part of Theorem 3.13. □

In view of the latter statement of Proposition 3.20, one way to approach the Legendre transform for differentiable convex functions, frequently used in mechanics, is to define p by (3.11) and attempt to solve this equation for x in terms of p and then express $p\cdot x - f(x)$ in terms of p. In case the function f is quadratic, the equation to be solved is linear and can be dealt with by techniques known from linear algebra. However, in general this approach is cumbersome since it requires solving non-linear equations and discussing whether the solution is unique. The definition of f^* given above avoids these problems altogether.

Example 3.21. We may calculate the Legendre transform of the function $f(x) = x^a/|a|$, defined for $x > 0$ where $a \neq 0$, by maximizing $g_p(x) = xp - x^a/|a|$.

Consider first the case $a > 1$. If $p \leqslant 0$ the supremum of g_p is 0. If $p > 0$ the maximum occurs when $p = ax^{a-1}/|a| = x^{a-1}$, i. e. $x = p^{1/(a-1)}$. Thus

$$f^*(p) = p^{1/(a-1)}p - a^{-1}p^{a/(a-1)} = p^b/b\,,$$

where $b = \frac{a}{a-1}$. Note that $a^{-1} + b^{-1} = 1$ and $a, b > 1$.

Next consider $a < 0$. If $p > 0$ then the supremum of g_p is infinite. For $p = 0$ the supremum is 0. For $p < 0$ the maximum occurs if $p = ax^{a-1}/|a| = -x^{a-1}$, i.e., $x = |p|^{1/(a-1)}$. Thus for $a < 0$, f^* is defined for $p \leqslant 0$ and

$$f^*(p) = |p|^{1/(a-1)}p - |p|^{a/(a-1)}/|a| = (-1 - |a|^{-1})|p|^{a/(a-1)} = -|p|^b/b\,,$$

where again $b = \frac{a}{a-1} = \frac{|a|}{|a|+1}$.

Finally consider $0 < a < 1$. If $p > 0$ then the supremum of $g_p(x)$ is again infinite. If $p \leqslant 0$ then the supremum is 0. Thus f^* is defined for $p \leqslant 0$ and $f^*(p) = 0$. In this case the Legendre transform is not very useful.

Example 3.22. The Legendre transform of an affine function $f(x) = \mu \cdot x + b$ is defined only for $p = \mu$ and in this case $f^*(\mu) = -b$.

Proposition 3.23. *Let $f : A \to \mathbb{R}$ be a function defined on a set $A \subset \mathbb{R}^k$, such that $D_{f^*} \neq \emptyset$. Then $A \subseteq D_{f^{**}}$ and*

$$f(x) \geqslant f^{**}(x) \;\; \textit{for all } x \in A\,.$$

Moreover, if A is open and $g : \mathcal{C} \to \mathbb{R}$ is any convex function such that $A \subseteq \mathcal{C}$ and $g(x) \leqslant f(x)$ for all $x \in A$, then $f^{**}(x) \geqslant g(x)$ for all $x \in A$, i. e., f^{**} is the largest convex function that is smaller than or equal to f on A.

Proof. The first statement follows straight away from the fact that for all $x \in A$ and all $p \in D_{f^*}$ we have $x \cdot p - f^*(p) \leqslant f(x)$ by the definition of f^*.

Assume now that A is open and let $x_0 \in A$. Then x_0 is an interior point of C, and since g is convex there is a supporting hyperplane for g at x_0, i. e., there exists $\mu_0 \in \mathbb{R}^k$ such that

$$g(x) \geqslant \mu_0 \cdot (x - x_0) + g(x_0)$$

for all $x \in \mathcal{C}$. In particular, we have for all $x \in A$ that

$$\mu_0 \cdot x_0 - g(x_0) \geqslant \mu_0 \cdot x - g(x) \geqslant \mu_0 \cdot x - f(x)\,.$$

Thus

$$\mu_0 \cdot x_0 - g(x_0) \geqslant f^*(\mu_0)$$

and consequently

$$f^{**}(x_0) \geqslant \mu_0 \cdot x_0 - f^*(\mu_0) \geqslant g(x_0).$$

□

Corollary 3.24. *Let $f : \mathcal{C} \to \mathbb{R}$ be a convex function defined on an open convex set $\mathcal{C} \subseteq \mathbb{R}^k$. Then $f^{**}(x) = f(x)$ for all x in $\mathcal{C}$.*

Proof. Using $g = f$ in the previous proposition we get that $f(x) \leqslant f^{**}(x)$ for all $x \in \mathcal{C}$. Together with Proposition 3.23 this proves that $f = f^{**}$ on $\mathcal{C}$. □

Proposition 3.25. *If $f : \mathcal{C} \to \mathbb{R}$ is a convex function defined on the open convex set $\mathcal{C} \subseteq \mathbb{R}^k$ such that f^{**} is strictly convex, then f^* has partial derivatives at all interior points p of D_{f^*} and $x(p) = \nabla f^*(p)$ is the unique point in $D_{f^{**}}$ at which f^{**} has a supporting hyperplane with normal vector $(p, -1)$.*

In particular, if $f : \mathbb{R}^k \to \mathbb{R}$ is strictly convex and p is an interior point of D_{f^*}, then $\nabla f^*(p)$ exists and is the unique solution to the equation

$$\nabla f(x) = p\,. \tag{3.12}$$

Proof. Let p_0 be an interior point of D_{f^*}. According to Theorem 3.13 we must show that f^* has a unique supporting hyperplane at p_0.

By Corollary 3.20 we know that the graph of the function $k(p) = f^*(p_0) + x_0 \cdot (p - p_0)$ is a supporting hyperplane for f^* at p_0 if and only if $x_0 \in D_{f^{**}}$ and $f^{**}(x_0) = p_0 \cdot x_0 - f^*(p_0)$. This implies that the graph of the function $h(x) = f^{**}(x_0) + p_0 \cdot (x - x_0)$ is a supporting hyperplane for f^{**} at x_0. Assume now that x_0' is another point in $D_{f^{**}}$ such that the graph of the function $h'(x) = f^{**}(x_0') + p_0 \cdot (x - x_0')$ is a supporting hyperplane at x_0'. Then the inequalities

$$f^{**}(x) \geqslant f^{**}(x_0) + p_0 \cdot (x - x_0) \quad \text{and} \quad f^{**}(x) \geqslant f^{**}(x_0') + p_0 \cdot (x - x_0')$$

hold for all $x \in D_{f^{**}}$. By inserting $x = x_0'$ into the former and $x = x_0$ into the latter one obtains

$$f^{**}(x_0') - f^{**}(x_0) = p \cdot (x_0' - x_0)\,.$$

Using this we then get

$$f^{**}(\alpha x_0 + (1-\alpha)x_0') \geqslant f^{**}(x_0) + (1-\alpha)p \cdot (x_0' - x_0) = \alpha f^{**}(x_0) + (1-\alpha) f^{**}(x_0')$$

for $0 < \alpha < 1$. But this contradicts strict convexity of f^{**} and therefore x_0 is uniquely determined by p. This proves uniqueness of the supporting hyperplane for f^* at p_0 and Theorem 3.13 implies $x_0 = \nabla f^*(p_0)$. This proves the first part of the proposition.

In case $\mathcal{C} = \mathbb{R}^k$, we obviously have $f = f^{**}$ by Corollary 3.24. Hence we conclude as before that $\nabla f^*(p)$ exists, if p is an interior point of D_{f^*}, and that f has a supporting hyperplane with normal vector $(p, -1)$ at the unique point $x(p) = \nabla f^*(p)$. But Theorem 3.13 tells us that this is equivalent to $x(p)$ being a solution to (3.12). This completes the proof. □

3.3 Hamilton's equations

We turn to the application of the Legendre transform in mechanics and consider a mechanical system with Lagrange function $L(q, v, t)$, which will be assumed to be defined on a set $A \times B \times \mathbb{R}$, where A and B are open subsets of $\mathbb{R}^k$. Convexity of L as a function of v will be a key property in order to apply the results of the previous section, hence B will be assumed to be a convex set. In fact, in most applications we have $B = \mathbb{R}^k$. Thus, for the mechanical models considered in Chapter 2, whether constrained or unconstrained, it is easily seen by inspection that the Lagrange function can be written in the form

$$L(q, v, t) = \|A(q, t)v + b(q, t)\|^2 + h(q, t) \tag{3.13}$$

where A is a non-singular square matrix function, b a vector function, and h a scalar function of coordinates and time. It follows that in all these cases the Lagrange function is a strictly convex function of $v \in \mathbb{R}^k$ for fixed q and t (see Exercise 3.16).

Instances of Lagrange functions not of the form (3.13) occur in the Lagrangian fomulation of relativistic mechanics. In particular, a standard form of L for a free relativistic particle of mass $m > 0$ is

$$L(v) = -mc\sqrt{c^2 - \|v\|^2}\,, \tag{3.14}$$

where c is the velocity of light. This function is independent of position x such that $A = \mathbb{R}^3$, while $B = \{v \in \mathbb{R}^3 \mid \|v\| < c\}$ is an open ball on which L is convex. Systems with Lagrangians of this and more general type can also be handeled in the Hamiltonian formalism as developed below, but in order to avoid technical complications we shall henceforth only consider the case $B = \mathbb{R}^k$.

Definition 3.26. *Given a Lagrange function $L : A \times \mathbb{R}^k \times \mathbb{R} \to \mathbb{R}$ which is a convex fucntion of v for fixed q and t, the Legendre transform w. r. t. v for fixed q and t will be denoted by H and is called the* **Hamilton function** *or* **Hamiltonian** *corresponding to L, that is*

$$H(q, p, t) = \sup_{v \in B} (v \cdot p - L(q, v, t))\ . \tag{3.15}$$

It will be assumed that the interior C of the set of points p such that the right hand side of (3.15) *is finite, is independent of q and t, and hence we consider H as a function defined on the open set $A \times C \times \mathbb{R}$.*

If L is a strictly convex C^1 function of v, then Proposition 3.25 implies that we may write

$$H(q, p, t) = v(q, p, t) \cdot p - L(q, v(p, q, t), t), \tag{3.16}$$

where $v(q, p, t)$ is determined as the unique solution to the equations

$$p_i = \frac{\partial L}{\partial v_i}(q, v, t), \quad i = 1, \ldots, k\,, \tag{3.17}$$

so p_i is the generalized momentum corresponding to the coordinate q_i.

In particular, we have seen in Chapter 2 that a Lagrangian for a system of N particles interacting via a potential V in an inertial system of coordinates is given by

$$L(x, v, t) = \sum_{i=1}^{N} \frac{1}{2} m_i \|v_i\|^2 - V(x),$$

where $q_i, v_i \in \mathbb{R}^3$, for $i = 1, \ldots, N$, and we have $k = 3N$. In this case $p_i = \nabla_{v_i} L(q, v) = m_i v_i$ is the standard momentum and the Hamilton function

$$H(q,p) = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + V(q)$$

is seen to be identical to the *energy function*.

Our goal is now to rewrite the Lagrange equations of motion for mechanical system in terms of the Hamiltonian.

Theorem 3.27. *Let the Lagrangian $L(q, v, t)$ be a C^2-function on $A \times \mathbb{R}^k \times \mathbb{R}$, where $A \subseteq \mathbb{R}^k$ is open, and assume the Hesse martix of L w. r. t. v*

$$\left(\frac{\partial^2 L}{\partial v_i \partial v_j} \right) \tag{3.18}$$

is positive definite at all points of $A \times \mathbb{R} \times \mathbb{R}$. If $\gamma(t)$, $t \in I$, is a solution to the Euler-Lagrange equations corresponding to L, then the curve $(q(t), p(t))$, $t \in I$, where

$$q_i/t) = \gamma_i(t) \quad \textit{and} \quad p_i(t) = \frac{\partial L}{\partial v_i}(\gamma(t), \dot{\gamma}(t), t) \tag{3.19}$$

for $i = 1, \ldots, k$, is a solution to **Hamilton's equations**

$$\boxed{\dot{q}_i(t) = \frac{\partial H}{\partial p_i}(q(t), p(t), t)\,, \qquad \dot{p}_i(t) = -\frac{\partial H}{\partial q_i}(q(t), p(t), t)\,, \quad i = 1, \ldots, k\,.} \tag{3.20}$$

Conversely, given a solution $(q(t), p(t))$, $t \in I$, to Hamilton's equations, then $\gamma(t) = q(t)$ is a solution to the Euler-Lagrange equations.

Remark 3.28. We shall henceforth use the notation $\frac{\partial L}{\partial v}$ for the gradient of L w. r. t. v and similarly for the gradients of H w. r. t. q and p. With this notation Hamilton's equations can be written as

$$\dot{q} = \frac{\partial H}{\partial p}\,, \qquad \dot{p} = -\frac{\partial H}{\partial q}\,.$$

Proof. By Corollary 3.10 L is a strictly convex function of v. As already noted it follows from Proposition 3.25 that for any $(q, p, t) \in A \times C \times \mathbb{R}$ there exists a unique $v(q, p, t) \in \mathbb{R}^k$ such that

$$\frac{\partial L}{\partial v}(q, v(q, p, t), t) = p\,, \tag{3.21}$$

and that

$$v(q, p, t) = \frac{\partial H}{\partial p}(q, p, t)\,. \tag{3.22}$$

On the other hand, $\frac{\partial L}{\partial v}(q, v, t)$ belongs to the domain of the Legendre transform of $L(q, v, t)$ with respect to v for all $v \in \mathbb{R}^k$ as a consequence of Proposition 3.20. Since the Hesse matrix (3.18) is regular, it follows from the *inverse mapping theorem* that the image of the function $v \to \frac{\partial L}{\partial v}(q, v, t)$ is an open subset of $\mathbb{R}^k$, which hence must be equal to C. In other words, for fixed q and t the function $v \to \frac{\partial L}{\partial v}(q, v, t)$ maps $\mathbb{R}^k$ bijectively onto C and its inverse function is $p \to \frac{\partial H}{\partial p}(q, p, t)$.

Assume now that $\gamma : I \to A$ is a solution to the Euler-Lagrange equations. Defining $(q(t), p(t))$ by (3.19) it follows from the equivalence of equations (3.21) and (3.22) that $q(t)$ satisfies the first of Hamilton's equations. Moreover, the second of Hamiltons equations takes the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v}(\gamma(t), \dot{\gamma}(t), t)\right) = -\frac{\partial H}{\partial q}(q(t), p(t), t)\,, \tag{3.23}$$

and hence follows from the Euler-Lagrange equations if we prove that

$$\frac{\partial H}{\partial q}(q, p, t) = -\frac{\partial L}{\partial q}(q, v, t). \tag{3.24}$$

In order to see this we first note that since $\frac{\partial L}{\partial v}$ is a C^1 function the inverse mapping theorem implies that $v(q, p, t)$ as given by (3.22) is a C^1 function. Applying the chain rule we hence obtain from (3.16)

$$\begin{aligned}\frac{\partial H}{\partial q_i}(q, p, t) &= \sum_{j=1}^{k} \frac{\partial v_j}{\partial q_i}(q, p, t)\, p_j - \sum_{j=1}^{k} \frac{\partial L}{\partial v_j}(q, v(q, p, t), t)\frac{\partial v_j}{\partial q_i}(q, p, t) - \frac{\partial L}{\partial q_i}(q, v(q, p, t), t) \\ &= -\frac{\partial L}{\partial q_i}(q, v(q, p, t), t),\end{aligned}$$

where we have used (3.21). This proves the first part og the theorem.

Assume next that $(q, p) : I \to A \times C$ is a solution to Hamilton's equations and set $\gamma(t) = q(t)$. From the equivalence of (3.21) and (3.22) and the first of Hamilton's equations we then get that $p(t)$ is given by (3.19) and hence the second Hamilton's equation has the form (3.23), which as we have seen is equivalent to the Euler-Lagrange equations. This completes the proof. □

In the rest of this section we shall assume that the Hamiltonian $H = H(p, q)$ does not depend on time t, and that H is a C^2 function defined on the open set $A \times \mathbb{R}^k \subseteq \mathbb{R}^{2k}$.

For such Hamiltonians H, it follows from the existence and uniqueness result for first order differential equations, briefly discussed below Theorem 1.15, that there exists a unique solution $(p(t), q(t))$ to Hamilton's equations defined on a maximal time interval around $t = 0$, for each choice of initial conditions $q(0) = q_0 \in A$ and $p(0) = p_0 \in \mathbb{R}^d$. The following discussion is based on the additional assumption that the solution $(p(t), q(t))$ is *globally defined*, i. e., defined for all times $t \in \mathbb{R}$, for any choice of (q_0, p_0). This is not the case for arbitrary Hamiltonians H, but one can impose quite natural conditions on H which ensure this property of solutions. For instance, if the *level set*

$$\mathcal{H}_E = \{(p, q) : H(p, q) = E\}$$

is closed and bounded for all $E \in \mathbb{R}$, then all solutions are globally defined. Assuming this to be the case, we can in a similar way as in Section 1.5 define the *Hamiltonian flow* $\Psi_t : A \times \mathbb{R}^k \to A \times \mathbb{R}^k$, $t \in \mathbb{R}$, by

$$\Psi_t(q_0, p_0) = (q(t), p(t)),$$

where $(q(t), p(t))$ is the solution to Hamilton's equations such that $(q(0), p(0)) = (q_0, p_0)$. It is a fact, that if H is a C^2 function on $A \times \mathbb{R}^k$ then $\Psi_t(q, p)$ is a C^2 function of $(q_0, p_0, t) \in A \times \mathbb{R}^k \times \mathbb{R}$ (see e. g. [4]).

From the uniqueness of solutions to Hamilton's equations with given initial conditions at time $t = 0$, we conclude that the following *group property* holds:

$$\Psi_{t+s} = \Psi_t \circ \Psi_s \quad \text{for all } t, s \in \mathbb{R}.$$

Furthermore, it is clear that $\Psi_{t=0}$ is the identity map on $A \times \mathbb{R}^k$. In particular, these facts show that Ψ_t is a *bijective map* on $A \times \mathbb{R}^k$ for any $t \in \mathbb{R}$, and the inverse map of Ψ_t is $\Psi_t^{-1} = \Psi_{-t}$.

As in Chapter 1 we call the space of initial values $A \times \mathbb{R}^k$ the *phase space* of the mechanical system under consideration. As compared to the definition given in Chapter 1, the velocity variables v_i have been replaced by the generalized momenta p_i. In the description of a particle system w.ṙ. t. an inertial system of coordinates this only means a rescaling of variables, since in that case $p_i = m_i v_i$, as we have seen.

A key property of the Hamiltonian flow is that it preserves volume in phase space. This result is known as *Liouville's theorem* whose precise formulation is as follows.

Theorem 3.29 (Liouville's Theorem)**.** *Suppose that the Hamiltonian H, defined on the open set $A \times \mathbb{R}^k \subseteq \mathbb{R}^{2k}$, satisfies the assumptions stated above and let Ψ_t, $t \in \mathbb{R}$, be the corresponding Hamiltonian flow. Then*

$$\det D\Psi_t(q,p) = 1, \quad \textit{for all } t \in \mathbb{R},\ q \in A,\ p \in \mathbb{R}^k, \tag{3.25}$$

where $D\Psi_t$ is the $2k \times 2k$ Jacobi matrix of the map Ψ_t on $A \times \mathbb{R}^k$.

Proof. Since $\Psi_{t=0}$ is the identity map, the identity (3.25) obviously holds for $t = 0$. Hence, it is sufficient to show that

$$\frac{d}{dt} \det D\Psi_t = 0 \quad \text{for all } t \in \mathbb{R}. \tag{3.26}$$

Recall that $\Psi_{s+t} = \Psi_t \circ \Psi_s$ for all $s, t \in \mathbb{R}$. Applying the chain rule, we get

$$D\Psi_{s+t} = (D\Psi_t \circ \Psi_s) D\Psi_s$$

and hence

$$\det(D\Psi_{s+t}) = \det(D\Psi_s \circ \Psi_t) \det(D\Psi_t).$$

Differentiating this identity w. r. t. s gives

$$\frac{d}{ds} \det(D\Psi_{s+t}) = \frac{d}{ds} \det(D\Psi_s \circ \Psi_t) \det(D\Psi_t)\,,$$

where we observe that on the left hand side $\frac{d}{ds}$ can be replaced by $\frac{d}{dt}$. Using this and setting $s = 0$ we obtain

$$\frac{d}{dt} \det(D\Psi_t) = \frac{d}{ds} \det(D\Psi_s \circ \Psi_t)\Big|_{s=0} \det(D\Psi_t)\,,$$

from which it follows that it suffices to prove (3.26) for $t = 0$. For this purpose let us first note that, as a consequence of Exercise (3.20), we have

$$\frac{d}{dt} \det(D\Psi_t)_{|t=0} = \operatorname{Tr} \frac{d}{dt} D(\Psi_t)_{|t0},$$

where Tr denotes the trace of a matrix. Furthermore, since Ψ is C^2, we can interchange the order of differentiation, so

$$\frac{d}{dt} D\Psi_t = D\frac{d}{dt}\Psi_t.$$

Next, we find

$$\frac{d}{dt}(\Psi_t)_{|t=0} = (\dot{q}(0), \dot{p}(0)) = \left(\frac{\partial H}{\partial p_1}, \ldots, \frac{\partial H}{\partial p_k}, -\frac{\partial H}{\partial q_1}, \ldots, -\frac{\partial H}{\partial q_k}\right),$$

where the last equation follows from Hamilton's equations and the fact that Ψ_t is its associated flow. Hence

$$\left(D\frac{d}{dt}\Psi_t\right)_{t=0} = \begin{pmatrix} \frac{\partial^2 H}{\partial q_1 \partial p_1} & \cdots & \frac{\partial^2 H}{\partial q_k \partial p_1} & \frac{\partial^2 H}{\partial p_1^2} & \cdots & \frac{\partial^2 H}{\partial p_k \partial p_1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 H}{\partial q_1 \partial p_k} & \cdots & \frac{\partial^2 H}{\partial q_k \partial p_k} & \frac{\partial^2 H}{\partial p_1 \partial p_k} & \cdots & \frac{\partial^2 H}{\partial p_k^2} \\ -\frac{\partial^2 H}{\partial q_1^2} & \cdots & -\frac{\partial^2 H}{\partial q_k \partial q_1} & -\frac{\partial^2 H}{\partial p_1 \partial q_1} & \cdots & -\frac{\partial^2 H}{\partial p_k \partial q_1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 H}{\partial q_1 \partial q_k} & \cdots & -\frac{\partial^2 H}{\partial q_k^2} & -\frac{\partial^2 H}{\partial p_1 \partial q_k} & \cdots & -\frac{\partial^2 H}{\partial p_k \partial q_k} \end{pmatrix}$$

and therefore

$$\mathrm{Tr}\frac{d}{dt} D(\Psi_t)_{|t=0} = \sum_{i=1}^{k} \left(\frac{\partial^2 H}{\partial q_i \partial p_i} - \frac{\partial^2 H}{\partial p_i \partial q_i}\right) = 0.$$

This proves (3.26) for $t = 0$ and hence concludes the proof of the theorem. □

Let us explain why Liouville's theorem, in the way stated above, implies the *"preservation of volume under the Hamiltonian flow"* as previously mentioned. Let $\Sigma \subset A \times \mathbb{R}^k$ be a set of finite volume

$$\mu(\Sigma) = \int_\Sigma d^k q \, d^k p.$$

More concretely, the reader may think of an open bounded set. Under the motion of the system starting at time $t = 0$ in a state in Σ, the system ends up after time t in a state $q(t), p(t))$ in $\Psi_t(\Sigma)$. Expressed differently, under the Hamiltonian flow Ψ the subset Σ of phase space flows after time t to $\Psi_t(\Sigma)$. Using the standard transformation rule for integrals in Euclidean space and Theorem 3.29, we have

$$\mu(\Psi_t(\Sigma)) = \int_{\Psi_t(\Sigma)} d^k q \, d^k p = \int_\Sigma \underbrace{|\det D\Psi_t|}_{=1} d^k q \, d^k p = \mu(\Sigma),$$

which expresses the fact that the Hamiltonian flow preserves the volume in phase space.

An important application of Liouville's theorem is in establishing the so-called *Poincaré recurrence* phenomenon: Under reasonable assumptions, a mechanical system always returns (after sufficiently long time) arbitrarily close to its initial state. Of course, the resolution of the seemingly paradoxical statement lies in the fact that "sufficiently long" can actually mean time scales longer than age of the universe.

Here is a precise formulation of the mentioned recurrence theorem.

Theorem 3.30 (Poincaré recurrence theorem)**.** *Consider a mechanical system whose motion obeys Hamilton's equations with Hamiltonian obeying the conditions of Theorem 3.29. Suppose that $\Sigma \subset A \times \mathbb{R}^k$ is open and has finite volume $\mu(\Sigma) < \infty$, and assume that Σ is invariant under Ψ_t, i. e. $\Psi_t(\Sigma) \subseteq \Sigma$ for all $t \in \mathbb{R}$.*

Then the following is true: For any time $T > 0$ and any neighborhood $U \subset \Sigma$, there exists a time $t \geqslant T$ and $x \in U$ such that $\Psi_t(x) \in U$.

Proof. Let $n = 0, 1, 2, 3, \ldots$ be a natural number. For notational convenience, we define

$$g^n(U) = \Psi_{t=nT}(U).$$

Note that $g^0(U) = U$ holds, since Ψ_0 is the identity map.

Now, observe that the sets

$$U, g^1(U), g^2(U), g^3(U), \cdots, g^n(U) \cdots \tag{3.27}$$

all have equal volume, $\mu(g^n(U)) = \mu(U) > 0$ for all $n \in \mathbb{N}$, as a consequence of the volume preserving property of Ψ_t. Suppose that these sets never intersect, i. e., $g^k(U) \cap g^l(U) = \emptyset$ for all $k \neq l$. Then

$$\mu(\Sigma) \geqslant \mu\left(\bigcup_{n=0}^{\infty} g^n(U)\right) = \sum_{n=0}^{\infty} \mu(U) = +\infty,$$

where the inequality follows from the invariance of Σ and the first equality from the disjointness of the sets $g^n(U)$. This contradicts our assumption $\mu(\Sigma) < \infty$.

Hence there must exist $k > l \geqslant 0$ and $l \geqslant 0$ such that

$$g^k(U) \cap g^l(U) \neq \emptyset.$$

This implies

$$g^{k-l}(U) \cap U \neq \emptyset,$$

thanks to the fact that $g^{n+m}(U) = g^n(g^m(U))$. (The reader is urged to check this detail.)

Finally, let $y \in g^{k-l}(U) \cap U$ and put $n = k - l$. Then $y = g^n(x)$ for some $x \in U$. That is, we have $x \in U$ and $y = g^n(x) = \Psi_{t=nT}(x) \in U$. Hence, the desired conclusion holds for $t = nT$ where $n \geqslant 1$ is some natural number. □

Example 3.31. For a system with Hamilton function as in Theorem (3.30) the set

$$\Sigma_E = \{(q, p) \in A \times \mathbb{R}^k | H(p, q) < E\}$$

is open since H is continuous. It is also invariant under the Hamiltonian flow as an immediate consequence of the theorem on conservation of energy proven below.

In many cases Σ_E is also a bounded set. Just to take one specific example, consider the k-dimensional harmonic oscillator whose Hamilton function defined on $\mathbb{R}^{2k}$ and given by

$$H(q, p) = \frac{\|p\|^2}{2m} + \frac{1}{2} k \|q\|^2,$$

where m and k are positive constants. In this case, Σ_E is empty for $E \leqslant 0$ while $(q, p) \in \Sigma_E$ for $E > 0$ implies $\|q\| < \sqrt{2mE}$ and $\|p\| < \sqrt{\frac{2E}{k}}$, and hence Σ_E is bounded. Now let $q_0, p_0)$

be any initial condition and choose $E > H(q_0, p_0)$. Given any neighborhood U of (q_0, p_0) no matter how small there exists by Theorem 3.30 for any $T > 0$, no matter how large, an x in $U \cap \Sigma_E$ and a $t \geqslant T$ such that $\psi_t(x) \in U \cap \Sigma_E$.

The reader may easily extend these arguments to systems consisting of a single particle in $\mathbb{R}^k$ of mass $m > 0$ in a potential $V(q)$ differentiable on $\mathbb{R}^k$ such that $V(q) \to \infty$ as $\|q\| \to \infty$.

As the final result of this section we next prove the announced extension of Theorem 2.5 on energy conservation.

Theorem 3.32 (Conservation of energy)**.** *If $H(q,p)$ is a C^1-function independent of time and $(q(t), p(t))$ is a solution to the corresponding Hamilton's equations, then*

$$\frac{d}{dt} H(q(t), p(t)) = 0.$$

Proof. By the chain rule we calculate

$$\begin{aligned} \frac{d}{dt} H(q(t), p(t)) &= \sum_{i=1}^{k} \frac{\partial H}{\partial q_i}(q(t), p(t))\dot{q}_i(t) + \frac{\partial H}{\partial p_i}(q(t), p(t))\dot{p}_i(t) \\ &= \sum_{i=1}^{k} \frac{\partial H}{\partial q_i}(q(t), p(t))\frac{\partial H}{\partial p_i}(q(t), p(t)) - \frac{\partial H}{\partial p_i}(q(t), p(t))\frac{\partial H}{\partial q_i}(q(t), p(t)) = 0. \end{aligned}$$

□

3.4 Noether's Theorem

The final topic to be discussed in the classical mechanics part of these notes is the notion of symmetries and conservation laws.

We consider a system described by a Lagrange function $L(q, v)$ independent of time and defined on $A \times \mathbb{R}^k$, where $A \subseteq \mathbb{R}^k$ is open. In this section a coordinate transformation $\psi : A \to A$ will simply be called a *transformation*. Recalling Definition 2.12 of coordinate transformations, this means that ψ is a transformation if it is a bijective C^1 mapping from A onto A and fulfills

$$\det D\psi(q) \neq 0,$$

for all $q \in A$. A *one-parameter family of transformations* is a collection of transformations ψ_s indexed by a parameter $s \in]-a, a[$, where $a > 0$. It is called a C^2 family if the function $(q, s) \to \psi_s(q)$ from $A \times]-a, a[$ to A is C^2. In this case we will denote the derivative with respect to s by $\psi_s'(q)$, whereas the Jacobian with respect to q is denoted by $D\psi_s(q)$. In view of the transformation rule (2.11) for Lagrange functions under coordinate transformations it is natural to define a symmetry of a mechanical system as follows.

Definition 3.33. *A transformation ψ is a* **symmetry** *of a mechanical system described by the Lagrange function L, and L is called* **invariant** *under ψ, if*

$$L(\psi(q), D\psi(q)v) = L(q, v) \quad \textit{for all } q \in A,\ v \in \mathbb{R}^k.$$

Example 3.34.

a) Let $A \in O(3)$ be a 3×3 orthogonal matrix. Then the linear transformation $\psi(q) = Aq$ is a symmetry of a system described by a Lagrangian function of the form

$$L(q, v) = \frac{1}{2}mv^2 - V(\|q\|),$$

where $m > 0$ and $V : [0, \infty) \to \mathbb{R}$ is a C^1 function. (See Exercise 3.21.)

b) The one-parameter family of transformations on $\mathbb{R}^k$ given by $\psi_s(q) = q + sv_o$, $s \in \mathbb{R}$, is a one-parameter family of symmetries of the free action $\mathcal{S}_0(\gamma) = \frac{1}{2}\int_{t_1}^{t_2} \|\dot\gamma(t)\|^2 dt$ for any fixed $v_0 \in \mathbb{R}^k$ (see Exercise 3.22). Since ψ_s is a translation in the direction of v_0 this expresses translation invariance of $\mathcal{S}_0$.

Theorem 3.35 (Noether's Theorem)**.** *If $\psi_s : A \to A$, $-a < s < a$, is a C^2 family of symmetries of a system described by a C^2 Lagrange function $L : A \times \mathbb{R}^k \to \mathbb{R}$, then the function*

$$I(q, v) = \nabla_v L(q, v) \cdot \psi_0'(q)$$

is an integral of the motion, or a conserved quantity.

This means that if $\gamma(t)$ is a solution to the Euler-Lagrange equations for L, then

$$\frac{d}{dt} I(\gamma(t), \dot\gamma(t)) = 0.$$

Proof. Using the Euler-Lagrange equations and the chain rule we calculate

$$\begin{aligned} \frac{d}{dt} I(\gamma(t), \dot\gamma(t)) &= \frac{d}{dt}\big(\nabla_v L(\gamma(t), \dot\gamma(t))\big)\psi_0'(\gamma(t)) + \nabla_v L(\gamma(t), \dot\gamma(t)) D\psi_0'(\gamma(t))\dot\gamma(t) \\ &= \nabla_q L(\gamma(t), \dot\gamma(t))\psi_0'(\gamma(t)) + \nabla_v L(\gamma(t), \dot\gamma(t)) D\psi_0'(\gamma(t))\dot\gamma(t)\,. \end{aligned} \tag{3.28}$$

On the other hand, differentiating both sides of the identity

$$L(\psi_s(q), D\psi_s(q)v) = L(q, v)\,,$$

which holds for all $s \in (-a, a)$ since ψ_s is a symmetry, yields

$$0 = \frac{d}{ds} L(\psi_s(q), D\psi_s(q)v)_{s=0} = \nabla_q L(q, v)\psi_0'(q) + \nabla_v L(q, v) D\psi_0'(q)v\,.$$

Inserting $(\gamma(t), \dot\gamma(t))$ for (q, v) in this identity it follows that the expression in the last line of (3.28) vanishes. This completes the proof. □

Example 3.36. Consider a mecanical system whose motion fulfills the Euler-Lagrange equations corresponding to a C^2 Lagrange function $L(q, v)$ defined on the open set $A \times \mathbb{R}^k \subseteq \mathbb{R}^{2k}$, and assume L is invariant under translations $\psi_(q) = q + sv_0$ (see Example 3.34 b)). Using that $\psi_0'(q) = v_0$ the integra of motion is

$$I(q, v) = \nabla_v L(q, v) \cdot v_0\,,$$

which is the component of the generalized momentum along v_0. If v_0 can be chosen arbitrarily in some subspace of $\mathbb{R}^k$ this means that component of the generalized momentum $p(t)$ along this subspace is constant.

3.5 Legendre transform in thermodynamics

The purpose of this section is to indicate briefly the use of the Legendre transform in thermodynamics and statistical mechanics.

It is an important property of thermodynamic functions that they are either convex or concave. As an example, the entropy $S(U,V)$ of a thermodynamic system is a monotone increasing concave function of total energy U at fixed volume V. The inverse function $U(S,V)$ which gives the total energy as a function of S and V is thus a convex function of S (see Exercise 3.6). The negative of the Legendre transform of this function w. r. t. S is called the *free energy*

$$F(T,V) = -U^*(T,V) = -\sup_S(TS - U(S,V)) = \inf_S(U(S,V) - TS).$$

The conjugate variable to S is called T, since it is indeed the temperature of the system. The free energy is the amount of work that the system can perform in a thermodynamic process at constant temperature. Not all the total energy U is available.

At the critical temperature of a phase transition the free energy may not be differentiable. At a boiling point of a liquid, for example, the temperature does not change while the liquid turns into vapor. Since, as we have seen in Example 1.26, the temperature is the derivative of U w. r. t. S at constant V the entropy increases linearly with the total energy during the phase transition. This is again reflected in a jump in the derivative of the free energy, see Figure 3.5.

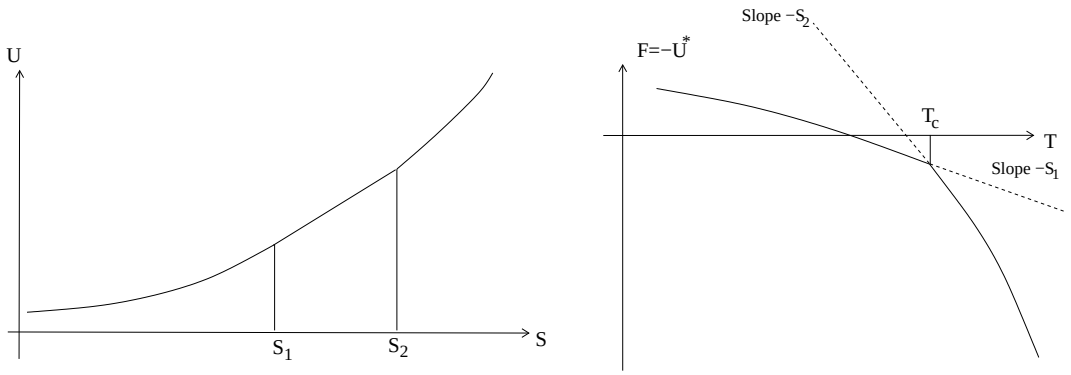

Figure 3.5: T_c is a critical temperature of a phase transition

The fact that the temperature and entropy are conjugate variables and that the free energy and the total energy are related by a Legendre transformation is an expression of what is called *equivalence of ensembles*.

We illustrate this for the ideal gas. In Section 1.5 we discussed the state defined by the Maxwell-Boltzmann probability distribution. Being a probabilistic state, it should be thought of as describing an ensemble of systems: observing one system at random from the ensemble the probability of finding the particles in that system with certain positions and momenta is determined by the Maxwell-Boltzmann distribution. The corresponding ensemble is referred to as the *canonical ensemble*, and we recall that in the Maxwell-Boltzmann distribution for N particles the states with momenta $p_1, \ldots, p_N$ have relative

weight

$$\exp\left(-\frac{\sum_{i=1}^{N}\frac{1}{2}\|p_i\|^2}{k_{\mathrm{B}}T}\right)$$

irrespective of the particle positions inside some finite volume V, assuming here that all particles have mass 1.

A different emsemble, called the *microcanonical ensemble* is defined by attributing equal weight to all states with total energy $\sum_{i=1}^{N}\frac{1}{2}\|p_i\|^2$ less than some U. Here U is as a free variable characterizing the distribution, whereas in the canonical ensemble it is the temperature that occurs as a free variable in the distribution of states.

Although different, the two ensembles provide equivalent descriptions of the ideal gas in the limit $N \to \infty$ and they are related by a Legendre transformation.

To understand this we must explain how the entropy and free energy are defined in terms of the distributions. We introduce the *partition functions* for the two ensembles as the normalization constants for the weights defining them, that is

$$Z_{\mathrm{can}}(T,V,N) = (N!)^{-1}V^N \int \exp\left(-\frac{\sum_{i=1}^{N}\frac{1}{2}\|p_i\|^2}{k_{\mathrm{B}}T}\right) d^{3N}p$$

$$Z_{\text{micro-can}}(U,V,N) = (N!)^{-1}V^N \int_{\sum_{i=1}^{N}\frac{1}{2}\|p_i\|^2<U} 1\, d^{3N}p\,,$$

where the factor $(N!)^{-1}$ comes from treating the particles as indistinguishable. The free energy and entropy are given by

$$F_N(T,V) = -k_{\mathrm{B}}T \ln Z_{\mathrm{can}}(T,V,N), \qquad S_N(U,V) = k_B \ln Z_{\text{micro-can}}(U,V,N). \tag{3.29}$$

These two functions are not directly related by a Legendre transformation for finite values of N. But the leading order term of $F_N(T,V)$ as $N \to \infty$ equals the Legendre transform of the corresponding leading order term of U w. r. t. S for fixed V as demonstrated in Exercise 3.27. This is a consequence of the probabilities in both cases concentrating on states with a fixed total energy. Moreover the leading order terms of the entropy and free energy coincide with the standard expressions as quoted in Exercise 3.26 (see also Exercise 1.22).

Equivalence of the micro-canonical and canonical ensembles for more general systems than the ideal gas can also be demonstrated, but this issue goes beyond the scope of these notes.

Exercises

Exercise 3.1. Show that the intersection of two convex subsets of $\mathbb{R}^k$ is convex.

Exercise 3.2. Determine which of the following functions of one variable are convex on their natural domain of definition:

$$x^2, \quad \frac{1}{1+x^2}, \quad \exp(x), \quad \ln(x), \quad \sqrt{1+x^2}, \quad |x-2|$$

Exercise 3.3. Show that the function

$$f(x) = \begin{cases} 0, & 0 \leqslant x < 1 \\ 1, & x = 1 \end{cases}$$

defined on $[0, 1]$ is convex and discontinuous.

Exercise 3.4. Let $a \in \mathbb{R}$, $b \in \mathbb{R}^k$, and let A be a positive semidefinite matrix. Show that the function

$$f(x) = a + \langle b, x\rangle + \frac{1}{2}\langle x, Ax\rangle$$

is convex on $\mathbb{R}^k$, and that it is strictly convex if A is positive definite.

Exercise 3.5. Show that a symmetric $k \times k$-matrix A is positive semi-definite, resp. positive definite, if and only if all its eigenvalues are non-negative, resp. positive.
Hint. Show first, that if B is a regular $k \times k$-matrix, then A is positive (semi-)definite if and only if the same holds for BAB^{-1}.

Exercise 3.6. Show that if $f : I \to \mathbb{R}$ is a monotone, strictly increasing, and concave function defined on an interval $I \subseteq \mathbb{R}$, then $f(I)$ is an interval and the inverse function $f^{-1} : f(I) \to \mathbb{R}$ is convex.

Exercise 3.7. Let $f : \mathcal{C} \to \mathbb{R}$ be a convex function of k variables, let $\alpha_1, \ldots, \alpha_m \geqslant 0$ be non-negative real numbers such that $\sum_{i=1}^m \alpha_i = 1$, and let $x_1, \ldots, x_m \in \mathcal{C}$. Show that **Jensen's inequality**

$$f\left(\sum_{i=1}^m \alpha_i x_i\right) \leqslant \sum_{i=1}^m \alpha_i f(x_i) \tag{3.30}$$

holds.
Hint. Note that for $m = 2$ Jensen's inequality coincides with (3.1). Show first for general m that the point

$$x_0 = \alpha_1 x_1 + \ldots + \alpha_m x_m \tag{3.31}$$

belongs to $\mathcal{C}$. This can be done by induction: For $m = 3$ one can rewrite

$$x_0 = \alpha_1 x_2 + (1 - \alpha_1)\left(\frac{\alpha_2}{1 - \alpha_1} x_2 + \frac{\alpha_3}{1 - \alpha_1} x_3\right),$$

assuming that $\alpha_1 \neq 1$. Noting that the coefficients of the linear combination in parenthesis are non-negative with sum 1, one concludes that the expression in parenthesis is in $\mathcal{C}$ and hence that $x_0 \in \mathcal{C}$. A similar rewriting for $m > 3$ combined with induction proves that $x_0 \in \mathcal{C}$ in general, and can also be used to prove (3.30).

We note that a linear combination as in (3.31), with non-negative coefficients adding up to 1, is called a *convex combination*. Geometrically, the convex combination x_0 is the average of the points $x_1, \ldots, x_m$ with corresponding weights $\alpha_1, \ldots, \alpha_m$.

Exercise 3.8. Use Jensen's inequality to show that the the arithmetic mean is bigger than or equal to the geometric mean:

$$\frac{x_1 + \ldots + x_m}{m} \geqslant (x_1 \cdots x_m)^{1/m} \quad \text{for } x_1, \ldots, x_m > 0\,.$$

Hint. Use that the exponential function e^x is convex.

Exercise 3.9. This exercise serves to show that a convex function $f : \mathcal{C} \to \mathbb{R}$ defined on an open convex set $\mathcal{C} \subset \mathbb{R}^k$ has supporting hyperplanes at all points in $\mathcal{C}$, even when the function does not have partial derivatives.

a) Show that for all $x_0 \in \mathcal{C}$ and $v \in \mathbb{R}^k$

$$\mu_{x_0}(v) = \lim_{t \to 0+} t^{-1}(f(x_0 + tv) - f(x_0)) = \inf_{t>0} t^{-1}(f(x_0 + tv) - f(x_0))$$

exists. *Hint.* Use the argument proving the existence of $\mu_\pm$ in Theorem 3.13.

b) Show that μ_{x_0} from a) satisfies

$$\mu_{x_0}(sv) = s\mu_{x_0}(v) \qquad \text{and} \quad \mu_{x_0}(v + w) \leqslant \mu_{x_0}(v) + \mu_{x_0}(w)$$

for all $s > 0$ and $v, w \in \mathbb{R}^k$. *Hint.* For the last inequality use that f is convex.

c) Show that the graph of $h(x) = f(x_0) + p \cdot (x - x_0)$ is a supporting hyperplane for f at x_0 if and only if

$$\mu_{x_0}(v) \geqslant p \cdot v$$

for all $v \in \mathbb{R}^k$.

We shall use induction on the dimension k to show that f has a supporting hyperplane at all points. The case $k = 1$ was treated in the proof of Theorem 3.13. We assume the result holds in $k - 1$ dimensions. We want to prove it in dimension k. Thus we have $p' \in \mathbb{R}^{k-1}$ such that for all $v' \in \mathbb{R}^{k-1}$ we have $\mu_{x_0}(v') \geqslant p' \cdot v'$. Let e be a unit vector in the k-th coordinate direction.

d) Use this induction hypothesis and the second result from question b) to show that for all $v', w' \in \mathbb{R}^{k-1}$

$$p' \cdot v' - \mu_{x_0}(v' - e) \leqslant -p' \cdot w' + \mu_{x_0}(w' + e).$$

Use this together with the first property in (b) to show that we can choose $p_k \in \mathbb{R}$ such that

$$\sup_{v' \in \mathbb{R}^{k-1}} \left(p' \cdot v' - \mu_{x_0}(v' - se)\right) \leqslant sp_k \leqslant \inf_{w' \in \mathbb{R}^{k-1}} \left(-p' \cdot w' + \mu_{x_0}(w' + se)\right),$$

for all $s > 0$.

e) Use the result of (d) to conclude that

$$(p' + p_k e) \cdot (v' + v_k e) \leqslant \mu_{x_0}(v' + v_k e)$$

for all $v' \in \mathbb{R}^{k-1}$ and all $v_k \in \mathbb{R}$. Conclude that the graph of $f(x_0) + p \cdot (x - x_0)$ is a supporting hyperplane for f at x_0 if $p = p' + p_k e$.

f) Use the construction of the supporting hyperplane in (e) to show that the hyperplane is unique if and only if f has partial derivatives at x_0.

Exercise 3.10. Show that the function

$$f(x) = a + |x - b|\,, \quad x \in \mathbb{R}\,,$$

is convex and find its Legendre transform. Here a and b are arbitrary real constants.

Exercise 3.11. Show that the function

$$f(x) = (|x| + 1)^2$$

defined on $\mathbb{R}$ is convex and determine its Legendre transform.

Exercise 3.12. Let $f : D \to \mathbb{R}$ be a convex function of k variables and denote its Legendre transform by $f^* : E \to \mathbb{R}$. Moreover, let $a \in \mathbb{R}$ and $b \in \mathbb{R}^k$ be fixed and denote by D' and E' the translated sets $D' = \{p + b \,|\, p \in D\}$, $E' = \{p + b \,|\, p \in E\}$.

a) Show that the Legendre transform of the function $f_1 : D \to \mathbb{R}$ defined by

$$f_1(x) = a + b \cdot x + f(x)\,, \quad x \in D$$

has domain of definition E' and is given by

$$f_1^*(p) = f^*(p - b) - a\,, \quad p \in E'\,.$$

b) Show that the Legendre transform of the function $f_2 : D' \to \mathbb{R}$ defined by

$$f_2(x) = f(x - b)\,, \quad x \in D'$$

has domain of definition E and is given by

$$f_2^*(p) = b \cdot p + f^*(p)\,, \quad p \in E\,.$$

Exercise 3.13. Determine the Legendre transform $\exp^*$ of the exponential function exp on $\mathbb{R}$.

Exercise 3.14. Let $f : \mathbb{R}^n \to \mathbb{R}$ be the quadratic function defined by

$$f(x) = \frac{1}{2} x^t A x\,, \quad x \in \mathbb{R}^n\,,$$

where A is a symmetric, positive definite matrix.

Show that the Legendre transform f^* of f is given by

$$f^*(p) = \frac{1}{2} p^t A^{-1} p\,, \quad p \in \mathbb{R}^n\,,$$

either by using that f is a convex C^2 function or by rewriting $x \cdot p - f(x)$ as the difference of two quadratic expressions.

Exercise 3.15. Show that if f is any function and x is in the domain of f and p is in the domain of its Legendre transform then the inequality

$$x \cdot p \leqslant f(x) + f^*(p)$$

holds. It is sometimes called Young's inequality.

As an application, prove that for all $x, y \geqslant 0$ and all $a, b > 1$ with $a^{-1} + b^{-1} = 1$ we have

$$xy \leqslant \frac{x^a}{a} + \frac{y^b}{b}\,.$$

Exercise 3.16. Show that if A is a non-singular $k \times k$-matrix and $b \in \mathbb{R}^k$, then the function $f(x) = \|Ax + b\|^2$ is strictly convex on $\mathbb{R}^k$. Calculate its Legendre transform.

Exercise 3.17. Consider the gravitational potential of two masses $m_1, m_2 > 0$ placed at the points $q_1, q_2 \in \mathbb{R}^3$:

$$V(q_1, q_2) = -Gm_1m_2\|q_1 - q_2\|^{-1}.$$

a) Write down the Lagrangian and the Hamiltonian for this two-body problem.

b) Write down Hamilton's equations for the system.

Exercise 3.18. Let $A : \mathbb{R}^3 \to \mathbb{R}^3$ be a C^2 vector field. We consider the magnetic field $B : \mathbb{R}^3 \to \mathbb{R}^3$ that has A as its vector potential, that is $B = \nabla \times A$. The Lagrangian for a particle of mass $m > 0$ and charge e moving in this magnetic field is

$$L(q, v) = \frac{1}{2}mv^2 + eA(q) \cdot v.$$

Find the corresponding Hamiltonian and write down Hamilton's equations.

Exercise 3.19. Consider two particles of masses $m_1, m_2 > 0$ at positions $q_1, q_2 \in \mathbb{R}^3$ and with velocities $v_1, v_2 \in \mathbb{R}^3$, whose motion is determined by the Euler-Lagrange equations for the Lagrangian

$$L(q_1, q_2, v_1, v_2) = \frac{1}{2}m_1v_1^2 + \frac{1}{2}m_2v_2^2 - V(q_2 - q_1),$$

where V is a C^2 function on $\mathbb{R}^3$.

a) Argue that the *center of mass*

$$Q = \frac{m_1q_1 + m_2q_2}{m_1 + m_2}$$

is a point between q_1 and q_2.

b) Define the coordinate transformation ψ with inverse

$$\psi^{-1}(q_1, q_2) = (Q, q_2 - q_1)$$

Determine ψ.

c) Find the Lagrangian in the new coordinates (Q, q) where $q = q_2 - q_1$.

d) Find the Hamiltonian in the new coordinates.

Exercise 3.20. Let $A(t)$ be a $k \times k$-matrix whose entries are C^1 functions of time t. Show that if $A(0) = I$, then

$$\left(\frac{d}{dt} \det A\right)(0) = \mathrm{Tr} A'(0).$$

Exercise 3.21. Prove that the linear transformation in Example 3.34 a) is really a symmetry of the given system.

Exercise 3.22. Verify the claim in Example 3.34 b).

Exercise 3.23. Consider a 2-dimensional system with Lagrangian

$$L(q, v) = \frac{1}{2} m v^2 - V(\|q\|)$$

where $m > 0$ and $V : [0, \infty) \to \mathbb{R}$ is a C^2 function.

a) Show that

$$\psi_s(x, y) = (\cos(s)x - \sin(s)y, \sin(s)x + \cos(s)y), \quad x, y, s \in \mathbb{R}$$

defines a C^2 family of symmetries of the system.

b) Determine the integral of the motion I corresponding to this continuous symmetry according to Noether's Theorem.

Exercise 3.24. Consider a single particle in a conservative force field whose corresponding potential $V(x)$ is rotationally invariant, i. e., it depends on $\|x\|$ only.

a) Write down the Lagrange function in spherical coordinates. (One may use Exercise 2.18 for this purpose.)

b) Determine the Hamilton function in spherical coordinates, find the cyclic coordinates and the corresponding conserved generalized momenta.

Exercise 3.25. Consider 3 particles with masses $m_1, m_2, m_3 > 0$ at positions $q_1, q_2, q_3 \in \mathbb{R}^3$ in an inertial system and interacting through a conservative force given by the potential

$$(q_1, q_2, q_3) \mapsto V(q_1 - q_2) + V(q_1 - q_3) + V(q_2 - q_3),$$

where $V : \mathbb{R}^3 \to \mathbb{R}$ is a C^2-function.

a) Write down the Lagrange function and show that $\psi : \mathbb{R}^9 \times \mathbb{R} \to \mathbb{R}^9$ given by

$$\psi_s(q_1, q_2, q_3) = (q_1 + (s, 0, 0), q_2 + (s, 0, 0), q_3 + (s, 0, 0))$$

for $s \in \mathbb{R}$ defines a C^2 family of symmetries of the problem.

b) Determine the integral of the motion given by Noether's Theorem.

Exercise 3.26. The entropy of an ideal gas of N particles, total energy $U > 0$, and volume $V > 0$, is

$$S(U,V) = Nk_{\mathrm{B}} \ln\left(C_0(V/N)(U/N)^{3/2}\right),$$

where k_{B} is Boltzmann's constant and $C_0 > 0$ is some constant. We assume N is kept fixed in this problem.

a) Determine the temperature $T(U)$ and pressure $P(U,V)$ such that

$$T(U)dS = dU + P(U,V)dV.$$

b) Determine the inverse function $U(S,V)$ of $U \mapsto S(U,V)$ and show that $U(S,V)$ is monotone increasing and convex as a function of S.

c) Determine the free energy $F(T,V)$ of the ideal gas defined as the Legendre transform of $S \mapsto U(S,V)$.

Exercise 3.27. This exercise serves to illustrate the equivalence of ensembles for the ideal gas.

a) If $f : [0,\infty) \to \mathbb{R}$ is a continuous function then using polar coordinates in all dimensions we obtain the formula

$$\int_{\mathbb{R}^n} f(\|x\|)d^n x = \omega_n \int_0^\infty f(r)r^{n-1}dr$$

for some constant ω_n. Using that

$$\int_{\mathbb{R}^n} e^{-\|x\|^2} d^n x = \left(\int_{-\infty}^{\infty} e^{-t^2} dt\right)^n = \pi^{n/2}$$

show that $\omega_n = 2\pi^{n/2}\Gamma(n/2)^{-1}$, where the Gamma function is given by $\Gamma(u) = \int_0^\infty e^{-u}u^{n-1}$. Check the formula in the case $n = 2$. Recall that $\Gamma(n) = (n-1)!$ for integer n.

b) Use the result of the previous question to calculate the free energy and entropy in (3.29).

c) Using the approximation in Stirling's formula $\lim_{u\to\infty} u^{-1}(\Gamma(u) - (u\ln(u) - u)) = 1$ we replace $\Gamma(u)$ by $u\ln(u) - u$ in the formula for the entropy above (also replace $\ln N!$ by $N\ln N - N$). Show that in this approximation the formula agrees with what was found in Exercise 1.20.

d) Show that in the large particle number approximation used in the previous question the Legendre transform of the total energy $U(S,V)$ as a function of entropy S is $-F(T,V)$ (minus the free energy) as a function of temperature T.

e) (Difficult) Show that this is a consequence of the probabilities in both cases concentrating on the set of states where the total energy is exactly U for the micro-canonical ensemble and $\frac{3}{2}Nk_{\mathrm{B}}T$ for the canonical ensemble.

Chapter 4

Hilbert Spaces

In the present chapter and the next the mathematical apparatus necessary for a proper discussion of quantum mechanics in Chapter 6 will be developed. The notion of Hilbert space turns out to be the right concept encompassing the structural properties of the state spaces of quantum mechanical systems, of which the so-called superposition principle may be the most prominent feature. Our goal in this chapter is to introduce the fundamental tools of Hilbert space theory. Although some of the examples and exercises are motivated by physical applications later on, the content is mainly mathematical and can be viewed as a further developement of the theory of inner product spaces known from linear algebra to the case of infinite dimensional vector spaces.

4.1 Inner product spaces

In the following both complex and real vector spaces will be treated. With $\mathbb{L}$ denoting either $\mathbb{R}$ or $\mathbb{C}$ let us recall that a vector space over $\mathbb{L}$ is a set E equipped with addition and scalar multiplication, which are maps $(x, y) \to x + y$ from $E \times E$ to E and $(\lambda, x) \to \lambda x$ from $\mathbb{L} \times E$ to E, respectively, satisfying the well known vector space axioms.

Examples of real vector spaces are known from linear algebra: they include $\mathbb{R}^k$ and the space $\mathcal{F}(M, \mathbb{R})$ of real functions defined on a set M with pointwise addition and multiplication by real numbers. Similarly, $\mathbb{C}^k$ and the set $\mathcal{F}(M, \mathbb{C})$ of complex valued functions on a set M with pointwise addition and multiplication by complex numbers are complex vector spaces. If M is a subset of $\mathbb{R}^k$ then the set $C(M, \mathbb{C})$ of continuous functions is a subspace of $\mathcal{F}(M, \mathbb{C})$, since the of two continuous functions is a continuous function and multiplication of a continuous function by a number is also a continuous function. Other interesting examples of subspaces of $\mathcal{F}(M, \mathbb{C})$ will be encountered in the following.

We turn to the definition of inner product spaces.

Definition 4.1. Let E be a vector space over $\mathbb{L}$ ($= \mathbb{R}$ or $\mathbb{C}$). An **inner product**, also called a **scalar product**, on E is a map $\langle \cdot \mid \cdot \rangle : E \times E \to \mathbb{L}$, satisfying the following conditions (where $\underline{0}$ denotes the null-vector in E):

i) $\langle\, x|x\,\rangle > 0$ for all $x \in E \setminus \{\underline{0}\}$

ii) $\langle\, x|y\,\rangle = \overline{\langle\, y|x\,\rangle}$,for all $x, y \in E$

iii) $\langle\, x+y|z\,\rangle = \langle\, x|z\,\rangle + \langle\, y|z\,\rangle$, for all $x, y, z \in E$

iv) $\langle\, x|\lambda y\,\rangle = \lambda\langle\, x|y\,\rangle$ for all $\lambda \in \mathbb{L}$ and $x, y \in E$

If $\langle\, \cdot \mid \cdot\,\rangle$ is an inner product on E the pair $(E, \langle\, \cdot \mid \cdot\,\rangle)$ is called an **inner product space**.

Note that $\langle\, x|x\,\rangle > 0$ in i) means that $\langle\, x|x\,\rangle$ is a real positive number. In the case $\mathbb{L} = \mathbb{R}$, complex conjugation in ii) is of course superfluous.

The last two conditions above express that the map $y \to \langle\, x|y\,\rangle$ from E into $\mathbb{L}$ is linear for each fixed $x \in E$. Combining this with ii) it follows that

$$\langle\, x+y|z\,\rangle = \langle\, x|z\,\rangle + \langle\, y|z\,\rangle \quad \text{and} \quad \langle\, \lambda x|y\,\rangle = \overline{\lambda}\langle\, x|y\,\rangle \tag{4.1}$$

for all $x, y, z \in E$ and $\lambda \in \mathbb{L}$. This is expressed by saying that the inner product is *conjugate linear* in the first variable. In the case $\mathbb{L} = \mathbb{C}$ a map from $E \times E \to \mathbb{C}$, which is linear in the second variable and conjugate linear in the first variable is often said to be a *sesqui-linear form* on E. Thus we have seen that an inner product on a complex vector space E is a sesquilinear form on E. Conversely, any sesqui-linear form, satisfying the property i), which is referred to as saying that the form is *positive definite*, is an inner product. To show this it only remains to show that ii) is satisfied. Note first that by the linearity in the second variable

$$\langle\, x \,|\, \underline{0}\,\rangle = 0 \quad \text{for all} \quad x \in E\,.$$

In particular, it follows that $\langle\, \underline{0}|\underline{0}) = 0$ and thus i) implies

$$\langle\, x|x\,\rangle = 0 \Leftrightarrow x = \underline{0}\,, \quad x \in E\,. \tag{4.2}$$

Together with i) this shows that $\langle\, x|x\,\rangle \in \mathbb{R}$ for all $x \in E$. Applying

$$\langle\, x+y|x+y\,\rangle = \langle\, x|x\,\rangle + \langle\, y|y\,\rangle + \langle\, x|y\,\rangle + \langle\, y|x\,\rangle$$

it then follows that $\langle\, x|y\,\rangle + \langle\, y|x\,\rangle \in \mathbb{R}$, which shows that $\mathrm{Im}\langle\, x|y\,\rangle = -\mathrm{Im}\langle\, y|x\,\rangle$ for all $x, y \in E$. Replacing x by ix and using the sesqui-linearity we find $i\langle\, x|y\,\rangle - i\langle\, y|x\,\rangle \in \mathbb{R}$, and consequently $\mathrm{Re}\langle\, x|y\,\rangle = \mathrm{Re}\langle\, y|x\,\rangle$. Hence, ii) holds as claimed.

Given an inner product the *norm* $\|x\|$ of $x \in E$ is defined by

$$\|x\| = \sqrt{\langle\, x|x\,\rangle}\,.$$

From i) above it follows that

$$\|x\| > 0 \ \text{ for } x \neq 0\,, \tag{4.3}$$

and from iv) and (4.1) follows that

$$\|\lambda x\|^2 = \langle\, \lambda x|\lambda x\,\rangle = \lambda\overline{\lambda}\langle\, x|x\,\rangle = |\lambda|^2\|x\|^2\,,$$

that is

$$\|\lambda x\| = |\lambda|\|x\|\,. \tag{4.4}$$

Below, as a consequence of the Cauchy-Schwarz' inequality in Theorem 4.10 we will show the *triangle inequality*

$$\|x+y\| \leqslant \|x\| + \|y\| \,, \quad x, y \in E \,. \tag{4.5}$$

Hence, the norm $\|\cdot\|$ defined here actually satisfies the conditions i)–iii) characterizing a norm as discussed in Section 2.1. In particular, it defines a natural distance between vectors x, y in E by

$$d(x,y) = \|x-y\| \,. \tag{4.6}$$

Of special importance in what follows is that the norm defines a notion of convergence of sequences in E:

Definition 4.2. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in an inner product space E, that is a function from $\mathbb{N}$ into E whose value at $n \in \mathbb{N}$ is denoted by x_n. We say that the sequence is **convergent** with **limit** $x \in E$, and write $x_n \to x$ for $n \to \infty$ or $\lim_{n\to\infty} x_n = x$, if

$$\|x_n - x\| \to 0 \quad \text{for } n \to \infty \,.$$

In addition, the important notion of a Cauchy sequence will be needed.

Definition 4.3. A sequence $(x_n)_{n\in\mathbb{N}}$ in an inner product space E is called a **Cauchy sequence**, if for all $\varepsilon > 0$ there is an $N \in \mathbb{N}$ such that

$$\|x_n - x_m\| < \varepsilon \quad \text{for all } n, m > N \,,$$

(which is sometimes written as $\|x_n - x_m\| \to 0$ as $n, m \to \infty$).

The proof of the following result relies only on the triangle inequality and is deferred to Exercise 4.1.

Proposition 4.4. *Every convergent sequence in an inner product space E is a Cauchy sequence.*

Hilbert spaces are inner product spaces for which the converse holds:

Definition 4.5. *If every Cauchy sequence in an inner product space E is convergent then E is called a* **complete** *inner product space or a* **Hilbert space**.

As is well known and easy to verify, the set of real numbers $\mathbb{R}$ with usual addition and multiplication of numbers is a real vector space. In fact, it is also easily seen that standard multiplication also furnishes an inner product on $\mathbb{R}$ and that the distance between two real numbers x and y w. r. t. this inner product is the usual distance $|x-y|$. It is a fundamental property of the real numbers that the resulting inner product space is also complete. Indeed, this property turns out to be equivalent to the so-called *supremum property* stating that every non-empty set of real numbers with an upper bound has a supremum, i. e., a least upper bound. We have made use of this property previously in these notes and will take it for granted as a defining property of the real numbers distinguishing them from the rational numbers. Hence $\mathbb{R}$ is a Hilbert space. As will be seen, this prime example of a Hilbert space serves as a building block for other Hilbert spaces. Among these we have already in Chapter 1 encountered $\mathbb{R}^k$ with standard inner product $\langle x|y \rangle = x \cdot y$. In the next example we discuss $\mathbb{R}^k$ and its complex analogue $\mathbb{C}^k$ in some more detail as

examples of finite dimensional Hilbert spaces. Later in this section it will be shown that all finite dimensional inner product spaces are complete. Important instances of infinite dimensional inner product spaces that are not Hilbert spaces are likewise provided in the following example.

Example 4.6.

a) The *canonical inner product on* $\mathbb{R}^k$ is assumed to be familiar to the reader from linear algebra, and is given by

$$\langle\, x|y\,\rangle = x_1y_1 + \ldots + x_ky_k \tag{4.7}$$

for $x = (x_1, \ldots, x_k),\ y = (y_1, \ldots, y_k) \in \mathbb{R}^k$. Evidently, the right hand side of (4.7) is symmetric in x and y and is linear in each variable. Moreover, since $\langle\, x|x\,\rangle = x_1^2 + \cdots + x_k^2$ is is clear that $\langle\, x|x\,\rangle = 0$ if and only if $x = \underline{0}$. Hence $\langle\, \cdot|\cdot\,\rangle$ is an inner product on $\mathbb{R}^k$ in the sense of Definition 4.1 and the corresponding distance is the standard Euclidean distance given by

$$\|x - y\| = \sqrt{(x_1 - y_1)^2 + \ldots + (x_k - y_k)^2}\,.$$

In Exercise 4.2 it is shown that $\mathbb{R}^k$ is complete and hence a Hilbert space, as a consequence of the fact that $\mathbb{R}$ is complete.

b) It is straight-forward to verify that $\mathbb{C}^k$ is a complex vector space with addition and multiplication by (complex) scalars defined by

$$x + y = (x_1 + y_1, \ldots, x_k + y_k)\,, \quad \lambda x = (\lambda x_1, \ldots, \lambda x_k)$$

for $x = (x_1, \ldots, x_k), y = (y_1, \ldots, y_k) \in \mathbb{C}^k\,, \lambda \in \mathbb{C}$. The *canonical inner product on* $\mathbb{C}^k$ is defined by

$$\langle\, x|y\,\rangle = \overline{x_1}y_1 + \ldots + \overline{x_k}y_k\,.$$

It is left as an exercise for he reader to show that the requirements i) – iv) in Definition 4.1 are fulfilled and we refer to Exercise 4.2 for a proof that $\mathbb{C}^k$ is complete w. r. t. this inner product.

c) On the vector space $C([a, b])$ of continuous complex functions on the interval $[a, b]$ a standard inner product is defined by

$$\langle\, f|g\,\rangle = \int_a^b \overline{f(x)}g(x)dx \tag{4.8}$$

for $f, g \in C([a, b])$. More generally, for any *positive* continuous function ρ defined on $[a, b]$ an inner product $\langle\, \cdot\,|\,\cdot\,\rangle_\rho$ is defined by

$$\langle\, f|g\,\rangle_\rho = \int_a^b \overline{f(x)}g(x)\rho(x)dx\,. \tag{4.9}$$

The reader should recall that the integral of a continuous complex function is obtained by integrating the real an imaginary part, that is if $f = \mathrm{Re}f + i\mathrm{Im}f$ then

$$\int_a^b f(x)dx = \int_a^b \mathrm{Re}f(x)dx + i\int_a^b \mathrm{Im}f(x)dx\,. \tag{4.10}$$

Then $f \to \int_a^b f(x)dx$ defines a linear map from $C([a,b])$ into $\mathbb{C}$, which immediately implies that $\langle \cdot \mid \cdot \rangle_\rho$ satisfies iii) and iv) in Definition 4.1. Property ii) follows from

$$\int_a^b \overline{f(x)}dx = \overline{\int_a^b f(x)dx}\,.$$

Finally, i) is a consequence of the fact that if f is continuous and $f \geq 0$ then $\int_a^b f(x)dx = 0$ if and only if $f = 0$, which follows by an argument similar to (but simpler than) the one given in the proof of Lemma 2.11.

Note that if $\underline{\rho}$ and $\overline{\rho}$ denote the maximum and the minimum of the positive continuous function ρ on $[a,b]$ such that $0 < \underline{\rho} \leqslant \rho \leqslant \overline{\rho} < +\infty$, then

$$\underline{\rho}\|f\| \leqslant \|f\|_\rho \leqslant \overline{\rho}\|f\|$$

for $f \in C([a,b])$, where $\|\cdot\|$ and $\|\cdot\|_\rho$ are the norms corresponding to the inner products (4.8) and (4.9), respectively.

d) Let $\ell_0(\mathbb{N})$ denote the set of complex sequences $(x_n)_{n\in\mathbb{N}}$ that vanish from a certain step, that is sequences having only a finite number of elements different from 0. The set $\ell_0(\mathbb{N})$ is then a subspace of the vector space of all complex sequences, which with our previous notation is the space $\mathcal{F}(\mathbb{N},\mathbb{C})$. On $\ell_0(\mathbb{N})$ we define an inner product by

$$\langle\, (x_n) | (y_n)\,\rangle = \sum_{n=1}^{\infty} \overline{x}_n y_n\,, \tag{4.11}$$

where the sum on the right has only finitely many terms different from 0 (and thus obviously is convergent). That this defines an inner product on $\ell_0(\mathbb{N})$ is seen as in the case of $\mathbb{C}^k$.

4.2 Orthogonality

Let now $(E, \langle \cdot \mid \cdot \rangle)$ be an inner product space. Two vectors $x, y \in E$ are called *orthogonal*, also written $x \perp y$, if $\langle\, x|y\,\rangle = 0$. More generally, a vector $x \in E$ is orthogonal to a subset $A \subseteq E$, also denoted $x \perp A$, if x is orthogonal to all vectors in A, and for two subsets A, B of E the notation $A \perp B$ means that all vectors in A are orthogonal to all vectors in B, in which case A and B are called orthogonal.

The *orthogonal complement* $A^\perp$ of A is defined to be the set of all vectors orthogonal to A, that is

$$A^\perp = \{x \in E \mid \langle\, x|y\,\rangle = 0 \ \text{ for all } \ y \in A\}\,. \tag{4.12}$$

We note that by iii) and iv) in Definition 4.1 $A^\perp$ is a subspace of E for any subset $A \subseteq E$. For the same reason

$$A^\perp = (\mathrm{span}A)^\perp\,. \tag{4.13}$$

Here $\mathrm{span}A$ is the subspace of E consisting of all linear combinations of vectors from A.

A family $(x_i)_{i\in I}$ of vectors from E, where I is any index set, is said to be an *orthogonal family*, if the vectors in the family are pairwise orthogonal, that is if $\langle\, x_i|x_j\,\rangle = 0$, whenever $i \neq j$. If, moreover, $\|x_i\| = 1$ for all $i \in I$ we say that the family is *orthonormal*.

A family $(x_i)_{i\in I}$ of vectors from E is said to be linearly independent if any finite subset of $(x_i)_{i\in I}$ is linearly independent.

Lemma 4.7. *Let $(x_i)_{i\in I}$ be an orthogonal family in E, such that $x_i \neq 0$ for all $i \in I$. Then $(x_i)_{i\in I}$ is a linearly independent family.*

Proof. Let $(x_{i_1}, \ldots, x_{i_n})$ be a finite subfamily of $(x_i)_{i\in I}$ and assume that $\lambda_1, \ldots, \lambda_n \in \mathbb{L}$ satisfy $\lambda_1 x_{i_1} + \ldots + \lambda_n x_{i_n} = 0$. Calculating the inner product of both sides of this equation with x_{i_j} we arrive at

$$\lambda_1 \langle\, x_{i_j} | x_{i_1} \,\rangle + \ldots + \lambda_n \langle\, x_{i_j} | x_{i_n} \,\rangle = 0\,.$$

Using that $\langle\, x_{i_j} | x_{i_k} \,\rangle = 0$ for $k \neq j$ this gives $\lambda_j \langle\, x_{i_j}, x_{i_k} \,\rangle = 0$. Hence $\lambda_j = 0$ because $\langle\, x_{i_j} | x_{i_j} \,\rangle \neq 0$, since $x_{i_j} \neq 0$. Since this holds for each $j = 1, \ldots, n$, the proof is complete. □

The following generalization of *Pythagoras' theorem* will be used repeatedly below.

Theorem 4.8. *Let $(x_1, \ldots, x_n)$ be a finite orthogonal family. Then*

$$\|\sum_{i=1}^{n} x_i\|^2 = \sum_{i=1}^{n} \|x_i\|^2\,.$$

Proof. We have

$$\begin{aligned}\|\sum_{i=1}^{n} x_i\|^2 = \left\langle \sum_{i=1}^{n} x_i \,\middle|\, \sum_{j=1}^{n} x_j \right\rangle &= \sum_{i,j=1}^{n} \langle\, x_i | x_j \,\rangle \\ &= \sum_{i=1}^{n} \langle\, x_i | x_i \,\rangle = \sum_{i=1}^{n} \|x_i\|^2\,,\end{aligned}$$

where the third equality is due to the fact that only the diagonal terms corresponding to $i = j$ contribute to the sum by the orthogonality assumption. □

The next theorem is also well known from elementary linear algebra.

Theorem 4.9. *Let $(e_1, \ldots, e_n)$ be a finite orthonormal family in E. For all vectors $x \in E$ there is a unique vector $u \in$ span$\{e_1, \ldots, e_n\}$ such that*

$$x - u \in \{e_1, \ldots, e_n\}^{\perp}.$$

It is given by

$$u = \sum_{i=1}^{n} \langle\, e_i | x \,\rangle e_i\,. \tag{4.14}$$

and may be characterized as the unique vector in span$\{e_1, \ldots, e_n\}$ with the shortest distance to x wrt. the norm $\|\cdot\|$. The vector u given by (4.14) *is called the* **orthogonal projection** *of x on the subspace span$\{e_1, \ldots, e_n\}$.*

Moreover, **Bessel's inequality**

$$\sum_{i=1}^{n} |\langle\, e_i | x \,\rangle|^2 \leq \|x\|^2 \tag{4.15}$$

holds for all $x \in E$.

Proof. All vectors $u \in \text{span}\{e_1, \ldots, e_n\}$ can be written in the form $u = \lambda_1 e_1 + \ldots + \lambda_n e_n$, where $\lambda_1, \ldots, \lambda_n \in \mathbb{L}$. Calculating the inner product with e_i on both sides of this equation we arrive at $\langle e_i|u\rangle = \lambda_i \langle e_i|e_i\rangle = \lambda_i$ since $\langle e_i|e_i\rangle = 1$. Hence

$$u = \sum_{i=1}^{n} \langle e_i|u\rangle e_i \tag{4.16}$$

for $u \in \text{span}\{e_1, \ldots, e_n\}$. But $x - u \in \{e_1, \ldots, e_n\}^{\perp}$ is equivalent to $\langle x - u|e_i\rangle = 0$ for all $i = 1, \ldots, n$. Thus $\langle x|e_i\rangle = \langle u|e_i\rangle \quad \text{for} \quad i = 1, \ldots, n$. Together with (4.16) this gives the first part of the theorem.

If $v \in \text{span}\{e_1, \ldots, e_n\}$ then $x - v = (x - u) + (u - v)$ where $(x - u) \perp (u - v)$. Thus

$$\|x - v\|^2 = \|x - u\|^2 + \|u - v\|^2 \geq \|x - u\|^2 ,$$

with equality if and only if $u = v$.

Bessel's inequality follows from Theorem 4.8 if we note that $x = u + (x - u)$ where $u \perp (x - u)$, since then

$$\|x\|^2 = \|u\|^2 + \|x - u\|^2 \geq \|u\|^2 = \sum_{i=1}^{n} |\langle e_i|x\rangle|^2 .$$

In the last step we have used that $\|\langle x|e_i\rangle e_i\|^2 = |\langle x|e_i\rangle|^2$. □

From Bessel's inequality we may now derive the **Cauchy-Schwarz inequality**:

Theorem 4.10. *If E is an inner product space then for all $x, y \in E$ we have*

$$|\langle x|y\rangle| \leq \|x\| \, \|y\|. \tag{4.17}$$

Proof. If $y = \underline{0}$ then both sides of the inequality are zero. If $y \neq \underline{0}$ then $\|\frac{1}{\|y\|} y\| = 1$ and it follows from (4.15) with $n = 1$ and $e_1 = \frac{1}{\|y\|} y$ that

$$\Big| \Big\langle x \Big| \frac{1}{\|y\|} y \Big\rangle \Big| \leq \|x\| ,$$

from which (4.17) follows after multiplication by $\|y\|$ on both sides. □

As the final result of this section we derive the triangle inequality (4.5) from Cauchy-Schwarz' inequality:

$$\begin{aligned} \|x + y\|^2 &= \langle x + y|x + y\rangle = \|x\|^2 + \langle x|y\rangle + \langle y|x\rangle + \|y\|^2 \\ &= \|x\|^2 + 2\text{Re}\langle x|y\rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2|\langle x|y\rangle| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \, \|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2 . \end{aligned}$$

4.3 Continuity of the inner product

In the following we will repeatedly make use of the fact that *the inner product* $\langle \cdot\,, \cdot \rangle : E \times E \to \mathbb{L}$ *is continuous*. This is seen as follows. Let $x_0, y_0 \in E$ be given and choose $x, y \in E$ such that $\|x - x_0\| \leq \delta$ and $\|y - y_0\| \leq \delta$ where $\delta > 0$ is given. Then

$$\begin{aligned} |\langle\, x|y\,\rangle - \langle\, x_0|y_0\,\rangle| &= |\langle\, x|y - y_0\,\rangle + \langle\, x - x_0|y_0\,\rangle| \\ &\leq |\langle\, x|y - y_0\,\rangle| + |\langle\, x - x_0|y_0\,\rangle| \\ &\leq \|x\|\,\|y - y_0\| + \|x - x_0\|\,\|y_0\| \\ &\leq \delta(\|x\| + \|y_0\|) \\ &\leq \delta(\|x_0\| + \delta + \|y_0\|)\,. \end{aligned} \tag{4.18}$$

Here we have used the Cauchy-Schwarz inequality and that $\|x\| = \|(x - x_0) + x_0\| \leq \|x - x_0\| + \|x_0\| \leq \|x_0\| + \delta$.

Since the last expression in (4.18) tends to 0 as $\delta \to 0$ we conclude that for all $\varepsilon > 0$ there is a $\delta > 0$ such that $|\langle\, x|y\,\rangle - \langle\, x_0|y_0\,\rangle| < \varepsilon$ if $\|x - x_0\| \leq \delta$ and $\|y - y_0\| \leq \delta$. This is exactly the definition of continuity of the inner product at (x_0, y_0).

The continuity may equivalently be stated as saying that for all sequences (x_n) and (y_n) in E we have that

$$\langle\, x_n|y_n\,\rangle \to \langle\, x_0|y_0\,\rangle \quad \text{as} \quad n \to \infty\,, \tag{4.19}$$

if $x_n \to x_0$ and $y_n \to y_0$ as $n \to \infty$.

Definition 4.11. A series $\sum_{n=1}^{\infty} x_n$ with terms in a vector space E equipped with a norm is said to be convergent, if the sequence of partial sums $(s_k)_{k\in\mathbb{N}}$ defined by

$$s_k = \sum_{n=1}^{k} x_n \tag{4.20}$$

is convergent in E as $k \to \infty$. The limit $\lim_{k\to\infty} s_k$ is then called the *sum* of the series and is also denoted by $\sum_{n=1}^{\infty} x_n$.

From (4.19) and iii) in Definition 4.1 it follows that

$$\left\langle \sum_{n=1}^{\infty} x_n \,\middle|\, y \right\rangle = \left\langle \lim_{k\to\infty} \sum_{n=1}^{k} x_n \,\middle|\, y \right\rangle = \lim_{k\to\infty} \sum_{n=1}^{k} \langle\, x_n|y\,\rangle = \sum_{n=1}^{\infty} \langle\, x_n|y\,\rangle \tag{4.21}$$

and likewise that

$$\left\langle y \,\middle|\, \sum_{n=1}^{\infty} x_n \right\rangle = \sum_{n=1}^{\infty} \langle\, y|x_n\rangle \tag{4.22}$$

for any convergent series $\sum_{n=1}^{\infty} x_n$ in E and all $y \in E$.

Another consequence of the continuity of the inner product is that $A^{\perp} = (\overline{A})^{\perp}$ for any subset $A \subseteq E$. Here $\overline{A}$ denotes the closure of the set A, that is the set of all limit points of sequences in A. In fact, if $x \in A^{\perp}$ and $y \in \overline{A}$ there is a sequence (y_n) in A such that

$y = \lim_{n\to\infty} y_n$, and so we conclude that $\langle\, x \,|\, y \,\rangle = \lim_{n\to\infty} \langle\, x \,|\, y_n \,\rangle = 0$, which shows that $x \perp y$. Since $y \in \overline{A}$ was arbitrary we have shown that $A^\perp \subseteq (\overline{A})^\perp$. The opposite inclusion follows immediately from $A \subseteq \overline{A}$. Together with (4.13) this implies that

$$A^\perp = (\mathrm{span} A)^\perp = (\overline{\mathrm{span} A})^\perp \,. \tag{4.23}$$

We call $\overline{\mathrm{span} A}$ the *closed subspace spanned* by A and will denote it by $\overline{\mathrm{span}} A$.

Similarly, it it seen in Exercise 4.11 that $A^\perp$ *is a closed subspace of* E *for any subset* $A \subseteq E$ (that is to say $\overline{A^\perp} = A^\perp$).

4.4 Hilbert spaces

It is well known from linear algebra that any finite dimensional inner product space E has orthonormal bases. If we let $(e_1, \ldots, e_n)$ denote such a basis. If $x \in E$ we denote by $\underline{x} = (x_1, \ldots, x_n) \in \mathbb{L}^n$ the coordinates of x w. r. t. this basis such that

$$x = x_1 e_1 + \ldots + x_n e_n \,.$$

According to Pythagoras' Theorem 4.8 we have

$$\|x\| = (|x_1|^2 + \ldots + |x_n|^2)^{1/2} \,.$$

It follows that the map $(x_1, \ldots, x_n) \mapsto x_1 e_1 + \ldots + x_n e_n$ is a linear isometry from $\mathbb{L}^n$ onto E and the inverse map is $x \mapsto ((x, e_1), \ldots, (x, e_n))$. Since the vector space $\mathbb{L}^n$ is known to be complete w. r. t. the standard norm, we conclude that E is also complete. Thus we have seen that any finite dimensional inner product space is a Hilbert space.

For general Hilbert spaces we will see that the completeness requirement ensures that several key properties of finite dimensional inner product spaces generalize to the infinite dimensional case.

We shall first consider two important examples of infinite dimensional Hilbert spaces.

Example 4.12.

a) The space $\ell_0(\mathbb{N})$ defined in Example 4.6 d) is not complete. In fact, let

$$x^n = (1, \tfrac{1}{2}, \tfrac{1}{3}, \ldots, \tfrac{1}{n}, 0, 0, \ldots)$$

for $n \in \mathbb{N}$. Then $x^n \in \ell_0(\mathbb{N})$ and

$$\|x^n - x^m\|^2 = \sum_{k=m+1}^{n} \frac{1}{k^2} \quad \text{if } m \le n \,,$$

and since the series $\sum\limits_{k=1}^{\infty} \frac{1}{k^2}$ is convergent it follows that the sequence $(x^n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\ell_0(\mathbb{N})$. The sequence $(x^n)_{n\in\mathbb{N}}$ is, however, obviously not convergent in $\ell_0(\mathbb{N})$. Thus $\ell_0(\mathbb{N})$ is not a Hilbert space.

Let us instead consider the larger subspace $\ell_2(\mathbb{N})$ of $\mathcal{F}(\mathbb{N},\mathbb{C})$ consisting of sequences of complex numbers $(x_n)_{n\in\mathbb{N}}$ that are square summable in the sense that

$$\sum_{n=1}^{\infty}|x_n|^2<+\infty\,.$$

That $\ell_2(\mathbb{N})$ is a subspace of $\mathcal{F}(\mathbb{N},\mathbb{C})$ is seen as follows. Let (x_n) and (y_n) be two sequences in $\ell_2(\mathbb{N})$ and recall the well-known inequality $|a+b|^2\le 2(|a|^2+|b|^2)$ for complex numbers a,b (which is actually Cauchy-Schwarz' inequality for the vectors $(1,1)$ and (a,b) in $\mathbb{C}^2$). This gives

$$\sum_{n=1}^{\infty}|x_n+y_n|^2\le 2\sum_{n=1}^{\infty}|x_n|^2+2\sum_{n=1}^{\infty}|y_n|^2<+\infty\,,$$

and thus $(x_n)+(y_n)\in\ell_2(\mathbb{N})$. Since we also have $\lambda(x_n)=(\lambda x_n)\in\ell_2(\mathbb{N})$ if $\lambda\in\mathbb{C}$ it follows that $\ell_2(\mathbb{N})$ is a subspace of $\mathcal{F}(\mathbb{N},\mathbb{C})$. Furthermore, the inequality $|ab|\le\frac{1}{2}(|a|^2+|b|^2)$ for $a,b\in\mathbb{C}$ implies

$$\sum_{n=1}^{\infty}|\overline{x_n}y_n|\leqslant\frac{1}{2}\sum_{n=1}^{\infty}(|x_n|^2+|y_n|^2)<\infty\,,$$

such that

$$\langle\,(x_n)\,|\,(y_n)\,\rangle=\sum_{n=1}^{\infty}\overline{x_n}y_n$$

is well defined for all $(x_n),(y_n)$ in $\ell_2(\mathbb{N})$, since the series above is absolutely convergent (see the definition of absolute convergence in Exercise 4.12). It is now clear that this provides an extension of the inner product (4.11) on $\ell_0(\mathbb{N})$ to an inner product on $\ell_2(\mathbb{N})$ which will be called the *canonical inner product on* $\ell_2(\mathbb{N})$.

It will now be shown that $\ell_2(\mathbb{N})$ with this inner product is a Hilbert space. Thus assume that $(x^n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\ell_2(\mathbb{N})$ and write $x^n=(x_1^n,x_2^n,\dots)$. For every $k\in\mathbb{N}$ the inequality $|x_k^n-x_k^m|\le\|x^n-x^m\|$ clearly holds and implies that the sequence $(x_k^n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{C}$. Since $\mathbb{C}$ is complete it follows that this sequence converges to a limit $x_k\in\mathbb{C}$. Setting $x=(x_1,x_2,\dots)$, we claim that $x\in\ell_2(\mathbb{N})$ and that $x^n\to x$ as $n\to\infty$.

In order to prove the claim, let $\varepsilon>0$ and choose $N\in\mathbb{N}$ such that

$$\sum_{k=1}^{K}|x_k^n-x_k^m|^2\le\sum_{k=1}^{\infty}|x_k^n-x_k^m|^2=\|x^n-x^m\|^2\le\varepsilon^2$$

for all $n,m\ge N$ and all $K\in\mathbb{N}$. Taking the limit $m\to\infty$ of the first sum then yields $\sum\limits_{k=1}^{K}|x_k^n-x_k|^2\le\varepsilon^2$ for $n\ge N$ and all $K\in\mathbb{N}$, and consequently

$$\|x^n-x\|^2=\sum_{k=1}^{\infty}|x_k^n-x_k|^2\le\varepsilon^2$$

for $n\ge N$. This shows that $x^N-x\in\ell_2(\mathbb{N})$ and hence that $x=x^N-(x^N-x)\in\ell_2(\mathbb{N})$, and also that $x^n\to x$ as $n\to\infty$ as claimed.

b) The space $C([a,b])$ with standard inner product (4.8) is not complete as we shall now discuss. The norm on this space is given by

$$\|f\|_\rho^2 = \int_a^b |f(x)|^2 \rho(x) dx\,.$$

Let f_n denote the function on $[0,2]$ which is equal to 0 on $[0,1]$, grows linearly from 0 to 1 on $[1, 1+\frac{1}{n}]$ and is equal to 1 on $[1+\frac{1}{n}, 2]$ (draw the graph!), It is easy to see that (f_n) is a Cauchy sequence in $C([0,2])$ w. r. t. the norm $\|\cdot\|$, but that it is not convergent in $C([0,2])$.

As in the discussion of ℓ_0 the space $C([a,b])$ can also be extended to a Hilbert space $L_2([a,b])$ consisting of square integrable functions on $[a,b]$, that is to say functions $f : [a,b] \to \mathbb{C}$ such that

$$\int_a^b |f(x)|^2\, dx < \infty\,.$$

Here, the integral refers to a generalization of the Riemann integral, called the Lebesgue integral. This also entails a generalization of what is meant by functions or rather what is meant by two functions being equal. The details of this is beyond the scope of these notes. Only the following three facts will be needed below:

i) Two functions f and g in $L_2([a,b])$ are considered to be equal if

$$\int_a^b |f(x) - g(x)|^2\, dx = 0\,,$$

and this is expressed by saying that f is equal to g *almost everywhere* on $[a,b]$, or that $f(x) = g(x)$ for *almost all* $x \in [a,b]$. This is, in particular, true for any two functions that differ at only finitely many points of the interval. This strictly speaking means that $L_2([a,b])$ is not a space of functions, but of classes of functions that agree almost everywhere. The proof that $L_2([a,b])$ is a vector space with inner product given by (4.8) can now be completed as for $\ell_2(\mathbb{N})$ above. That $L_2([a,b])$ is a Hilbert space is one of the fundamental results in measure theory known as the *Riesz-Fischer Theorem*.

ii) $L_2([a,b])$ is a minimal extension of $C([a,b])$ in the sense that the closure $\overline{(}C([a,b])$ in $L_2([a,b])$ is equal to all of $L_2([a,b])$. This is also expressed by saying that $C([a,b])$ is *dense* in $L_2([a,b])$, because it means that every $f \in L_2([a,b])$ can be approximated arbitrarily well by elements in $C([a,b])$.

iii) In contrast to the Riemann integral the Lebesgue integral is defined on an arbitrary interval $I \subseteq \mathbb{R}$ and for continuous positive functions it agrees with the improper Riemann integral. Thus, the Hilbert space $L_2(I)$ can be defined for all intervals I, including the real axis $\mathbb{R}$. Note, however, that if I is an infinite interval or if it is not closed, then $L_2(I)$ does not contain all continuous functions. Instead it holds that the subspace $C_0(\mathbb{R})$ consisting of all continuous functions vanishing outside a closed and bounded subinterval of I is contained as a dense subspace in $L_2(I)$. In Section 5.5 it will used that the same holds for the subspace $C_0^\infty(\mathbb{R})$ consisting of C^∞ functions vanishing outside a closed and bounded subinterval of I.

Finally, we also remark that for an arbitrary positive continuous weight function ρ on $[a,b]$ the space $C([a,b])$ with inner product (4.9) can be extended to a Hilbert space $L_2([a,b],\rho)$, and more generally a Hilbert space $L_2(I,\rho)$ with inner product

$$\langle\, f|g\,\rangle = \int_I \overline{f(x)}g(x)dx$$

is defined for any interval I and any continuous positive *weight function* ρ on I.

For more details on measure theory and Lebesgue integration the interested reader may consult [11] or volume I of [13].

In the following H will denote a Hilbert space with inner product $\langle\,\cdot\,|\,\cdot\,\rangle$. Any subspace X of H is an inner product space with the inner product defined as the restriction of $\langle\,\cdot\,|\,\cdot\,\rangle$ to $X\times X$. Then $X\subseteq H$ *is a Hilbert space if and only if* X *is a closed subspace of* H, that is if $X=\overline{X}$ (see Exercise 4.13). In particular, all finite-dimensional subspaces of H are closed.

We shall make use of the following extension of Pythagoras' theorem to infinite orthogonal sets.

Theorem 4.13. *Let* $(x_i)_{i\in\mathbb{N}}$ *be an orthogonal family in a Hilbert space* H*. Then* $\sum_{i=1}^{\infty} x_i$ *is convergent in* H *if and only if*

$$\sum_{i=1}^{\infty}\|x_i\|^2 < +\infty\,,$$

and

$$\|\sum_{i=1}^{\infty} x_i\|^2 = \sum_{i=1}^{\infty}\|x_i\|^2\,. \tag{4.24}$$

Proof. Since H is a Hilbert space $\sum_{i=1}^{\infty} x_i$ is convergent in H if and only if the sequence of partial sums (s_n) is a Cauchy sequence. This means that for all $\varepsilon>0$ there is an $N\in\mathbb{N}$ such that

$$\|s_n - s_m\|^2 = \|\sum_{i=m+1}^{n} x_i\|^2 = \sum_{i=m+1}^{n}\|x_i\|^2 \le \varepsilon^2 \tag{4.25}$$

for all $n>m\ge N$ where we have used Theorem 4.8. Since $\mathbb{R}$ is also complete we have on the other hand that $\sum_{i=1}^{\infty}\|x_i\|^2$ is convergent if and only if the sequence (r_n) given by

$$r_n = \sum_{i=1}^{n}\|x_i\|^2$$

is a Cauchy sequence in $\mathbb{R}$. This means that for all $\varepsilon>0$ there is an $N\in\mathbb{N}$ such that

$$|r_n - r_m| = \sum_{i=m+1}^{n}\|x_i\|^2 < \varepsilon^2 \tag{4.26}$$

for all $n>m\ge N$. The first claim in the theorem follows by comparing (4.25) and (4.26).

Finally, the continuity of $x \mapsto \|x\|$ and Theorem 4.8 imply

$$\|\sum_{i=1}^{\infty} x_i\|^2 = \|\lim_{n\to\infty} \sum_{i=1}^{n} x_i\|^2 = \lim_{n\to\infty} \|\sum_{i=1}^{n} x_i\|^2 = \lim_{n\to\infty} \sum_{i=1}^{n} \|x_i\|^2 = \sum_{i=1}^{\infty} \|x_i\|^2 \,,$$

which shows that (4.24) holds. □

In order to extend the notion of orthonormal basis to the infinite dimensional case, recall first that an orthonormal basis in a finite dimensional inner product space H can be characterized as an orthonormal family that spans H. If H is infinite dimensional it is instead required that an orthonormal basis spans a dense subspace, that is a subspace whose closure equals the whole of H. Since, as noted above, any finite dimensional subspace of an inner product space is closed, the two requirements coincide for finite dimensional spaces and the following definition of an orthonormal basis is, indeed, an extension of the finite dimensional version.

Definition 4.14. An **orthonormal basis** for A Hilbert space H is an orthonormal family $(e_i)_{i\in I}$ in H such that $\overline{\text{span}}\{e_i \mid i \in I\} = H$.

It should be emphasized that for infinite dimensional vector spaces this notion of orthonormal basis is different from the purely algebraic notion of a basis, where it is required that the basis spans the whole vector space.

Since $H^\perp = \{\underline{0}\}$ it follows from (4.23) that

$$\{e_i \mid i \in I\}^\perp = \{\underline{0}\} \,,$$

for any orthonormal basis $(e_i)_{i\in I}$ for H. This means any orthonormal basis is a maximal orthonormal family in H in the sense that there is no vector $e \in H$ with $\|e\| = 1$, such that e together with $(e_i)_{i\in I}$ is an orthonormal family. The converse that any maximal orthonormal family in H is an orthonormal basis will be shown in Theorem 4.16 below.

Note that the definition of orthonormal basis implies that any orthonormal family $(e_i)_{i\in I}$ is an orthonormal basis for $\overline{\text{span}}\{e_i \mid i \in I\}$.

Any Hilbert space has an orthonormal basis. This claim, however, relies on what is known as *the axiom of choice*, which will not be discuss here. In the following we shall restrict attention to Hilbert spaces that have an orthonormal basis which is either finite or countable, where countable means that the basis vectors can be labeled by natural numbers (or equivalently by integers). Such Hilbert spaces are called *separable*.

It is known from elementary linear algebra that in finite dimensional vector spaces all bases have the same number of elements equal to the dimension of the space. In Exercise 4.21 it is demonstrated that in an infinite dimensional separable Hilbert space all orthonormal bases are infinite and have countably many elements.

Example 4.15. In the space $\ell_2(\mathbb{N})$ discussed in Example 4.12 a) an orthonormal basis $(\varepsilon_i)_{i\in\mathbb{N}}$ is obtained by letting ε_i be the sequence with all elements equal to 0 except element number i which is set to 1, that is

$$(\varepsilon_i)_j = \begin{cases} 1 & \text{for} \quad j = i \\ 0 & \text{for} \quad j \neq i \,. \end{cases}$$

That this is an orthonormal basis follows by observing that it is obviously an orthonormal family and that $\operatorname{span}\{\varepsilon_i \mid i \in \mathbb{N}\} = \ell_0(\mathbb{N})$ whose closure is $\ell_2(\mathbb{N})$ (see Exercise 4.18). We call $(\varepsilon_i)_{i\in\mathbb{N}}$ the *canonical orthonormal basis* for $\ell_2(\mathbb{N})$.

In Section 4.6 a particularly useful orthonormal basis for the space $L_2([-a,a])$, $a>0$, will be discussed. By translation one can then obtain an orthonormal basis for $L_2([a,b])$ for any $a<b$. In Exercise 4.20 a construction of a different type of orthonormal basis for $L_2([0,1])$ is described.

4.5 Orthonormal expansions

Our goal in this section is to generalize the familiar expansion of vectors w. r. t. orthonormal bases for finite dimensional Hilbert spaces to the case of infinite dimensional separable Hilbert spaces. Actually, all results in this section generalize to non-separable Hilbert spaces but we shall not discuss this here.

Let $(e_i)_{i\in\mathbb{N}}$ be an orthonormal family in H (assumed to be infinite dimensional) and consider a vector $x \in H$. According to Bessel's inequality (4.15) $\sum\limits_{i=1}^{n} |(x,e_i)|^2 \leq \|x\|^2$ for all $n \in \mathbb{N}$. Since the terms $\|x_i\|^2$ are non-negative this implies the following generalized version of Bessel's inequality:

$$\sum_{i=1}^{\infty} |\langle\, e_i|x\,\rangle|^2 \leq \|x\|^2 \,. \tag{4.27}$$

By Theorem 4.13 together with (4.27) we therefore conclude that the series $\sum\limits_{i=1}^{\infty}\langle\, e_i|x\,\rangle e_i$ is convergent in H. If we write

$$u = \sum_{i=1}^{\infty}\langle\, e_i|x\,\rangle e_i \tag{4.28}$$

it is clear that $u \in \overline{\operatorname{span}}\{e_i \mid i \in \mathbb{N}\}$. From Definition 4.11 of convergence of a series the vector u could in principle depend on the order in which we perform the sum in (4.28), i. e., on the chosen ordering of the basis vectors $e_1, e_2, e_3, \ldots$. That this is not the case is a consequence of the following key result.

Theorem 4.16. *Let $(e_i)_{i\in\mathbb{N}}$ be an orthonormal family in the separable Hilbert space H. To any vector $x \in H$ there is a unique vector $u \in \overline{\operatorname{span}}\{e_i \mid i \in \mathbb{N}\}$ such that*

$$x - u \in \{e_i \mid i \in \mathbb{N}\}^{\perp} \,,$$

and u is given by (4.28)*. In particular, u does not depend on the order of the summation and we also write*

$$u = \sum_{i\in\mathbb{N}}\langle\, e_i|x\,\rangle e_i \,.$$

Among all vectors in $\overline{\operatorname{span}}\{e_i \mid i \in \mathbb{N}\}$ the vector u has the shortest distance to x.

Proof. The vector u given by (4.28) satisfies that for each $j \in \mathbb{N}$

$$\langle\, e_j|u\,\rangle = \left\langle e_j \,\middle|\, \sum_{i=1}^{\infty}\langle\, e_i|x\,\rangle e_i \right\rangle = \sum_{i=1}^{\infty}\langle\, e_i|x\,\rangle\langle\, e_j|e_i\,\rangle = \langle\, e_j|x\,\rangle \,.$$

This shows that $\langle e_j | x-u \rangle = 0$ for all $j \in \mathbb{N}$, that is $x-u \in \{e_j \mid j \in \mathbb{N}\}^\perp$, and therefore proves the existence of u. The uniqueness is demonstrated below.

If $v \in \overline{\text{span}}\{e_i \mid i \in \mathbb{N}\}$ is arbitrary then $u-v \in \overline{\text{span}}\{e_i \mid i \in \mathbb{N}\}$ and hence by Pythagoras' Theorem

$$\|x-v\|^2 = \|x-u\|^2 + \|u-v\|^2 \geqslant \|x-u\|^2.$$

This proves the last claim of the theorem. If v moreover, fulfills that $x-v \in \{e_i \mid i \in \mathbb{N}\}^\perp$ then $u-v = (x-v)-(x-u) \in \{e_i \mid i \in \mathbb{N}\}^\perp = (\overline{\text{span}}\{e_i \mid i \in \mathbb{N}\})^\perp$. In particular, $\langle u-v | u-v \rangle = 0$ and hence $u=v$ showing that u is unique. □

The following result on orthonormal bases is of fundamental importance in many applications of Hilbert space theory.

Theorem 4.17. *For an orthonormal family $(e_i)_{i\in\mathbb{N}}$ in a separable Hilbert space H the following statements are equivalent*

(i) *$(e_i)_{i\in\mathbb{N}}$ is an orthonormal basis in H.*

(ii) *$\{e_i \mid i \in \mathbb{N}\}^\perp = \{\underline{0}\}$.*

(iii) **The orthonormal expansion**

$$x = \sum_{i\in\mathbb{N}} \langle e_i \,|\, x \rangle e_i$$

holds for all $x \in H$.

(iv) **Parseval's identity**

$$\|x\|^2 = \sum_{i=1}^{\infty} |\langle e_i \,|\, x \rangle|^2 \tag{4.29}$$

holds for all $x \in H$.

Proof. It has previously been shown that (i) ⇒ (ii). The implication (ii) ⇒ (iii) follows from Theorem 4.16 since $x = u$ with the notation used there. That (iii) ⇒ (iv) follows immediately from (4.24).

Finally, assume that (iv) holds and let $x \in H$. Then

$$\|x - \sum_{i=1}^{n} \langle e_i | x \rangle e_i\|^2 = \|x\|^2 - \sum_{i=1}^{n} |\langle e_i | x \rangle|^2 \to 0 \quad \text{for } n \to \infty\,,$$

which shows that $span\{e_i \mid i \in I\}$ is dense in H and hence (iv) ⇒ (i). □

A further consequence of Theorem 4.16 is the **projection theorem**:

Theorem 4.18. *Let X be a closed subspace of a separable Hilbert space H. Then for each $x \in H$ there exist unique vectors $u \in X$ and $v \in X^\perp$ such that*

$$x = u + v\,. \tag{4.30}$$

Proof. Since X is a closed subspace of the separable Hilbert space H then X is a separable Hilbert space (see Exercise 4.19). Thus, there exists an orthonormal basis $(e_i)_{i\in I}$ for X, where I is finite or equal to $\mathbb{N}$. Hence $X = \overline{\text{span}}\{e_i \mid i \in I\}$. The claim is now an immediate consequence of Theorems 4.9 and 4.16. □

As seen previously, the vector u in (4.30) may also be characterized as the vector in X with the shortest distance to x.

In general, if two subspaces V and W of a vector space E satisfy that any vector $x \in E$ can be uniquely written as $x = v + w$, where $v \in V$ and $w \in W$ then V and W are called *complementary subspaces* of E, which is also expressed by writing $E = V \oplus W$ or saying that E is the *direct sum* of V and W.

Hence Theorem 4.18 can be rephrased as saying that

$$H = X \oplus X^{\perp}$$

for any closed subspace X of H. The vector u in (4.30) is called the *orthogonal projection* of x onto X and as shown above it can be calculated as

$$u = \sum_{i\in I} \langle\, e_i | x \,\rangle e_i \,, \tag{4.31}$$

where $(e_i)_{i\in I}$ is an orthonormal basis for X.

Since $X^{\perp}$ is also a closed subspace of H we have $H = X^{\perp\perp} \oplus X^{\perp}$. If we use that any vector $x \in X^{\perp\perp}$ may be written uniquely as $x = u + v$, with $u \in X$ and $v \in X^{\perp}$ and the obvious inclusion $X \subseteq X^{\perp\perp}$ (why?) we see that, in fact, $u = x$ and $v = 0$. Thus

$$X = X^{\perp\perp}$$

for every closed subspace X of H. It then follows that v in (4.30) is the orthogonal projection of x onto $X^{\perp}$.

Proposition 4.19. *Let X be a closed subspace in a separable Hilbert space H. Assume $(e_i)_{i\in I}$ is an orthonormal basis for X and $(e_j)_{j\in J}$ is an orthonormal basis for $X^{\perp}$. Then the combined family $(e_i)_{i\in I\cup J}$ is an orthonormal basis for H.*

More generally, let X_n, $n = 1, 2, 3, \ldots$ be a sequence of closed subspaces of H fulfilling

i) $X_n \perp X_m$, if $n \neq m$.

ii) The linear span of all $X_n, n = 1, 2, 3, \ldots$ is a dense subspace of H.

Then the combination of any sequence of orthonormal bases for the individual subspaces X_n into one family (in any order) forms an orthonormal basis for H.

Proof. We refer to Exercise 4.22 for the argument. □

4.6 Fourier series

The celebrated theory of Fourier series which has its origin in J. Fourier's 1807 analysis of the heat equation may be conveniently formulated in terms of orthonormal expansions. We shall now discuss this very briefly.

The relevant Hilbert space is $H = L_2([-a,a])$, where $a > 0$, with inner product

$$\langle\, f|g\,\rangle = \frac{1}{2a}\int_{-a}^{a} \overline{f(\theta)}g(\theta)\,d\theta\,.$$

We have included the factor $\frac{1}{2a}$ for convenience. The key result about Fourier series is stated in the next theorem. We will give a proof based on the Stone-Weierstrass theorem (see e. g. [12], Chapter 5).

Theorem 4.20. *Let the functions $e_n \in C([-a,a])$ be given by*

$$e_n(\theta) = e^{in\frac{\pi}{a}\theta}\,,\ \ \theta \in [-a,a]\,,\ n \in \mathbb{Z}\,.$$

Then $(e_n)_{n\in\mathbb{Z}}$ is an orthonormal basis in $H = L_2([-a,a])$.

Proof. That $(e_n)_{n\in\mathbb{Z}}$ is an orthonormal family is seen from the calculation

$$\begin{aligned}\langle\, e_n|e_m\,\rangle &= \frac{1}{2a}\int_{-a}^{a} \overline{e_n(x)}e_m(x)dx \\ &= \frac{1}{2a}\int_{-a}^{a} e^{i(m-n)\frac{\pi}{a}\theta}\,d\theta \\ &= \begin{cases} \frac{1}{2a}\left[\frac{a}{i(m-n)\pi}e^{i(m-n)\frac{\pi}{a}\theta}\right]_{-a}^{a} = 0 & \text{for}\quad n \neq m \\ \frac{1}{2a}\,[x]_{-a}^{a} = 1 & \text{for}\quad n = m\,, \end{cases}\end{aligned}\,.$$

Here we have used that e_n is a periodic function of $\theta \in \mathbb{R}$ with period $2a$.

It remains to prove that $\overline{\text{span}}\{e_n \mid n \in \mathbb{Z}\} = H$. To conclude this we note that it is enough to show that $C([-a,a]) \subseteq \overline{\text{span}}\{e_n \mid n \in \mathbb{Z}\}$ since $\overline{C([-a,a])} = H$ as already remarked. It is thus enough to show that for all functions $f \in C(-a,a)$ and all $\varepsilon > 0$ there is an $f_1 \in \text{span}\{e_n \mid n \in \mathbb{Z}\}$ such that

$$\|f - f_1\| < \varepsilon\,. \tag{4.32}$$

To see this we first choose a function (why is this possible?) $f_2 \in C([-a,a])$ such that $f_2(-a) = f_2(a) = 0$ and

$$\|f - f_2\| < \frac{\varepsilon}{2}\,. \tag{4.33}$$

Then f_2 may be extended to a continuous periodic function on $\mathbb{R}$ with period $2a$. It then follows from the Stone-Weierstrass theorem that there is a function $f_1 \in \text{span}\{e_n \mid n \in \mathbb{Z}\}$, such that

$$|f_2(\theta) - f_1(\theta)| < \frac{\varepsilon}{\sqrt{4a}}\,,\ \ \theta \in [-a,a]\,.$$

Hence

$$\|f_2 - f_1\| < \frac{\varepsilon}{2}\,. \tag{4.34}$$

From (4.33), (4.34), and the triangle inequality we arrive at (4.32). □

If for $f \in L_2([-a,a])$ we define *the Fourier coefficients* $c_n(f)$ by

$$c_n(f) = \frac{1}{2a}\int_{-a}^{a} f(\theta)e^{-in\frac{\pi}{a}\theta}\,d\theta\,, \tag{4.35}$$

the above theorem and Theorem 4.19 imply that

$$f(\theta) = \sum_{n\in\mathbb{Z}} c_n(f)e^{in\frac{\pi}{a}\theta}\,. \tag{4.36}$$

This series is called the *Fourier series* for f. We emphasize that it converges in the Hilbert space $L_2([-a,a])$, that is with respect to the L_2-norm $\|\cdot\|$. More precisely, this means that

$$\|\sum_{n=-N}^{N} c_n(f)e_n - f\| \;\to\; 0 \;\text{ for } N\to\infty\,.$$

It is important to realize that this is not the same as uniform or pointwise convergence. For a discussion of uniform and pointwise convergence of Fourier series we refer the reader to more advanced texts on the subject.

Exercises

Exercise 4.1. Prove Proposition 4.4. *Hint.* Use the triangle inequality (4.5).

Exercise 4.2.

a) Show that a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^k$ (or $\mathbb{C}^k$) is convergent, respectively a Cauchy sequence, w. r. t. the standard norm if and only if each coordinate sequence

$$(x_n^i)_{n\in\mathbb{N}}\,,\quad i=1,\dots,k,$$

is convergent, respectively a Cauchy sequence, in $\mathbb{R}$ (or $\mathbb{C}$), where we use the notation $x_n = (x_n^1,\dots,x_n^k)$.

b) Use a) to conclude that $\mathbb{R}^k$ and $\mathbb{C}^k$ are complete for any $k \geqslant 2$, assuming that $\mathbb{R}$ is complete.

Exercise 4.3. Consider the finite dimensional complex Hilbert space $H = \mathbb{C}^k$, with standard inner product.

Show that $((1,0,\dots,0),(0,1,0\dots,0),\dots,(0,\dots,0,1))$ is an orthonormal basis for H.

Exercise 4.4. Show that in a complex inner product space E the *polarization identity* holds:

$$\langle\, x|y\,\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2 + i\|x-iy\|^2 - i\|x+iy\|^2)\,,\quad x,y\in H\,.$$

Exercise 4.5. Let E be a complex inner product space and let $n \in \mathbb{N}$, $a \in \mathbb{C}$ be such that $a^n = 1$ and $a^2 \neq 1$. Show that the *generalized polarization identity*

$$\langle\, x|y\,\rangle = \frac{1}{n}\sum_{\nu=0}^{n-1} a^\nu \|a^\nu x + y\|^2 \,.$$

holds for all $x, y \in E$.

Exercise 4.6. Show that $(\sin n\theta)_{n\in\mathbb{N}}$ is an orthogonal family in $C([0,\pi])$ with inner product given by (4.8).

Exercise 4.7. Determine $a_1, a_2, a_3 \in \mathbb{C}$ such that

$$\int_0^\pi |\cos\theta - \sum_{n=1}^{3} a_n \sin n\theta|^2 d\theta$$

is as small as possible.

Exercise 4.8. Consider the Hilbert space $H = L_2([-1,1])$ with the inner product

$$\langle\, f|g\,\rangle = \int_{-1}^{1} \overline{f(t)} g(t) dt$$

and the elements $f_1, f_2, g \in H$ given by $f_1(t) = \sqrt{3/2}\, t$, $f_2(t) = \sqrt{5/2}\, t^2$, and $g(t) = 1$. Show that f_1, f_2 are orthonormal and find the closest vector to g in $\operatorname{span}\{f_1, f_2\}$.

Exercise 4.9. Let the polynomials $p_0(x), p_1(x), \ldots$ be given such that $p_n(x)$ is a polynomial of degree n in the variable x with the coefficient of x^n being 1 and $(p_0(x), p_1(x), \ldots)$ being an orthogonal family in $C([0,1])$ with inner product given by (4.8) . Find $p_0(x), p_1(x)$ and $p_2(x)$.

Exercise 4.10. Show that $(\sin(n-\frac{1}{2})\theta)_{n\in\mathbb{N}}$ is an orthogonal family in $C([0,\pi])$ with inner product given by (4.8).

Exercise 4.11. Let H be a Hilbert space. Show that for any subset $A \subseteq H$ the subset $A^\perp$ is a closed subspace of H.

Exercise 4.12. A series $\sum_{n=1}^\infty x_n$ with terms in a vector space E eaquipped with a norm $\|\cdot\|$ is called *absolutely convergent* if $\sum_{n=1}^\infty \|x_n\| < \infty$.

Show that any absolutely convergent series with terms in a Hilbert space H is convergent in H.

Exercise 4.13. Let H be a Hilbert space. Show that a subspace X of H is a Hilbert space if and only if X is closed. Show also that the closure of a subspace of H is itself a subspace (here the continuity of addition and of scalar multiplication must be used).

Exercise 4.14. Let H be a Hilbert space and let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal family in H.

a) Show that the series $\sum_{n=1}^\infty \frac{1}{n} e_n$ is convergent in H and determine for which $\alpha \in \mathbb{R}$ the series $\sum_{n=1}^\infty n^\alpha e_n$ is convergent in H.

b) Determine the orthogonal projection of the vectors $e_1 \pm 2e_2$ on the subspace spanned by the vector $\sum_{n=1}^{\infty} n^{-1} e_n$. You may use that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.

Exercise 4.15. Show that $\lim_{n\to\infty} \int_0^\pi \log\theta \sin n\theta d\theta = 0$.
Hint. Use Bessel's inequality (4.27) for the orthogonal family in Exercise 4.6.

Exercise 4.16. Let $(e_i)_{i\in\mathbb{N}}$ be an orthonormal basis for the Hilbert space H. Show that the following generalization of Parseval's equation holds for all $x, y \in H$:

$$\langle\, x|y\,\rangle = \sum_{i=1}^{\infty} \langle\, x|e_i\,\rangle\langle\, e_i|y\,\rangle\,.$$

Exercise 4.17. Consider the inner product space $\ell_0(\mathbb{N})$ defined in Example 4.6 d). Let

$$X = \left\{ (x_n)_{n\in\mathbb{N}} \in \ell_0(\mathbb{N}) \mid \sum_{n=1}^{\infty} x_n \frac{1}{n} = 0 \right\}\,.$$

Show that X is a closed subspace of $\ell_0(\mathbb{N})$ and that $X \oplus X^\perp \neq \ell_0(\mathbb{N})$.

Exercise 4.18. Argue that $\ell_0(\mathbb{N})$ is dense in $\ell_2(\mathbb{N})$. In other words, show that any sequence x in $\ell_2(\mathbb{N})$ is the limit of a convergent sequence $(x^n)_{n\in\mathbb{N}}$ in $\ell_0(\mathbb{N})$.
Hint. Let x^n be the sequence whose elements coincide with those of x up to the n-th element and vanish afterwards.

Exercise 4.19.

a) Show that a Hilbert space H is separable if and only if it has a countable dense subset.

Hint. For the if-part one may use the Gram-Schmidt procedure to construct an orthonormal basis.

b) Let H be a separable Hilbert space. Show that any infinite subset M of H has a countable dense subset.

Hint. Let $\{x_i \mid i \in \mathbb{N}\}$ be a countable dense subset of H. For $i, n \in \mathbb{N}$ let $x_{i,n}$ denote any arbitrarily chosen point in $\{x \in M \mid \|x - x_i\| \leqslant \frac{1}{n}\}$ assuming that this set is non-empty. Otherwise let $x_{i,n}$ be an arbitrarily chosen point in M. Now show that $\{x_{i,n} \mid i, n \in \mathbb{N}\}$ is a countable dense subset of M.

c) Show that a closed subspace X of a separable Hilbert space is itself a separable Hilbert space. *Hint.* Use the characterization of separability found in question a) and the result of b).

Exercise 4.20. (Difficult) Consider the Hilbert space $H = L_2([0,1])$ with the inner product $\langle\, f|g\,\rangle = \int \overline{f(\theta)} g(\theta) d\theta$. Let $F_0 : \mathbb{R} \to \mathbb{R}$ be given by $F_0(t) = 1$ if $0 \leqslant t < 1$ and $F(t) = -1$ if $1 \leqslant t < 2$. Define f_n, $n = 0, 1, 2, \ldots$ by

$$F_n(t) = 2^{n-1} F_0(2^n t).$$

Let $f_{n,k} \in H$, $n = 0, 1, 2, \ldots$, $k = 0, 1, \ldots, 2^{n-1} - 1$ be given by

$$f_{n,k}(\theta) = f_n\left(\theta - k2^{n-1}\right), \quad \theta \in [0.1]$$

a) Show that the family $(f_{n,k}|n = 0, 1, 2, \dots,\ k = 0, 1, \dots, 2^{n-1} - 1)$ is a countable orthonormal family.

b) Show that the subspace

$$\operatorname{span}\{f_{n,k} \mid n = 0, 1, 2, \dots, N,\ k = 0, 1, \dots, 2^{n-1} - 1\}$$

consists of all functions that are constant on the intervals $[k2^{-N}, (k + 1)2^{-N})$, $k = 0, 1, \dots, 2^{-N} - 1$.

c) Show that

$$C^1([a, b]) \subseteq \overline{\operatorname{span}}\{f_{n,k} \mid n = 0, 1, 2, \dots,\ k = 0, 1, \dots, 2^{n-1} - 1\}.$$

Hint. Use the Mean Value Theorem.

d) Conclude from the information $L_2([0, 1]) = \overline{C^1([a, b])}$ that $(f_{n,k}|n = 0, 1, 2, \dots,\ k = 0, 1, \dots, 2^{n-1} - 1)$ is an orthonormal basis for $L_2([0, 1])$.

Exercise 4.21. Show that if H is an infinite dimensional separable Hilbert space then all orthonormal bases are infinite and have countably many elements.
Hint. From the characterization of separability in Exercise 4.19 we know that H has a countable dense set D. Given an orthonormal basis $(e_1, e_2, \dots)$ show that the open balls

$$B_i = \{x \in H \mid \|x - e_i\| < 1\}, \quad i \in \mathbb{N},$$

centered at basis vectors are disjoint. Argue that these balls have non-empty intersection with D and then that the orthonormal basis is at most countable.

Exercise 4.22. Prove Proposition 4.19.

Chapter 5

Operators on Hilbert spaces

In this chapter we first recall in the first two sections some basic facts known from elementary linear algebra about matrix representations of linear mappings defined on finite dimensional real or complex Hilbert spaces. Linear mappings defined on infinite dimensional Hilbert spaces are introduced in Section 5.3, including some illustrative examples. As is usual, we generally use the name *linear operator* or just *operator* instead of linear mapping in the following. For the sake of technical simplicity the main focus is on continuous (also called *bounded*) operators, although many operators relevant in physics, such as differential operators, are actually not bounded and will also be discussed to the extent needed.

The *adjoint* of an operator is defined and the basic properties of the adjoint operation are established. This allows the introduction of *self-adjoint* operators (corresponding to symmetric or Hermitian matrices) which together with *diagonalizable operators* (corresponding to diagonalizable matrices) are the subject of Section 5.4. The relevance of self-adjoint operators for quantum mechanics is that they represent physical quantities and may be regarded as corresponding to real functions on phase space in classical mechanics. This will be explained in more detail in the next chapter.

In the final Section 5.5 *unitary operators* (corresponding to orthogonal matrices) are introduced and the Fourier transformation is discussed as an important example.

5.1 Operators on finite dimensional real Hilbert spaces

In this section H denotes a real Hilbert space of dimension $\dim H = N < \infty$ with inner product $\langle \cdot \,|\, \cdot \rangle$, and $\alpha = (e_1, \ldots, e_N)$ denotes an orthonormal basis for H. For any given vector $x \in H$ its coordinates w. r. t. α will be denoted by $x_1, \ldots, x_N$, that is $x = x_1 e_1 + \ldots + x_N e_N$, and $\underline{x}$ will denote the corresponding coordinate column vector,

$$\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} .$$

It is a fact to be noted that every linear mapping A from H to a second Hilbert space H', whose norm we denote by $\|\cdot\|'$, is continuous. Indeed, for $x = x_1e_1 + \ldots + x_Ne_N \in H$ it holds that

$$Ax = x_1Ae_1 + \ldots + x_NAe_N \tag{5.1}$$

and therefore

$$\begin{aligned}\|Ax\|' &= \|x_1Ae_1 + \ldots + x_NAe_N\|' \\ &\leq |x_1|\,\|Ae_1\|' + \ldots + |x_N|\,\|Ae_N\|' \\ &\leq (\|Ae_1\|' + \ldots + \|Ae_N\|')\max\{|x_1|, \ldots, |x_N|\} \\ &\leq (\|Ae_1\|' + \ldots + \|Ae_N\|')\|x\|\,,\end{aligned}$$

where the triangle inequality has been used and also

$$\|x\|^2 = |x_1|^2 + \ldots + |x_N|^2 \geq (\max\{|x_1|, \ldots, |x_N|\})^2.$$

This shows that A is *bounded* in the sense that there exists a constant $c \geqslant 0$, such that

$$\|Ax\|' \leqslant c\,\|x\| \quad \text{for all } x \in H\,. \tag{5.2}$$

The *norm* $\|A\|$ of a bounded operator $A : H \to H'$ is by definition the smallest number c for which the bound (5.2) holds. Equivalently,

$$\|A\| = \sup\{\|Ax\|' \mid \|x\| = 1\}\,, \tag{5.3}$$

see Exercise 5.4.

As shown in Exercise 5.5 an operator $A : H \to H'$, where H and H' are arbitrary Hilbert spaces of finite or infinite dimension, is continuous if and only if it is bounded. Thus, we have shown the previously announced fact that if H is of finite dimension, then every operator $A : H \to H'$ is bounded and hence continuous.

For the sake of simplicity let us now assume that $H = H'$. As is well known from linear algebra a linear operator $A : H \to H$ is represented w. r. t. the basis α by an $N \times N$-matrix $\underline{\underline{A}}$ in the sense that the relation between the coordinate set for a vector $x \in H$ and its image $y = Ax$ is given by

$$\underline{y} = \underline{\underline{A}}\,\underline{x}\,. \tag{5.4}$$

Moreover, matrix multiplication is defined in such a way that if the operators A and B on H are represented by the matrices $\underline{\underline{A}}$ and $\underline{\underline{B}}$ w. r. t. the same basis α, then the composition AB is represented by the matrix product $\underline{\underline{A}}\,\underline{\underline{B}}$ w. r. t. α. Since the basis α is orthonormal the matrix elements a_{ij}, $1 \leq i, j \leq N$, of $\underline{\underline{A}}$ are given by the formula

$$a_{ij} = \langle\, e_i | Ae_j \,\rangle \tag{5.5}$$

as a consequence of the following calculation:

$$y_i = \langle\, e_i | Ax \,\rangle = \langle\, e_i | x_1Ae_1 + \ldots + x_NAe_N \,\rangle = \sum_{j=1}^{N} \langle\, e_i | Ae_j \,\rangle x_j\,. \tag{5.6}$$

Now let A^* denote the operator on H represented by the transpose $\underline{\underline{A}}^t$ of $\underline{\underline{A}}$. Using equation (5.5) it follows that

$$\langle\, e_j|Ae_i\,\rangle = a_{ji} = \langle\, e_i|A^*e_j\,\rangle = \langle\, A^*e_j|e_i\,\rangle\,,$$

where in the last step the symmetry of the inner has been product. Exploiting the linearity properties of the inner product this gives for $x = \sum_{j=1}^{N} x_j e_j$ and $z = \sum_{i=1}^{N} z_i e_i$ in H:

$$\langle\, x|Az\,\rangle = \sum_{i,j=1}^{N} x_j z_i \langle\, e_j|Ae_i\,\rangle = \sum_{i,j=1}^{N} x_j z_i \langle\, A^*e_i|e_j\,\rangle = \langle\, A^*x|z\,\rangle\,. \tag{5.7}$$

We claim that the validity of this identity for all $x, z \in H$ implies that A^* is uniquely determined by A and does not depend on the choice of orthonormal basis α. Indeed, if the operator B on H satisfies $\langle\, x|Az\,\rangle = \langle\, Bx|z\,\rangle$ for all $x, z \in H$, then $\langle\, A^*x - Bx|z\,\rangle = 0$ for all $x, z \in H$, and hence $A^*x - Bx \in H^\perp = \{\underline{0}\}$ for all $x \in H$. This shows that $A^*x = Bx$ for all $z \in H$, that is $A^* = B$. The operator A^* is called the *adjoint operator* of A. If $A = A^*$, we say that A is *self-adjoint*. It follows from the definition of A^* that *the self-adjoint operators on a real finite dimensional Hilbert space are precisely those operators that are represented by symmetric matrices w. r. t. an arbitrary orthonormal basis for H.*

It is known from linear algebra, that every symmetric $N \times N$-matrix $\underline{\underline{A}}$ can be diagonalized by an orthogonal matrix, that is there exists an orthogonal $N \times N$-matrix $\underline{\underline{O}}$ such that

$$\underline{\underline{O}}^{-1}\,\underline{\underline{A}}\,\underline{\underline{O}} = \underline{\underline{D}}\,, \tag{5.8}$$

where $\underline{\underline{D}} = \Delta(\lambda_1, \ldots, \lambda_N)$ is an $N \times N$ diagonal matrix with the eigenvalues $\lambda_1, \ldots, \lambda_N$ of $\underline{\underline{A}}$ in the diagonal. That $\underline{\underline{O}}$ is orthogonal means that the columns (and therefore also the rows) of $\underline{\underline{O}}$ form an orthonormal basis for $\mathbb{R}^N$, which is equivalent to

$$\underline{\underline{O}}^t\,\underline{\underline{O}} = \underline{\underline{I}}\,, \tag{5.9}$$

where I denotes the $N \times N$ identity matrix. This also means that $\underline{\underline{O}}$ is orthogonal if and only if $\underline{\underline{O}}$ is invertible and

$$\underline{\underline{O}}^t = \underline{\underline{O}}^{-1}\,. \tag{5.10}$$

Equation (5.8) expresses that the columns of $\underline{\underline{O}}$ are eigenvectors for $\underline{\underline{A}}$. Therefore, we conclude that for every symmetric matrix $\underline{\underline{A}}$ there exists an orthonormal basis for $\mathbb{R}^N$ (with standard inner product) consisting of eigenvectors for $\underline{\underline{A}}$.

Let now $\underline{\underline{A}}$ represent a self-adjoint operator A w. r. t. the basis α as above, and let the orthogonal matrix $\underline{\underline{O}} = (o_{ij})$ be chosen according to (5.8). Let $O : H \to H$ denote the operator represented by the matrix $\underline{\underline{O}}$ w. r. t. α. Furthermore, let $D : H \to H$ denote the operator represented by $\underline{\underline{D}}$ w.r.t. α, that is

$$De_i = \lambda_i e_i\,, \qquad i = 1, \ldots, N\,. \tag{5.11}$$

Then equation (5.8) is equivalent to

$$O^{-1}AO = D\,, \tag{5.12}$$

since the matrix representing $O^{-1}AO$ w. r. t. α is $\underline{\underline{O}}^{-1}\underline{\underline{A}}\,\underline{\underline{O}}$.

Similarly, equations (5.9) and (5.10) are equivalent to

$$O^*O = 1 \tag{5.13}$$

and

$$O^* = O^{-1}\,, \tag{5.14}$$

respectively, where 1 denotes the identity operator on H. An operator O fulfilling (5.14) is called an *orthogonal operator*. Thus *orthogonal operators are precisely those operators that are represented by orthogonal matrices w. r. t. an arbitrary orthonormal basis.*

Setting

$$f_i = Oe_i\,, \qquad i = 1, \ldots, N\,,$$

it follows that

$$\langle\, f_i | f_j \,\rangle = \langle\, Oe_i | Oe_j \,\rangle = \langle\, O^*Oe_i | e_j \,\rangle = \langle\, e_i | e_j \,\rangle = \delta_{ij}\,, \tag{5.15}$$

which means that $(f_1, \ldots, f_N)$ is an orthonormal basis for H. In other words, *orthogonal operators map orthonormal bases to orthonormal bases.*

It now follows from (5.11) and (5.12) that

$$Af_i = AOe_i = ODe_i = O(\lambda_i e_i) = \lambda_i Oe_i = \lambda_i f_i\,. \tag{5.16}$$

A vector $x \in H \setminus \{0\}$ with the property that the image Ax is proportional to x, i. e., such that there exists $\lambda \in \mathbb{R}$ fulfilling

$$Ax = \lambda x\,, \tag{5.17}$$

is called an *eigenvector* for A, and λ is called the corresponding *eigenvalue*. Thus, equations (5.15) and (5.16) imply that *for every self-adjoint operator A on a finite dimensional real Hilbert space there exists an orthonormal basis consisting of eigenvectors for A.* We say that such a basis diagonalizes A, since the matrix representing A w. r. t. this basis is the diagonal matrix $\underline{\underline{D}}$ whose diagonal entries are the eigenvalues of A.

5.2 Operators on finite dimensional complex Hilbert spaces

In this section H denotes a finite dimensional complex Hilbert space with inner product $\langle\, \cdot\, | \cdot \,\rangle$ and $\alpha = (e_1, \ldots, e_N)$ again denotes an orthonormal basis for H.

By the same argument as in the previous section (see (5.1)) every operator $A : H \to H$ is bounded. By the same calculations as those leading to (5.6) A is represented w. r. t. α by the complex matrix $\underline{\underline{A}} = (a_{ij})$, where

$$a_{ij} = \langle\, e_i | Ae_j \,\rangle\,, \qquad 1 \le i, j \le N\,.$$

Let A^* denote the operator represented w. r. t. α by the matrix $\underset{=}{A}^*$, called the *Hermitian conjugate* of $\underset{=}{A}$, obtained from $\underset{=}{A}$ by transposition and complex conjugation, that is

$$\langle\, e_j|Ae_i\,\rangle = a_{ji} = \langle\, \overline{e_i|A^*e_j}\,\rangle = \langle\, A^*e_j|e_j\,\rangle\,.$$

For $x = \sum_{j=1}^{N} x_j e_j$ and $z = \sum_{i=1}^{N} z_i e_i$ in H it then follows that

$$\langle\, x|Az\,\rangle = \sum_{i,j=1}^{N} \overline{x}_j z_i \langle\, e_j|Ae_i\,\rangle = \sum_{i,j=1}^{N} \overline{x}_j z_i \langle\, A^*e_j|e_i\,\rangle = \langle\, A^*x|z\,\rangle\,.$$

By a similar argument as in the previous section it follows that the operator A^* is uniquely determined by A. It is called the adjoint operator of A, and A is called self-adjoint if $A = A^*$. It follows, in particular, that for a self-adjoint operator A it holds that $\underset{=}{A}^* = \underset{=}{A}$. Matrices for which this relation holds are called *Hermitian*. Notice that addition and multiplication of complex matrices are defined in the same way and satisfy the same basic laws of calculation as for real matrices. Moreover, well known concepts such as rank and determinant are defined in an analogous manner. Results and proofs can be transferred directly from the real to the complex case. It suffices here to mention that *a quadratic matrix is regular, or invertible, if and only if its determinant is different from zero*. Moreover, for a quadratic matrix $\underset{=}{A}$ the relation

$$\det \underset{=}{A}^* = \overline{\det \underset{=}{A}}\,,$$

follows immediately from $\det \underset{=}{A}^t = \det \underset{=}{A}$.

The complex counterpart of an orthogonal operator is called a *unitary operator*. More specifically, $U : H \to H$ is called unitary if it fulfills

$$U^*U = 1\,, \tag{5.18}$$

that is

$$U^* = U^{-1}\,. \tag{5.19}$$

Letting $\underset{=}{U}$ denote the matrix representing U w. r. t. α, equation (5.18) is equivalent to

$$\underset{=}{U}^*\underset{=}{U} = \underset{=}{I}\,. \tag{5.20}$$

A matrix $\underset{=}{U}$ fulfilling (5.20) is called *unitary* and it is easily seen that this is equivalent to the statement that the columns of $\underset{=}{U}$ form an orthonormal basis for $\mathbb{C}^N$. In particular, a real unitary matrix is the same thing as an orthogonal matrix.

Evidently, equation (5.18) implies

$$\langle\, Ux|Uy\,\rangle = \langle\, U^*Ux|y\,\rangle = \langle\, x|y\,\rangle\,, \quad x, y \in H\,, \tag{5.21}$$

which is expressed by saying that any unitary operator U *preserves the inner product*. Setting $x = y$ yields

$$\|Ux\| = \|x\|\,, \qquad x \in H\,, \tag{5.22}$$

and thus U is an isometry. The following result shows, among other things, that (5.22) is in fact equivalent to U being unitary.

Theorem 5.1. *Let $U : H \to H$ be a linear operator on a finite dimensional Hilbert space H. The following four statements are equivalent.*

(i) *U is unitary.*

(ii) *U preserves the inner product.*

(iii) *U is an isometry.*

(iv) *U maps an orthonormal basis onto an orthonormal basis for H.*

Proof. Above it has been shown (i) $\Rightarrow$ (ii) $\Rightarrow$ iii).
(iii) $\Rightarrow$ (ii): This follows by using the polarization identity for the two sesquilinear forms $\langle\, x|y\,\rangle$ and $\langle\, Ux|Uy\,\rangle$. Indeed, if U is an isometry, it follow that

$$\langle\, Ux|Uy\,\rangle = \tfrac{1}{4}\sum_{\nu=0}^{3} i^\nu \|U(x+i^\nu y)\|^2 = \tfrac{1}{4}\sum_{\nu=0}^{3} i^\nu \|x+i^\nu y\|^2 = \langle\, x|y\,\rangle$$

for arbitrary $x, y \in H$.

(ii) $\Rightarrow$ (i): The relation $\langle\, Ux|Uy\,\rangle = \langle\, x|y\,\rangle$ can be rewritten as $\langle\, x|U^*Uy\,\rangle = \langle\, x|y\,\rangle$. This holds for all $x, y \in H$ if and only if $U^*U = 1$. This proves that (i) and (ii) are in fact equivalent.

We have now demonstrated that (i), (ii) and (iii) are equivalent.
(ii) $\Rightarrow$ (iv): Setting $f_i = Ue_i$ condition b) implies

$$\langle\, f_i|f_j\langle = \rangle Ue_i|Ue_j\,\rangle = \langle\, e_i|e_j\,\rangle = \delta_{ij}\,, \quad 1 \le i, j \le N\,,$$

which shows that $(f_1, \ldots, f_N)$ is an orthonormal basis for H.
(iv) $\Rightarrow$ (iii): For $x = x_1e_1 + \ldots + x_Ne_N \in H$ condition d) and Pythagoras' theorem imply

$$\|Ux\|^2 = \|x_1Ue_1 + \ldots + x_NUe_N\|^2 = |x_1|^2 + \ldots + |x_N|^2 = \|x\|^2\,,$$

which proves the claim.

This finishes the proof of the theorem. □

In relation to Theorem 5.1 it should be noted that for any pair of orthonormal bases $(e_1, \ldots, e_N)$ and $(f_1, \ldots, f_N)$ for H there is exactly one operator U, which maps the first basis onto the second, that is

$$Ue_i = f_i\,, \qquad i = 1, \ldots, N\,,$$

and it is given by

$$U(x_1e_1 + \ldots + x_Ne_N) = x_1f_1 + \ldots + x_Nf_N\,.$$

This operator is unitary by Theorem 5.1.

Eigenvectors and eigenvalues for an operator $A : H \to H$ are defined as in the previous section (cf. (5.17)), except that the eigenvalue now may assume complex values instead of real values only. Moreover, A is called diagonalizable if there exists an orthonormal basis for H consisting of eigenvectors for A. Equivalently, this means that there exists an

orthonormal basis $(f_1, \ldots, f_N)$ w. r. t. which A is represented by a diagonal matrix. Indeed, if $(f_1, \ldots, f_N)$ is an orthonormal basis consisting of eigenvectors such that $Af_i = \lambda_i f_i$, $i = 1, \ldots, N$, then

$$a_{ij} = \langle\, f_i | Af_j \,\rangle = \langle\, f_i | \lambda_j f_j \,\rangle = \lambda_j \delta_{ij}\,, \qquad 1 \le i, j \le N\,.$$

Conversely, if $(f_1, \ldots, f_N)$ is an orthonormal basis such that

$$a_{ij} = \langle\, f_i | Af_j \,\rangle = \lambda_j \delta_{ij}\,, \qquad 1 \le i, j \le N\,,$$

then

$$Af_j = \sum_{i=1}^{N} \langle\, Af_j | f_i \,\rangle f_i = \sum_{i=1}^{N} \lambda_j \delta_{ij} f_i = \lambda_j f_j$$

for $1 \le j \le N$.

In view of the discussion in the previous section it is natural to ask if every self-adjoint operator A on H is diagonalizable. The answer is yes, and the same holds for unitary operators, as will be seen in Section 5.4.

We end this section by looking at a couple of simple examples.

Example 5.2. a) Let H be a 2-dimensional complex Hilbert space and let $\alpha = (e_1, e_2)$ be an orthonormal basis for H. Consider the operator A on H, which is represented w. r. t. α by the matrix

$$\underline{\underline{A}} = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}.$$

Then A is self-adjoint since $\underline{\underline{A}}^* = \underline{\underline{A}}$. The eigenvalues of A can be determined by solving the characteristic equation $\det(\underline{\underline{A}} - \lambda \underline{\underline{I}}) = 0$, since this condition ensures (as in the case of real matrices) that the system of linear equations $(\underline{\underline{A}} - \lambda \underline{\underline{I}})\underline{x} = 0$ has a non-trivial solution $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, which is equivalent to stating that the vector $x = x_1 e_1 + x_2 e_2$ satisfies $Ax = \lambda x$.

The characteristic equation is

$$\det \begin{pmatrix} 1-\lambda & i \\ -i & 1-\lambda \end{pmatrix} = (1-\lambda)^2 - 1 = 0\,,$$

which gives $\lambda = 0$ or $\lambda = 2$. For $\lambda = 0$ the equation $(\underline{\underline{A}} - \lambda \underline{\underline{I}}) \binom{x_1}{x_2} = 0$ is equivalent to $x_1 + i\, x_2 = 0$, whose solutions are of the form $t \binom{1}{i}$, $t \in \mathbb{C}$.

For $\lambda = 2$ a similar calculation gives the solutions $t \binom{1}{-i}$, $t \in \mathbb{C}$.

Two normalized eigencolumns corresponding to $\lambda = 0$ and $\lambda = 2$, respectively, are then

$$\underline{x}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{and} \quad \underline{x}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

The matrix

$$\underline{\underline{U}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$$

is seen to be unitary, and so A is diagonalizable since

$$\underset{=}{U}^{-1}\underset{=}{A}\,\underset{=}{U} = \underset{=}{U}^{*}\underset{=}{A}\,\underset{=}{U} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} .$$

The unitary operator represented by $\underset{=}{U}$ w. r. t. α, maps the basis α onto the orthonormal basis $\big(\frac{1}{\sqrt{2}}(e_1 + ie_2), \frac{1}{\sqrt{2}}(e_1 - ie_2)\big)$ consisting of eigenvectors for A.

b) The real matrix

$$\underset{=}{O} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

is seen at to be orthogonal by inspection. It has no real eigenvalues and is therefore not diagonalizable over $\mathbb{R}$.

On the other hand, considering it as a complex matrix representing an operator O on a 2-dimensional complex Hilbert space H w. r. t. an orthonormal basis $\alpha = (e_1, e_2)$, then O is unitary. Its eigenvalues are found by solving

$$\det \begin{pmatrix} \frac{1}{\sqrt{2}} - \lambda & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} - \lambda \end{pmatrix} = 0 .$$

This gives $\lambda = \frac{1}{\sqrt{2}}(1 \pm i)$. Corresponding eigencolumns for $\underset{=}{O}$ are as in the previous example $\underline{x}_1 = \frac{1}{\sqrt{2}}\binom{1}{i}$ and $\underline{x}_2 = \frac{1}{\sqrt{2}}\binom{1}{-i}$.

Since

$$\underset{=}{U}^{*}\underset{=}{O}\,\underset{=}{U} = \begin{pmatrix} \frac{1+i}{\sqrt{2}} & 0 \\ 0 & \frac{1-i}{\sqrt{2}} \end{pmatrix} ,$$

we conclude that O is diagonalizable and that $\big(\frac{1}{\sqrt{2}}(e_1 + ie_2), \frac{1}{\sqrt{2}}(e_1 - ie_2)\big)$ is an orthonormal basis for H that diagonalizes O.

5.3 Operators on infinite dimensional Hilbert spaces.

In the remainder of this chapter H, H_1 and H_2 denote arbitrary separable Hilbert spaces over $\mathbb{L}$, where $\mathbb{L}$ denotes either $\mathbb{R}$ or $\mathbb{C}$. Results are mostly formulated for the case where the Hilbert spaces are infinite dimensional, but all results hold with obvious modifications for finite dimensional Hilbert spaces as well.

We shall use the notation $\mathcal{B}(H_1, H_2)$ for the set of bounded operators from H_1 to H_2 and set $\mathcal{B}(H) = \mathcal{B}(H, H)$. See (5.2) for the definition of a bounded operator. In Exercise 5.4 it is shown that $\mathcal{B}(H_1, H_2)$ is a vector space and that the norm defined by (5.3) fulfills the standard requirements (4.3)–(4.5).

Given an operator $A \in \mathcal{B}(H)$ and an orthonormal basis $(e_i)_{i\in\mathbb{N}}$ for H the matrix

$$(a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} & \cdots \\ \vdots & & \vdots & \\ a_{n1} & \cdots & a_{nn} & \cdots \\ \vdots & & \vdots & \end{pmatrix} \tag{5.23}$$

that represents A w. r. t. $(e_i)_{i\in\mathbb{N}}$ is defined by the same formula as in the finite dimensional case

$$a_{ij} = \langle\, e_i | Ae_j \,\rangle\,. \tag{5.24}$$

The coefficients of the orthonormal expansion

$$Ax = \sum_{i=1}^{\infty} y_i e_i\,,$$

for a vector $x = \sum\limits_{j=1}^{\infty} x_j e_j \in H$, are then given by

$$y_i = \langle\, e_i | Ax \,\rangle = \left\langle\, e_i \Big| \sum_{j=1}^{\infty} x_j Ae_j \,\right\rangle = \sum_{j=1}^{\infty} x_j \langle\, e_i | Ae_j \,\rangle\,. \tag{5.25}$$

Here the second equality follows from linearity and continuity of A:

$$\begin{aligned} Ax &= A\big(\lim_{N\to\infty} \sum_{j=1}^{N} x_j e_j\big) = \lim_{N\to\infty} A\big(\sum_{j=1}^{N} x_j e_j\big) \\ &= \lim_{N\to\infty} \sum_{j=1}^{N} x_j Ae_j = \sum_{j=1}^{\infty} x_j Ae_j\,, \end{aligned} \tag{5.26}$$

and similarly the last equality follows from linearity and continuity of $\langle\, e_i | x \,\rangle$ as a function of x. From (5.24) and (5.25) it is seen that

$$y_i = \sum_{j=1}^{\infty} a_{ij} x_j\,,$$

which is the infinite dimensional version of (5.4) and, in particular, shows that the matrix (5.23) determines the operator A uniquely.

In the infinite dimensional case, matrix representations are generally of rather limited use since calculations involve infinite series and can only rarely be performed explicitly. Moreover, the notion of determinant is not immediately generalizable to the infinite dimensional case, which implies that the standard method for determining eigenvalues and eigenvectors is not available any more. The coordinate independent operator point of view to be developed in the following will turn out more advantageous. It is useful first to introduce a few concrete examples of operators on infinite dimensional Hilbert spaces.

Example 5.3.

a) Let $H = \ell_2(\mathbb{N})$ and define for $x = (x_n)_{n\in\mathbb{N}} \in \ell_2(\mathbb{N})$ the sequence

$$Ax = (\frac{1}{n}x_n)_{n\in\mathbb{N}}\,. \tag{5.27}$$

Since

$$\sum_{n=1}^{\infty}\frac{1}{n^2}|x_n|^2 \leqslant \sum_{n=1}^{\infty}|x_n|^2 = \|x\|^2\,, \tag{5.28}$$

it follows that $Ax \in \ell_2(\mathbb{N})$ so that equation (5.27) defines a mapping A from H to H. It is easily seen that A is linear (verify this!), and from (5.28) it then follows that $\|A\| \leqslant 1$. In fact, $\|A\| = 1$, because equality holds in (5.28) for the sequence $(1, 0, 0, \dots)$.

More generally, let $a = (a_n)_{n\in\mathbb{N}}$ be a *bounded* (complex) sequence of numbers and set $\|a\|_u = \sup\{|a_n| \mid n \in \mathbb{N}\}$. For $x = (x_n)_{n\in\mathbb{N}} \in \ell_2(\mathbb{N})$ define the sequence

$$M_a x = (a_n x_n)_{n\in\mathbb{N}}\,. \tag{5.29}$$

Noting that

$$\sum_{n=1}^{\infty}|a_n x_n|^2 \leqslant \|a\|_u^2\|x\|^2\,,$$

it follows that $M_a x \in \ell_2(\mathbb{N})$. Hence, equation (5.29) defines a mapping M_a from H to H. It is easily seen that M_a is linear and the previous inequality then shows that $\|M_a\| \leqslant \|a\|_u$. In fact, $\|M_a\| = \|a\|_u$, because $M_a e_n = a_n\varepsilon_n$, where ε_n denotes the n'th vector in the canonical orthonormal basis for $\ell_2(\mathbb{N})$, and as a consequence we get $\|M_a\| \geqslant \|M_a\varepsilon_n\| = |a_n|$ for all $n \in \mathbb{N}$, which yields $\|M_a\| \geqslant \|a\|_u$.

Viewing $(a_n)_{n\in\mathbb{N}}$ and $x = (x_n)_{n\in\mathbb{N}}$ in (5.29) as functions on $\mathbb{N}$, the operator M_a is defined by multiplication by the function $(a_n)_{n\in\mathbb{N}}$. For this reason it is called the *multiplication operator defined by the sequence* $(a_n)_{n\in\mathbb{N}}$. Note, that w. r. t. the canonical orthonormal basis the operator M_a is represented by the diagonal matrix $\Delta(a_1, a_2, \dots)$ (verify this!).

b) Let $H = L_2([a, b])$, where $[a, b]$ is a closed bounded interval, and let $f : [a, b] \to \mathbb{C}$ be a continuous function. Since f is continuous it is bounded and

$$|f(x)g(x)| \leqslant \|f\|_u|g(x)|\,, \quad x \in [a, b]\,, \tag{5.30}$$

where

$$\|f\|_u = \max\{|f(x)| \mid x \in [a, b]\}$$

is the *uniform norm of f*. It follows that $f \cdot g \in H$ if $g \in H$, and so

$$M_f g = f \cdot g\,, \quad g \in H\,, \tag{5.31}$$

defines a mapping M_f from H to H. Since M_f is clearly linear in g and (5.30) implies

$$\|M_f g\| \leqslant \|f\|_u\|g\|\,,$$

it holds that M_f is a bounded operator on H, and the norm of M_f fulfills $\|M_f\| \leq \|f\|_u$. It is shown in Exercise 5.7 that, in fact, $\|M_f\| = \|f\|_u$ holds. The operator M_f is called the *multiplication operator* defined by the function f.

c) Let $H = \ell_2(\mathbb{N})$ and set

$$T(x_1, x_2, x_3, \ldots) = (0, x_1, x_2, x_3, \ldots)$$

for $(x_1, x_2, x_3, \ldots) \in \ell_2(\mathbb{N})$. It is evident that $T(x_1, x_2, x_3, \ldots) \in \ell_2(\mathbb{N})$ and that

$$\|T(x_1, x_2, x_3, \ldots)\| = \|(0, x_1, x_2, x_3, \ldots)\| \,.$$

Furthermore, it is easily verified that the mapping $T : H \to H$ so defined is linear. Hence, T is a bounded operator on H with norm 1, in fact an isometry. T is sometimes called the *right shift operator* on $\ell_2(\mathbb{N})$.

d) Let $H = L_2([-\pi, \pi])$ with the standard inner product normalized by the factor $\frac{1}{2\pi}$ as in Section 4.6, and let $(e_n)_{n\in\mathbb{Z}}$ denote the orthonormal basis, where $e_n(\theta) = e^{in\theta}$. Setting $D = -i\frac{d}{d\theta}$, we have

$$De_n = ne_n \,, \tag{5.32}$$

and D acts, of course, linearly on the subspace $\operatorname{span}\{e_n \mid n \in \mathbb{Z}\}$, whose closure is $L_2([-\pi, \pi])$. However, D is not a bounded operator on this space, since $\|D\| \geqslant \|De_n\| = |n|$ for all $n \in \mathbb{Z}$. As a consequence, D cannot be extended to a bounded operator on the whole space H. This is a general feature of so-called differential operators. As will be discussed at the end of Section 5.4, D has a natural extension to a larger domain, which is the appropriate domain for applications in quantum mechanics.

e) Let again $H = L_2(I)$ where $I = [a, b]$ is a closed bounded interval and let $\varphi : I \times I \to \mathbb{C}$ be continuous. For $x \in I$ and $f \in H$ we define

$$(\phi f)(x) = \int_I \varphi(x, y) f(y) dy \,. \tag{5.33}$$

In order to see that $(\phi f)(x)$ is well defined by this formula it is useful to note that the continuous function $\varphi_x : y \to \varphi(x, y)$ belongs to $L_2(I)$ for every $x \in I$, and that the right hand side of (5.33) is the inner product of $\overline{\varphi}_x$ and f in H. Applying the Cauchy-Schwarz inequality then yields

$$|(\phi f)(x)|^2 \leqslant \int_I |\varphi(x, y)|^2 dy \cdot \|f\|^2 \,, \tag{5.34}$$

which implies that (5.33) defines a function ϕf on I.

To show that ϕf belongs to H it suffices to prove that it is continuous. For this purpose, notice first that f is integrable since

$$\int_a^b |f(x)| dx \leqslant (b - a)^{\frac{1}{2}} \left(\int_a^b |f(x)|^2 dx \right)^{\frac{1}{2}}$$

as a consequence of Cauchy-Schwarz' inequality again. Let now $\epsilon > 0$ and choose $K > \int_I |f(x)| dx$. Since φ is continuous and I is closed and bounded there exists[1] a $\delta > 0$, such that $|\varphi(x, y) - \varphi(x', y')| < \frac{\epsilon}{K}$ for all $(x, y), (x', y') \in I \times I$ fulfilling $|x - x'|, |y - y'| < \delta$.

[1] It is known from basic analysis that any continuous complex valued function f on a closed and bounded subset of $\mathbb{R}^k$ is uniformly continuous in the sense that for all $\varepsilon > 0$ there is a $\delta > 0$ such that for all x, y in the domain of f, $\|x - y\| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$.

In particular, $|\varphi(x,y) - \varphi(x',y)| < \frac{\epsilon}{K}$ for all $y \in I$ and all $x, x' \in I$, such that $|x - x'| < \delta$. From this follows that

$$|(\phi f)(x) - (\phi f)(x')| \leq \int_I |(\varphi(x,y) - \varphi(x',y))f(y)|dy \leqslant \tfrac{\epsilon}{K} \int_I |f(x)|dx < \epsilon\,,$$

when $|x - x'| < \delta$. This shows that ϕf is continuous.

It now follows from (5.34) that $\phi f \in H$ and that

$$\|\phi f\|^2 \leqslant \int_I \int_I |\varphi(x,y)|^2 dydx \cdot \|f\|^2\,. \tag{5.35}$$

Since ϕf clearly depends linearly on f, we have shown that (5.33) defines a bounded operator ϕ on $L_2(I)$, whose norm fulfills

$$\|\phi\| \leqslant \big(\int_I \int_I |\varphi(x,y)|^2 dydx\big)^{\frac{1}{2}}\,.$$

In general, equality does not hold here.

The operator ϕ is called an *integral operator* and the function φ is called its *kernel*. Operators of this kind play an important role in the theory of differential equations.

We now aim at defining, for an arbitrary operator $A \in \mathcal{B}(H_1, H_2)$, the adjoint operator $A^* \in \mathcal{B}(H_2, H_1)$. It is not convenient to make use of a matrix representation of A, since it is not clear in the infinite dimensional case that the Hermitian conjugate matrix represents a bounded everywhere defined operator. Instead, we take (5.7) as the defining equation for A^* (see Theorem 5.5 below). First, however, some preparatory work is needed.

The *dual space* to the Hilbert space H, which is denoted by H^*, is defined as

$$H^* = \mathcal{B}(H, \mathbb{L})\,.$$

The elements of H^* are thus continuous linear functions from H to $\mathbb{L}$. These are also called *continuous linear forms* on H. Let us define, for $y \in H$, the function $\ell_y : H \to \mathbb{L}$ by

$$\ell_y(x) = \langle\, y|x\,\rangle\,, \tag{5.36}$$

where $\langle\,\cdot\,|\,\cdot\,\rangle$ is the inner product on H. Then ℓ_y is linear. The Cauchy-Schwarz inequality gives

$$|\ell_y(x)| \leq \|x\|\,\|y\|\,, \quad y \in H\,,$$

from which it follows that $\ell_y \in H^*$ and that

$$\|\ell_y\| \leq \|y\|\,. \tag{5.37}$$

The following theorem tells us that all continuous linear forms on H are of (or can be represented in) the form (5.36).

Theorem 5.4 (Riesz' representation theorem)**.** *Let H be a Hilbert space over $\mathbb{L}$. The mapping $y \to \ell_y$, where $\ell_y \in H^*$ is given by* (5.36)*, is a bijective isometry from H onto H^*, that is for every $\ell \in H^*$ there exists exactly one vector $y \in H$ such that $\ell = \ell_y$, and the norm identity*

$$\|\ell_y\| = \|y\| \tag{5.38}$$

holds.

Proof. As already noted $y \to \ell_y$ is a well defined mapping from H into H^*. From (5.36) follows that

$$\ell_{y+z} = \ell_y + \ell_z \text{ and } \ell_{\lambda y} = \overline{\lambda}\ell_y \tag{5.39}$$

for $y, z \in H$ and $\lambda \in \mathbb{L}$, i. e., the mapping $y \to \ell_y$ is conjugate linear.

That it is isometric can be seen as follows. If $y = 0$ then $\ell_y = 0$, so $\|\ell_y\| = \|y\| = 0$. For $y \neq 0$, set $x = \|y\|^{-1}y$ such that $\|x\| = 1$. Since

$$|\ell_y(x)| = |\langle\, y|\|y\|^{-1}y\,\rangle| = \|y\|$$

it follows that $\|\ell_y\| \geq \|y\|$. Together with (5.37) this proves (5.38). That (5.38) is equivalent to $y \to \ell_y$ being an isometry now follows from (5.39), since

$$\|\ell_y - \ell_z\| = \|\ell_{y-z}\| = \|y - z\|$$

for $y, z \in H$.

Knowing that the mapping $y \to \ell_y$ is isometric and conjugate linear, it is obviously also injective. It remains to show that for every $\ell \in H^*$ there exists a $y \in H$ such that $\ell = \ell_y$. Evidently, $\ell_{\underline{0}} = 0$ so it can be assumed that $\ell \neq 0$. Then $X = \ell^{-1}(\{\underline{0}\}) \neq H$, and since $\{\underline{0}\}$ is a closed subspace of $\mathbb{L}$ and ℓ is continuous, it follows that X is a closed subspace of H. Hence $H = X \oplus X^\perp$ by Theorem 4.18 and $X^\perp \neq \{\underline{0}\}$.

In fact $X^\perp$ is a one-dimensional subspace of H. In order to see this, choose $e \in X^\perp \setminus \{\underline{0}\}$ with $\|e\| = 1$, and let $z \in X^\perp$ be arbitrary. Then $\ell(e) \neq 0$ and

$$\ell(z) = \frac{\ell(z)}{\ell(e)}\ell(e) = \ell\left(\frac{\ell(z)}{\ell(e)}e\right)$$

which shows that $z - \frac{\ell(z)}{\ell(e)}e \in X$. On the other hand, it also holds that $z - \frac{\ell(z)}{\ell(e)}e \in X^\perp$, since $X^\perp$ is a subspace. Using $X \cap X^\perp = \{\underline{0}\}$ it follows that $z = \frac{\ell(z)}{\ell(e)}e$, which shows that e is a basis for $X^\perp$.

Every vector in H can therefore be written in the form $x + \lambda e$, where $x \in X$ and $\lambda \in \mathbb{L}$. The action of ℓ can then be written as

$$\ell(x + \lambda e) = \ell(x) + \lambda\ell(e) = \lambda\ell(e) = \langle\, \overline{\ell(e)}e \,|\, x + \lambda e\,\rangle\,,$$

which shows that $\ell = \ell_y$ where $y = \overline{\ell(e)}e$. Here, the linearity of ℓ as well as the fact that $x \perp e$ have been used. □

Dirac introduced in his classic monograph [5] the notation $\langle\, y|$ for ℓ_y and called it a *bra*-vector, whereas a vector x in H was called a *ket*-vector and denoted by $|x\rangle$. With this notation (5.36) takes the form

$$\langle\, y|(|x\,\rangle) = \langle\, y|x\,\rangle\,,$$

and by Theorem 5.4 the mapping

$$|x\rangle \to \langle x|$$

from H to H^* is bijective. In the following, we shall make use of the Dirac notation when advantageous, in particular in Chapter 6.

Notice that (5.38) is equivalent to

$$\|x\| = \sup\{|\langle\, x|y\,\rangle| \;\mid \|y\| \le 1\} \tag{5.40}$$

for $x \in H$. For later use we remark that, as a consequence, the norm of an operator $A \in \mathcal{B}(H_1, H_2)$ is given by

$$\|A\| = \sup\{|\langle\, Ax|y\,\rangle| \;\mid x \in H_1, y \in H_2\,,\, \|x\|\,,\, \|y\| \le 1\}\,. \tag{5.41}$$

We are now ready to introduce the *adjoint* A^* of a bounded operator A.

Theorem 5.5. *Let H_1 and H_2 be Hilbert spaces and $A \in \mathcal{B}(H_1, H_2)$. There exists a unique operator $A^* \in \mathcal{B}(H_2, H_1)$ which fulfills*

$$\langle\, Ax|y\,\rangle = \langle\, x|A^*y\,\rangle\,, \quad x \in H_1\,,\, y \in H_2\,. \tag{5.42}$$

Moreover,

$$\|A\| = \|A^*\|\,. \tag{5.43}$$

Proof. For any given $y \in H_2$ the mapping $x \to \langle\, y|Ax\,\rangle$ belongs to H_1^*, being a composition of A and ℓ_y. By Theorem 5.4 there exists a unique vector $z \in H_1$, such that

$$\langle\, y|Ax\,\rangle = \langle\, z|x\,\rangle\,, \quad x \in H_1\,. \tag{5.44}$$

Since z depends only on y for the given operator A, a mapping $A^* : H_2 \to H_1$ is defined by setting $A^*y = z$. By taking the complex conjugate of both sides of (5.44) it follows that A^* satisfies (5.42).

That A^* is a linear mapping can be seen as follows. For given $y, z \in H_2$ the defining relation

$$\langle\, Ax|y+z\,\rangle = \langle\, x|A^*(y+z)\,\rangle\,, \quad x \in H_1 \tag{5.45}$$

holds, while on the other hand

$$\begin{aligned} \langle\, Ax|y+z\,\rangle &= \langle\, Ax|y\,\rangle + \langle\, Ax|z\,\rangle \\ &= \langle\, x|A^*y\,\rangle + \langle\, x|A^*z\,\rangle \\ &= \langle\, x|A^*y + A^*z\,\rangle\,, \quad x \in H_1\,. \end{aligned} \tag{5.46}$$

Since A^* is uniquely determined by (5.42), it follows by comparing (5.45) and (5.46) that

$$A^*(y+z) = A^*y + A^*z\,.$$

Similarly,

$$A^*(\lambda y) = \lambda A^*y$$

for $\lambda \in \mathbb{L}$ and $y \in H_1$ follows from

$$\langle\, Ax|\lambda y\,\rangle = \langle\, x|A^*(\lambda y)\,\rangle$$

and

$$\langle\, Ax|\lambda y\,\rangle = \lambda\langle\, Ax|y\,\rangle = \lambda\langle\, x|A^*y\,\rangle = \langle\, x|\lambda A^*y\,\rangle$$

for $x \in H_1$.

That A^* is bounded and that $\|A\| = \|A^*\|$ follows immediately from (5.41) and (5.42). □

In case $H_1 = H_2 = H$ an operator $A \in \mathcal{B}(H)$ is called *self-adjoint* if $A = A^*$, which by (5.42) means that

$$\langle\, Ax|y\,\rangle = \langle\, x|Ay\,\rangle\,, \quad x, y \in H\,.$$

Self-adjoint operators play a particularly important role in operator theory, similar to that of symmetric matrices in linear algebra, as will be further discussed in the next section.

Example 5.6. a) Let M_a be the multiplication operator on $\ell_2(\mathbb{N})$ defined in Example 5.3 a). In order to determine the adjoint operator one may start by calculating the left hand side of (5.42) and rewrite it in the form of the right hand side as follows, for $x = (x_n)_{n\in\mathbb{N}}$ and $y = (y_n)_{n\in\mathbb{N}}$:

$$\langle\, M_a x \,|\, y\,\rangle = \sum_{n=1}^{\infty} \overline{a_n x_n}\, y_n = \sum_{n=1}^{\infty} \overline{x_n}\,\overline{a_n}\, y_n = \langle\, x \,|\, M_{\overline{a}}\, y\,\rangle\,,$$

where $\overline{a} = (\overline{a_n})_{n\in\mathbb{N}}$. This shows that

$$M_a^* = M_{\overline{a}}\,.$$

In particular, M_a is self-adjoint if and only if a is a real sequence.

b) Let M_f be the multiplication operator on $L_2([a,b])$ defined in Example 5.3 b). In a similar way as in a) one gets for $g, h \in L_2([a,b])$ that

$$\langle\, M_f g \,|\, h\,\rangle = \int_a^b \overline{f(x)g(x)}h(x)dx = \int_a^b \overline{g(x)}\,\overline{f(x)}h(x)dx = \langle\, g \,|\, M_{\overline{f}} h\,\rangle\,,$$

from which it follows that

$$M_f^* = M_{\overline{f}}\,,$$

and in particular M_f is self-adjoint if and only if f is a real function.

c) Let ϕ be the integral operator on $L_2(I)$ defined in Example 5.3 e). By interchanging the order of integrations (Fubini's theorem) at one stage one gets for $f, g \in L_2(I)$ that

$$\begin{aligned}\langle\, \phi f|g\,\rangle &= \int_I \overline{\Big(\int_I \phi(x,y)f(y)dy\Big)}g(x)dx = \int_I\int_I \overline{f(y)\phi(x,y)}g(x)dydx \\ &= \int_I\int_I \overline{f(y)\phi(x,y)}g(x)dxdy = \int_I \Big(\overline{f(x)}\int_I \overline{\phi(y,x)}g(y)dy\Big)dx\,.\end{aligned}$$

From this calculation one reads off the action of ϕ^* to be given by

$$(\phi^* g)(x) = \int_I \overline{\phi(y,x)}g(y)dy\,.$$

In other words, ϕ^* is the integral operator on $L_2(I)$ whose kernel is $\overline{\varphi(y,x)}$.

d) Let T denote the right shift operator on $\ell_2(\mathbb{N})$ defined in Example 5.3 c). For $x = (x_n)_{n\in\mathbb{N}}$ and $y = (y_n)_{n\in\mathbb{N}}$ in $\ell_2(\mathbb{N})$ one gets

$$\langle\, Tx|y\,\rangle = \overline{x_1}y_2 + \overline{x_2}y_3 + \overline{x_3}y_4 \ldots = \langle\, x|(y_2, y_3, y_4, \ldots)\,\rangle\,,$$

from which it follows that

$$T^*y = (y_2, y_3, y_4, \ldots) \, .$$

The operator T^* is called the *left shift operator* on $\ell_2(\mathbb{N})$. Note that $T^*T = 1$ and $TT^* \neq 1$ (see Exercise 5.8). It is not possible for an operator on a finite dimensional Hilbert space to fulfill both of these identities (why?).

e) Consider the differential operator D defined in Example 5.3 d). By partial integration one gets for functions $f, g \in \text{span}\{e_n \mid n \in \mathbb{Z}\}$ (with notation as in Example 5.3)

$$\begin{aligned} \langle \, Df | g \, \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{-i \frac{df}{d\theta}} \, g(\theta) \, d\theta \\ &= \Big[i \overline{f(\theta)} g(\theta) \Big]_{-\pi}^{\pi} - i \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{f(\theta)} \, \frac{dg}{d\theta} \, d\theta = \langle \, f | Dg \, \rangle \, , \end{aligned}$$

where it is used that the first term after the second equality sign vanishes since f and g are periodic with period 2π. Below we explain how to define the adjoint for unbounded operators such as D, in general. It will transpire that despite the identity above, D is not self-adjoint, but it can be extended to an operator $\bar{D}$ such that $\bar{D} = \bar{D}^*$, as will be further discussed is Example 5.19.

The following useful properties of adjoint operators are worth noting:

i) $(A + B)^* = A^* + B^*$, for $A, B \in \mathcal{B}(H_1, H_2)$

ii) $(\lambda A)^* = \overline{\lambda} A^*$, for $A \in \mathcal{B}(H_1, H_2)$

iii) $(BA)^* = A^* B^*$, for $A \in \mathcal{B}(H_1, H_2), B \in \mathcal{B}(H_2, H_3)$

iv) $A^{**} = A$, for $A \in \mathcal{B}(H_1, H_2)$,

where H_1, H_2 and H_3 are (separable) Hilbert spaces. Properties i) and ii) can be shown in a way similar to the linearity of A^* in the proof of Theorem 5.5 and are left for the reader. Property iii) follows similarly by comparing

$$\langle \, BAx | y \, \rangle = \langle \, x | (BA)^* y \, \rangle$$

and

$$\langle \, BAx | y \, \rangle = \langle \, B(Ax) | y \, \rangle = \langle \, Ax | B^* y \, \rangle = \langle \, x | A^* (B^* y) \, \rangle = \langle \, x | A^* B^* y \, \rangle$$

for $x \in H_1$, $y \in H_3$.

By complex conjugation of both sides of (5.42) one gets

$$\langle \, A^* y | x \, \rangle = \langle \, y | Ax \, \rangle \, , \quad y \in H_2 \, , \, x \in H_1 \, .$$

Comparison with (5.42) finally shows that $A^{**} = A$.

We end this section with a brief discussion of adjoints of unbounded operators. With the exception of the operator D in Example 5.3 d) we have up to now discussed operators defined everywhere on a Hilbert space. An operator A from H_1 into H_2 is called *densely*

defined if its domain of definition $\mathcal{D}(A)$ is a subspace whose closure equals H_1 or, equivalently, whose orthogonal complement is $\{\underline{0}\}$. If A is densely defined and bounded, i. e., fulfills (5.2) for all $x \in \mathcal{D}(A)$, then it can be extended uniquely to a bounded operator defined everywhere on H_1, see Exercise 5.9. In this case the adjoint of A can be defined simply as the adjoint of its extension to H_1. For unbounded densely defined operators, such as D in Example 5.3 d), this method is not applicable. Nevertheless, a proper definition of the adjoint A^* can be given as follows.

As the domain of A^* one takes the subspace

$$\mathcal{D}(A^*) = \{y \in H_2 \mid x \to \langle\, y|Ax\,\rangle \text{ is a bounded linear form on } \mathcal{D}(A)\}\,.$$

If $y \in \mathcal{D}(A^*)$ the extension result quoted above implies that the linear form $x \to \langle\, y|Ax\,\rangle$ has a unique extension to a bounded linear form on H_1. Hence, by Theorem 5.4, there exists a unique vector A^*y in H_1 such that

$$\langle\, Ax|y\,\rangle = \langle\, x|A^*y\,\rangle\,, \quad x \in \mathcal{D}(A)\,. \tag{5.47}$$

This defines A^* on $\mathcal{D}(A^*)$. That $\mathcal{D}(A^*)$ is a subspace of H_2 and that A^* is linear on this subspace is shown in the same way as for bounded operators. Note that it may happen that A^* is not densely defined.

With this definition a densely defined operator A from H into H is called self-adjoint if $A = A^*$, that is if

$$\mathcal{D}(A^*) = \mathcal{D}(A) \quad \text{and} \quad \langle\, Ax|y\,\rangle = \langle\, x|Ay\,\rangle \text{ for } x, y \in \mathcal{D}(A)\,.$$

As seen in Example 5.6 e) the operator D with domain $\operatorname{span}\{e_n \mid n \in \mathbb{Z}\}$ satisfies the latter of these requirements, which is expressed by saying that D is *symmetric*. However, it does not satisfy the first requirement concerning domains. Although D is densely defined, since $\{e_n \mid n \in \mathbb{Z}\}$ is an orthonormal basis for $L_2([-\pi, \pi])$, the domain of D^* is bigger as will be further discussed in Example 5.19. Thus D is symmetric but not self-adjoint.

5.4 Diagonalizable operators and self-adjoint operators.

This section is devoted to a discussion of some basic notions and results relating to diagonalizable operators on separable Hilbert spaces. A rather straight-forward generalization of the finite dimensional version of the notion of a diagonalizable operator discussed in Sections 5.1 and 5.2 will be used. It is a basic result from linear algebra that every self-adjoint operator on a finite dimensional real Hilbert space is diagonalizable. The corresponding result for the infinite dimensional case (the spectral theorem) requires a further extension of the notion of a diagonalizable operator than we offer in these notes. In technical terms the main focus will be on operators with discrete spectrum (consisting of the eigenvalues of the operator), although examples of operators with continuous spectrum will also occur.

Definition 5.7. Let $A \in \mathcal{B}(H)$. A scalar $\lambda \in \mathbb{L}$ is called an *eigenvalue* of A, if there exists a vector $x \neq \underline{0}$ in H, such that

$$Ax = \lambda x\,. \tag{5.48}$$

A vector $x \neq \underline{0}$ fulfilling (5.48) is called an *eigenvector* for A corresponding to the eigenvalue λ, and the subspace of all vectors fulfilling (5.48) (that is the set of eigenvectors and the null-vector) is called the *eigenspace* corresponding to λ and is denoted by $E_\lambda(A)$.

The *kernel* or *null-space* $N(A)$ for an operator $A \in \mathcal{B}(H_1, H_2)$ is defined as

$$N(A) = \{x \in H_1 \mid Ax = \underline{0}\}\,.$$

which is the same thing as $E_0(A)$. By continuity of A it follows that $N(A) = A^{-1}(\{\underline{0}\})$ is a closed subspace of H_1, since $\{\underline{0}\}$ is a closed subspace of H_2. This also follows from

$$N(A) = (A^*(H_2))^\perp\,,$$

which is a direct consequence of the definition of A^*.

In particular, it follows that $\lambda \in \mathbb{L}$ is an eigenvalue of $A \in \mathcal{B}(H)$ if and only if

$$N(A - \lambda) \neq \{\underline{0}\}\,,$$

and the eigenspace $E_\lambda(A) = N(A - \lambda)$ is a closed subspace of H.

Definition 5.8. An operator $A \in \mathcal{B}(H)$ is called (unitarily) *diagonalizable*, if there exists an orthonormal basis $(f_i)_{i\in\mathbb{N}}$ for H consisting of eigenvectors for A. Equivalently, this means that A is represented by a diagonal matrix w. r. t. $(f_i)_{i\in\mathbb{N}}$, which can be seen as in the previous section. The basis $(f_i)_{i\in\mathbb{N}}$ is then said to *diagonalize* A.

The following characterization of bounded diagonalizable operators is useful.

Theorem 5.9. *An operator $A \in \mathcal{B}(H)$ is diagonalizable if and only if there exists an orthonormal basis $(f_i)_{i\in\mathbb{N}}$ for H and a bounded sequence $(\lambda_i)_{i\in\mathbb{N}}$ in $\mathbb{L}$, such that*

$$Ax = \sum_{i=1}^{\infty} \lambda_i \langle\, f_i | x \,\rangle f_i\,, \quad x \in H\,. \tag{5.49}$$

In that case $\lambda_1, \lambda_2, \lambda_3, \ldots$ are exactly the eigenvalues of A (possibly with repetitions) and the eigenspace corresponding to a given eigenvalue $\lambda \in \mathbb{L}$ is given by

$$E_\lambda(A) = \overline{\operatorname{span}}\{f_i \mid i \in \mathbb{N}\,, \lambda_i = \lambda\}\,. \tag{5.50}$$

Moreover,

$$\|A\| = \sup\{|\lambda_i| \mid i \in \mathbb{N}\}\,. \tag{5.51}$$

Proof. Suppose $A \in \mathcal{B}(H)$ is diagonalizable and let $(f_i)_{i\in\mathbb{N}}$ be an orthonormal basis consisting of eigenvectors for A, such that $Af_i = \lambda_i f_i$, $i \in \mathbb{N}$. The image of $x = \sum_{i=1}^{\infty} \langle\, f_i | x \,\rangle f_i \in H$ is then given by (see (5.26))

$$Ax = \sum_{i=1}^{\infty} \langle\, f_i | x \,\rangle A f_i = \sum_{i=1}^{\infty} \lambda_i \langle\, f_i | x \,\rangle f_i$$

as desired. Since

$$|\lambda_i| = \|\lambda_i f_i\| = \|Af_i\| \leq \|A\| \, \|f_i\| = \|A\|$$

for $i \in \mathbb{N}$, it follows that

$$\sup\{|\lambda_i| \mid i \in \mathbb{N}\} \leq \|A\| \,. \tag{5.52}$$

Assume, conversely, that A is given by (5.49), where $(f_i)_{i\in\mathbb{N}}$ is an orthonormal basis, and the sequence $(\lambda_i)_{i\in\mathbb{N}}$ is bounded, such that

$$\sup\{|\lambda_i| \mid i \in \mathbb{N}\} \equiv M < +\infty \,.$$

Then

$$\|Ax\|^2 = \sum_{i=1}^{\infty} |\lambda_i \langle \, f_i | x \, \rangle|^2 \leq M^2 \|x\|^2 \,,$$

which shows that

$$\|A\| \leq M = \sup\{|\lambda_i| \mid i \in \mathbb{N}\} \,. \tag{5.53}$$

Inserting $x = f_i$ into (5.49) it is seen that

$$Af_i = \lambda_i f_i \,.$$

Hence $\lambda_1, \lambda_2, \ldots$ are eigenvalues of A.

That there are no other eigenvalues is seen as follows. Suppose

$$Ax = \lambda x \,,$$

where $\lambda \in \mathbb{L}$ and $x \in H \setminus \{\underline{0}\}$. Then

$$\sum_{i=1}^{\infty} \lambda_i \langle \, f_i | x \, \rangle f_i = \lambda x = \sum_{i=1}^{\infty} \lambda \langle \, f_i | x \, \rangle f_i \,,$$

which implies

$$\lambda_i \langle \, f_i | x \, \rangle = \lambda \langle \, f_i | x \, \rangle \text{ for all } i \in \mathbb{N} \,.$$

Since $x \neq \underline{0}$ there exists $i_0 \in \mathbb{N}$, such that $\langle \, f_{i_0} | x \, \rangle \neq 0$, and this implies $\lambda = \lambda_{i_0}$ and that $\langle \, f_i | x \, \rangle = 0$ for $\lambda_i \neq \lambda_{i_0}$. This proves (5.50).

Finally, (5.51) follows from (5.52) and (5.53). □

Example 5.10. Let H_0 be a closed subspace of a Hilbert space H and let Px denote the orthogonal projection of a vector $x \in H$ onto H_0, that is $Px \in H_0$ is determined by $x - Px = (1 - P)x \in H_0^{\perp}$. It is clear from this characterization of Px that it depends linearly on x, and hence P is a linear operator on H called the *orthogonal projection onto* H_0. An expression for Px is obtained from Theorem 4.16 or equation (4.31) as

$$Px = \sum_{i\in I} \langle \, f_i | x \, \rangle f_i \,, \qquad x \in H \,, \tag{5.54}$$

where $(f_i)_{i\in I}$ is any orthonormal basis for H_0.

Letting $(f_j)_{j\in J}$ be an orthonormal basis for $H_0^{\perp}$ the two orthonormal bases together form an orthonormal basis $(f_i)_{i\in I\cup J}$ for H by Proposition 4.19 (and since H is separable it can be assumed that $I \cup J = \mathbb{N}$). Together with (5.54) this shows that P has the form (5.49) with $\lambda_i = 1$ for $i \in I$ and $\lambda_i = 0$ for $i \in J$. In particular, the only eigenvalues of P are 1 and 0, provided $I \neq \emptyset$ and $J \neq \emptyset$, and the corresponding eigenspaces are H_0 and $H_0^{\perp}$, respectively. (Clearly, $P = 0$ if $I = \emptyset$ whereas $P = 1$ for $J = \emptyset$).

Returning to the diagonalizable operator A given by (5.49) and noting that the orthogonal projection P_i onto the subspace $H_i = \operatorname{span}\{f_i\}$ is given by

$$P_i x = \langle\, f_i | x \,\rangle f_i \,, \qquad x \in H \,, \tag{5.55}$$

it follows that the representation (5.49) of A can be rewritten in the form

$$Ax = \sum_{i=1}^{\infty} \lambda_i P_i x \,, \qquad x \in H \,, \tag{5.56}$$

which is expressed by writing

$$A = \sum_{i=1}^{\infty} \lambda_i P_i \,. \tag{5.57}$$

It is worth noting that the series (5.57) does generally not converge w. r. t. the norm on $\mathcal{B}(H)$, i. e. the operator norm. In fact, this is the case only if $\lambda_i \to 0$ for $i \to \infty$. Otherwise, (5.57) only makes sense when interpreted as (5.56).

The operator A given by (5.57) is represented by the diagonal matrix $\Delta(\lambda_1, \lambda_2, \dots)$ w. r. t. $(f_i)_{i\in\mathbb{N}}$. Hence the adjoint operator A^* is represented by the diagonal matrix $\Delta(\overline{\lambda}_1, \overline{\lambda}_2, \dots)$ w. r. t. $(f_i)_{i\in\mathbb{N}}$. That is

$$A^* = \sum_{i=1}^{\infty} \overline{\lambda}_i P_i \,. \tag{5.58}$$

In particular, A is self-adjoint if and only if its eigenvalues are real. This is e. g. the case for orthogonal projections, cf. Example 5.10.

At this point a brief remark on Dirac notation is appropriate. Given a ket-vector $|x\rangle$ and a bra-vector $\langle y|$, the linear operator on H defined by

$$|z\rangle \to \langle y|z\rangle |x\rangle$$

is naturally denoted by $|x\rangle\langle y|$. In particular, if $|e\rangle$ is a unit vector in H, the orthogonal projection onto the subspace spanned by $|e\rangle$ is $|e\rangle\langle e|$. More generally, if A is a diagonalizable operator, the form (5.49) is written in Dirac notation as

$$A = \sum_{i=1}^{\infty} \lambda_i |f_i\rangle\langle f_i| \,.$$

As a special case, the expression (5.54) for the orthogonal projection P onto a closed subspace H_0 of H takes the form

$$P = \sum_{i=1}^{\infty} |f_i\rangle\langle f_i| \,,$$

where $(f_i)_{i\in\mathbb{N}}$ is an arbitrary orthonormal basis for H_0. When applied to $|x\rangle$, this is just another way of expressing formula (4.31) for the orthogonal projection:

$$P|x\rangle = \sum_{i=1}^{\infty} |f_i\rangle\langle f_i|x\rangle \,.$$

If the eigenvalues λ_i are all different in the preceding discussion, in particular in (5.57), then P_i is the orthogonal projection onto the eigenspace $E_{\lambda_i}(A)$, which in this case is one-dimensional. This latter property is also expressed by saying that the eigenvalues are *non-degenerate*. In general, however, the eigenvalues are not non-degenerate and it may be advantageous to combine the terms in (5.57) corresponding to identical eigenvalues into one term. This can be done by letting μ_i, $i = 1, 2, 3, \ldots$ denote the sequence of *different* eigenvalues of A in some order and defining Q_i to be the orthogonal projection onto the eigenspace $E_{\mu_i}(A)$. Using Proposition 4.19 it follows by similar arguments as given above that

$$A = \sum_i \mu_i Q_i \qquad \text{and} \qquad \sum_i Q_i = 1\,, \tag{5.59}$$

when interpreted in the same fashion as equation (5.57). The orthogonal projections Q_i are called the *spectral projections* of A.

In the remainder of this section various versions (but not the most general one) of the so-called spectral theorem will be discussed, Its main virtue is to ensure diagonalizability of important classes of operators, of which special focus will be on self-adjoint operators.

That the eigenvalues of a self-adjoint operator are real, holds generally (and not only for diagonalizable operators). Likewise, the eigenspaces corresponding to different eigenvalues are orthogonal, which for diagonalizable operators is evident from (5.50) in Theorem 5.9. These facts are contained in the following lemma.

Lemma 5.11. *Let $A \in \mathcal{B}(H)$ be self-adjoint. Then*

a) Every eigenvalue of A is real.

b) If λ_1 and λ_2 are two different eigenvalues of A, then the two corresponding eigenspaces are orthogonal.

c) If H_0 is a subspace of H such that $A\,H_0 \subseteq H_0$, then $A(H_0^\perp) \subseteq H_0^\perp$.

Proof. a) Assume $Ax = \lambda x$, $x \in H \setminus \{\underline{0}\}$. Then

$$\lambda \|x\|^2 = \langle\, x|\lambda x\,\rangle = \langle\, x|Ax\,\rangle = \langle\, Ax|x\,\rangle = \langle\, \lambda x|x\,\rangle = \overline{\lambda}\|x\|^2\,,$$

which implies that $\lambda = \overline{\lambda}$, since $\|x\| \neq 0$.

b) Assume $Ax_1 = \lambda_1 x_1$ and $Ax_2 = \lambda_2 x_2$. Since $\lambda_1, \lambda_2 \in \mathbb{R}$ by a) it follows that

$$\begin{aligned}(\lambda_1 - \lambda_2)\langle\, x_1|x_2\,\rangle &= \langle\, \lambda_1 x_1|x_2\,\rangle - \langle\, x_1|\lambda_2 x_2\,\rangle \\ &= \langle\, Ax_1|x_2\,\rangle - \langle\, x_1|Ax_2\,\rangle \\ &= \langle\, x_1|Ax_2\,\rangle - \langle\, x_1|Ax_2\,\rangle = 0\,,\end{aligned}$$

which gives $\langle\, x_1|x_2\,\rangle = 0$, since $\lambda_1 - \lambda_2 \neq 0$. Hence $x_1 \perp x_2$.

c) Let $x \in H_0^\perp$ and $y \in H_0$ be arbitrary. Then

$$\langle\, Ax|y\,\rangle = \langle\, x|Ay\,\rangle = 0$$

since $Ay \in H_0$. This shows that $Ax \in H_0^\perp$, and the claim is proven. □

A subspace H_0 as in Lemma 5.11 c) is called an *invariant subspace* for A. Thus c) above can be expressed by saying that the orthogonal complement to an invariant subspace for a self-adjoint operator $A \in \mathcal{B}(H)$ is also an invariant subspace for A. If H_0 is closed such that $H = H_0 \oplus H_0^\perp$, this means that A can be split into two operators $A_1 \in \mathcal{B}(H_0)$ and $A_2 \in \mathcal{B}(H_0^\perp)$ in the sense that

$$Ax = A_1 x_1 + A_2 x_2$$

for $x = x_1 + x_2$, $x_1 \in H_0$, $x_2 \in H_0^\perp$. We then write

$$A = A_1 \oplus A_2 \, .$$

These remarks suffice to prove the finite dimensional versions of the **spectral theorem**.

Theorem 5.12. *Every self-adjoint operator $A : H \to H$ on a finite dimensional complex Hilbert space H is diagonalizable.*

Proof. Choose an arbitrary orthonormal basis α for H and let $\underline{\underline{A}}$ be the matrix representing A w. r. t. α. By the fundamental theorem of algebra the characteristic polynomial

$$p(\lambda) = \det(\underline{\underline{A}} - \lambda I)$$

has a root, and hence A has an eigenvalue. Call it λ_1 and let e_1 be a corresponding normalised eigenvector. The subspace $H_0 = \mathrm{span}\{e_1\}$ is then clearly invariant under A. Writing accordingly $A = A_1 \oplus A_2$ as above, it is clear that A_2 is a self-adjoint operator on the subspace $H_0^\perp$ of one dimension lower than H. Hence the proof can be completed by induction. □

The assumption that H is a complex vector space is essential in the argument above, since the fundamental theorem of calculus only garantees the existence of complex roots. However, by Lemma 5.12 a) the roots are real for any symmetric real matrix and the argument carries through also in the real case. For completeness, we state the result in the following theorem.

Theorem 5.13. *Every self-adjoint operator A on a finite dimensional real Hilbert space H is diagonalizable.*

Theorem 5.12 is easily generalizable to so-called *normal operators*, i. e., operators commuting with their adjoint.

Theorem 5.14. *Suppose A is an operator on a finite dimensional complex Hilbert space such that*

$$AA^* = A^*A$$

Then A is diagonalizable.

Proof. Write

$$A = U + iV \text{ where } U = \frac{1}{2}(A + A^*) \text{ and } V = \frac{1}{2i}(A - A^*)\,.$$

Then U and V are self-adjoint and they commute,

$$UV = VU\,,$$

since A and A^* commute. By Theorem 5.12, U is diagonalizable, and we can write

$$U = \lambda_1 P_1 + \cdots + \lambda_k P_k\,,$$

where $\lambda_1, \ldots, \lambda_k$ are the different eigenvalues of U and P_i is the orthogonal projection onto $E_{\lambda_i}(U)$. Since V commutes with U it follows that each eigenspace $E_{\lambda_i}(U)$ is invariant under V (see Exercise 5.10), and evidently V is self-adjoint when restricted to this subspace. By Theorem 5.12 there exists an orthonormal basis of eigenvectors for V for each of the eigenspaces $E_{\lambda_i}(U)$. Together these bases form an orthonormal basis consisting of eigenvectors for both U and V and hence they are also eigenvectors for A. □

Since unitary operators clearly are normal the theorem can be applied to conclude that:

Corollary 5.15. *Any unitary operator on a finite dimensional complex Hilbert space is diagonalizable.*

It is worth noting that the corresponding result does not hold for orthogonal operators on a real Hilbert space. For instance, a rotation in the plane through an angle different from 0 and π has no eigenvalues at all (see also Example 5.2 b)).

An example of an infinite dimensional analogue of Theorems 5.12 and 5.13 is the following *spectral theorem for Hilbert-Schmidt operators*, which turns out to by quite usefull in various contexts, but whose proof is beyond the scope of these notes. Before stating the theorem let us define the class of operators in question.

Definition 5.16. An operator $A : H \to H$ on a Hilbert space H is said to be a **Hilbert-Schmidt operator** if

$$\sum_{i\in\mathbb{N}} \|Ae_i\|^2 < \infty$$

for any orthonormal basis $(e_i)_{i\in\mathbb{N}}$ for H.

It can be shown that if the condition holds for one orthonormal basis then it holds for all orthonormal bases. Important examples of Hilbert-Schmidt operators are provided by the integral opertors as defined in Example 5.3 e) for which one can show that

$$\sum_{i\in\mathbb{N}} \|Ae_i\|^2 = \int_a^b \int_a^b |\varphi(x,y)|^2 dxdy\,.$$

The announced spectral theorem can then be formulated as follows.

Theorem 5.17. *Every self-adjoint Hilbert-Schmidt operator A on a separable Hilbert space H is diagonalizable.*

The difficult step in the proof of this result is to establish the existence of an eigenvalue. Having done so, the proof can be completed in a similar way as in the finite dimensional case. Details of the proof can be found in e. g. [6].

We end this section by two examples of which the first contains some further discussion of multiplication operators providing prototypes of self-adjoint operators that cannot be diagonalized in the sense of Definition 5.8, and the second one discusses self-adjointness of the unbounded differential operator of Example 5.3 d).

Example 5.18. Let $H = L_2([0,3])$, and let $f : [0,3] \to \mathbb{R}$ denote the function

$$f(x) = x\,, \qquad x \in [0,3]\,.$$

Since f is real, M_f is a self-adjoint operator by Example 5.6 b).

The operator M_f has no eigenvalues, which is seen as follows. Assume $M_f g = \lambda g$ for some $\lambda \in \mathbb{R}$ and some function $g \in H$. This means that $(f-\lambda)g = 0$ almost everywhere. But since $f - \lambda$ is $\neq 0$ almost everywhere, it follows that $g = 0$ almost everywhere, that is $g = \underline{0} \in H$, which shows that λ is not an eigenvalue.

In particular, it follows that M_f is not diagonalizable in the sense of Definition 5.8.

Let us next consider the multiplication operator M_{f_1} defined by the function $f_1 : [0,3] \to \mathbb{R}$, where

$$f_1(x) = \begin{cases} x & \text{for } 0 \leqslant x \leqslant 1 \\ 1 & \text{for } 1 \leqslant x \leqslant 2 \\ x-1 & \text{for } 2 \leqslant x \leqslant 3\,. \end{cases}$$

Note that f_1 is constant and equal to 1 on the interval $[1,2]$ and strictly increasing outside this interval. Since f_1 is real M_{f_1} is self-adjoint. We claim that 1 is the only eigenvalue of M_{f_1}. Indeed, if $\lambda \neq 1$ then $f_1 - \lambda \neq 0$ almost everywhere in $[0,3]$, and it follows as above that λ is not an eigenvalue. On the other hand, $f_1 - 1 = 0$ on $[1,2]$ and $f_1 - 1 \neq 0$ outside $[1,2]$. It follows that $M_{f_1} g = g$, if and only if $g = 0$ almost everywhere in $[0,3] \setminus [1,2]$. The set of functions fulfilling this requirement is an infinite dimensional closed subspace of H, that can be identified with $L_2([1,2])$, and which hence equals the eigenspace $E_1(M_{f_1})$.

Clearly, M_{f_1} is not diagonalizable in the sense of Definition 5.8.

The primary lesson to be drawn from this example is that self-adjoint operators in general are not diagonalizable in the sense of Definition 5.8. However, as mentioned previously, it is a principal result of analysis, called the *spectral theorem for self-adjoint operators*, that they are diagonalizable in a generalized sense. It is outside the scope of this course to formulate and even less to prove this result. Interested readers are referred to e. g. [13].

Example 5.19. Consider again the operator D defined in Example 5.3 d). Keeping the same notation, it follows from equation (5.32) that D is given on the domain $\mathrm{span}\{e_n \mid n \in \mathbb{Z}\}$ by

$$Df = \sum_{n\in\mathbb{Z}} n \langle\, e_n | f \,\rangle\, e_n\,. \tag{5.60}$$

This formula can be used to define an extension $\bar{D}$ of D by extending the domain of definition to the largest subspace of $L_2([-\pi,\pi])$ on which the right hand side is convergent, i. e., to the set

$$\mathcal{D}(\bar{D}) = \{f \in L_2([0,2\pi]) \mid \sum_{n\in\mathbb{Z}} n^2 |\langle\, e_n | f\,\rangle|^2 < \infty\}\,,$$

on which $\bar{D}f$ is given by the right hand side of (5.60). Defining the adjoint $\bar{D}^*$ as at the end of Section 5.3 it is then not difficult to show that $\bar{D}$ is a self-adjoint operator, see Exercise 5.17.

5.5 The Fourier transformation as a unitary operator.

The Fourier transformation is an indispensable tool in both physics and mathematics. In this section we give a brief introduction to the Fourier transformation in one variable with special emphasis on its interpretation as an operator on $L_2(\mathbb{R})$. The discussion will be further extended to several variables in Chapter 6.

First, we need a generalization of the notion of a unitary operator as defined in Section 5.2 in the finite dimensional case.

Definition 5.20. Let H_1 and H_2 be Hilbert spaces. An operator $U \in \mathcal{B}(H_1, H_2)$ is unitary if it is bijective and

$$U^{-1} = U^*$$

or, equivalently, if

$$U^*U = \mathrm{id}_{H_1} \quad \text{and} \quad UU^* = \mathrm{id}_{H_2}\,.$$

Note that, contrary to the finite dimensional case, the relation $U^*U = \mathrm{id}_{H_1}$ does not imply the relation $UU^* = \mathrm{id}_{H_2}$, or vice versa, if H_1 and H_2 are of infinite dimension, see Exercise 5.8. As a consequence, Theorem 5.1 needs to be modified as follows in order to cover the infinite dimensional case.

Theorem 5.21. *Let $U : H_1 \to H_2$ be a bounded linear operator. The following four statements are equivalent.*

a) U is unitary.

b) U is surjective and preserves the inner product.

c) U is surjective and isometric.

d) U maps an orthonormal basis for H_1 into an orthonormal basis for H_2.

The proof is identical to that of Theorem 5.1 with minor modifications and is left to the reader.

Example 5.22. Let H be a Hilbert space with orthonormal basis $(e_i)_{i\in\mathbb{N}}$. The mapping $C : H \to \ell_2(\mathbb{N})$ which maps the vector x to its coordinate sequence w. r. t. $(e_i)_{i\in\mathbb{N}}$, that is

$$Cx = (\langle\, e_i | x\,\rangle)_{i\in\mathbb{N}}$$

is unitary. Indeed, C is an isometry by Parseval's equality (4.27), and C is surjective because, if $a = (a_i)_{i\in\mathbb{N}} \in \ell_2(\mathbb{N})$ then $Cx = a$, where $x = \sum_{i=1}^{\infty} a_i e_i$.

Thus every separable Hilbert space can be mapped onto $\ell_2(\mathbb{N})$ by a unitary operator. We say that every separable Hilbert space is unitarily isomorphic to $\ell_2(\mathbb{N})$.

The Fourier transformation provides an additional, highly non-trivial, example of a unitary operator, which we turn to now without giving detailed proofs. For a full account the reader is referred to volume II of [13].

If f is an integrable function on $\mathbb{R}$, its *Fourier transform* $\hat{f}$ is the function on $\mathbb{R}$ defined by

$$\hat{f}(p) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ipx}\, dx\,, \quad p \in \mathbb{R}\,. \tag{5.61}$$

The first fundamental result about the Fourier transform is that, if f is a C^∞ function vanishing outside some bounded interval, then $\hat{f}$ is a C^∞ function that is both integrable and square integrable, and

$$\int_{\mathbb{R}} |f(x)|^2\, dx = \int_{\mathbb{R}} |\hat{f}(p)|^2\, dp\,. \tag{5.62}$$

It is clear that the set $C_0^\infty(\mathbb{R})$ of C^∞ functions that vanish outside some bounded interval is a subspace of $L_2(\mathbb{R})$ and that the mapping $f \to \hat{f}$ is linear on this subspace. As mentioned in Example 4.12 b), the closure of $C_0^\infty(\mathbb{R})$ equals $L_2(\mathbb{R})$. Hence, by the extension result of Exercise 5.9, the mapping $f \to \hat{f}$ has a unique extension to an operator $\mathcal{F} : L_2(\mathbb{R}) \to L_2(\mathbb{R})$, which moreover is isometric by (5.62). It is called the *Fourier transformation on* $L_2(\mathbb{R})$.

The second result of fundamental importance is the *inversion theorem*, which states that for $f \in C_0^\infty(\mathbb{R})$ the identity

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(p) e^{ipx}\, dp\,, \quad x \in \mathbb{R} \tag{5.63}$$

holds. In operator language this equation can also be written as

$$\mathcal{F} \circ \bar{\mathcal{F}}(f) = f\,, \tag{5.64}$$

where $\bar{\mathcal{F}}$ is defined by the same procedure as $\mathcal{F}$ except that p is replaced by $-p$ in (5.61), i. e.

$$\bar{\mathcal{F}}(f)(p) = \mathcal{F}(f)(-p)\,.$$

By continuity of $\mathcal{F}$ and $\bar{\mathcal{F}}$ equation (5.64) holds for all $f \in L_2(\mathbb{R})$. In particular, $\mathcal{F}$ is surjective and by Theorem 5.20 it follows that $\mathcal{F}$ is a unitary operator on $L_2(\mathbb{R})$, whose adjoint is $\bar{\mathcal{F}}$.

The outcome of the preceding discussion can be stated as the following main result of this section.

Theorem 5.23. *The Fourier transformation $\mathcal{F}$ defined for $f \in C_0^\infty(\mathbb{R})$ by*

$$\mathcal{F}(f)(p) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ipx}\, dx\,, \quad p \in \mathbb{R}\,,$$

extends uniquely to a unitary operator on $L_2(\mathbb{R})$, whose adjoint is given by

$$\mathcal{F}^*(f)(p) = \mathcal{F}(f)(-p)\,, \quad p \in \mathbb{R}\,.$$

Exercises

Exercise 5.1. Let H be a 3-dimensional complex Hilbert space and let $A : H \to H$ be the operator represented by the matrix

$$\begin{pmatrix} 2 & 3+i & 2i \\ 0 & 2 & 0 \\ i & 1-i & 2 \end{pmatrix}$$

w. r. t. an orthonormal basis $\alpha = (e_1, e_2, e_3)$. Find the matrix that represents A^* w. r. t. α, and determine $A^*(e_1 + ie_2 - e_3)$.

Exercise 5.2. Let A be the linear operator on $\mathbb{C}^3$ represented by the matrix

$$\begin{pmatrix} 2 & 0 & -i \\ 0 & 2 & 0 \\ i & 0 & 2 \end{pmatrix}$$

w. r. t. the canonical basis for $\mathbb{C}^3$. Show that A is self-adjoint, and find its eigenvalues as well as an orthonormal basis for $\mathbb{C}^3$ which diagonalizes A.

Exercise 5.3. Let H and α be as in Exercise 5.1 and let U be the linear operator on H represented by the matrix

$$\begin{pmatrix} \frac{1}{6}+\frac{i}{2} & \frac{1}{6}-\frac{i}{2} & -\frac{2}{3} \\ \frac{1}{6}-\frac{i}{2} & \frac{1}{6}+\frac{i}{2} & -\frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \end{pmatrix} .$$

w. r. t. α.

Show that U is unitary.

Show that 1, -1, i are eigenvalues of U, and find an orthonormal basis for H which diagonalizes U.

Exercise 5.4. Let $A : E \to E'$ be a bounded operator between inner product spaces as defined by equation (5.2).

Show that the norm $\|A\|$ is given by (5.3). Show also the following properties of the norm:

$$\begin{aligned} \|A\| &> 0 \text{ if } A \neq 0\,, \\ \|\lambda A\| &= |\lambda| \|A\|\,, \\ \|A + B\| &\leqslant \|A\| + \|B\|\,, \\ \|CA\| &\leqslant \|C\| \|A\|\,, \end{aligned}$$

where $\lambda \in \mathbb{L}$ and the operators $B : E \to E'$, $C : E' \to E''$ are bounded (E'' being an inner product space).

Exercise 5.5. Show that a linear operator $A : E \to E'$ between inner product spaces is bounded if and only if it is continuous.

Hint. Show first by using linearity that A is continuous if and only if it is continuous at $\underline{0}$. To show that continuity implies boundedness it may be useful to argue by contradiction.

Exercise 5.6. Let H be a finite dimensional Hilbert space over $\mathbb{L}$ and define the following three subsets of operators on H:

$$\begin{aligned} GL(H) &= \{A \in \mathcal{B}(H) \mid A \text{ is invertible, i.e. } A \text{ is bijective}\}\,, \\ O(H) &= \{O \in \mathcal{B}(H) \mid O \text{ is orthogonal}\}\,, \text{ if } \mathbb{L} = \mathbb{R}\,, \\ U(H) &= \{U \in \mathcal{B}(H) \mid U \text{ is unitary}\}\,, \text{ if } \mathbb{L} = \mathbb{C}\,. \end{aligned}$$

Show that

a) If $A, B \in GL(H)$ then $AB \in GL(H)$ and $A^{-1} \in GL(H)$,

b) $O(H) \subseteq GL(H)$ and $U(H) \subseteq GL(H)$,

c) If $A, B \in O(H)$ then $AB \in O(H)$ and $A^{-1} \in O(H)$, and similarly if $O(H)$ is replaced by $U(H)$.

Show likewise the corresponding statements for $GL(n)$, $O(n)$ and $U(n)$, denoting the sets of invertible, orthogonal and unitary $n \times n$-matrices, respectively.

$GL(H)$, $O(H)$ and $U(H)$ are called the *general linear*, the *orthogonal* and the *unitary group* over H, respectively.

Exercise 5.7. Show that the norm of the multiplication operator in Example 5.3 b) is given by

$$\|M_f\| = \|f\|_u\,.$$

Exercise 5.8. Let H be an infinite dimensional Hilbert space and let $(e_i)_{i\in\mathbb{N}}$ be an orthonormal basis for H.

Show that there exists exactly one operator $T \in \mathcal{B}(H)$ such that

$$Te_i = e_{i+1}\,, \quad i \in \mathbb{N}\,,$$

and that T is an isometry, that is $\|Tx\| = \|x\|$, $x \in H$.

Find T^* and show that

$$T^*T = 1 \quad \text{and} \quad TT^* \neq 1\,.$$

Exercise 5.9. Let A be a densely defined operator from H_1 to H_2 and assume A is bounded.

Show that A has a unique extension to a bounded operator $\bar{A}$ defined everywhere on H_1 in the following two steps.

a) Since A is densely defined there exists for every $x \in H_1$ a sequence (x_i) in $\mathcal{D}(A)$ that converges to x. Show that the sequence (Ax_i) is convergent in H_2 and that its limit only depends on x and not on the choice of sequence (x_i) converging to x.

b) With notation as in a) set

$$\bar{A}x = \lim_{i\to\infty} Ax_i$$

and show that $\bar{A}$ is a bounded linear operator that extends A.

Exercise 5.10. Assume that U and V are two commuting bounded operators on a Hilbert space H. Show that each eigenspace $E_\lambda(U)$ is invariant under V, i. e., if x is an eigenvector for U with eigenvalue λ or equals $\underline{0}$, then the same holds for Vx.

Exercise 5.11. Let H be a Hilbert space and let f be a function given by a power series

$$f(z) = \sum_{n=0}^{\infty} c_n z^n\,, \qquad |z| < \rho\,,$$

where $\rho > 0$ is the corresponding radius of convergence[2]. Show that for an operator $A \in \mathcal{B}(H)$ with $\|A\| < \rho$ the series

$$\sum_{n=0}^{\infty} |c_n|\, \|A^n\|$$

is convergent (in $\mathbb{R}$) and that the series

$$\sum_{n=0}^{\infty} c_n A^n x$$

is convergent in H for all $x \in H$. The sum is then denoted by $f(A)x$, that is

$$f(A)x = \sum_{n=0}^{\infty} c_n A^n x\ .$$

Verify that $f(A)$ is a bounded linear operator on H.

Assume now that H is finite dimensional and that $\alpha = (e_1, \ldots, e_n)$ is an orthonormal basis for H, and let $\underline{\underline{A}}$ be the matrix representing A w. r. t. α. Show that the series

$$\sum_{n=0}^{\infty} c_n (\underline{\underline{A}}^n)_{ij}$$

is convergent in $\mathbb{C}$ for all $1 \leq i, j \leq n$, and that the so defined matrix $f(\underline{\underline{A}})$, where

$$(f(\underline{\underline{A}}))_{ij} = \sum_{n=0}^{\infty} c_n (\underline{\underline{A}}^n)_{ij}\ ,$$

represents the operator $f(A)$ w. r. t. α.

Exercise 5.12. Let A and f be given as in Exercise 5.11. Show that if A is diagonalizable, then $f(A)$ is also diagonalizable and that

$$f(A) = \sum_{i \in I} f(\lambda_i) P_i\,,$$

if A is given by (5.57).

[2] A series of the form $\sum_{n=0}^{\infty} c_n z^n$, where c_n are complex numbers and z is a complex variable is called a power series. There exists a number $\rho \geqslant 0$ called the radius of convergence of the power series such that the series is absolutely convergent for $|z| < \rho$ and is divergent for $|z| > \rho$.

Exercise 5.13. Let H be a complex Hilbert space and let $A \in \mathcal{B}(H)$. Define e^A as in Exercise 5.11 by using the power series for the exponential function. Show that

$$e^{(t+s)A} = e^{tA}e^{sA}\,, \quad s,t \in \mathbb{C}\,,$$

and conclude that e^A is invertible, and that

$$(e^A)^{-1} = e^{-A}\,.$$

Does the identity $e^{A+B} = e^A e^B$ hold for arbitrary $A, B \in \mathcal{B}(H)$?

Exercise 5.14. Let H be a Hilbert space. A function $x : \mathbb{R} \to H$ with values in H is called differentiable at $t_0 \in \mathbb{R}$ if there exists a vector $a \in H$, such that

$$\|(t-t_0)^{-1}(x(t)-x(t_0)) - a\| \to 0 \ \text{ as } \ t \to t_0\,.$$

As usual, the vector a is then denoted by $x'(t_0)$ or by $\frac{dx}{dt}(t_0)$:

$$\lim_{t\to t_0}(t-t_0)^{-1}(x(t)-x(t_0)) = x'(t_0) \ \text{ in } \ H\,.$$

Show, using notation as in Exercise 5.13 that the function $y(t) = e^{tA}x$ is differentiable on $\mathbb{R}$ for all $x \in H$ and fulfills

$$\frac{dy}{dt} = Ay\,.$$

In case A is self-adjoint this differential equation is a special case of the Schrödinger equation describing the time-evolution of a quantum mechanical system as will be discussed in more detail in Chapter 6.

Exercise 5.15. Let σ_j, $j = 1,2,3$ denote the so-called Pauli matrices given by

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\,, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\,.$$

Verify that

$$\sigma_j^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\,, \quad j = 1,2,3\,,$$

and use this to calculate $e^{i\theta\sigma_j}$ defined as in Exercises 5.11 and 5.13 for $\theta \in \mathbb{C}$ and $j = 1,2,3$.

Exercise 5.16. Show that if $P \in \mathcal{B}(H)$ fulfills $P^* = P$ and $P^2 = P$, then P is the orthogonal projection onto $X = P(H)$.

Exercise 5.17. Let H be a Hilbert space with orthonormal basis $(e_i)_{i\in\mathbb{N}}$ and let $(\lambda_i)_{i\in\mathbb{N}}$ be a sequence in $\mathbb{L}$, not necessarily bounded.

a) Show that the equation

$$Ax = \sum_{i=1}^{\infty} \lambda_i \langle\, e_i | x \,\rangle e_i \tag{5.65}$$

defines a densely defined operator with domain

$$\mathcal{D}(A) = \{x \in H \mid \sum_{i=1}^{\infty} |\lambda_i \langle\, e_i | x \,\rangle|^2 < \infty\}\,.$$

b) Show that the adjoint of the operator A in a) has the same domain of definition as A and is given by the formula

$$A^*x = \sum_{i=1}^{\infty} \overline{\lambda}_i \langle\, e_i | x \,\rangle e_i \,. \tag{5.66}$$

c) Deduce from a) and b) that A is self-adjoint if and only if λ_i is real for all $i \in \mathbb{N}$.

Exercise 5.18.

a) Find the Fourier transform of the function

$$g(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \,. \end{cases}$$

b) Show for any integrable function f on $\mathbb{R}$ that the inequality

$$|\widehat{f}(p)| \leqslant \int_{\mathbb{R}} |f(x)| dx$$

holds for all $p \in \mathbb{R}$.

c) Show that if the function f on $\mathbb{R}$ is integrable, real and fulfills $f(x) = f(-x)$ for all $x \in \mathbb{R}$, then $\widehat{f}$ is real, and if moreover $f(x) \geqslant 0\,, x \in \mathbb{R}$, then $\widehat{f}$ assumes its maximal value at $p = 0$. What is this value?

Exercise 5.19. Let $f : \mathbb{R} \to \mathbb{C}$ be integrable and define the function f_a by $f_a(x) = f(x-a)\,, x \in \mathbb{R}$. What is $\widehat{f_a}(p)$ expressed in terms of $\widehat{f}$?

Exercise 5.20. Let H_1 and H_2 be two Hilbert spaces with inner products denoted by $\langle \cdot \,|\, \cdot \rangle_1$ and $\langle \cdot \,|\, \cdot \rangle_2$, respectively. The Cartesian product $H_1 \times H_2$ is then a vector space with addition and scalar multiplication defined by

$$(u_1, u_2) + (v_1, v_2) = (u_1 + v_1, u_2 + v_2)\,, \qquad \lambda(u_1, u_2) = (\lambda u_1, \lambda u_2)$$

for $u_1, v_1 \in H_1$, $u_2, v_2 \in H_2$ and $\lambda \in \mathcal{L}$.

a) Show that an inner product is defined on $H_1 \times H_2$ by

$$\langle\, (u_1, u_2) | (v_1, v_2) \,\rangle = \langle\, u_1 | v_1 \,\rangle_1 + \langle\, u_2 | v_2 \,\rangle_2\,, \quad u_1, v_1 \in H_1,\ u_2, v_2 \in H_2\,,$$

and that with this inner product it is a Hilbert space.

b) Show that in the Hilbert space $H_1 \times H_2$ from a) it holds that

$$(H_1 \times \{\underline{0}\})^{\perp} = \{\underline{0}\} \times H_2\,.$$

The subspace $H_1 \times \{\underline{0}\}$ can naturally be identified with H_1 and similarly $\{\underline{0}\} \times H_2$ can be identified with H_2. Recalling the definition of direct sum in Section 4.5 the previous equality can equivalently be stated as

$$H_1 \times H_2 = H_1 \oplus H_2\,.$$

It is therefore customary to use the notation $u_1 \oplus u_2$ for (u_1, u_2).

Exercise 5.21. A linear map $A : \mathcal{D} \to H$ from a subspace $\mathcal{D}$ of a Hilbert space H to the Hilbert space itself is said to be a *closed operator* if its graph

$$\mathcal{G}(A) = \{u \oplus Au \ \mid \ u \in \mathcal{D}\}$$

is a closed subspace of $H \otimes H$.

a) Show that a bounded operator $A \in \mathcal{B}(H)$ is a closed operator.

b) Let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis for H and $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence of complex numbers. Consider the mapping $A : \mathcal{D} \to H$ given by

$$\mathcal{D} = \left\{ x \in H \ \middle| \quad \sum_{n=1}^{\infty} |\lambda_n \langle\, e_n | x \,\rangle|^2 < \infty \right\},$$

$$Ax = \sum_{n=1}^{\infty} \lambda_n \langle\, e_n | x \,\rangle e_n, \quad \text{for } x \in \mathcal{D}\,.$$

Show that A is a closed operator.

Exercise 5.22. Let $\bar{D}$ be the unbounded operator on $H = L_2([-\pi, \pi])$ defined in Example 5.19 and let $(e_n)_{n\in\mathbb{Z}}$ denote the orthonormal basis for H given by

a) Find the Fourier coefficients $\langle\, e_n | g \,\rangle$ for the function $g(\theta) = \theta$, $\theta \in [-\pi, \pi]$, for all $n \in \mathbb{Z}$ and show that $g \in C^1([-\pi, \pi])$ but $g \notin \mathcal{D}(\bar{D})$.

b) Use the result in a) and Parseval's identity to show that the infinite series $\sum_{n=1}^{\infty} n^{-2}$ is convergent and determine its value.

c) Show that the function $h(\theta) = |\theta|$, $\theta \in [-\pi, \pi]$, satisfies

$$h \in \mathcal{D}(\bar{D}), \quad h \notin C^1([-\pi, \pi]).$$

d) Let

$$C^1_{=}([-\pi, \pi]) = \{f \in C^1[-\pi, \pi] \mid f(-\pi) = f(\pi)\},$$

and show by integration by parts that if $f \in C^1_{=}([-\pi, \pi])$ then

$$\langle\, e_n | -if' \,\rangle = n \langle\, e_n | f \,\rangle\,.$$

e) Use the result of d) to show that $C^1_{=}([-\pi, \pi]) \subseteq \mathcal{D}(\bar{D})$ and that $\bar{D}f = -if'$ for all $f \in C^1_{=}([-\pi, \pi])$.

Chapter 6

Quantum mechanics

In this chapter we use the mathematical formalism developed in the two preceding chapters to first set up the fundamental principles of Quantum Mechanics and second to analyse in some detail some rather simple concrete systems. Moreover, some general phenomena implied by the formalism and its interpretation, such as the implementation of symmetries and their relation to conservation laws, will be discussed.

6.1 The quantum mechanical formalism

The realization, by Heisenberg, that the position and momentum of a quantum mechanical particle cannot be measured simultaneously renders phase space as introduced in Chapters 1 and 3 inappropriate for quantum mechanics and inevitably implies that the notion of state for a quantum mechanical system is profoundly different from that of classical mechanics. We shall not in these notes try to derive the new mathematical formalism needed for the description of quantum phenomena from more basic principles, but rather be content with the fact that so far it has been found satisfactory and consistent with observed phenomena ever since it was proposed by Dirac [5] and von Neumann [9] at the beginning of the 1930'ies.

The formalism is conveniently described in the form of the following three postulates concerning the state space, the observable quantities, and the time development of quantum mechanical systems.

As we have seen in Chapter 3 the states of a classical mechanical system are labelled by points in phase space. In quantum mechanics the states are vectors in a complex Hilbert space.

Postulate 1. (The state space) The state space for a quantum mechanical system is a Hilbert space H over the complex numbers, such that states are represented by non-zero vectors, and two such vectors φ and ψ represent the same state if and only if $\varphi = c\psi$ for some scalar c. By normalization one can also say that states are represented by unit vectors in H, where two unit vectors represent the same state if they deviate by a phase factor $e^{i\alpha}$.

The vector space structure of H encompasses the so-called *superposition principle*: if φ and ψ represent two different states of the system, then any non-trivial linear combination $c_1\varphi + c_2\psi$ is also a possible state of the system, called a *superposition* of the two states. This principle, responsible for observed interference phenomena, has no analogue in classical mechanics.

Postulate 2. (Observables) Measurable physical quantities associated with the system are represented by self-adjoint (generally unbounded) operators on H, which are called the *observables* of the system. If the observable A is diagonalizable with (different) eigenvalues $\lambda_1, \lambda_2, \ldots$, then these are the possible values that the corresponding physical quantity can assume, and the probability of measuring the value λ_i in the normalized state ψ equals $\langle\, \psi | P_i \psi \,\rangle$, where P_i is spectral projection onto the eigenspace $E_{\lambda_i}(A)$.

A similar interpretation holds for more general observables. For instance, the possible results of a measurement of a quantity represented by a general self-adjoint operator A is an element in the *spectrum* of A defined as the set of $\lambda \in \mathbb{C}$ for which $(A - \lambda)$ does not have a bounded inverse. The reader is referred to Exercises 6.4 and 6.5 for details on the spectra of diagonalizable operators and multiplication operators, respectively. For the sake of simplicity the main focus in these notes will be on diagonalizable observables A, in which case

$$A = \sum_{i=1}^{\infty} \lambda_i P_i \,, \quad \text{and} \quad \sum_{i=1}^{\infty} P_i = 1 \,,$$

as a consequence of (5.59) (with notation P_i instead of Q_i for the spectral projections). A normalized state ψ can be written as

$$\psi = \sum_{i=1}^{\infty} P_i \psi \,,$$

and since eigenspaces corresponding to different eigenvalues are orthogonal, Pythagoras' theorem implies

$$\sum_{i=1}^{\infty} \|P_i \psi\|^2 = \|\psi\|^2 = 1 \,, \tag{6.1}$$

which shows that the numbers $\|P_i\psi\|^2$ can, indeed, be interpreted as probabilities. It is part of the interpretation of the formalism that if a measurement of the observable A yields the result λ_i, then the system is in the state $P_i\psi$ after the performance of the measurement.

The average or expectation value of (the physical quantity represented by) A, when repeatedly measured for the system prepared in the state ψ, equals

$$\sum_{i=1}^{\infty} \lambda_i |\langle e_i | \psi \rangle|^2 = \langle \psi | A | \psi \rangle \,,$$

where the equality holds if ψ is in the domain $\mathcal{D}(A)$ of A. Here, the last expression for the average value holds for arbitrary observables A, diagonalizable or not, as originally proposed by Max Born in 1926.

As in the case of classical mechanics the notion of state can be extended by considering probabilistic averages of states. Given states represented by normalized vectors $\psi_1, \psi_2 \ldots$

in H and real positive numbers (probabilities) $\lambda_1, \lambda_2 \ldots \geqslant 0$ with $\sum_{j=1}^{\infty} \lambda_j = 1$ one obtains a probabilistically mixed state by defining the expectation value of any observable A to be

$$\sum_{j=1}^{\infty} \lambda_j \langle \psi_j | A \psi_j \rangle .$$

A state of this form is called a *mixed state* whereas a state represented by just one normalized ψ is called a *pure state*. Mixed states are relevant for quantum thermodynamics, for instance as positive temperature states. They will not be discussed further here.

The third postulate deals with the time evolution of a quantum system, which will be defined as an analogue of the flow on state space in classical mechanics. Clearly, it must be assumed that time evolution preserves the total probability. In view of Postulates 1 and 2 this means that it must map unit vectors to unit vectors. Together with the requirement of time reversibility it can be argued that time evolution must be represented by unitary operators in the following sense.

Postulate 3 (The time evolution). The time evolution of the system from time t_1 to time t_2 is given by a *unitary operator*

$$U(t_2, t_1) : H \to H \, ,$$

in the sense that if the system is in the state ψ at time t_1 then it will be in state $U(t_2, t_1)\psi$ at time t_2, and the function $U(t', t)$ satisfies the flow relation

$$U(t_3, t_2) U(t_2, t_1) = U(t_3, t_1)$$

for all $t_1, t_2, t_3 \in \mathbb{R}$. Moreover, $U(t', t)\psi$ is assumed to be a continuous function from $\mathbb{R} \times \mathbb{R}$ to H for all $\psi \in H$.

It is difficult to study this general type of time evolution and usually further assumptions on differentiability are required. We will focus on evolutions that are time translation invariant, which amounts to assuming that the evolution only depends on the time elapsed and not on when it started, that is to say $U(t_2, t_1) = U(t_2 - t_1, 0)$. Using the notation $U(t) = U(t, 0)$, Postulate 3 takes the following form in this case.

Postulate 3'. The time evolution of the system is given by a function $U(t)$, $t \in \mathbb{R}$, taking values in the unitary operators on H such that a state represented by $\psi \in H$ at time t_1 evolves to the state represented by $U(t_2 - t_1)\psi$ at time t_2. Moreover, $U(t)$ forms a *one-parameter group* in the sense that

$$U(t + s) = U(t)U(s), \quad U(0) = 1 \tag{6.2}$$

for all $t, s \in \mathbb{R}$, and it is *strongly continuous*, i. e., $U(t)\psi$ is a continuous function of t for all $\psi \in H$.

We state without proof the following important result on the structure of strongly continuous unitary one-parameter groups.

Theorem 6.1 (Stone's Theorem)**.** *Let $U(t)$, $t \in \mathbb{R}$, be a strongly continuous unitary one-parameter group on a Hilbert space H. Then*[1]

$$\mathcal{D} = \{\phi \in H \mid U(t)\phi \text{ is differentiable at } t = 0\}$$

is a dense subspace of H and the operator $A : \mathcal{D} \to H$ defined by

$$A\psi = i\frac{dU(t)\psi}{dt}\Big|_{t=0}, \quad \psi \in \mathcal{D}$$

is a self-adjoint (generally unbounded) operator called the generator *of the one-parameter group.*

Conversely, given a (densely defined) self-adjoint operator $A : \mathcal{D}(A) \to H$ then there exists a strongly continuous unitary one-parameter group such that A is its generator.

This result leads directly to the quantum analogue of the Hamiltonian in classical mechanics.

Definition 6.2. *If A is the generator of the time evolution of Postulate 3' then $\mathcal{H} = \hbar A$, where $\hbar = 1.055\times10^{-34} Js$ is the reduced Planck constant, is called the* **Hamilton operator** *or* **Hamiltonian** *of the system, that is*

$$\mathcal{H}\psi = i\hbar\frac{dU(t)\psi}{dt}\Big|_{t=0}, \quad \psi \in \mathcal{D}(\mathcal{H}). \tag{6.3}$$

The physical quantity represented by the Hamiltonian is by definition the total energy of the system.

Assuming that the relation (6.3) holds, then $\psi(t) = U(t)\psi$ fulfills

$$\psi(t) - \psi(t_0) = U(t_0)(U(t - t_0)\psi - \psi) = (U(t - t_0) - 1)U(t_0)\psi\,.$$

The first equality immediately shows that $\psi(t)$ is differentiable at $t = t_0$. Using the second equality, it then follows that $U(t_0)\psi \in \mathcal{D}(\mathcal{H})$ and that

$$i\hbar\frac{d\psi(t)}{dt}\Big|_{t=t_0} = U(t_0)\mathcal{H}\psi = \mathcal{H}U(t_0)\psi = \mathcal{H}\psi(t)\,.$$

Thus, $\psi(t)$ satisfies the **Schrödinger equation**

$$\boxed{i\hbar\frac{d\psi(t)}{dt} = \mathcal{H}\psi(t)\,.} \tag{6.4}$$

In quantum mechanics this equation is the analogue of Hamilton's equations in classical mechanics, and in both cases the Hamiltonian plays the role as generator of the time evolution.

Although Stone's Theorem will not be proven here the fact that the Hamiltonian $\mathcal{H}$ is symmetric is a consequence of the following simple calculation for $\phi, \psi \in \mathcal{D}(\mathcal{H})$:

$$0 = \frac{d}{dt}\langle U(t)\varphi|U(t)\psi\rangle = \frac{i}{\hbar}(\langle \mathcal{H}U(t)\varphi|U(t)\psi\rangle - \langle U(t)\varphi|\mathcal{H}U(t)\psi\rangle)\,.$$

[1] See Exercise 5.14 for the definition of differentiability of vector valued functions

Setting $t = 0$ then gives

$$\langle \mathcal{H}\varphi|\psi\rangle = \langle \varphi|\mathcal{H}\psi\rangle\,,$$

showing that $\mathcal{H}$ is symmetric.

It is worth noting that the previous calculation still holds if $\psi(t)$ is a solution of the Schrödinger equation with $\mathcal{H}$ selfadjoint and time-dependent. It hence shows that any such solution has constant norm and so defines a unitary time evolution provided a solution exists for any initial value $\psi_0 \in H$. This would then correspond to a general unitary time evolution $U(t', t)$ as described in Postulate 3.

If $\mathcal{H}$ is time independent the corresponding unitary one-parameter group is written as

$$U(t) = e^{-\frac{i}{\hbar}\mathcal{H}t}\,. \tag{6.5}$$

In case $\mathcal{H}$ is bounded, the right hand side can be defined as a unitary one-parameter group with generator $-\frac{i}{\hbar}H$ in terms of the power series expansion of the exponential function as demonstrated in Exercises 5.13 and 5.14. If $\mathcal{H}$ is unbounded the situation is more complicated and the standard way to define the right hand side of (6.5) is by use of the general spectral theorem for self-adjoint operators mentioned towards the end of Section 5.4.

A state of a quantum system is called an *equilibrium state* if it does not change with time. If such a state is represented by the normalized vector ψ it must satisfy

$$U(t)\psi = e^{-i\hbar^{-1}\alpha(t)}\psi$$

for some function $\alpha : \mathbb{R} \to \mathbb{R}$. Using that $U(t)$ is a strongly continuous unitary one-parameter group it follows (see Exercise 6.2) that $\alpha(t)$ can be chosen to be of the form $\alpha(t) = Et$ for all $t \in \mathbb{R}$ for some $E \in \mathbb{R}$. It follows that $\psi \in \mathcal{D}(\mathcal{H})$ with

$$\mathcal{H}\psi = E\psi\,, \tag{6.6}$$

that is ψ is an eigenvector of $\mathcal{H}$ with eigenvalue E. Thus the state represented by ψ corresponds to a state with well-defined energy E.

Another aspect worth stressing is an inherent ambiguity in the choice of state space and observables for a given system that results from the application of unitary transformations.

Suppose a given state space H and associated observables for the system in question have been determined, and let $U : H \to H'$ be a unitary operator from H onto a Hilbert space H'. Define, for any observable A on H, the operator A' on H' by

$$A' = UAU^*\,.$$

Then A' is self-adjoint (see Exercise 6.3 for this and the exact definition of UAU^* if A is unbounded). Since U preserves the inner product by Theorem 5.21, a second equivalent description, or *representation*, of the same system is obtained by letting H' be the state space, such that $U\psi \in H'$ represents the same state as $\psi \in H$, and by letting A' represent the same physical quantity as A. Of course, the explicit form of the observables depends on the choice of representation.

A common method for obtaining a representation for a given system is by applying a *quantization procedure* to the corresponding classical system (if it exists). A standard representation obtained in this way is the *Schrödinger representation*, where the state space

is chosen to be the space of square integrable functions, also called *wave functions*, on the configuration space of the classical system, the observable representing a coordinate is the corresponding multiplication operator and the observable corresponding to a (generalized) momentum is the corresponding derivative operator multiplied by $\frac{\hbar}{i}$. The Schrödinger representation for a single particle is defined more precisely in Example 6.3 below and in Section 6.3, where the so-called *momentum representation* will be introduced as well.

Finally, we mention that the preceding postulates are usually supplemented by a certain technical condition called *irreducibility*, which ensures that sufficiently many observables are available. One way of formulating this is to require that the state space H cannot be split into a non-trivial direct sum

$$H = H_1 \oplus H_2 \,,$$

where H_1 and H_2 are invariant under all observables. This condition may be regarded as the replacement of the property possessed by classical systems that any state can be characterized by the values assumed in the state by a certain collection of observables, such as the coordinates and the associated (generalized) momenta. We shall, however, not make essential use of this condition below.

Example 6.3 (*Schrödinger representation)***.** In the following sections we shall discuss in detail a particle moving in k-dimensional Euclidean space. The state space for such a particle in the Schrödinger representation will be chosen to be the space $L_2(\mathbb{R}^k)$ of square integrable functions of k variables. It is a fundamental result of integration theory in several variables, that $L_2(\mathbb{R}^k)$ is a Hilbert space with inner product given by

$$\langle\, f|g\,\rangle = \int_{\mathbb{R}^k} \overline{f(x)} g(x) dx \,, \quad f, g \in L_2(\mathbb{R}^k) \,, \tag{6.7}$$

and remarks i)-iii) of Example 4.12 b) pertaining to the case $k = 1$ extend in a straightforward way to the multi-variable case. In particular, the space $C_0^\infty(\mathbb{R}^k)$ of C^∞ functions vanishing outside some ball of finite (but arbitrary) radius is dense in $L_2(\mathbb{R}^k)$.

The observable corresponding to the i'th coordinate x_i of the position of the particle is the unbounded self-adjoint multiplication operator (see Exercise 6.5) M_{x_i} with domain

$$\mathcal{D}(M_{x_i}) = \left\{ \psi \in L_2(\mathbb{R}^k) \,\middle|\, \int_{\mathbb{R}^k} |x_i \psi(x)|^2 dx < \infty \right\} .$$

The operator M_{x_i} has no eigenvalues (why?). It is thus not possible to talk about a particle with a definite position. Given an interval $[a, b] \subset \mathbb{R}$, $a < b$, there exist normalized functions $\psi \in L_2(\mathbb{R}^k)$ that are zero outside $[a, b] \times \mathbb{R}^{k-1}$. Since

$$\langle \psi | M_{x_1} | \psi \rangle \in [a, b] \,,$$

for such functions it is natural to interpret these functions as describing states for which the particle's first coordinate belongs to $[a, b]$. For a general normalized function $\psi \in L_2(\mathbb{R}^k)$ the integral

$$\int_{x_1 \in I} |\psi(x)|^2 dx$$

is interpreted as the probability of finding the particle with x_1 in the interval $I = [a, b]$. Indeed, defining the operator $P_I^1 : L_2(\mathbb{R}^k) \to L_2(\mathbb{R}^k)$ by

$$P_I^1\psi(x) = \begin{cases} \psi(x), & x_1 \in I \\ 0, & x_1 \notin I \end{cases} \tag{6.8}$$

it is easy to see that P_I^1 is an orthogonal projection (see Exercise 6.6) and that

$$\langle \psi \,|P_I^1|\, \psi \rangle = \int_{x_1 \in I} |\psi(x)|^2 dx \,.$$

Since the projection P_I^1 has eigenvalues 0 and 1, the average value is precisely the probability of measuring the eigenvalue 1, that is of finding the particle with first coordinate in $[a, b]$. Similar remarks of course apply to the other coordinates and we hence are able to conclude that the function $|\psi(x)|^2$ can be interpreted as the probability density for finding the particle in some region of space.

To complete the specification of the Schrödinger representation for a single particle the observable representing its momentum must be defined. This will be done in Section 6.3.

6.2 Transformations and symmetries in quantum mechanics

It is natural to define transformations of a quantum system with state space H to be bijective maps on states with the following property: if the states represented by the unit vectors ϕ and ψ are transformed to states represented by the unit vectors ϕ' and ψ', respectively, then

$$|\langle \phi' | \psi' \rangle|^2 = |\langle \phi | \psi \rangle|^2.$$

It is a fundamental theorem of Wigner and Bargmann that any such transformation can be implemented by a unitary or an anti-unitary operator V on H, where an anti-unitary operator means a conjugate linear bijective isometry. To say that the transformation is implemented by V means, more precisely, that the state represented by ψ is transformed to the state represented by $V\psi$ for all states ψ. Note that, if V implements a given transformation then the same holds for $e^{i\theta}V$ for any $\theta \in \mathbb{R}$. In these notes a transformation is simply defined to be a unitary or anti-unitary operator on the state space H. Although one can in general not avoid anti-unitary transformations, for instance in relation to the time-reversal transformation (see Exercise 6.7), we shall focus on unitary transformations in the following. A symmetry is defined to be a transformation that preserves the time evolution. This is expressed more precisely in the next definition.

Definition 6.4. *A* **transformation** *of a quantum mechanical system with state space H is a unitary (or anti-unitary) operator $V : H \to H$. If $\theta \in \mathbb{R}$ then V and $e^{i\theta}V$ represent the same transformation.*

A **symmetry** *of a quantum mechanical system with state space H and unitary time evolution $U(t)$, $t \in \mathbb{R}$, is a unitary operator V such that*

$$U(t)V = VU(t) \tag{6.9}$$

for all t. We say that V commutes with the time evolution.

The set of symmetries of a quantum system form a group (see Definition 1.2), whose structure depends on the system. Hence, we are interested in groups of transformations and their representations, defined as follows.

Definition 6.5. *If G is a group and H is a complex Hilbert space, then a mapping $\rho : G \to \mathcal{B}(H)$ is called a* **unitary representation** *of G if $\rho(g)$ is a unitary operator and*

$$\rho(gh) = \rho(g)\rho(h)$$

for all $g \in G$. It is called a **projective (unitary) representation** *if there exists a function $\omega : G \times G \to \mathbb{R}$ such that*

$$\rho(gh) = e^{i\omega(g,h)}\rho(g)\rho(h). \tag{6.10}$$

The representations natural to quantum mechanics are the projective representations since multiplying a unitary by a complex number of modulus one does not change the transformation it represents.

We note that the strongly continuous unitary one-parameter groups in Stone's Theorem are, in fact, unitary representations of the additive group of real numbers. Stone's Theorem thus characterizes all such representations if they additionally satisfy the strong continuity property. This allows us to formulate a quantum mechanical version of Noether's theorem:

Theorem 6.6 (*Quantum Noether's Theorem*)**.** *Assume that $V : \mathbb{R} \to \mathcal{B}(H)$ is a strongly continuous unitary one-parameter group of symmetries of a quantum system with state space H and unitary time evolution $U(t)$, $t \in \mathbb{R}$. Then the generator A of V is a conserved quantity, that is a self-adjoint operator fulfilling*

$$U(t)^* A U(t) = A,\ t \in \mathbb{R}\,. \tag{6.11}$$

That the relation (6.11) specifies that A represents a conserved quantity may be understood as follows. If the system is in a state represented by ψ at time $t = 0$, then it is in a state represented by $U(t)\psi$ at time t. Consequently, the expected value of A at time t is

$$\langle U(t)\psi, AU(t)\psi\rangle = \langle \psi, U(t)^* AU(t)\psi\rangle = \langle \psi, A\psi\rangle,$$

and hence is independent of time.

Proof of Theorem 6.6. Since $V(s)$ is a symmetry we have

$$U(t)^* V(s) U(t) = V(s)$$

for all $s, t \in \mathbb{R}$. Letting both sides of this equation act on an element ψ in the domain of A and differentiate w. r. t. to s it follows that $U(t)\psi$ is in the domain of A and that $U(t)^* AU(t)\psi = A\psi$. □

Conservation of energy in quantum mechanics follows immediately as a special case of this result.

Corollary 6.7. *The total energy of a quantum mechanical system represented by the Hamiltonian $\mathcal{H}$ is a conserved quantity.*

Proof. From the one-parameter property (6.2) we find that

$$U(t)U(s) = U(t+s) = U(s)U(t).$$

Hence $U(s)$ is for all $s \in \mathbb{R}$ a symmetry of system. □

Our goal in Section 6.4 is to discuss the implementation of Galilei transformations in quantum mechanics. In order to do this it is convenient first to discuss the momentum representation for a quantum particle moving in $\mathbb{R}^k$. The momentum representation is closely connected to the Fourier transform from Section 5.5.

6.3 The momentum representation

In Section 5.5 the Fourier transform in one variable was introduced as a unitary operator on $L_2(\mathbb{R})$. In this section, the generalization to k variables is given and at the same time variables are subject to a scaling relevant to quantum mechanics.

The space of *rapidly decreasing functions* of k variables $\mathcal{S}(\mathbb{R}^k)$ consists of C^∞ functions on $\mathbb{R}^k$ whose derivatives all decay faster that any power at infinity. More precisely,

$$\mathcal{S}(\mathbb{R}^k) = \{ f \in C^\infty(\mathbb{R}^k) \quad | \quad \|x\|^N \partial_{x_1}^{n_1} \cdots \partial_{x_k}^{n_k} f(x) \text{ is a bounded function of } x \\ \text{for all } n_1, \ldots, n_k, N = 0, 1, 2, \ldots \}\,.$$

The set $\mathcal{S}(\mathbb{R}^k)$ contains the space $C_0^\infty(\mathbb{R}^k)$ of C^∞ functions vanishing outside some bounded set. Since the latter set is dense in $L_2(\mathbb{R}^k)$ it follows that $\mathcal{S}(\mathbb{R}^k)$ is also dense in $L_2(\mathbb{R}^k)$. Moreover, it is clear that if $f \in \mathcal{S}(\mathbb{R}^k)$ then all partial derivatives of f to any order also belong to $\mathcal{S}(\mathbb{R}^k)$.

The $\hbar$-Fourier transform of a function $f \in \mathcal{S}(\mathbb{R}^k)$ is defined by

$$\mathcal{F}_\hbar f(p) = (2\pi\hbar)^{-k/2} \int_{\mathbb{R}^k} e^{-i\hbar^{-1} p\cdot x} f(x) dx, \qquad p \in \mathbb{R}^k\,.$$

For $\hbar = 1$ and $k = 1$ this definition coincides with the Fourier transform defined in Section 5.5. Hence, the notaton $\mathcal{F} = \mathcal{F}_{\hbar=1}$ and $\mathcal{F}f = \widehat{f}$ will be used in the following. It is a fundamental result that the isometry property (5.62) for rapidly decreasing functions extends to the multi-variable case. By a change of variables (see Exercise 6.8) it then follows that

$$\int_{\mathbb{R}^k} |\mathcal{F}_\hbar f(p)|^2 dp = \int_{\mathbb{R}^k} |f(x)|^2 dx\,, \quad f \in \mathcal{S}(\mathbb{R}^k)\,, \tag{6.12}$$

and thus $\mathcal{F}_\hbar$ can be extended as in the one-variable case to an isometry on all of $L_2(\mathbb{R}^k)$.

The following lemma is instrumental for the use of Fourier analysis in the theory of differential equations.

Lemma 6.8. *The Fourier transformation $\mathcal{F}_\hbar$ maps $\mathcal{S}(\mathbb{R}^k)$ bijectively onto itself and the following formulas hold for $f \in \mathcal{S}(\mathbb{R}^k)$:*

$$\mathcal{F}_\hbar(\hbar\partial_{x_j} f)(p) = (ip_j)\mathcal{F}_\hbar f(p) \tag{6.13}$$

$$\mathcal{F}_\hbar(x_j f)(p) = i\hbar\partial_{p_j}\mathcal{F}_\hbar f(p)\,, \tag{6.14}$$

for $j = 1, \ldots, k$.

The first formula follows by integration by parts, and the second by differentiating both sides of the formula defining $\mathcal{F}_\hbar f$ and then differentiating under the integral sign on the right-hand side. See Exercise 6.9 for the details of the proof in the case $k = 1$, $\hbar = 1$.

It follows as in Section 5.5 that the extension of $\mathcal{F}_\hbar$ to $L_2(\mathbb{R}^k)$ is a unitary operator whose inverse $\mathcal{F}_\hbar^*$ is given by the multi-variable inversion formula

$$(\mathcal{F}_\hbar^* f)(x) = (\mathcal{F}_\hbar^{-1} f)(x) = (2\pi\hbar)^{-k/2} \int_{\mathbb{R}^k} e^{i\hbar^{-1} p\cdot x} f(x) dx, \qquad p \in \mathbb{R}^k\,,$$

for $f \in \mathcal{S}(\mathbb{R}^k)$.

As a consequence of (6.13) we obtain

$$\frac{\hbar}{i} \mathcal{F}_\hbar \partial_{x_j} f(p) = p_j \mathcal{F}_\hbar f(p)\,,$$

for $j = 1, \ldots, k$. Applying the inverse Operator $\mathcal{F}_\hbar^*$ on both sides yields

$$\frac{\hbar}{i} \partial_{x_j} f = \mathcal{F}_\hbar^* \circ M_{p_j} \circ \mathcal{F}_\hbar f\,, \quad f \in \mathcal{S}(\mathbb{R}^k)\,.$$

Since M_{p_j} is a self-adjoint operator by Exercise 6.5 with domain

$$\mathcal{D}(M_{p_j}) = \left\{ f \in L_2(\mathbb{R}^k) \,\middle|\, \int_{\mathbb{R}^k} |p_j^2 |f(p)|^2\, dp < \infty \right\}\,, \tag{6.15}$$

this formula can be used to extend the operator $-i\partial_{x_j}$ to a self-adjoint operator D_j given by

$$\hbar D_j = \mathcal{F}_\hbar^* \circ M_{p_j} \circ \mathcal{F}_\hbar \tag{6.16}$$

whose domain is

$$\mathcal{D}(D_j) = \{ f \in L_2(\mathbb{R}^k) \mid \int_{\mathbb{R}^k} p_j^2 |\mathcal{F}_\hbar f(p)|^2\, dp < \infty \}\,,$$

see Exercise 6.3.

By definition the self-adjoint operator $\hbar D_j$ is the observable representing the momentum of a particle along the x_j-axis in the Schrödinger representation as announced previously in Section 6.1. Since M_{p_j} has no eigenvalues, the same holds for D_j and we conclude that there are no states in H in which the particle has a well defined position or a well defined momentum.

An equivalent description of the particle is obtained by applying the Fourier transform $\mathcal{F}_\hbar$ to the Schrödinger representation. The Hilbert space $H' = \mathcal{F}_\hbar H$ is again $L_2(\mathbb{R}^k)$, that will be distinguished from H by denoting the corresponding variable by p instead of x. The momentum observable now becomes

$$\mathcal{F}_\hbar(\hbar D_j)\mathcal{F}_\hbar^* = M_{p_j}$$

by (6.16). The position observable, on the other hand, becomes

$$\mathcal{F}_\hbar \circ M_{x_j} \circ \mathcal{F}_\hbar^* = -\hbar D_j\,,$$

which is obtained in the same way as (6.16) by using the expression for $\mathcal{F}_\hbar^*$. The so obtained representation is commonly referred to as the *momentum representation*, since the probability density for a particle in the normalized state $\mathcal{F}_\hbar\psi$ for having momentum p now equals $|\mathcal{F}_\hbar\psi(p)|^2$ by the same arguments as previously used in Example 6.3 for the position observable in the Schrödinger representation.

For use in the next section it will now be shown that the multiplication operators in the momentum representation can be characterized as the operators that commute with translations in the Schrödinger representation, also called translation invariant operators.

Theorem 6.9. *Let $\mathcal{T}_a : L_2(\mathbb{R}^k) \to L_2(\mathbb{R}^k)$ for $a \in \mathbb{R}^k$ be the translation operator defined by*

$$\mathcal{T}_a\psi(x) = \psi(x-a)\,. \tag{6.17}$$

Then a bounded operator $A : L_2(\mathbb{R}^k) \to L_2(\mathbb{R}^k)$ satisfies $A\mathcal{T}_a = \mathcal{T}_a A$ for all $a \in \mathbb{R}^k$ if and only if $\mathcal{F}_\hbar A\mathcal{F}_\hbar^*$ is a multiplication operator, in other words, there exists a bounded function $g : \mathbb{R}^k \to \mathbb{C}$ such that

$$A = \mathcal{F}_\hbar^* M_{g(p)} \mathcal{F}_\hbar.$$

Proof. Observe first that $\mathcal{T}_a$ is an isometry; in fact, it is a unitary operator with inverse $\mathcal{T}_a^{-1} = \mathcal{T}_{-a}$. Moreover, for $\psi \in \mathcal{S}(\mathbb{R}^k)$

$$\begin{aligned}(\mathcal{F}_\hbar\mathcal{T}_a\psi)(p) &= (2\pi\hbar)^{-k/2}\int_{\mathbb{R}^k} e^{-i\hbar^{-1}p\cdot x}\psi(x-a)dx = (2\pi\hbar)^{-k/2}\int_{\mathbb{R}^k} e^{-i\hbar^{-1}p\cdot(x+a)}\psi(x)dx \\ &= e^{-i\hbar^{-1}p\cdot a}\mathcal{F}_\hbar\psi(p)\,,\end{aligned}$$

that is

$$\mathcal{T}_a = \mathcal{F}_\hbar^* M_{\exp(-i\hbar^{-1}p\cdot a)}\mathcal{F}_\hbar\,,$$

since both sides are bounded operators and $\mathcal{S}(\mathbb{R}^k)$ is dense in $L_2(\mathbb{R}^k)$. Setting $A' = \mathcal{F}_\hbar A\mathcal{F}_\hbar^*$ it follows that the relation $\mathcal{T}_a A = A\mathcal{T}_a$ is equivalent to

$$A' M_{\exp(-i\hbar^{-1}p\cdot a)} = M_{\exp(-i\hbar^{-1}p\cdot a)}A'. \tag{6.18}$$

It is immediate that if A' is of the form $A' = M_{g(p)}$ for any function g, then it satisfies this relation. It remains to prove the converse, that any operator A' satisfying this relation must be of the form $A' = M_{g(p)}$.

As a first step let us show that the relation

$$A' M_{f(p)} = M_{f(p)} A' \tag{6.19}$$

holds for any function $f \in \mathcal{S}(\mathbb{R}^k)$. Indeed, set $\tilde{f} = \mathcal{F}_\hbar^* f$ and apply $\mathcal{F}_\hbar$ on both sides to obtain

$$f(p) = \mathcal{F}_\hbar\tilde{f}(p) = (2\pi\hbar)^{-k/2}\int_{\mathbb{R}^k} e^{-i\hbar^{-1}p\cdot a}\tilde{f}(a)da\,, \quad p \in \mathbb{R}^k\,. \tag{6.20}$$

Multiplying both sides of this equation by $\psi \in L_2(\mathbb{R}^k)$ gives

$$M_{f(p)}\psi = (2\pi\hbar)^{-k/2}\int_{\mathbb{R}^k} \tilde{f}(a)(M_{\exp(-i\hbar^{-1}p\cdot a)}\psi)da$$

interpreted as an integral with values in $L_2(\mathbb{R}^k)$. The integral on the right hand side of (6.20) can be approximated by a Riemann sum uniformly in p (see Exercise 6.10). From this follows that the last integral above can be approximated in L_2-norm by a Riemann sum with values in $L_2(\mathbb{R}^k)$. Using the linearity and continuity of A' this implies that

$$A'M_{f(p)}\psi) = (2\pi\hbar)^{-k/2}\int_{\mathbb{R}^k}\tilde{f}(a)(A'M_{\exp(-i\hbar^{-1}p\cdot a)})\psi da\,.$$

Next, by invoking (6.18) one gets

$$\begin{aligned} A'M_{f(p)}\psi &= (2\pi\hbar)^{-k/2}\int_{\mathbb{R}^k}\tilde{f}(a)(A'M_{\exp(-i\hbar^{-1}p\cdot a)})\psi da \\ &= (2\pi\hbar)^{-k/2}\int_{\mathbb{R}^k}\tilde{f}(a)(M_{\exp(-i\hbar^{-1}p\cdot a)}A')\psi da \\ &= (M_{f(p)}A')\psi\,, \end{aligned}$$

which is exactly (6.19).

To conclude the proof consider the function $h(p) = e^{-p^2}$ which clearly belongs to $\mathcal{S}(\mathbb{R}^k)$ and is $\neq 0$ for all p. Define $g(p) = h(p)^{-1}(A'h)(p)$, $p \in \mathbb{R}^k$. We claim that $A' = M_{g(p)}$. In order to see this, let $\psi \in C_0^\infty(\mathbb{R}^k)$ be arbitrary and notice that ψ/h also belongs to $C_0^\infty(\mathbb{R}^k)$. Using (6.19) this implies

$$A'\psi = A'M_{\psi/h}h = M_{\psi/h}A'h = M_{\psi/h}(gh) = M_g\psi.$$

Since this is true for all $\psi \in C_0^\infty(\mathbb{R}^k)$, which is a dense subset of $L_2(\mathbb{R}^k)$, it holds for all $\psi \in L_2(\mathbb{R}^k)$, and hence $A' = M_{g(p)}$. □

It is common to write $g(\hbar D)$ as a short hand for the operator $\mathcal{F}_\hbar^* M_{g(p)}\mathcal{F}_\hbar$, thus indicating that such operators can be thought of as functions of the partial derivative operators. These are examples of the type of operators called *pseudo-differential operators*.

6.4 Galilei invariance in quantum mechanics

As formulated in Chapter 1 the principle of relativity states that the fundamental equations of motion of a classical mechanical system are invariant under the group of inertial transformations, the Galilei group. In the quantum mechanical formalism the analogous statement would be that the Galilei group acts as a group of transformations on the system, in the sense that there exists a unitary projective representation of the Galilei group on the state space. A more elaborate explanation of this statement in the case of a single particle moving in $\mathbb{R}^3$ is as follows. The way it is defined in Section 1.2, the Galilei group acts on space-time and not on the state space of the classical particle. However, since the state of the particle at any time is assumed to be determined via the equations of motion by the state at time $t = 0$, the state space can be identified with the set of space-time motions $\{(x(t), t) \mid t \in \mathbb{R}\}$, say, such that the state (x_0, v_0) is identified with the graph $\{(x(t), t) \mid t \in \mathbb{R}\}$ of the motion with initial condition $(x(0), \dot{x}(0)) = (x_0, v_0)$. In this way, the Galilei group acts on the state space. More precisely, the group element $\psi[a, s, v, A]$

acts on the state $\{(x(t),t) \mid t \in \mathbb{R}\}$ according to (1.9)

$$\begin{aligned}\psi[a,s,v,A]\{(x(t),t) \mid t \in \mathbb{R}\} &= \{(a+vt+Ax(t),t+s) \mid t \in \mathbb{R}\} \\ &= \{(a-vs+vt'+Ax(t'-s),t') \mid t' \in \mathbb{R}\}\,. \quad (6.21)\end{aligned}$$

For $s=0$ this gives

$$\psi[a,0,v,1]\{(x(t),t) \mid t \in \mathbb{R}\} = \{(a+vt+Ax(t),t) \mid t \in \mathbb{R}\}\,.$$

The requirement that the right hand side of this equation is (the graph of) a motion, together with the assertion that the motion is uniquely determined by the initial condition, immediately implies that all motions are of the form $x(t) = x_0 + v_0 t$. Inserting this into (6.21), the group action in terms of initial conditions is given by

$$\psi[a,s,v,A](x_0,v_0) = (a-vs+A(x_0-v_0 s), v+Av_0)\,.$$

In particular, for time translations where $a = v = \underline{0}$ and $A = I$, it follows that

$$T_{-s}(x_0,v_0) = (x_0+v_0 s, v_0)\,,$$

which is exactly the motion in state space with initial condition (x_0,v_0), when s is interpreted as time.

The purpose of this section is to carry out similar arguments for a quantum particle moving in $\mathbb{R}^3$. To be specific, we show that there is a natural unitary projective representation of the Galilei group on the state space in the Schrödinger representation, and that under certain natural assumptions the requirement that it acts as a group of transformations determines the time evolution of the free particle uniquely up to a trivial phase factor. The problem of determining all possible unitary projective representations of the Galilei group will not be pursued here, and some technical calculations in this section will be left as exercises to the reader. In a first reading it may be wise to skip these exercises.

Recall that the Galilei group is generated by space translations S_a, boosts B_v, $O(3)$-transformations R_A, and time translations T_s. In order to construct the desired unitary projective representation ρ, we start by defining the unitary operators $\rho(S_a) = \mathcal{T}_a$, $\rho(B_v) = \mathcal{B}_v$, and $\rho(R_A) = \mathcal{R}_A$, for $a, v \in \mathbb{R}^3$ and $A \in O(3)$, on the Schrödinger state space $H = L_2(\mathbb{R}^3)$. As already indicated by the chosen notation, the natural candidate for $\mathcal{T}_a$ is the translation operator on H defined by (6.17) of the previous section. As remarked in the proof of Theorem 6.9 these are indeed unitary operators, and for a normalized wave function ψ the probability density $|\mathcal{T}_a\psi|^2$ has the same value at $x+a$ as the value of ψ at x for any $x \in \mathbb{R}^3$. That is to say, the action of $\mathcal{T}_a$ is to translate the probability density for the position of the particle by the vector a. In the following we shall stick to this choice of $\mathcal{T}_a$. By a similar argument we define $\mathcal{R}_A$ by

$$(\mathcal{R}_A\psi)(x) = \psi(A^{-1}x)\,, \quad x \in \mathbb{R}^3\,. \quad (6.22)$$

Since $|\det A| = 1$ it follows by changing variables from x to $y = A^{-1}x$ in the integral defining the L_2-norm of the $\mathcal{R}_A\psi$, that $\mathcal{R}_A$ is a unitary operator on H, and its inverse is obviously $\mathcal{R}_{A^{-1}}$ (see Exercise 6.11).

To obtain $\mathcal{B}_v$ it is convenient to turn to the momentum representation and note that the boost B_v should have the effect of increasing the momentum of the particle by mv,

if $m > 0$ denotes the particle mass. In other words, a natural candidate for $\mathcal{B}_v$ is the translation operator $\mathcal{T}_{mv}$ in momentum space. Recalling from the proof of Theorem 6.9 that the translation operator $\mathcal{T}_a$ in the Schrödinger representation corresponds to the unitary multiplication operator $M_{e^{-i\hbar^{-1}a\cdot p}}$ in the momentum representation, we are led to assert that $\mathcal{B}_v$ is a multiplication operator in the Schrödinger representation given by

$$(\mathcal{B}_v\psi)(x) = e^{i\hbar^{-1}mv\cdot x}\psi(x)\,, \quad x \in \mathbb{R}^k\,. \tag{6.23}$$

Finally, the natural operator representing a time tranlation T_s is the time evolution operator $U(-s)$ from Postulate 3' by analogy with the classical case discussed above. The operator $\rho(\psi[a,s,v,A])$ representing a general element of the Galilei group of the form (1.13) will be denoted by $\mathcal{U}(a,s,v,A)$ and is defined by

$$\mathcal{U}(a,s,v,A) = U(-s)\mathcal{B}_v\mathcal{T}_a\mathcal{R}_A\,. \tag{6.24}$$

In order to show that this defines a projective representation of the Galilei group we need to show that (6.10) holds. Taking into account the explicit form (1.11) of the multiplication law for the Galilei group, this means that

$$\begin{aligned}&\mathcal{U}(a_1,s_1,v_1,A_1)\mathcal{U}(a_2,s_2,v_2,A_2) = \\ &\qquad e^{i\omega(a_1,s_1,v_1,A_1,a_2,s_2,v_2,A_2)}\mathcal{U}(a_1+v_1s_2+A_1a_2, s_1+s_2, v_1+A_1v_2, A_1A_2)\end{aligned} \tag{6.25}$$

for all $a_1,a_2,v_1,v_2 \in \mathbb{R}^3$, $s_1,s_2 \in \mathbb{R}$, and $A_1,A_2 \in O(3)$, and some real function ω. There are two separate issues involved in this task: it still remains to define the one-parameter group $U(t)$, and only after this has been done can (6.25) be verified.

Disregarding time translations in the first instance, it follows from the definitions of $\mathcal{T}_a$, $\mathcal{B}_v$ and $\mathcal{R}_A$ and from equation (6.24) that the action of $\mathcal{U}(a,0,v,A)$ on a state $\psi \in \mathcal{S}(\mathbb{R}^3)$ is given by

$$\mathcal{U}(a,0,v,A)\psi(x) = e^{i\hbar^{-1}mv\cdot x}\psi(A^{-1}(x-a))\,, \tag{6.26}$$

and by a straight-forward calculation which is left as an exercise for the reader (see Exercise 6.15) it is seen that (6.25) is indeed satisfied for $s_1 = s_2 = 0$ with

$$\omega(a_1,0,v_1,A_1,a_2,0,v_2,A_2) = -\hbar^{-1}ma_1\cdot(A_1v_2)\,. \tag{6.27}$$

For later use let us also determine the form of $\mathcal{U}(a,0,v,A)$ in the momentum representation. Applying the Fourier transformation on both sides of (6.26) gives

$$\begin{aligned}\mathcal{F}_\hbar\mathcal{U}(a,0,v,A)\psi(p) &= (2\pi\hbar)^{-3/2}\int_{\mathbb{R}^3} e^{-i\hbar^{-1}(x\cdot p - mv\cdot x)}\psi(A^{-1}(x-a))dx \\ &= (2\pi\hbar)^{-3/2}\int_{\mathbb{R}^3} e^{-i\hbar^{-1}(A\cdot x+a)\cdot(p-mv)}\psi(x)dx \\ &= e^{-i\hbar^{-1}(p-mv)\cdot a}\mathcal{F}_\hbar\psi(A^{-1}(p-mv))\,,\end{aligned}$$

where $A^t = A^{-1}$ is used in the last step. Hence, if $\phi \in \mathcal{S}(\mathbb{R}^3)$ is a state in the momentum representation, it follows that the operator $\mathcal{F}_\hbar\mathcal{U}(a,0,v,A)\mathcal{F}_\hbar^*$ takes the form

$$\mathcal{F}_\hbar\mathcal{U}(a,0,v,A)\mathcal{F}_\hbar^*\phi(p) = e^{-i\hbar^{-1}((p-mv)\cdot a)}\phi(A^{-1}(p-mv))\,. \tag{6.28}$$

Now let us return to the problem of including time translations. Setting $s_2 = 0$, $a_1 = v_1 = v_2 = \underline{0}$ and $A_1 = A_2 = I$ in (6.25) gives

$$\mathcal{U}(a_1, s_1, \underline{0}, I)\mathcal{U}(a_2, s_2, \underline{0}, I) = e^{i\alpha}\mathcal{U}(a_1 + a_2, s_1 + s_2, \underline{0}, I)\,,$$

where the phase factor $e^{i\alpha}$ depends on a_1, a_2, s_1, s_2. If α vanishes identically this implies that $U(s) = \mathcal{U}(\underline{0}, -s, 0, A)$ and $\mathcal{T}_a = \mathcal{U}(a, 0, 0, I)$ commute for all s and a. In the following we assume this to be the case. Notice that this means that space translations are assumed to be symmetries of the system and by Noether's theorem it implies that the generators of translations, which are the components of the momentum, are conserved.

On the other hand, the boost operators do not commute with time translations, which is seen from the following calculation using (6.25):

$$\begin{aligned} \mathcal{B}_v U(s) &= \mathcal{U}(0, 0, v, I)\mathcal{U}(0, -s, 0, I) = e^{i\omega'}\mathcal{U}(-vs, -s, v, I) \\ &= e^{i\alpha'}U(s)\mathcal{B}_v\mathcal{T}_{-vt} \neq U(t)\mathcal{B}_v\,, \end{aligned} \tag{6.29}$$

where α' is some phase depending on v and t, and $vt \neq \underline{0}$ is assumed in the last step. Incidentally, this implies that the phase ω in (6.25) cannot vanish identically, no matter how $\mathcal{B}_v$ is defined, as long as $\mathcal{T}_a$ represents translation by a. In fact, otherwise it could be concluded that

$$\mathcal{T}_a\mathcal{B}_v = \mathcal{U}(a, 0, \underline{0}, I)\mathcal{U}(\underline{0}, 0, v, I) = \mathcal{U}(a, 0, v, I)$$

and

$$\mathcal{B}_v\mathcal{T}_a = \mathcal{U}(\underline{0}, 0, v, I)\mathcal{U}(a, 0, \underline{0}, I) = \mathcal{U}(a, 0, v, I)\,,$$

and hence $\mathcal{B}_v$ would commute with $\mathcal{T}_a$ for all v, a. But, as seen in Theorem 6.9, any operator commuting with translations is a multiplication operator in the momentum representation. Hence, both B_v and $U(s)$ would be multiplication operators in the momentum represention. In particular they would commute, which contradicts (6.29).

We are now ready to state the main result of this section.

Theorem 6.10. *Let $\mathcal{T}_a, \mathcal{R}_A, \mathcal{B}_v$ be defined by* (6.17), (6.22),(6.23)*, and let $U(t)$ be a strongly continuous unitary one-parameter group on $L_2(\mathbb{R}^3)$, which commutes with $\mathcal{T}_a, a \in \mathbb{R}^3$, and has the property that $\mathcal{U}(a, s, v, A)$ defined by* (6.24) *obeys* (6.25)*.*

Then there exists a constant e_0 such that

$$U(t) = \mathcal{F}_\hbar^* M_{\exp(-it\hbar^{-1}(\frac{p^2}{2m} + e_0))}\mathcal{F}_\hbar\,, \tag{6.30}$$

and (6.25) *is then satisfied with*

$$\omega(a_1, s_1, v_1, A_1, a_2, s_2, v_2, A_2) = -\hbar^{-1}m(\tfrac{1}{2}s_2v_1^2 + (a_1 + s_2v_1)\cdot(A_1v_2))\,. \tag{6.31}$$

Proof. First, note that by (6.25)

$$\mathcal{U}(a + vt, 0, v, A)U(t) = e^{-i\omega_2(a,v,A,t)}\mathcal{U}(a, -t, v, A)\,,$$

where $\omega_2(a, v, A, t) = -\omega(a + vt, 0, v, A, 0, -t, 0, I)$. Since $\mathcal{U}(a, -t, v, A) = U(t)\mathcal{U}(a, 0, v, A)$ it follows that

$$U(t)\mathcal{U}(a, 0, v, A) = e^{i\omega_2(a,v,A,t)}\mathcal{U}(a + vt, 0, v, A)U(t) \tag{6.32}$$

for all $a, v \in \mathbb{R}^3$, $A \in O(3)$, and $t \in \mathbb{R}$.

As already noted, since $U(t)$ commutes with $\mathcal{T}_a$ it has the form $U(t) = \mathcal{F}_\hbar^* M_{g(p,t)} \mathcal{F}_\hbar$ for some function $g(p)$ by Theorem 6.9. Using that $U(t)$ is a strongly continuous unitary one-parameter group, one concludes that $g(p)$ is purely imaginary and proportional to t such that $U(t) = \mathcal{F}_\hbar^* M_{\exp(-i\hbar^{-1}h(p)t)} \mathcal{F}_\hbar$, where $h(p)$ is a real-valued function (see Exercise 6.2).

The relation (6.32) can now be written

$$M_{\exp(-i\hbar^{-1}h(p)t)} \mathcal{F}_\hbar \mathcal{U}(a,0,v,A) \mathcal{F}_\hbar^* = e^{i\omega_2(a,v,A,t)} \mathcal{F}_\hbar \mathcal{U}(a+vt,0,v,A) \mathcal{F}_\hbar^* M_{\exp(-i\hbar^{-1}h(p)t)} \,.$$

According to (6.28) this is equivalent to

$$-th(p) - (p - mv) \cdot a = \hbar\omega_2(a,v,A,t) - (p-mv) \cdot (a+vt) - th(A^{-1}(p-mv)) \,,$$

that is

$$th(p) = -\hbar\omega_2(a,v,A,t) + (p-mv) \cdot vt + th(A^{-1}(p-mv)) \,. \tag{6.33}$$

Inserting $p = mv$ gives

$$\omega_2(a,v,A,t) = -t\hbar^{-1}(h(mv) - h(0)).$$

Now insert this and $p = 0$ into (6.33) to arrive at

$$h(0) = (h(mv) - h(0)) - mv \cdot v + h(-A^{-1}mv) \,.$$

From this identity it is easily deduced that $h(-A^{-1}mv)$ is independent of A and so $h(-A^{-1}mv) = h(mv)$ by either choosing $A = -I$ or to be a rotation that rotates v to $-v$. Thus, the resulting form of h is

$$h(mv) = \frac{1}{2} m \|v\|^2 + h(0) \,.$$

Setting $e_0 = h(0)$ it is seen that (6.30) holds and that

$$\omega_2(a,t,v,A) = -\tfrac{1}{2}\hbar^{-1} mt \|v\|^2 \tag{6.34}$$

It is straight-forward to check that (6.33) now holds for all a, v, p, A, and t. Hence (6.32) holds. It is also a straight-forward calculation to check that (6.32) together with (6.27) imples that (6.25) holds with ω given by (6.31). We leave this for the reader to check in Exercise 6.16. □

We have thus concluded the construction of a projective unitary representation (6.26) of the Galilei group for a single particle in $\mathbb{R}^3$ with no internal degrees of freedom. Projective unitary representations of the Galilei group corresponding to particles with internal degrees of freedom such as spin and to multi-particle systems can also be accomplished. Although we shall not here discuss how and in what sense one may obtain all possible projective unitary representations of the Galilei group, it is worth while noting that the basic philosophy is that different representations represent different physical systems.

The value e_0 may be interpreted as the internal energy of the particle. For a particle without internal degrees of freedom it may consistently be set to 0, which results in the free time evolution

$$U(t) = \mathcal{F}_\hbar^* M_{\exp(t\|p\|^2/2mi\hbar)} \mathcal{F}_\hbar \,, \tag{6.35}$$

to be discussed further in the next section.

6.5 The free particle

From (6.35) follows that the time evolution in the momemtum representation is given by unitary multiplication operators

$$\hat{U}(t) = \mathcal{F}_\hbar U(t) \mathcal{F}_\hbar^* = M_{\exp(t\|p\|^2/2mi\hbar)}\,, \quad t \in \mathbb{R}\,,$$

from which we read off that the Hamiltonian in the momentum representation is the self-adjoint operator

$$\hat{\mathcal{H}} = M_{\|p\|^2/2m}\,,$$

that is multiplication by the classical energy of a particle with mass m and momentum p. It is precisely for this reason that it is natural to interpret m as being the mass and p as being the momentum variable.

In the Schrödinger representation the Hamiltonian is

$$\mathcal{H} = \mathcal{F}_\hbar^* M_{\|p\|^2/2m} \mathcal{F}_\hbar\,,$$

which is a self-adjoint operator by Exercises 6.3 and 6.5 with domain

$$\mathcal{D}(\mathcal{H}) = \left\{ \psi \in L_2(\mathbb{R}^3) \;\middle|\; \int_{\mathbb{R}^3} \|p\|^2 |\mathcal{F}_\hbar \psi(p)|^2 dp < \infty \right\}.$$

Using Lemma 6.8 it follows in the same way as discussed for the partial derivative operators in Section 6.3 that $\mathcal{H}$ is a self-adjoint extension of the operator

$$-\frac{\hbar^2}{2m}(\partial_{x_1}^2 + \partial_{x_2}^2 + \partial_{x_2}^2) = -\frac{\hbar^2}{2m}\Delta\,,$$

where $\Delta = \partial_{x_1}^2 + \partial_{x_2}^2 + \partial_{x_2}^2$ is the Laplace operator originally defined on $C_0^2(\mathbb{R}^3)$. Hence we write

$$\mathcal{H} = \frac{\hbar^2}{2m}(D_1^2 + D_2^2 + D_3^2)\,.$$

In view of the preceding discussion it is natural to define the free quantum mechanical particle in $\mathbb{R}^k$ in the Schrödinger representation by asserting its state space to be $L_2(\mathbb{R}^k)$ and its Hamiltonian to be

$$\mathcal{H} = \frac{\hbar^2}{2m}(D_1^2 + \cdots + D_k^2)$$

with domain

$$\mathcal{D}(\mathcal{H}) = \left\{ \psi \in L_2(\mathbb{R}^k) \;\middle|\; \int_{\mathbb{R}^k} \|p\|^2 |\mathcal{F}_\hbar \psi(p)|^2 dp < \infty \right\}.$$

Of course, this operator is again a self-adjoint extension of the operator

$$-\frac{\hbar^2}{2m}\Delta = -\frac{\hbar^2}{2m}(\partial_{x_1}^2 + \cdots + \partial_{x_k}^2)$$

defined on $C_0^2(\mathbb{R}^k)$.

It is clear that $\mathcal{H}$ is a positive operator in the sense that all its expectation values are positive since

$$\langle \psi | \mathcal{H} | \psi \rangle = \frac{1}{2m} \int_{\mathbb{R}^k} \|p\|^2 |\mathcal{F}_\hbar \psi(p)|^2 dp > 0,$$

if $\psi \neq 0$. The positivity can also be seen explicitly in the Schrödinger representation at least for functions $\psi \in \mathcal{S}(\mathbb{R}^k)$. Indeed, for such functions one gets by using integration by parts that

$$-\int_{\mathbb{R}^k} \overline{\psi(x)} \Delta\psi(x) dx = \int_{\mathbb{R}^k} |\nabla\psi(x)|^2 dx \tag{6.36}$$

(see Exercise 6.17).

The Schrödinger equation in the momentum representation takes the form

$$i\hbar \frac{d}{dt}\phi(t,p) = \frac{1}{2m}\|p\|^2 \phi(t,p)\,.$$

The solution to this equation is of course

$$\phi(t,p) = e^{-i\frac{p^2}{2m\hbar}t}\phi_0(p)\,,$$

provided the initial value $\phi(0,\cdot) = \phi_0$ belongs to $\mathcal{D}(M_{p^2/2m})$.

The solution to the Schrödinger equation

$$i\frac{d}{dt}\psi(t) = \frac{\hbar^2}{2m}(D_1^2 + \cdots + D_k^2)\psi(t)$$

in the Schrödinger representation with initial value ψ_0 is obtained by applying the inverse Fourier transformation to the solution in the momentum representation with initial value $\mathcal{F}_\hbar\psi_0$:

$$\psi(t) = \mathcal{F}_\hbar{}^*(e^{\frac{\|p\|^2}{2mi\hbar}t}\mathcal{F}_\hbar\psi_0)\,.$$

Note that two integrations are involved on the right, one for each of the two Fourier transformations. It turns out that one of the integrations can be performed explicitly, yielding the result

$$\psi(t,x) = \Big(\frac{m}{2\pi\hbar it}\Big)^{\frac{k}{2}} \int_{-\infty}^{\infty} e^{im\frac{\|x-y\|^2}{2\hbar t}}\psi_0(y)\,dy\,,$$

valid for $\psi_0 \in \mathcal{S}(\mathbb{R}^k)$. This formula can e. g. be used to analyze the spreading in time of the wavefunction due to the uncertainty of momentum as can be found in most text books on introductory quantum mechanics.

6.6 The free particle on a circle

Contrary to the free particle moving on the unbounded real line, a particle constrained to a finite region of space can have eigenstates of definite momentum. A simple example of this phenomenon is the free particle whose configuration space is S^1, the unit circle in $\mathbb{R}^2$ centered at the origin. Since functions on S^1 can be identified with functions on $(-\pi,\pi]$, when parametrizing the circle by the angular variable θ, such that

$$(x,y) = (\cos\theta, \sin\theta)\,,$$

the state space in the Schrödinger representation is taken to be

$$H = L_2((-\pi,\pi]) = L_2([-\pi,\pi])$$

with inner product normalized as in Section 4.6. The two coordinate variables are represented by the multiplication operators $M_{\cos}$ and $M_{\sin}$ and the momentum is represented by the differential operator $\hbar\bar{D}$ defined in Example 5.19. The two multiplication operators are self-adjoint by Example 5.6 b) and $\bar{D}$ is self-adjoint by Exercise 5.17.

The Hamiltonian is defined by analogy with the free particle on the line as

$$\mathcal{H} = \frac{\hbar^2}{2m}\bar{D}^2\,,$$

where m is called the mass of the particle. Since e_n (with notation as in Section 4.6) is an eigenvector for $\bar{D}$ with eigenvalue n it follows that e_n is likewise an eigenvector for $\mathcal{H}$ with eigenvalue

$$E_n = \frac{\hbar^2 n^2}{2m}\,.$$

Thus $\mathcal{H}$ is an unbounded diagonalizable operator given by

$$\mathcal{H} = \frac{\hbar^2}{2m}\sum_{n\in\mathbb{Z}} n^2 |e_n\rangle\langle e_n|$$

and with domain

$$\mathcal{D}(\mathcal{H}) = \left\{ \psi \in L_2([-\pi,\pi]) \,\middle|\, \sum_{n\in\mathbb{Z}} n^4 |\langle e_n|\psi\rangle|^2 < \infty \right\}\,.$$

The possible values of the (kinetic) energy of the particle are $E_n\,,\ n = 0,1,2,3\ldots$, and for each positive energy value E_n there correspond two momentum eigenstates $e_{\pm n}$ of opposite momenta $\pm\hbar n$.

The time evolution operator $e^{-\frac{i}{\hbar}\mathcal{H}t}$ has the same eigenvectors e_n as $\mathcal{H}$ with corresponding eigenvalues $e^{-\frac{i}{\hbar}E_n t}$ (see Exercise 5.12 for the case where $\mathcal{H}$ is bounded). This means that $e^{-\frac{i}{\hbar}\mathcal{H}t}$ is diagonalizable and given by

$$e^{-\frac{i}{\hbar}\mathcal{H}t} = \sum_{n\in\mathbb{Z}} e^{-\frac{i\hbar}{2m}n^2 t} |e_n\rangle\langle e_n|\,,$$

which also shows explicitly that it is unitary. Indeed, it can be shown that the solution to the Schrödinger equation

$$i\frac{d\psi(t)}{dt} = \frac{\hbar}{2m}\bar{D}^2\psi(t) \tag{6.37}$$

with initial condition $\psi(0) = \psi_0$ is given by

$$\psi(t) = \sum_{n\in\mathbb{Z}} e^{-\frac{i\hbar}{2m}n^2 t}\langle e_n|\psi_0\rangle e_n\,, \tag{6.38}$$

provided $\psi_0 \in \mathcal{D}(\mathcal{H})$. Note that while $\psi(t)$ is well defined by (6.38) for all $\psi_0 \in H$, the Schrödinger equation (6.37) only makes sense if $\psi(t) \in \mathcal{D}(\mathcal{H})$ for all $t \in \mathbb{R}$. Hence the above statement requires showing that if $\psi_0 \in \mathcal{D}(\mathcal{H})$, then $\psi(t)$ defined by (6.38) belongs to $\mathcal{D}(\mathcal{H})$ for all $t \in \mathbb{R}$ and provides a solution to (6.37) with initial condition $\psi(0) = \psi_0$. We refrain from giving the detailed argument here (see, however, Exercise 5.14 for the case where $\mathcal{H}$ is bounded).

As mentioned, the states $e_n\,,\ n \in \mathbb{Z}$, are characterized by having well defined momenta $\hbar n$. However, the position operators $M_{\cos}$ and $M_{\sin}$ have no eigenvalues, and the same holds for M_θ. Consequently, states in which the particle has a well defined position do not exist. However, for any normalized state ψ represented by a wave function that vanishes (almost everywhere) outside an interval $I = [\theta_1, \theta_2]$ contained in $[-\pi, \pi]$ it is clear that

$$\langle \psi | M_\theta | \psi \rangle \in I\,,$$

and it follows by arguments similar to those for the free particle on the real line that the (closed) subspace $L_2(I)$ of functions in H vanishing outside I is naturally interpreted as consisting of states for which the particle is localized in the interval I. Obviously, the function in $L_2(I)$ closest to a given function $\psi \in H$ is the function coinciding with ψ on I, so the orthogonal projection P_I onto $L_2(I)$ is given by

$$P_I\psi(\theta) = \begin{cases} \psi(\theta) & \text{if } \theta \in I \\ 0 & \text{if } \theta \notin I\,. \end{cases}$$

In particular, the probability that the particle in a state ψ is located in the angular interval I is

$$\|P_I\psi\|^2 = \int_I |\psi(\theta)|^2 d\theta\,.$$

which means that $|\psi(\theta)|^2$ is the probability density for finding the particle at $(\cos\theta, \sin\theta)$.

6.7 The harmonic oscillator

The one-dimensional harmonic oscillator is the system consisting of one particle constrained to a line and acted on by a linear force. Thus the state space is the same as for the free particle, $H = L_2(\mathbb{R})$, and the position and momentum operators are also the same. The Hamiltonian, on the other hand, is obtained by adding the potential, now regarded as a multiplication operator, which is quadratic to the free particle Hamiltonian. More precisely we define

$$\mathcal{H}_0 = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}kx^2\,, \tag{6.39}$$

where k is a positive constant. Since both operators in the above sum are unbounded, appropriate attention has to be paid to the domain of $\mathcal{H}_0$. Our strategy will be to take as the domain of $\mathcal{H}_0$ the space $\mathcal{S}(\mathbb{R})$ of rapidly decreasing functions and to find an orthonormal basis of eigenvectors $\Omega_n\,,\ n = 0, 1, 2, 3, \ldots$, in this space. Once this has been accomplished $\mathcal{H}_0$ can be extended to a self-adjoint operator $\mathcal{H}$ in a similar way as done for the operator D in Example 5.19.

First, let us note that since both $\frac{d}{dx}$ and M_{x^2} evidently map $\mathcal{S}(\mathbb{R})$ into itself, the same holds for $\mathcal{H}$. Second, let us introduce the two operators

$$A = \frac{d}{dx} + cx \quad \text{and} \quad A^\dagger = -\frac{d}{dx} + cx\,, \tag{6.40}$$

where

$$c = \frac{\sqrt{mk}}{\hbar}\,. \tag{6.41}$$

For reasons that will become clear below $A^\dagger$ is called a *raising operator* and A a *lowering operator*. Given any $f \in \mathcal{S}(\mathbb{R})$ a simple calculation yields

$$(A^\dagger A f)(x) = -\frac{d^2 f}{dx^2} + c^2 x^2 f(x) - c f(x)\,,$$

that is

$$\mathcal{H}_0 = \frac{\hbar^2}{2m}(A^\dagger A + c) \quad \text{on } \mathcal{S}(\mathbb{R})\,. \tag{6.42}$$

In addition, one easily verifies the relation

$$A^\dagger A = A A^\dagger - 2c \quad \text{on } \mathcal{S}(\mathbb{R})\,,$$

which is commonly written as

$$[A, A^\dagger] = 2c\,, \tag{6.43}$$

where

$$[A, A^\dagger] = AA^\dagger - A^\dagger A$$

is the *commutator* of A and $A^\dagger$. From (6.42) it is seen that a state $\Omega \in \mathcal{S}(\mathbb{R})$ which satisfies

$$A\Omega = \Omega' + c\Omega = 0 \tag{6.44}$$

is an eigenstate for $\mathcal{H}_0$ with eigenvalue $E_0 = \frac{\hbar^2}{2m}c$. But (6.44) is a linear first order differential equation whose solutions are proportional to $e^{-cx^2/2}$. Since $c > 0$, this function belongs to $\mathcal{S}(\mathbb{R})$ (see Exercise 6.20). In particular, it is square integrable and there exists a constant $a > 0$ such that the function

$$\Omega_0(x) = a\, e^{-cx^2/2} \tag{6.45}$$

has L_2-norm 1.

The key to obtain more eigenstates for $\mathcal{H}_0$ is to observe that if $\Omega \in \mathcal{S}(\mathbb{R})$ is an eigenstate for $A^\dagger A$ with eigenvalue λ then, as a consequence of (6.43), the state $A^\dagger\Omega$, provided it is non-zero, is an eigenstate of $A^\dagger A$ with eigenvalue $\lambda + 2c$:

$$A^\dagger A(A^\dagger\Omega) = A^\dagger(A^\dagger A + 2c)\Omega = A^\dagger(\lambda + 2c)\Omega = (\lambda + 2c)A^\dagger\Omega\,.$$

Since $\mathcal{S}(\mathbb{R})$ is evidently invariant under $A^\dagger$, we conclude by induction that the states

$$\Omega_n = a_n (A^\dagger)^n \Omega_0\,, \quad n = 0, 1, 2, 3, \ldots\,, \tag{6.46}$$

all of which are non-zero as will be seen below, are eigenstates for $\mathcal{H}_0$ with corresponding eigenvalues

$$E_n = \frac{\hbar^2}{2m}(2nc + c) = \hbar\omega(n + \frac{1}{2})\,, \quad n = 0, 1, 2, 3, \ldots\,, \tag{6.47}$$

where

$$\omega = \sqrt{\frac{k}{m}}$$

is the frequency of the classical harmonic oscillator. Here the coefficients $a_n > 0$ are chosen such that Ω_n has norm 1. It will now be shown that $E_n\,,\ n = 0, 1, 2, \ldots$, is the exhaustive list of energy levels for $\mathcal{H}_0$ (and its extension $\mathcal{H}$) by establishing the following result:

Theorem 6.11. *The set $(\Omega_n)_{n=0,1,2,\dots}$ is an orthonormal basis for $L_2(\mathbb{R})$.*

Proof. Note first that, for $f, g \in \mathcal{S}(\mathbb{R})$,

$$\begin{aligned}
\langle f|Ag\rangle &= \int_{\mathbb{R}} \overline{f(x)}(g'(x) + xg(x))dx \\
&= \lim_{R\to\infty} \int_{-R}^{R} \overline{f(x)}g'(x)dx + \int_{\mathbb{R}} x\overline{f(x)}g(x)dx \\
&= \lim_{R\to\infty} \left\{ [\overline{f(x)}g(x)]_{-R}^{R} - \int_{-R}^{R} \overline{f'(x)}g(x)dx \right\} + \int_{\mathbb{R}} x\overline{f(x)}g(x)dx \\
&= \int_{\mathbb{R}} (\overline{-f'(x) + xf(x)})g(x)dx \\
&= \langle A^\dagger f|g\rangle \,.
\end{aligned} \tag{6.48}$$

This shows that the domain of the adjoint of the (unbounded) operator A contains $\mathcal{S}(\mathbb{R})$ and that A^* coincides with $A^\dagger$ on this space.

Combining (6.42) and (6.48) it follows that

$$\langle f, \mathcal{H}_0 g\rangle = \langle \mathcal{H}_0 f, g\rangle \,, \quad f, g \in \mathcal{S}(\mathbb{R}) \,, \tag{6.49}$$

showing that $\mathcal{H}_0$ is symmetric on $\mathcal{S}(\mathbb{R})$. Since the eigenfunctions Ω_n belong to $\mathcal{S}(\mathbb{R})$, the proof of Lemma 5.11 b) still applies to show that Ω_n is orthogonal to Ω_m for $n \neq m$. Hence $(\Omega_n)_{n=0,1,2,\dots}$ is an orthonormal set in $L_2(\mathbb{R})$.

It remains to show that

$$\{\Omega_n \mid \ n = 0, 1, 2, \dots\}^\perp = \{\underline{0}\} \,.$$

Defining the function ϕ_p for fixed $p \in \mathbb{R}$ by

$$\varphi_p(x) = e^{ixp}\Omega_0(x) \,, \quad x \in \mathbb{R} \,,$$

it suffices to establish the following two facts:

1) $\quad \phi_p \in \overline{\text{span}}\{\Omega_n \mid n = 0, 1, 2, \dots\}$ for all $p \in \mathbb{R}$
2) $\quad \{\varphi_p \mid p \in \mathbb{R}\}^\perp = \{\underline{0}\} \,.$

Indeed, if 1) and 2) hold, then

$$\{\Omega_n \mid \ n = 0, 1, 2, \dots\}^\perp = \overline{\text{span}}\{\Omega_n \mid \ n = 0, 1, 2, \dots\}^\perp \subseteq \{\varphi_p \mid p \in \mathbb{R}\}^\perp = \{\underline{0}\}$$

as desired.

From the definition of Ω_n it follows that it has the form

$$\Omega_n(x) = P_n(x)e^{-cx^2/2} \,,$$

where P_n is a polynomial of order n. As a matter of fact, this is obviously the case for $n = 0$ and then it follows easily by induction for all n, since

$$A^\dagger(P_n(x)e^{-cx^2/2}) = (2cxP_n(x) - P_n'(x))e^{-cx^2/2} \,,$$

where $2cxP_n(x)$ is a polynomial of order $n+1$ while $P_n'(x)$ is of order $n-1$, if $P_n(x)$ is of order n. In particular, it follows that $\Omega_n \neq 0$ for all $n = 0, 1, 2, \dots$ as claimed previously, and that

$$\operatorname{span}\{\Omega_n \mid n = 0, 1, 2, \dots\} = \{Q\Omega_0 \mid Q \text{ polynomial}\}\,.$$

In order to establish 1) above it therefore suffices to show that for any given $p \in \mathbb{R}$ and any $\varepsilon > 0$ there exists a polynomial Q such that

$$\|\varphi_p - Q\Omega_0\|^2 = \int_{\mathbb{R}} a^2 |e^{ixp} - Q(x)|^2 e^{-cx^2} dx < \varepsilon^2\,. \tag{6.50}$$

Let $T_n(x)$ be the n'th Taylor polynomial for exp,

$$T_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}\,.$$

Note that $T_n(ix) = C_n(x) + iS_n(x)$ where C_n and S_n are the corresponding Taylor polynomials for cos and sin, respectively, and Taylors formula with remainder for sin and cos then implies for each $R > 0$ the existence of an n such that

$$|e^{ixp} - T_n(ixp)| < \frac{\varepsilon}{2a\sqrt{R}} \qquad \text{for } x \in [-R, R]\,.$$

Moreover, the bound

$$|T_n(ipx)| \leqslant \sum_{k=0}^{n} \frac{|px|^k}{k!} \leqslant \exp(|xp|)$$

trivially holds for all x and p and implies

$$|e^{ixp} - T_n(ixp)|^2 e^{-cx^2} \leqslant 4e^{2|xp| - cx^2}\,.$$

Since the last function is integrable (verify this!) it is possible to choose $R > 0$ such that

$$\int_{|x|>R} 4e^{2|xp| - cx^2} < \frac{\varepsilon^2}{2a^2}\,.$$

With this choice of R and n one finally obtains

$$\begin{aligned}
\|\varphi_p - T_n(ixp)\Omega_0\|^2 &= \int_{-R}^{R} a^2 |e^{ixp} - T_n(ixp)|^2 e^{-cx^2} dx + \int_{|x|>R} a^2 |e^{ixp} - T_n(ixp)|^2 e^{-cx^2} dx \\
&\leqslant \int_{-R}^{R} a^2 \left(\frac{\varepsilon}{2a\sqrt{R}}\right)^2 dx + \int_{|x|>R} 4a^2 e^{2|xp| - cx^2}\, dx \\
&< \frac{\varepsilon^2}{2} + \frac{\varepsilon^2}{2} = \varepsilon^2\,.
\end{aligned}$$

Thus, setting $Q(x) = T_n(ixp)$ the estimate (6.50) holds and 1) is proven.

In order to verify 2), assume that $f \in L_2(\mathbb{R})$ is orthogonal to all $\varphi_p\,,\ p \in \mathbb{R}$, that is

$$\int_{\mathbb{R}} f(x)\Omega_0(x) e^{ixp} dx = 0 \quad \text{for all } p \in \mathbb{R}\,.$$

By definition of the Fourier transformation this means that $\mathcal{F}(f\Omega_0) = 0$. Since $\mathcal{F}$ is an isometry this implies $f\Omega_0 = 0$ and hence $f = 0$, since Ω_0 is a positive function. This proves 2) and thereby the theorem. □

Having found an orthonormal basis of eigenvectors for $\mathcal{H}_0$, this operator can be extended to a self-adjoint operator $\mathcal{H}$ by the same procedure as used in Example 5.19. This means that $\mathcal{H}$ is defined by the formula

$$\mathcal{H}\psi = \hbar\omega \sum_{n=0}^{\infty} (n + \frac{1}{2}) \langle \Omega_n | \psi \rangle \, \Omega_n \tag{6.51}$$

on the domain

$$\mathcal{D}(\mathcal{H}) = \left\{ \psi \in H \;\middle|\; \sum_{n=0}^{\infty} (n + \frac{1}{2})^2 \, |\langle \Omega_n | \psi \rangle|^2 \; < \; \infty \right\},$$

or in Dirac notation

$$\mathcal{H} = \hbar\omega \sum_{n=0}^{\infty} (n + \frac{1}{2}) \, |\Omega_n\rangle\langle\Omega_n| \, .$$

To see that this is really an extension of $\mathcal{H}_0$, let $\psi \in \mathcal{S}(\mathbb{R})$ and use (6.49) to write

$$\hbar\omega(n + \frac{1}{2})\langle \Omega_n | \psi \rangle = \langle \mathcal{H}_0 \Omega_n | \psi \rangle = \langle \Omega_n | \mathcal{H}_0 \psi \rangle \, .$$

Since $\mathcal{H}_0\psi \in L_2(\mathbb{R})$ it follows that

$$\hbar^2\omega^2 \sum_{n=0}^{\infty} (n + \frac{1}{2})^2 \, |\langle \Omega_n | \psi \rangle|^2 = \|\mathcal{H}_0\psi\|^2 < \infty$$

and hence $\psi \in \mathcal{D}(\mathcal{H})$. This shows that $\mathcal{S}(\mathbb{R}) \subseteq \mathcal{D}(\mathcal{H})$ and

$$\mathcal{H}\psi = \hbar\omega \sum_{n=0}^{\infty} (n + \frac{1}{2}) \, \langle \Omega_n | \psi \rangle \, \Omega_n = \sum_{n=0}^{\infty} \langle \Omega_n | \mathcal{H}_0 \psi \rangle \Omega_n = \mathcal{H}_0\psi.$$

The energy levels of the oscillator are given by (6.47). The time evolution operator $e^{-i\frac{t}{\hbar}\mathcal{H}}$ is unitary and given by

$$e^{-i\frac{t}{\hbar}\mathcal{H}} = \sum_{n=0}^{\infty} e^{-i\omega(n+\frac{1}{2})t} \, |\Omega_n\rangle\langle\Omega_n| \, .$$

Thus the solution to the Schrödinger equation (6.4) with initial value ψ_0 is in this case given by

$$\psi(t) = \sum_{n=0}^{\infty} e^{-i\omega(n+\frac{1}{2})t} \, \rangle \langle \Omega_n | \psi_0 \rangle \Omega_n \, , \tag{6.52}$$

provided $\psi_0 \in \mathcal{D}(\mathcal{H})$.

Remark 6.12. The eigenfunctions Ω_n have independent interest and are well known from classical analysis. In case $c = 1$ they are called *Hermite functions* and the polynomials P_n are, up to normalization, called the *Hermite polynomials*. The first few of them are given in Exercise 6.22. A detailed analysis of the Hermite functions allows one to carry out the sum in (6.52) obtaining an integral formula for the solution (see e. g. [8] p. 19.)

It is a consequence of Lemma 6.8 that if $c = 1$ then

$$\mathcal{F} \circ \mathcal{H}_0 = \mathcal{H}_0 \circ \mathcal{F} .$$

In other words, $\mathcal{F}$ commutes with $\mathcal{H}_0$ on $\mathcal{S}(\mathbb{R})$. Applying both sides to Ω_n implies

$$\mathcal{H}_0(\mathcal{F}\Omega_n) = \hbar\omega(n + \frac{1}{2})\mathcal{F}\Omega_n .$$

Since the eigenstate for $\mathcal{H}_0$ corresponding to eigenvalue $\hbar\omega(n+\frac{1}{2})$ is unique up to a constant multiple there exists a constant c_n such that

$$\hat{\Omega}_n = c_n\Omega_n .$$

This formula shows that the Fourier transformation is a diagonalizable operator with the Hermite functions as eigenfunctions and corresponding eigenvalues equal to c_n. One can show (see Exercise 6.21) that $c_n = (-i)^n$, that is the Fourier transformation has four different eigenvalues $\pm 1, \pm i$.

6.8 Stability and the hydrogen atom

One of the early triumphs of quantum mechanics was to explain the stability of the hydrogen atom. Stability is an important notion in physics. This last section is devoted to a brief introductory discussion of this topic.

As seen in Section 6.4 the kinetic energy of a particle is a positive quantity if normalized in such a way that the kinetic energy is zero if the particle is at rest, or more precisely nearly zero if the particle is nearly at rest, since in quantum mechanics it cannot be completely at rest.

The potential energy, however, may be negative and so the total energy of a system may be negative. Through interactions one system can transfer energy to other systems without violating conservation of energy. A system which may possess an arbitrarily large negative energy could in this way be a source of an infinite amount of energy and would be unstable.

Definition 6.13. *A quantum mechanical system with state space H and Hamiltonian $\mathcal{H}$ is said to be* **stable** *if*

$$E_0 = E_0(\mathcal{H}) = \inf\{\langle\psi|\mathcal{H}|\psi\rangle \ \mid \psi \in \mathcal{D}(\mathcal{H}) \text{ normalized } \} > -\infty.$$

The quantity E_0 is called the ground state energy of the system and if there is a normalized $\psi \in \mathcal{D}(\mathcal{H})$ such that $E_0 = \langle\psi|\mathcal{H}|\psi\rangle$ then we say that ψ is a ground state. (A ground state is not necessarily unique.)

Theorem 6.14. *If ψ is a normalized ground state of a stable quantum mechanical system with Hamiltonian $\mathcal{H}$ and with ground state energy $E_0 = E_0(\mathcal{H})$, then ψ is an eigenvector of $\mathcal{H}$ with eigenvalue E_0, that is*

$$\mathcal{H}\psi = E_0\psi\,.$$

Moreover, E_0 is smallest eigenvalue of $\mathcal{H}$.

Proof. For any $\phi \in \mathcal{D}(\mathcal{H})$ the function

$$E(t) = \frac{\langle \psi + t\phi | \mathcal{H} | \psi + t\phi \rangle}{\langle \psi + t\phi | \psi + t\phi \rangle}.$$

is well defined in a sufficiently small interval around 0 such that $\psi + t\phi \neq \underline{0}$, and it fulfills $E(t) \geqslant E(0) = E_0$. Since $E(t)$ is the ratio of two polynomials of degree at most two it is differentiable and $E'(0) = 0$. Using that $\langle \psi | \psi \rangle = 1$ and $\langle \psi | \mathcal{H} | \psi \rangle = E_0$ a straight-forward calculation gives

$$E'(0) = \langle \psi | \mathcal{H} - E_0 | \phi \rangle + \langle \phi | \mathcal{H} - E_0 | \psi \rangle = 0\,.$$

Replacing ϕ by $i\phi$ (which is also in $\mathcal{D}(\mathcal{H})$) in this identity yields

$$0 = i\langle \psi | \mathcal{H} - E_0 | \phi \rangle - i\langle \phi | \mathcal{H} - E_0 | \psi \rangle\,.$$

These two last identities now imply $\langle \phi | \mathcal{H} | \psi \rangle = E_0 \langle \phi | \psi \rangle$ for all $\phi \in \mathcal{D}(\mathcal{H})$. Since $\mathcal{D}(\mathcal{H})$ is dense in H it follows that $\mathcal{H}\psi = E_0\psi$.

If λ is another eigenvalue of $\mathcal{H}$ corresponding to a normalized eigenvector $\psi_\lambda \in \mathcal{D}(\mathcal{H})$ then $\lambda = \langle \psi_\lambda | \mathcal{H} | \psi_\lambda \rangle \geqslant E_0$. □

Note that E_0 is called the ground state energy even if there is no state corresponding to this energy, i. e. no ground state (see Exercise 6.18).

The definition of stability given above assumes that the Hamiltonian of the system is known. In many cases one only has knowledge of an operator $\widetilde{\mathcal{H}}$ on some dense domain $\mathcal{D}(\widetilde{\mathcal{H}})$ on which it is not necessarily self-adjoint, but where

$$E_0(\widetilde{\mathcal{H}}) = \inf\{\langle \psi | \widetilde{\mathcal{H}} | \psi \rangle \;\; | \; \psi \in \mathcal{D}(\widetilde{\mathcal{H}}) \text{ normalized } \} > -\infty.$$

If $\widetilde{\mathcal{H}}$ is not self-adjoint it is not the Hamiltonian of a quantum mechanical system. We will nevertheless say that $\widetilde{\mathcal{H}}$ is stable in this case. It is a fundamental result which will not be proven here that in this case there is a self-adjoint extension $\mathcal{H}$, known as the *Friedrichs extension*, of $\widetilde{\mathcal{H}}$ such that

$$E_0(\mathcal{H}) = E_0(\widetilde{\mathcal{H}}).$$

That $\mathcal{H}$ is a self-adjoint extension of $\widetilde{\mathcal{H}}$ means that it is a self-adjoint operator $\mathcal{H} : \mathcal{D}(\mathcal{H}) \to H$ such that

$$\mathcal{D}(\widetilde{\mathcal{H}}) \subseteq \mathcal{D}(\mathcal{H}), \quad \text{and} \quad \widetilde{\mathcal{H}}\phi = \mathcal{H}\phi, \quad \text{for all } \phi \in \mathcal{D}(\widetilde{\mathcal{H}}).$$

The Friedrichs extension $\mathcal{H}$ is the Hamiltonian of a stable quantum mechanical system. It is possible that there is a ground state in the domain of $\mathcal{H}$ but that it is not in the smaller domain of $\widetilde{\mathcal{H}}$. Thus, although the ground state energy is determined by the operator $\widetilde{\mathcal{H}}$ this may not be the case with the ground state itself.

The state space of the hydrogen atom, where the nucleus is considered as infinitely heavy, is the same as for the free particle, namely $L_2(\mathbb{R}^3)$. The operator

$$\widetilde{\mathcal{H}} = -\frac{\hbar^2}{2m}\Delta - \frac{e^2}{\|x\|} \tag{6.53}$$

defined on the space $C_0^2(\mathbb{R}^3)$ of twice continuously differentiable functions on $\mathbb{R}^3$) that vanish outside a bounded set, is taken as the starting point for defining the Hamiltonian. Here m is the mass of the electron and $-e < 0$ is its charge. The first term in $\mathcal{H}$ is the kinetic energy operator and the second term is the (Coulomb) potential energy due to the attraction of the electron with charge $-e$ to the nucleus with charge $+e$.

In Exercise 6.23 the reader is asked to verify that $\widetilde{\mathcal{H}}$ is well defined on $C_0^2(\mathbb{R}^3)$. It will be shown below that $\widetilde{\mathcal{H}}$ is a densely defined symmetric operator, which therefore has a self-adjoint Friedrichs extension $\mathcal{H}$.

In the preceding considerations the fact that the electron has an internal degree of freedom called its *spin* has not been taken into account. To do so would require replacing the state space $L_2(\mathbb{R}^3)$ by the Hilbert space

$$H = L_2(\mathbb{R}^3) \times L_2(\mathbb{R}^3)$$

of square integrable functions with two components representing spin components. However, for the discussion here the spin is not important and we will focus on showing that $\widetilde{\mathcal{H}}$ is stable and determine its ground state energy, which will also be the ground state energy of hydrogen.

Classically, the hydrogen atom would be described by the Hamilton function

$$H(p,q) = \frac{p^2}{2m} - \frac{e^2}{\|q\|}, \quad (q,p) \in \mathbb{R}^3 \times \mathbb{R}^3 .$$

Clearly, this function can become arbitrarily negative and thus the hydrogen atom is not stable classically. In quantum mechanics, on the other hand, it is stable as shown in the subsequent theorem. In physics textbooks the stability of hydrogen is often explained from the *Heisenberg uncertainty relation*. It is in principle correct that the stability of hydrogen is related to the uncertainty principle, but the stability does not follow from the standard form of the Heisenberg uncertainty relations (see Exercise 6.25).

Theorem 6.15. *Let $\Omega \subseteq \mathbb{R}^k$ be an open set and $V \in C(\Omega)$. Assume that $0 < \psi \in C^2(\Omega)$ and that $(-\Delta - V)\psi(x) = \lambda\psi(x)$ for some $\lambda \in \mathbb{R}$ and all $x \in \Omega$. Then, for all functions ϕ in the set $C_0^1(\Omega)$ of C^1 functions on Ω that vanish outside a closed and bounded set inside Ω, the following inequality holds:*

$$\int_\Omega |\nabla\phi(x)|^2 dx - \int_\Omega V(x)|\phi(x)|^2 dx \geqslant \lambda \int_\Omega |\phi(x)|^2 dx$$

Proof. Given $\phi \in C_0^1(\Omega)$ one can write $\phi = f\psi$, where $f \in C_0^1(\Omega)$ (why?). By using the

same integration by parts as in (6.36) one then obtains

$$\begin{aligned}
&\int_\Omega |\nabla\phi(x)|^2 dx - \int_\Omega V(x)|\phi(x)|^2 dx \\
&= \int_\Omega \left[\psi^2|\nabla f|^2 + |f|^2|\nabla\psi|^2 + \left(\overline{f}\nabla f + f\overline{\nabla f}\right)\psi\nabla\psi\right] dx - \int_\Omega V|f\psi|^2 dx \\
&\geqslant \int_\Omega \left[|f|^2|\nabla\psi|^2 + \left(\overline{f}\nabla f + f\overline{\nabla f}\right)\psi\nabla\psi\right] dx - \int_\Omega V|f\psi|^2 dx \\
&= \int_\Omega \left[|f|^2|\nabla\psi|^2 + \nabla(|f|^2)\psi\nabla\psi\right] dx - \int_\Omega V|f\psi|^2 dx \\
&= \int_\Omega \left[|f|^2\psi(-\Delta - V)\psi\right] dx = \lambda\int_\Omega |\phi(x)|^2 dx\,,
\end{aligned}$$

where the last step again follows by integration by parts. □

Recalling the identity

$$\langle\phi| - \Delta - V|\phi\rangle = \int |\nabla\phi(x)|^2 dx - \int V(x)|\phi(x)|^2 dx\,,$$

which holds if $\phi \in C_0^2(\mathbb{R}^k)$ by (6.36), the previous theorem loosely speaking states that for an operator of the form $-\Delta - V$ (a Schrödinger operator) an eigenfunction which is positive must correspond to the lowest eigenvalue.

Corollary 6.16. *The hydrogen atom is a stable quantum mechanical system and its ground state energy is* $E_0(\widetilde{\mathcal{H}}) = -\frac{me^4}{2\hbar^2}$.

Proof. Consider the function $\psi(x) = e^{-me^2\|x\|/\hbar^2}$ defined on the set $\Omega = \mathbb{R}^3 \setminus \{0\}$. For any $x \in \Omega$ a straightforward calculation shows that $(-\frac{\hbar^2}{2m}\Delta - e^2\|x\|^{-1})\psi(x) = -\frac{me^4}{2\hbar^2}\psi(x)$. The fact that $\langle\phi|\widetilde{\mathcal{H}}|\phi\rangle \geqslant -\frac{me^4}{2\hbar^2}$ for any normalized $\phi \in C_0^2(\Omega)$ follows from the previous theorem by using again (6.36) to get

$$\langle\phi|\widetilde{\mathcal{H}}|\phi\rangle = \frac{\hbar^2}{2m}\int_\Omega |\nabla\phi(x)|^2 dx - xe^2\int_\Omega \|x\|^{-1}|\phi(x)|^2 dx\,.$$

It is not difficult to show that any function $\phi \in C_0^2(\mathbb{R}^3)$ can be obtained as a limit in $L_2(\mathbb{R}^3)$ of functions $\phi_n \in C_0^2(\Omega)$ in such a way that, as $n \to \infty$,

$$\frac{\hbar^2}{2m}\int |\nabla\phi_n(x)|^2 dx - e^2\int \|x\|^{-1}|\phi_n(x)|^2 dx \to \frac{\hbar^2}{2m}\int |\nabla\phi(x)|^2 dx - e^2\int \|x\|^{-1}|\phi(x)|^2 dx\,.$$

Hence $\langle\phi|\widetilde{\mathcal{H}}|\phi\rangle \geqslant -\frac{me^4}{2\hbar^2}$ for all normalized $\phi \in C_0^2(\mathbb{R}^3)$, and it follows that $E_0(\widetilde{\mathcal{H}}) \geqslant -\frac{me^4}{2\hbar^2}$, which concludes the proof that hydrogen is stable.

To establish $E_0(\widetilde{\mathcal{H}}) \leqslant -\frac{me^4}{2\hbar^2}$ one constructs in a similar way a sequence of normalized functions $\psi_n \in C_0^2(\Omega)$ converging to $\psi/\|\psi\|$ in $L_2(\mathbb{R}^3)$ such that

$$\frac{\hbar^2}{2m}\int |\nabla\psi_n(x)|^2 dx - e^2\int \|x\|^{-1}|\psi_n(x)|^2 dx \to -\frac{me^4}{2\hbar^2}$$

as $n \to \infty$. The details of the construction of ϕ_n and ψ_n are left to the reader (see Exercise 6.26). □

One may, in fact, completely determine the spectrum of the hydrogen Hamiltonian to be

$$\left\{-\frac{me^4}{2\hbar^2 n^2}\middle| n = 1, 2, \ldots\right\} \cup [0, \infty) .$$

The details of the proof, however, go beyond the scope of these notes.

Exercises

Exercise 6.1. Show that if a map $U : (-a, a) \to \mathcal{B}(H)$ for $a > 0$ maps into the unitary operators on a Hilbert space H and satisifes (6.2) for $s, t, s + t \in (-a, a)$ then U can be extended to a map on all of $\mathbb{R}$ satisfying (6.2) for all $s, t \in \mathbb{R}$.

Exercise 6.2. Assume that $U(t)$ is a strongly continuous unitary one-parameter group on a Hilbert space H and that $U(t)\psi = e^{i\alpha(t)}\psi$ for some $\psi \in H$ and $\alpha : \mathbb{R} \to \mathbb{R}$, with $\alpha(0) = 0$.

a) Show that if α is continuous then $\alpha(t+s) = \alpha(t) + \alpha(s)$ for all $t, s \in \mathbb{R}$ and conclude that for all $t \in \mathbb{R}$ we must have $\alpha(t) = \alpha(1)t$.

b) Show that α need not be continuous, but that one can always find one α with the above property which is continuous.

Exercise 6.3. Let A be a (possibly unbounded) self-adjoint operator on H_1 and let $U \in \mathcal{B}(H_1, H_2)$ be unitary, where H_1 and H_2 are Hilbert spaces.

Show that the operator UAU^* with domain of definition

$$\mathcal{D}(UAU^*) = \{y \in H_2 \mid U^*y \in \mathcal{D}(A)\}$$

is a selfadjoint operator on H_2. Show also that A and UAU^* have the same eigenvalues and that

$$E_\lambda(UAU^*) = UE_\lambda(A)$$

for all λ.

Exercise 6.4. For the operator in Exercise 5.17 show that

$$\operatorname{spec}(A) = \overline{\{\lambda_1, \lambda_2, \ldots\}}.$$

is the spectrum of A, i. e., that $\lambda \notin \operatorname{spec}(A)$ if and only if $(A - \lambda)$ has a bounded inverse. That means that there exists an operator $B \in \mathcal{B}(H)$ such that $(A-\lambda)Bx = x$ for all $x \in H$ and $B(A - \lambda)x = x$ for all $x \in \mathcal{D}(A)$.
Hint. Argue that B must be the operator $Bx = \sum_{i=1}^{\infty}(\lambda_i - \lambda)^{-1}(x, e_i)e_i$ and that this operator is bounded if and only if $\lambda \notin \operatorname{spec}(A)$.

Exercise 6.5. Let $H = L_2(\mathbb{R}^k)$ and $f \in C(\mathbb{R}^k)$ be a continuous function.

a) Show that

$$(M_f\psi)(x) = f(x)\psi(x), \quad \psi \in L_2(\mathbb{R}^k)$$

defines a possibly unbounded densely defined operator with domain

$$\mathcal{D}(M_f) = \left\{\psi \in L_2(\mathbb{R}^k) \middle| \int_{\mathbb{R}^k} |f(x)\psi(x)|^2 dx < \infty\right\}.$$

b) Show that the adjoint of the operator M_f in (a) is $M_f^* = M_{\overline{f}}$ and in particular that it has the same domain of definition as M_f.

Hint. That $\mathcal{D}(M_f) \subseteq \mathcal{D}(M_f^*)$ and that M_f^* coincides with $M_{\overline{f}}$ on $\mathcal{D}(M_f)$ follows by using the Cauchy-Schwarz inequality for the integral $\int_{\mathbb{R}^k} \overline{f(x)\psi(x)}\phi(x)dx$. To obtain $\mathcal{D}(M_f^*) \subseteq \mathcal{D}(M_f)$ one can assume $\psi \in \mathcal{D}(M_f^*)$ and then insert for ϕ in the integral just mentioned the function that equals $f\psi$ on the box $[-R,R]^k$ and vanishes elsewhere, and which belongs to $\mathcal{D}(M_f)$ (why?). One can then deduce that $\int_{[-R,R]^k} |f(x)\psi(x)|^2 dx$ is bounded by a constant for all $R > 0$ and consequently that $\psi \in \mathcal{D}(M_f)$.

c) Deduce that M_f is self-adjoint if and only if f is real.

d) Show that the set

$$\operatorname{spec}(M_f) = \overline{f(\mathbb{R}^k)}$$

is the spectrum of M_f, i. e., that $\lambda \notin \operatorname{spec}(M_f)$ if and only if $(M_f - \lambda)$ has a bounded inverse. That means that there exists an operator $B \in \mathcal{B}(L_2(\mathbb{R}^k))$ such that $(M_f - \lambda)Bx = x$ for all $x \in L_2(\mathbb{R}^k)$ and $B(M_f - \lambda)x = x$ for all $x \in \mathcal{D}(M_f)$.

Hint. Argue that B must be the operator $B = M_{(f-\lambda)^{-1}}$ and that this operator is bounded if and only if $\lambda \notin \operatorname{spec}(M_f)$.

Exercise 6.6. Show that the map (6.8) is an orthogonal projection and determine the eigenspaces corresponding to the eigenvalues 0 and 1.

Exercise 6.7.

a) Show that if $V : H \to H$ is a unitary or anti-unitary operator on a Hilbert space H and if $U(s)$, $s \in \mathbb{R}$, is a strongly continuous unitary one-parameter group with generator A then $V^*U(s)V$, $s \in \mathbb{R}$, is a strongly continuous unitary one-parameter group with generator $\pm V^*AV$ ($+$ if V is unitary and $-$ if V is anti-unitary).

b) A time reversal transformation for a quantum mechanical system with state space H and unitary time evolution $U(t)$ is a unitary or anti-unitary operator $J : H \to H$ such that $J^*U(t)J = U(-t)$, $t \in \mathbb{R}$.

Show for such a time-reversal operator J that $J^*\mathcal{H}J = \pm\mathcal{H}$.

Show also that if one insists on the energy of free particles being non-negative after time reversal then $J^*\mathcal{H}J = \mathcal{H}$. Show that this requirement implies that J must be anti-unitary.

Exercise 6.8. Show by a change of variables that the identity (6.12) holds for all $\hbar > 0$, if it holds for $\hbar = 1$.

Exercise 6.9.

a) Prove the identity (6.13) for $\hbar = k = 1$ by using partial integration (see e. g. the proof of equation (6.48)).

b) Show the identity (6.14) for $\hbar = k = 1$ by writing

$$\left| \frac{\hat{f}(p) - \hat{f}(p_0)}{p - p_0} + i\widehat{xf}(p_0) \right| = \left| \int_{-\infty}^{\infty} \left(\frac{e^{-ixp} - e^{-ixp_0}}{p - p_0} + ixe^{-ixp_0} \right) f(x)dx \right|$$
$$\leqslant \int_{-\infty}^{\infty} \left| \frac{e^{-ix(p-p_0)} - 1}{p - p_0} + ix \right| |f(x)|dx$$

and split the integral into a contribution I_1 from $[-R, R]$ and a contribution I_2 from $\mathbb{R} \setminus [-R, R]$. Then estimate I_1 by using the inequality

$$\left| \frac{e^{-ix(p-p_0)} - 1}{p - p_0} + ix \right| \leqslant |p - p_0| R e^R \,, \quad |x| \leqslant R \,,\ |p - p_0| \leqslant 1 \,,$$

which can be obtained by Taylor expanding the exponential function, and estimate I_2 by using

$$\left| \frac{e^{-ix(p-p_0)} - 1}{p - p_0} + ix \right| \leqslant 3|x| \,, \quad x \in \mathbb{R} \,,$$

which can be obtained from the mean value theorem applied to cos and sin. By choosing $R > 0$ large and p close enough to p_0 one then shows that the first expression above can be made arbitrarily small in a similar way as in the estimates used for proving 1) in the proof of Theorem 6.11.

c) Use a) and b) to conclude that $\mathcal{F}$ maps $\mathcal{S}(\mathbb{R})$ into itself.

Exercise 6.10. Let $f \in \mathcal{S}(\mathbb{R})$ and show that the integral

$$\int_{-\infty}^{\infty} e^{-ipx} f(x)dx$$

can be approximated by appropriately chosen Riemann sums

$$F_N(p) = \sum_{j=1}^{N} e^{-ipx_j} f(x_j)(x_{j+1} - x_j)$$

in such a way that

$$\lim_{N \to \infty} \sup_{p \in \mathbb{R}} \left| F_N(p) - \int_{-\infty}^{\infty} e^{-ipx} f(x)dx \right| = 0.$$

Exercise 6.11. Show that the map $\rho : O(3) \to \mathcal{B}(L_2(\mathbb{R}^3))$ given by

$$(\rho(A)\psi)(x) = \psi(A^{-1}x)$$

is a unitary representation of the orthogonal group.
Hint. Use the transformation theorem for integrals and the fact, that $\det A = \pm 1$.

Exercise 6.12.

a) Show that the set S_n of all bijective maps of $\{1, 2, \ldots, n\}$ forms a group with multiplication being composition of maps and inverse being the inverse map. The group S_n is called the *permutation group* and its elements are called permutations.

b) Find the number of elements in S_n.

c) For $\sigma \in S_n$ define the matrix A^σ by

$$A^\sigma_{ij} = \begin{cases} 1, & \text{if } i = \sigma(j) \\ 0, & \text{otherwise} \end{cases}$$

The number $(-1)^\sigma = \det A^\sigma$ is called the sign of the permutation σ. Show that $(-1)^\sigma \in \{-1, 1\}$ and that $(-1)^{\sigma\tau} = (-1)^\sigma (-1)^\tau$ for all $\sigma, \tau \in S_n$.

d) Show that the maps $S_n \ni \sigma \mapsto U^\pm_\sigma \in \mathcal{B}(\mathbb{C}^n))$ given by

$$U^\pm_\sigma(x_1, \ldots, x_n) = (\pm 1)^\sigma (x_{\sigma^{-1}(1)}, \ldots, x_{\sigma^{-1}(n)})$$

are unitary representations of the group S_n. We here use the notation that $(+1)^\sigma = 1$.

Exercise 6.13. Show that if ρ is a unitary representation of a group G and $e \in G$ is the neutral element satisfying that $eg = ge = g$ for all $g \in G$ then $\rho(e) = 1$.

Exercise 6.14. Let the operators $\mathcal{T}_a, \mathcal{R}_A$ and $\mathcal{B}_v$ be defined by (6.17), (6.22) and (6.23) and let $\mathcal{U}(a, s, v, A)$ be given by (6.24).

a) Show that $\mathcal{T}_{\underline{0}} = \mathcal{B}_{\underline{0}} = \mathcal{R}_0 = 1$ and that

$$U(-s) = \mathcal{U}(0, s, 0, I),\ \mathcal{B}_v = \mathcal{U}(0, 0, v, I),\ \mathcal{T}_a = \mathcal{U}(a, 0, 0, I),\ \mathcal{R}_A = \mathcal{U}(0, 0, 0, A).$$

b) What should the function ω in (6.25) satisfy in order for (6.24) to be consistent with (6.25)?

Exercise 6.15. Show that the unitary operators in (6.26) satisfy

$$\mathcal{U}(a_1, 0, v_1, A_1)\mathcal{U}(a_2, 0, v_2, A_2) = e^{-im\hbar^{-1}A_1 v_2 \cdot a_1}\mathcal{U}(a_1 + A_1 a_2, 0, v_1 + A_1 v_2, A_1 A_2)$$

and thus defines a projective representation of the part of the Galilei group that excludes time translations.

Exercise 6.16. Complete the calculation of ω at the end of the proof of Theorem 6.10.

Exercise 6.17. Show the formula (6.36) for all functions $\psi \in \mathcal{S}(\mathbb{R}^k)$.

Exercise 6.18. Show that a free particle in $\mathbb{R}^k$ is stable with ground state energy zero, but that there is no ground state.

Exercise 6.19. Consider a free particle of mass m moving on the circle S^1. Determine the time evolution if the initial normalized state is represented by $\psi_0 = 2^{-1/2}e_1 + 2^{-1/2}e_{-2}$.

Exercise 6.20.

a) Verify that Ω_0 belongs to $\mathcal{S}(\mathbb{R})$, by using appropriate asymptotic properties of exp.

b) Show that $Q\Omega_0$ belongs to $\mathcal{S}(\mathbb{R})$ for any polymomial Q.

Exercise 6.21.

a) Show by using Lemma 6.8 that

$$\mathcal{F} \circ A^\dagger = -i\, A^\dagger \circ \mathcal{F}$$

and deduce that

$$c_n = (-i)^n c_0\,,$$

where c_n is defined as in Remark 6.12.

b) Use that $\mathcal{F}$ is isometric to argue that $|c_n| = 1$ and then use the form of Ω_0 to conclude that

$$c_0 = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-x^2/2} dx = 1,$$

and hence that $c_n = (-i)^n\,,\ n = 0, 1, 2, \ldots$.

Exercise 6.22.

a) Show that Ω_n is an even function if n is even and an odd function if n is odd.

b) The Hermite polynomial H_n is defined as follows:

Let

$$\Omega_n(x) = P_n(x)e^{x^2/2}$$

be the n'th (normalized) eigenfunction for the harmonic oscillator with $c = 1$. Then

$$H_n(x) = b_n P_n(x) = 2^n x^n + \ldots\,,$$

that is the normalization factor b_n is chosen such that the highest order term in $H_n(x)$ has coefficient 2^n.

Show that

$$\begin{aligned} H_0(x) &= 1 \\ H_1(x) &= 2x \\ H_2(x) &= 4x^2 - 2 \\ H_3(x) &= 8x^3 - 12x\,. \end{aligned}$$

Exercise 6.23.

a) Use polar coordinates to show that if $\psi \in C_0^2(\mathbb{R}^3)$ then $\|x\|^{-1}\psi(x) \in L_2(\mathbb{R}^3)$.

b) Show that the hydrogen operator (6.53) is well defined as an operator $\widetilde{\mathcal{H}} : C_0^2(\mathbb{R}^3) \to L_2(\mathbb{R}^3)$.

Exercise 6.24.

a) Show that the inequality

$$\langle\psi|\hbar^2 D^2|\psi\rangle + mk\langle\psi|x^2|\psi\rangle \geqslant \hbar\sqrt{mk}$$

holds for all normalized wave functions $\psi \in \mathcal{S}(\mathbb{R})$ and all constants $m, k \geqslant 0$.

Hint. Think of $\sqrt{mk}$ as a variable quantity.

b) Use a) to derive the Heisenberg uncertainty relation

$$\langle\psi|\hbar^2 D^2|\psi\rangle\langle\psi|x^2|\psi\rangle \geqslant \frac{1}{4}\hbar^2$$

for any normalized wave function $\psi \in \mathcal{S}(\mathbb{R})$.

Exercise 6.25. If $\psi \in \mathcal{S}(\mathbb{R}^3)$ the Heisenberg uncertainty relation (see Exercise 6.24) states that

$$\left(\int_{\mathbb{R}^3} \|x\|^2|\psi(x)|^2 dx\right)\left(\int_{\mathbb{R}^3} \overline{\psi(x)}(-\Delta\psi(x))dx\right) \geqslant \frac{1}{4}.$$

a) Let $\widetilde{\mathcal{H}}$ be the hydrogen operator (6.53). Use the uncertainty relation to show the following inequality holds all $\psi \in \mathcal{S}(\mathbb{R}^3)$:

$$\langle\psi|\widetilde{\mathcal{H}}|\psi\rangle \geqslant \frac{\hbar^2}{8m}\left(\int_{\mathbb{R}^3} \|x\|^2|\psi(x)|^2 dx\right)^{-1} - e^2\int_{\mathbb{R}^3} \|x\|^{-1}|\psi(x)|^2 dx. \tag{6.54}$$

b) Given any $C > 0$ find a normalized function $\psi \in \mathcal{S}(\mathbb{R}^3)$ such that the right hand side of (6.54) is smaller than $-C$, thus demonstrating that the inequality (6.54) cannot be used to prove stability of hydrogen.

Exercise 6.26. Complete the details of the proof of Corollary 6.16.

Exercise 6.27. Let $U(t)$ be a strongly continuous unitary one-parameter group on a Hilbert space H and let V be a unitary operator on H. Assume that

$$U(t)V = \exp(i\theta(t))VU(t) \tag{6.55}$$

for all $t \in \mathbb{R}$ and some function $\theta : \mathbb{R} \to \mathbb{R}$.

a) Show that there must be an $e_0 \in \mathbb{R}$ such that (6.55) is satisfied with $\theta(t) = e_0 t$.

Hint. Use the approach of Exercise 6.2.

b) Show that the generator A of $U(t)$ satisfies $V^*AV = A - e_0$.

c) Show that if $U(t)$ is the time evolution of a stable quantum mechanical system (see Definition 6.13) then $e_0 = 0$. Show in particular, that in this case time translations commute with space translations.

d) Give an example of a strongly continuous unitary one-parameter group $U(t)$ and a unitary V satisfying (6.55) with $\theta \neq 0$.

Hint. Think of Galilei transformations.

Exercise 6.28. Let Ω_0 be the ground state of the harmonic oscillator defined in (6.45). Define

$$\Omega_{a,v} = e^{i\hbar^{-1}mv\cdot x}\Omega_0(x-a).$$

(Compare with the Galilei transformation in (6.26).) Argue that $\Omega_{a,v} \in \mathcal{S}(\mathbb{R})$ and show that it is a unit eigenvector of the operator A in (6.40). What is the eigenvalue?

Index

Bibliography

[1] R. Abraham and J. Marsden, *Foundations of Mechanics*. Americam Mathematical Society, 1978 (second edition 2008).

[2] V. I. Arnol'd, *Mathematical Methods of Classical Mechanics*, second edition. Springer Verlag, New York, 1997.

[3] C. Berg, *Metriske rum*. Lecture notes, Department of Mathematical Sciences, University of Copenhagen 1997.

[4] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*. McGraw-Hill Book Company, New York, 1955.

[5] P. A. M. Dirac, *The Principles of Quantum Mechanics*. Oxford University Press, 1930 (fourth edition 1982).

[6] B. Durhuus, *Hilbert rum med anvendelser*. Lecture notes, Department of Mathematical Sciences, University of Copenhagen 1997.

[7] G. Galilei, *Dialogo sopra i due massimi sistemi del mondo*. First edition 1632.

[8] J. Glimm and A. Jaffe, *Quantum Physics: a functional integral point of view*. Springer Verlag, New York, 1981 (second edition 1987.

[9] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*. Springer Verlag, Berlin, 1932.

[10] I. Newton, *Philosophiae Naturalis Principia Mathematica*. First edition 1687.

[11] W. Rudin, *Real and Complex analysis*. McGraw-Hill Book Company, New York, 1970.

[12] W. Rudin, *Functional Analysis*. McGraw-Hill Book Company, New York, 1991 (first edition 1973).

[13] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I, II*. Academic Press, New York, 1975 and 1980.