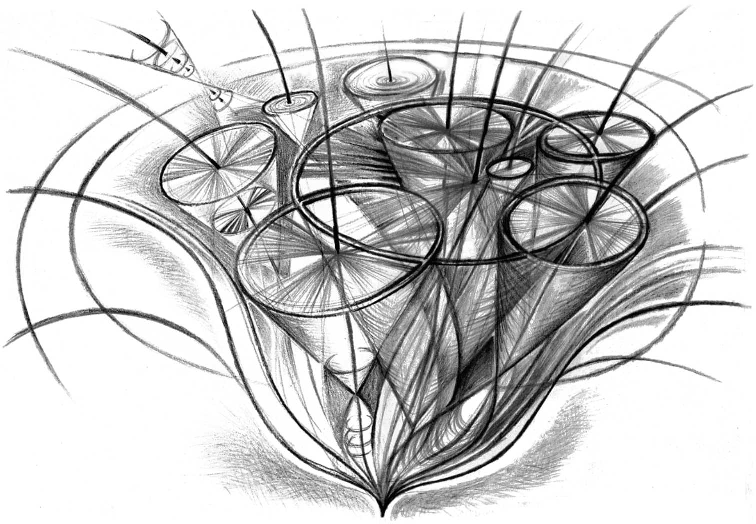
SELECTED TOPICS IN THEORETICAL PHYSICS I



Petr Kulhánek

Theoretical Mechanics
Quantum Theory
Mathematics for Physics

AGA 2026

ISBN: 978-80-906638-2-4

Contents

Foreword

Dear Readers,

This text was developed over many years based on my lectures at the Czech Technical University in Prague, where, after 1989, the restrictions of the former regime were finally lifted, and I was able to begin teaching. At that time, a four-semester course in theoretical physics for master's and doctoral students was established at the Faculty of Electrical Engineering. After several revisions, this course is still in use today and is loosely followed by a three-semester course on plasma theory at the Faculty of Nuclear and Physical Engineering. In 2011, the textbook *Introduction to Plasma Theory* was published, and only then, in 2016, did the book *Selected Topics in Theoretical Physics* appear. Four years later, I decided to publish a complete textbook based on the entire lecture series–which is also available as recorded videos on YouTube (in Czech only)–consisting of three volumes: *Selected Topics in Theoretical Physics I*, *II*, and *III*. You are currently holding the first volume of this trilogy in your hands. In it, you will find the fundamentals of theoretical mechanics, which serves as a launching pad for the study of quantum mechanics, statistical physics, electromagnetism, relativity, plasma theory, and other branches of physics. The text includes classic sections on Lagrange and Hamilton equations, conservation laws, Poisson brackets, and canonical transformations. A section on adiabatic invariants, which is important in plasma theory, is also included. In Chapter 1.4 the topics covered are illustrated using several important problems. For example, motion in a rotating frame is derived, equations of motion are found in the case of energy dissipation, and Lagrange points are calculated in the so-called restricted three-body problem. In Section 1.4.6, the reader will find a method for solving the inverse problem in certain cases, i.e., finding the Lagrangian of the problem using the equations of motion. The chapter concludes with an overview of Lagrangian equations for field problems, ending with the Lagrangian formulation of Maxwell equations.

The second part of this textbook is devoted to quantum theory. The relatively standard sections on the structure and interpretation of quantum theory, angular momentum, spin, time evolution, and simple examples of finding the spectrum of the Hamiltonian operator are supplemented by current topics concerning the boundary between the quantum and classical worlds, the refutation of the existence of hidden parameters, the EPR paradox, and Bell inequalities. The topics are mostly addressed using Dirac notation, which is explained in Chapter 3.4 of this book. The conclusion of the section on the fundamentals of quantum theory concerns the Klein–Gordon and Dirac equations. This constitutes a minimum foundation that will enable the reader to tackle more advanced textbooks on quantum theory.

The final section of the textbook is devoted to the mathematical fundamentals necessary for studying physical phenomena. Readers may consult this chapter whenever they lack the mathematical background required to understand the physics chapters. It introduces Einstein summation convention, covers the basics of complex analysis, linear vector spaces, Hilbert spaces, the Lie algebra is defined, the basics of handling covariant and contravariant tensor components are demonstrated, conics are described, Dirac notation is explained, certain special functions are discussed, and many other useful topics in mathematics for physicists are covered.

In this textbook, variables are always represented in italics. Functions, abbreviations, digits, and various mathematical operations are shown in regular typeface. Vectors, tensors, and composite objects are set in bold typeface. In exceptional cases where ambiguity or confusion might arise, arrows are placed above vectors and tensors. Roman indices denote the position of a quantity, coordinate axes, etc. Greek letters denote the components of four-vectors, for example A_α, where $\alpha = 0$ (time component), 1, 2, 3 (spatial components).

Since the number of letters in the alphabet is limited, some quantities are denoted by the same symbol. However, their meaning can be easily inferred from the context. The list of symbols included at the end of the book can also be helpful. When reading, do not skip the notes; they often contain important insights necessary for understanding the phenomenon being discussed. In the book, you will find them on a gray background. Illustrative examples are separated from the rest of the text at the beginning and end by a black semicircle. Important relationships are marked on the left side with a black triangle. Hopefully, these marks will help the reader navigate this difficult study text more easily.

What can I say in conclusion? I would like to thank the many students who studied the texts on which this book is based and carefully identified any typos and ambiguities. In particular, I would like to thank Ing. et Ing. Petr Endel, Ing. Radek Beňo, RNDr. David Břeň, Ph.D., Ing. Miroslav Horký, Ph.D., David Maňas, Ing. Antonín Krpenský, and many others. A big thank you goes to Daniel Handl, who selflessly filmed the theoretical physics course, and Ing. Jan Sláma, who was instrumental in filming additional parts of the course at the Faculty of Electrical Engineering (FEE) of the Czech Technical University (CTU) in Prague. Last but not least, my thanks go to Ing. arch. Ivan Havlíček, who created the introductory graphics for the individual chapters, the cover, and some of the illustrations. To future students, I would like to wish, above all, sound judgment, without which the study of natural phenomena would be meaningless. In today's fast-paced world, however, there are other necessary factors as well: a quiet place to study, sufficient time, and adequate family and financial support. I hope that the readers of this book will find everything they need to successfully study the textbook they have just opened.

Petr Kulhánek, Prague 2026

Readers can find hypertext links to the complete recordings (in Czech language only) of the lectures accompanying this textbook in Czech on the aldebaran.cz website under the "*Study*" section. There, you can also download the current electronic version of this textbook and other supplementary materials for the lectures.

Introduction

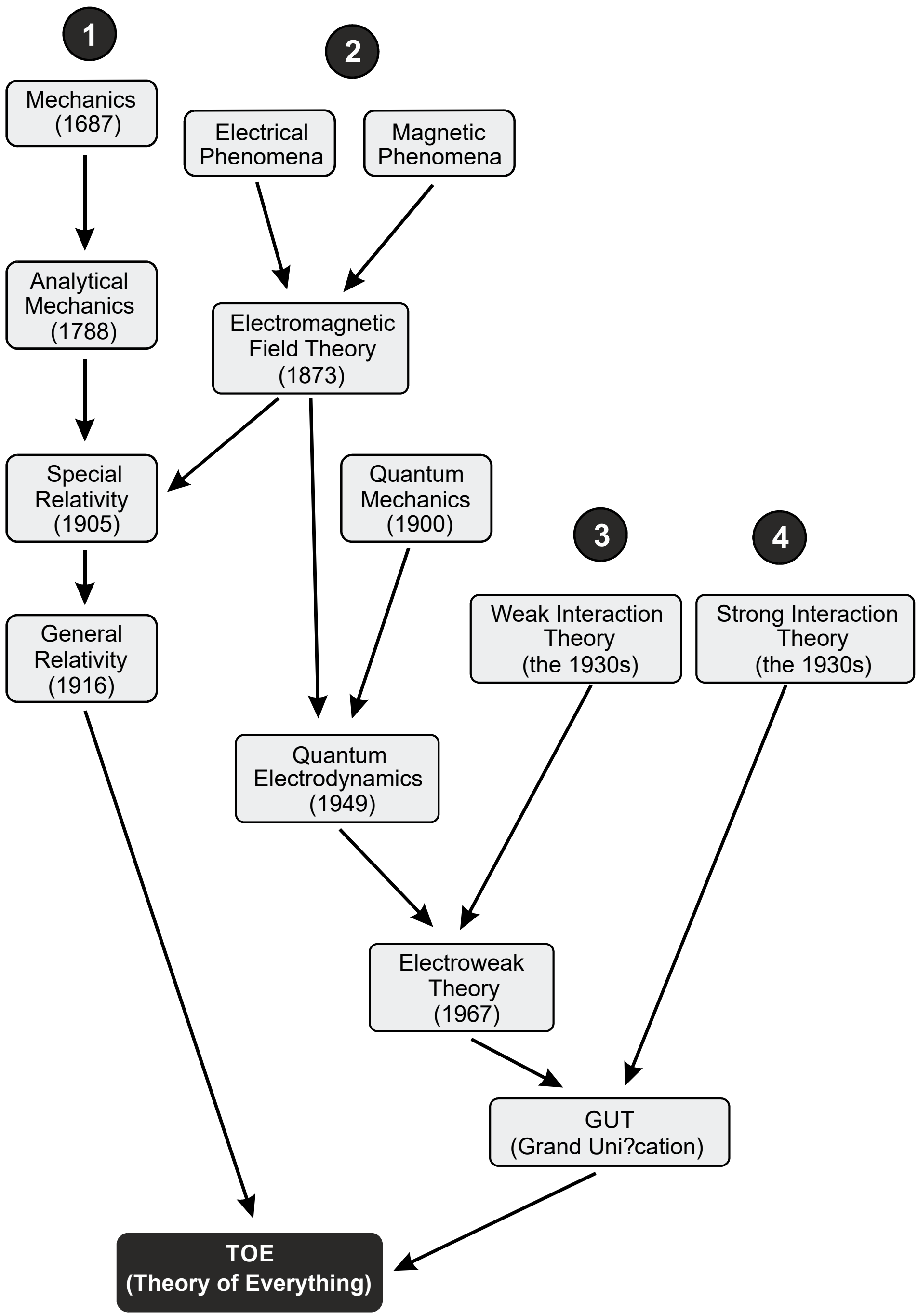
Not so long ago, physicists were divided into two main groups–experimentalists and theorists. Members of each group knew they couldn’t do without the members of the other group. The result was a fruitful collaboration filled with apparent rivalry and amusing anecdotes. With the advent of computer technology, everything changed. Gradually, a third group emerged, one dedicated to numerical simulations. Today, we cannot imagine physics without them. Numerical simulations allow for the initial verification of new theories without costly experiments. When processing experimental data, they help identify the processes hidden behind the measured data. Currently, physics consists of three inseparable components: theory, experiment, and numerical simulations. As its title suggests, this textbook is dedicated to selected chapters from theoretical physics. Its first part serves as an introduction to theoretical mechanics and quantum theory.

Over the centuries, physics has exhibited two fundamental trends. The first is a gradual division into ever-narrowing subfields. This development is linked to deepening knowledge and is a natural path in any scientific discipline. Specialists in increasingly narrow fields are gradually emerging; they develop their own scientific language, and the ability of experts from previously related areas of physics to communicate with one another is steadily deteriorating. On the other hand, there is a deeper understanding of the connections between the various branches of physics and their gradual unification into more universal theories. Perhaps one day we will succeed in unifying the physical view of all fundamental natural interactions into a single theory, which we now call the *Theory of Everything* (TOE).

Mechanics, as a scientific discipline within physics, emerged in the 17th century. The first well-known scientific experiments were conducted by Galileo Galilei (1564–1642). The theoretical framework of classical mechanics, as a tool for predicting the motion of bodies in a given force field, was proposed by Isaac Newton (1642–1727) in his *Principia* of 1687. In the 18th century, Joseph Louis Lagrange (1736–1813) completed the development of classical mechanics, formulating mechanical problems independently of the choice of coordinate system using the calculus of variations.

In the 19th century, scientists successfully began to gradually understand electrical and magnetic phenomena. A number of prominent physicists contributed to these experiments, including Hans Oersted (1777–1881), André Ampère (1775–1836), Michael Faraday (1791–1867), Heinrich Hertz (1857–1894), Oliver Heaviside (1850–1925), and others. This entire period culminated in the realization that electrical and magnetic phenomena share the same nature and a common origin. In 1873, James Clerk Maxwell (1831–1879) published *A Treatise on Electricity and Magnetism*, which contained equations that unified classical electrodynamics into a single framework.

At the end of the 19th century, many physicists succumbed to the illusion that physics as a science was complete. The laws of mechanics were known on the one hand, and the laws of electricity and magnetism on the other. It seemed that all natural phenomena were the result of these two scientific disciplines, and that the future lay solely in applying known laws to unknown situations. This was, of course, a grave mistake, which quickly became apparent at the beginning of the 20th century, when it was impossible to explain new physical phenomena using the knowledge available at the time.



Integration trends in physics

It turned out that neither classical mechanics nor classical electrodynamics could satisfactorily describe the world at the atomic level. The consequence of this was an inability to clarify the behavior of electrons in the atomic shell, explain blackbody radiation, understand the photoelectric effect, and reconcile the manifestations of objects in the microworld, which sometimes exhibited particle-like and at other times wave-like properties. Quantum mechanics was born, in which $ab \neq ba$, and non-commutativity became a newly discovered feature of nature at the microscopic level. Quantum mechanics brought with it a whole range of hard-to-imagine phenomena – the quantization of energy and angular momentum, wave-particle duality, the uncertainty principle, the ambiguity of the act of measurement, and the probabilistic interpretation of results leading to the indeterminism of quantum physics.

And that was just the beginning. The discovery of the spin of elementary particles in 1925 marked another significant leap forward in humanity's understanding of nature. It is a consequence of relativistic physics, which developed in parallel with quantum mechanics from the early 20th century onward. The combination of quantum mechanics and special relativity led to Dirac equation, which became the basis for the quantum description of the electron's motion. Paul Adrien Maurice Dirac (1902–1984) proposed his equation in 1928 and, in the same year, used it to deduce the existence of the positron, the antiparticle of the electron. The positron was not experimentally discovered until four years later by Carl Andersson (1905–1991). For his work, Dirac was awarded the Nobel Prize in Physics for 1933. Between 1946 and 1949, the first quantum field theory was completed – the quantum theory of the electromagnetic field, which we now call *quantum electrodynamics* (QED). For its formulation, Richard Feynman (1918–1988), Shin-Itiro Tomonaga (1906–1979), and Julian Schwinger (1918–1994) were awarded the 1965 Nobel Prize in Physics. Quantum electrodynamics is the quantum analogue of Maxwell equations. Electromagnetic interaction is caused by field particles – in this case, photons – which exchange charged particles with one another. The classical concept of force loses its meaning. Feynman succeeded in interpreting the complex equations using illustrative graphs, which we now call Feynman diagrams. On a similar basis, the current quantum theory of weak and strong interactions was later developed. A fundamental feature of these theories is the so-called gauge symmetries, which determine how a given interaction acts on elementary particles.

Since the early 1960s, efforts had been underway to unify the electromagnetic and weak interactions into a single theory. Steven Weinberg (1933), Abdus Salam (1926–1996), and Sheldon Glashow (1932) succeeded in doing so. For their work, they were awarded the 1979 Nobel Prize in Physics. The weak interaction field particles they predicted – W^+, W^-, and Z^0 – were discovered at the turn of 1983 and 1984 at the European Organization for Nuclear Research (CERN). Their discoverers, Carlo Rubbia (1934) and Simon van der Meer (1925–2011), were awarded the Nobel Prize that same year (1984).

As early as the 1930s, Japanese physicist Hideki Yukawa (1907–1981) contributed to our understanding of the strong interaction. He was awarded the 1949 Nobel Prize in Physics for his work. The current quantum field theory of the strong interaction is called *quantum chromodynamics* (QCD), and Frank Wilczek (1951), David Gross (1941), and David Politzer (1949) were awarded the 2004 Nobel Prize in Physics for formulation of the theory and, in particular, for the discovery of the asymptotic freedom of the strong interaction between quarks and gluons.

Quantum mechanics achieved extraordinary successes throughout the 20th century. A simple theory describing mechanical processes gradually evolved into a field quantum theory capable of successfully describing three of the four fundamental interactions of nature. This journey was, of course, not without its difficulties and problems, but it culminated in today's *Standard Model* of elementary particles and interactions. Without quantum theory and a deep understanding of the laws of the microworld, we would have neither computers nor any other electronics today.

At the beginning of the 20th century, however, another, no less successful theory emerged – general relativity. Maxwell electrodynamics implied that the speed of light in a vacuum should be a universal constant and that it should not be added to the speed of the source of electromagnetic waves. At first glance, this result contradicted classical mechanics, in which the source's velocity is added to the signal's velocity. A series of experiments confirmed the correctness of electrodynamics. It was therefore necessary to reformulate mechanics so that it would be consistent with Maxwell electrodynamics. Albert Einstein succeeded in doing this in 1905 within the framework of the so-called *special theory of relativity*. The price paid for unifying the two theories was high. Time, along with space, ceased to be absolute. The length of a moving rod and the time interval between two events actually depend on the observer's choice of coordinate system.

Einstein's efforts to generalize special relativity to non-inertial coordinate systems led, in 1915, to the development of *general relativity* – a completely new theory of gravity that describes this interaction through the curvature of space-time. The new theory is based on two key ideas:

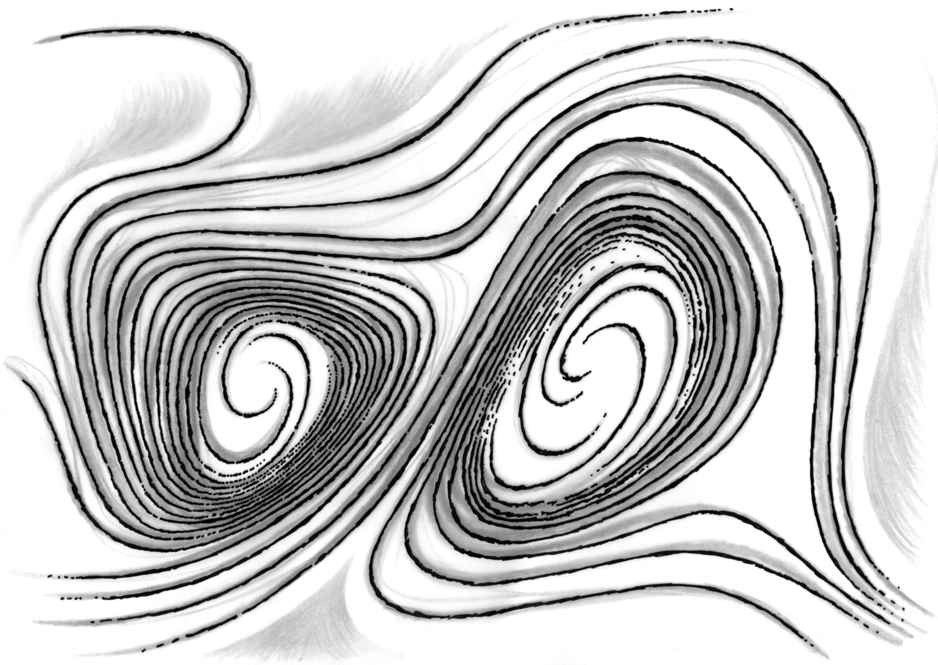
- Every object curves the spacetime around it simply by its presence;
- Every object moves through this curved spacetime along the straightest possible paths – known as geodesics.

This new understanding of time and space was truly revolutionary. Physical objects themselves play a role in the creation of time and space; without them, time and space would not exist. The question of what the universe would look like without physical objects ceases to make sense.

In a certain sense, twentieth-century physics has thus become somewhat schizophrenic. Three of the four fundamental interactions are described using exchange (field) particles within the framework of quantum field theory. And one interaction, gravity, is described using the curved world of general relativity. Solving many physical puzzles has brought even greater mysteries. Is there a unified theory of all four interactions? Is it possible to combine quantum theory and general relativity into a single theory? We do not yet know the answers to these questions. Various string theories, in which particles are understood as one-dimensional oscillating entities in a multidimensional world, have achieved great success, but whether this is a step in the right direction or not is not clear at this time. In 2010, Dutch physicist Erik Verlinde proposed a hypothesis suggesting that gravity might not be a real force, but merely a statistical manifestation of the increase in entropy in the microcosm. It is difficult to predict whether this ambitious idea will find support in future experiments or whether it is a dead end.

If you are interested in the fundamental properties of nature and their theoretical description, you should first and foremost begin by studying classical mechanics, which is closely linked to quantum mechanics. Further study of field problems, in turn, is not possible without a knowledge of statistical physics. Therefore, this textbook could serve as a springboard for you to understand the problems of modern physics and to further explore the strange laws of the world around us.

1. Theoretical Mechanics



1.1 Integral Principles in Mechanics

Einstein summation convention, the differential, and Lagrange increment theorem are widely used in theoretical mechanics. If the reader is not familiar with these mathematical fundamentals, he should first carefully read Chapter 3.1, where these concepts are explained. The reader can find further information on the study of theoretical mechanics in textbooks [3]–[7].

1.1.1 Basic Concepts in Mechanics

Mechanical system

We refer to a mechanical system as any collection of particles or bodies that we choose to describe (an electron, an atom, the Earth, the solar system, etc.).

Cartesian coordinates

Cartesian coordinates are based on three perpendicular and straight axes. For coordinates, we use the notation $\mathbf{x} \equiv \mathbf{r} \equiv (x_1, x_2, x_3) \equiv (x, y, z)$, and similarly $\mathbf{F} \equiv (F_1, F_2, F_3) \equiv (F_x, F_y, F_z)$. The equation of motion for a point mass is given by $m\, \mathrm{d}^2\mathbf{x}/\mathrm{d}t^2 = \mathbf{F}$.

Generalized coordinates

We consider generalized coordinates to be any parameters describing motion (angles, distances, areas). We denote them by $\mathbf{q} = (q_1, q_2, \ldots)$.

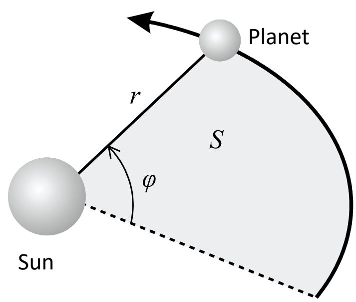


Fig. 1.1: Coordinates for a planet orbiting the Sun

◖ **Example 1.1: Motion of a planet around the Sun**

$q_1 = r(t)$ – distance from the Sun
$q_2 = \varphi(t)$ – angle between the radius vector and a given reference line
$q_3 = S(t)$ – area swept out by the radius vector ◗

Generalized velocities

We define the generalized velocity as the time derivative of the generalized coordinate.

◖ **Example 1.2**

$v_r = \mathrm{d}r/\mathrm{d}t$	Radial velocity
$v_\varphi = \mathrm{d}\varphi/\mathrm{d}t$	Angular velocity
$v_S = \mathrm{d}S/\mathrm{d}t$	Areal velocity
$v_x = \mathrm{d}x/\mathrm{d}t$	Velocity x-component

◗

Constraints

A body or some of its parts may not be able to move completely freely. In that case, we say that there are constraints in the system. Examples are shown in the following figure:

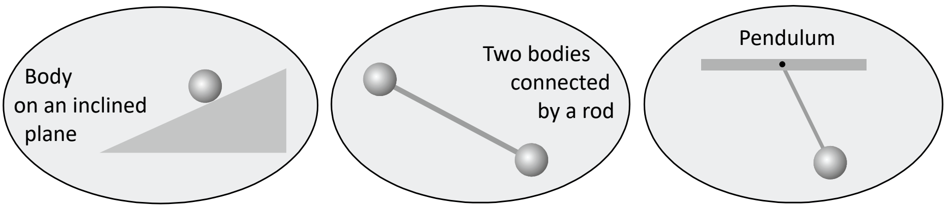


Fig. 1.2: Constraints in the system

Degrees of freedom

By degrees of freedom, we mean the number of independent variables (parameters) that can fully describe the motion of a system (denoted by f).

◖ **Example 1.3**

Free point mass	$f = 3$,
N free point masses	$f = 3N$,
Point mass on an inclined plane	$f = 2$,
Two point masses connected by a rod	$f = 5$,
Spatial pendulum	$f = 2$,
Plane pendulum	$f = 1$.

◗

For a system of N point masses with R constraints, $f = 3N - R$ holds. We always choose the generalized coordinates as a set of independent parameters that fully describe the system, i.e., there are exactly f of them:

$$\mathbf{q} \equiv (q_1 , q_2 , \ldots q_f) .$$

Configuration space

The configuration space is an f-dimensional space in which we represent the values of generalized coordinates. A point in the configuration space is called a *configuration*. The time evolution of a system's configuration $\mathbf{q}(t)$ is called a *trajectory*.

State of the system

In classical mechanics, at a given time t_0 , the state of the system under consideration is completely determined by the configuration $\mathbf{q} \equiv (q_1, q_2 , ... q_f)$ and the tendency (generalized velocities) $\mathbf{v} \equiv (v_1 , v_2 , ..., v_f)$.

Real and virtual trajectories:

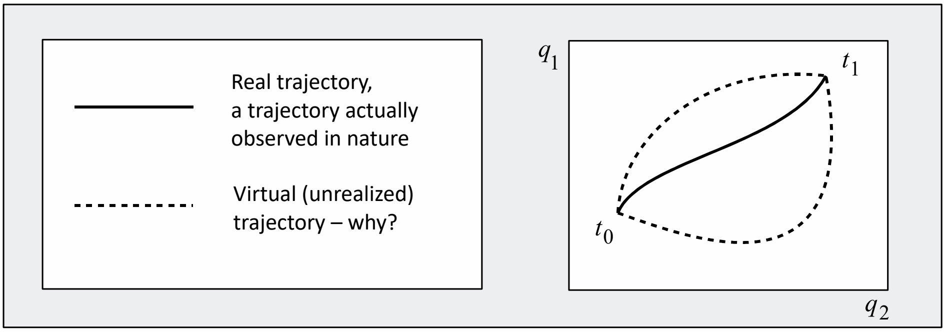


Fig. 1.3: Real and virtual trajectories

1.1.2 Integral Principles

◖ **Example 1.4.** Let's imagine that a person is drowning in a pond. Between the rescuer and the pond lies a swampy strip where it is very difficult to move, as well as a strip of plowed land and a field. The rescuer must choose the optimal route to reach the drowning person as quickly as possible (this route may not be the shortest line between the drowning person and the rescuer):

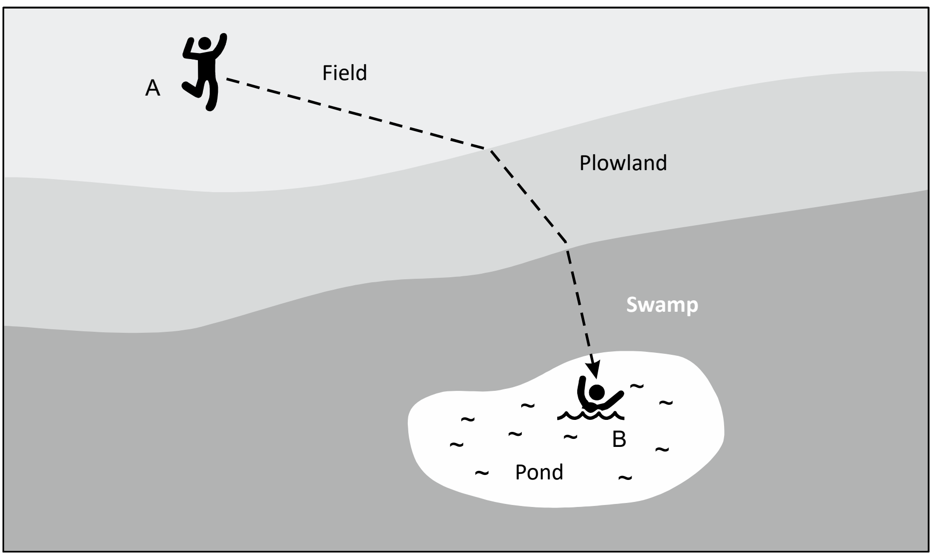


Fig. 1.4: What is the fastest way to reach a drowning person?

We will determine the total time the rescuer will be moving as follows:

$$v = \frac{\mathrm{d}l}{\mathrm{d}t} \quad \Rightarrow \quad \mathrm{d}t = \frac{\mathrm{d}l}{v} \quad \Rightarrow$$

$$T = \int_{t_A}^{t_B} \frac{\mathrm{d}l}{v} = \int_{t_A}^{t_B} \frac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{v(x,y)} = \int_{x_A}^{x_B} \frac{\sqrt{1+y'^2}}{v(x,y)} \mathrm{d}x \,.$$

We assume that we know the spatial dependence of the velocity $v(x, y)$. This is determined by the type of terrain (field, plowed field, swamp). We are now looking for a curve $y(x)$ such that the previous integral takes on a minimum value. Variational calculus deals with solving problems of this type. ◗

◖ **Example 1.5: Brachistochrone.** Let's solve the following problem. A body is to slide down an inclined plane of general shape between two points A and B, which are at different heights. The task is to find the equation of the inclined plane such that the body reaches point B in the shortest possible time. The name of the curve comes from Greek (βραχιστος = shortest, χρονος = time).

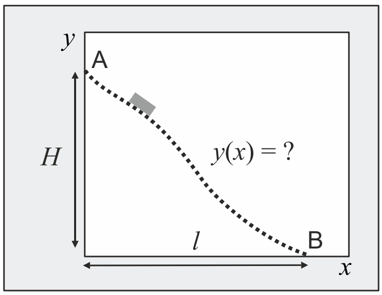


Fig. 1.5: Brachistochrone

The calculation is similar to the previous one:

$$v = \frac{\mathrm{d}l}{\mathrm{d}t} \quad \Rightarrow \quad \mathrm{d}t = \frac{\mathrm{d}l}{v} \quad \Rightarrow$$

$$T = \int_{t_A}^{t_B} \frac{\mathrm{d}l}{v(y)} = \int_{t_A}^{t_B} \frac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{v(y)} = \int_{x_A}^{x_B} \frac{\sqrt{1+y'^2}}{v(y)} \mathrm{d}x \,.$$

We determine the speed using the law of conservation of energy

$$mgy + \frac{1}{2}mv^2 = mgH \,.$$

The resulting travel time is

$$T = \int_{x_A}^{x_B} \sqrt{\frac{1+y'^2}{2g(H-y)}} \, \mathrm{d}x \,. \tag{1.1}$$

We must now find the curve $y(x)$ for which the integral (1.1) attains its minimum – this is again a typical problem in the calculus of variations. You will find the solution at the end of Section 1.2.3. The fundamental laws of mechanics, the theory of the electromagnetic field, and other branches of physics can also be formulated using variational principles. In this section, we will examine one of the integral principles of mechanics – the so-called *Hamilton principle*. ◗

1.1.3 Hamilton Principle of Least Action

Both of the introductory examples led to the optimization of an integral of the form

$$T(x_A, x_B) = \int_{x_A}^{x_B} F\left(x, y(x), y'(x)\right) \mathrm{d}x \,. \tag{1.2}$$

The integrand consists of the independent variable x, the target function $y(x)$, and its first derivative $y'(x)$. The result of the optimization should be the target trajectory or curve $y(x)$. In the introductory example, the rescuer chose a trajectory that minimized the total time. All other trajectories (so-called *virtual* – unrealized) are, in principle, possible, but they last longer. The situation is similar in the example with the sliding body. Integrals of the type described above are called *functionals*. A functional is a mapping in which we assign a number to a function (in our case, the total time).

The basic idea of the integral principles in mechanics is similar. Of all the possible trajectories, only the one that is in some way more favorable than the others was realized. The criterion of favorability is considered in a manner similar to the examples, except that time is now the independent variable, as we are looking for the curve $\mathbf{q}(t)$.

Hamilton principle

Let us assume that there exists a function of time t, generalized coordinates, and their first derivatives (i.e., the state)

► $$L(t, q_1, \dots, q_f, \dot{q}_1, \dots, \dot{q}_f\,) \,,$$

such that, of all possible dependencies $q_k(t) = f_k(t)$, the one realized in nature is that for which the integral

► $$S(t_A, t_B) \;\equiv\; \int_{t_A}^{t_B} L(t, q_1, \dots, q_f, \dot{q}_1, \dots, \dot{q}_f)\, \mathrm{d}t \tag{1.3}$$

has an extreme (minimum). We call the function $L(t, \mathbf{q}, \mathrm{d}\mathbf{q}/\mathrm{d}t)$ the *Lagrangian*, and the integral $S(t_A, t_B)$ the *action integral*. Hamilton principle is a fundamental axiom of the theory.

1.1.4 Lagrange Equations

Let's introduce *virtual displacement*

$$\begin{aligned} \delta q_k &= q_{k,\mathrm{virt}}(t) - q_{k,\mathrm{real}}(t) \quad \text{or} \\ \delta \mathbf{q} &= \mathbf{q}_{\mathrm{virt}}(t) - \mathbf{q}_{\mathrm{real}}(t) \end{aligned} \tag{1.4}$$

as the infinitesimal difference between a virtual (imaginary) trajectory and a real (actual) trajectory. Points on both trajectories correspond at the same time (so-called *isochronous variation*). Let us list the basic properties of virtual displacements:

► 1) $\delta\mathbf{q}(t_A) = \delta\mathbf{q}(t_B) = 0\,,$ (1.5)

► 2) $\delta\dot{\mathbf{q}} = \dfrac{\mathrm{d}}{\mathrm{d}t}\delta\mathbf{q}\,.$ (1.6)

The first property states that both the virtual and real trajectories begin and end at the same point in configuration space. The second property states that the operations of differentiation d/dt and variation δ are interchangeable.

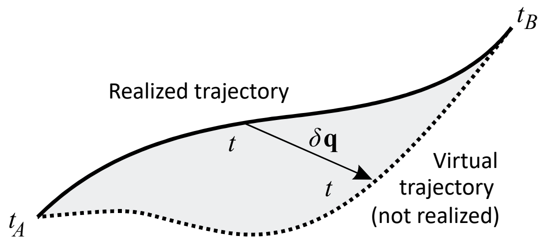


Fig. 1.6: Definition of virtual displacement

Note: Constraints are included in the system via the choice of generalized coordinates – their total number is equal to the number of degrees of freedom. Virtual displacements are displacements consistent with the constraints at a given time.

Let us now derive the necessary conditions for the extreme of the action integral:

$$\delta\int_{t_A}^{t_B} L(t,\mathbf{q},\dot{\mathbf{q}})\,\mathrm{d}t = 0 \quad\Rightarrow\quad \int_{t_A}^{t_B}\delta\, L(t,\mathbf{q},\dot{\mathbf{q}})\,\mathrm{d}t = 0 \quad\Rightarrow\quad \int_{t_A}^{t_B}\left(\frac{\partial L}{\partial q_k}\delta q_k + \frac{\partial L}{\partial \dot{q}_k}\delta\dot{q}_k\right)\mathrm{d}t = 0\,,$$

where, due to isochronism, we have omitted the differentiation with respect to time. We now integrate per partes the second term using (1.6):

$$\int_{t_A}^{t_B}\left(\frac{\partial L}{\partial q_k}\delta q_k - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_k}\right)\delta q_k\right)\mathrm{d}t \;+\; \left[\frac{\partial L}{\partial \dot{q}_k}\delta q_k\right]_{t_A}^{t_B} \;=\; 0\,.$$

The last term is zero according to (1.5), and therefore

$$\int_{t_A}^{t_B}\left(\frac{\partial L}{\partial q_k} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k}\right)\delta q_k\,\mathrm{d}t \;=\; 0\,.$$

This equality must hold for any two times t_A and t_B and for any virtual displacement δq_k. Since the δq_k are independent (the number of generalized coordinates equals the number of degrees of freedom of the system), the term in parentheses in the previous equation must necessarily be zero for every k, i.e.:

► $$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} \;-\; \frac{\partial L}{\partial q_k} \;=\; 0\,; \qquad k = 1,\ldots,f\,.$$ (1.7)

These equations represent the necessary conditions for the extremality of the action integral and are called Lagrange equations. From a mathematical point of view, they are ordinary second-order differential equations for the extremal trajectory $q_k(t)$; $k = 1 \dots f$, which is realized in nature.

Note 1: Lagrange equations are the equations of motion for our system in generalized coordinates. Their form does not depend on the choice of coordinate system. Newton's equations must be a special case in the Cartesian coordinate system.

Note 2: The equation must be supplemented with initial conditions at time $t_0 = t_A$

► $$q_k(t_0) = q_{k0} \,, \qquad \dot{q}_k(t_0) = \dot{q}_{k0} \,, \tag{1.8}$$

i.e., specify the state at some initial time t_0.

Note 3: The Lagrangian is not uniquely determined; for example, if two Lagrangians differ by the total time derivative of arbitrary function, then the same equations will hold for both Lagrangians, and the same physical solution will result:

$$\tilde{L} = L + \mathrm{d}f/\mathrm{d}t\,; \qquad f = f(q,\dot{q}) \qquad \Rightarrow$$

$$\delta \int_{t_A}^{t_B} \tilde{L}\,\mathrm{d}t = \delta \int_{t_A}^{t_B} L\,\mathrm{d}t + \delta \int_{t_A}^{t_B} \frac{\mathrm{d}f}{\mathrm{d}t}\mathrm{d}t = 0 + \delta \int_{A}^{B} \mathrm{d}f = \delta\left[f(B) - f(A)\right] = 0\,. \tag{1.9}$$

Thus, if Hamilton variational principle holds for the original Lagrangian, it also holds for the new one (shifted by df/dt). This can be used when modifying the Lagrangian of interest (see Example 1.27 at the end of Section 1.4.6).

Note 4: Hamilton principle, as stated, applies only to non-dissipative systems, i.e., systems in which there are no heat losses.

Note 5: Lagrange equations are only necessary conditions for the extremality of the action integral; they are not sufficient ones.

Note 6: In the case of the first two examples, where the goal is not to find the time-dependent trajectory but rather a general solution for the extremality of the functional (1.2), the necessary conditions are Euler equations

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial F}{\partial y'} - \frac{\partial F}{\partial y} = 0\,.$$

Note 7: In mathematics, the necessary conditions for a functional to attain a minimum are called Euler equations; in physics, the necessary conditions for the extreme of the action integral are called Lagrange equations. These equations are sometimes simply referred to collectively as the Euler–Lagrange equations.

The most important task in this field of science is choosing the correct Lagrangian. If we select a specific form of the Lagrangian, we can solve the corresponding Lagrangian equations and compare these solutions with the experimental trajectory. If they do not match, the chosen Lagrangian is incorrect. The choice of the Lagrangian function is one of the fundamental axioms of the theory under construction. As a rule, a suitable scalar function is chosen for L (its value does not depend on the choice of coordinates). For simple mechanical problems, we know of two important scalar functions: kinetic and potential energy. In the simplest case, the Lagrangian could be a linear combination of these: $L = \alpha T + \beta V$. Indeed, it can be shown that for the choice $\alpha = 1$, $\beta = -1$, we obtain the correct equations of motion; in the Cartesian coordinate system, these are Newton equations (see Example 1.6 in the following section). Therefore

$$L(t,\mathbf{q},\dot{\mathbf{q}}) = T(\mathbf{q},\dot{\mathbf{q}}) - V(t,\mathbf{q}) . \tag{1.10}$$

Potential energy depends on position. For more complex systems, breaking down the Lagrangian into kinetic and potential energy is quite difficult and, moreover, unnecessary. The sole task of mechanics is to choose the correct Lagrangian for a given system so that the solutions to the corresponding Lagrangian equations match the observed trajectories. Conversely, as we will see later, based on various symmetries of the system, one can use the Lagrangian to define quantities such as energy, momentum, angular momentum of the system, etc.

A suitable Lagrangian can also be found for relativistic mechanics, the motion of charged particles, the theory of the electromagnetic field, the general relativity, and other branches of physics. From this, equations describing the problem are derived – for example, Maxwell equations in the theory of the electromagnetic field.

1.1.5 Simple Examples

◖ **Example 1.6: A point mass in a potential field** V. A point mass has three degrees of freedom; we choose the following generalized coordinates:

$$q_1 = x ; \qquad q_2 = y ; \qquad q_3 = z ,$$

then

$$\begin{aligned} T(\dot{x},\dot{y},\dot{z}) &= \frac{1}{2} m\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right); \\ V(x,y,z) &\cdots \text{ given function;} \\ L(\mathbf{x},\dot{\mathbf{x}}) &= T - V = \frac{1}{2} m\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) - V(x,y,z). \end{aligned}$$

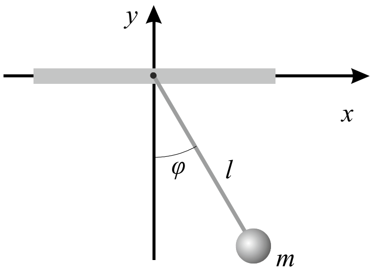
The corresponding Lagrange equations take the form

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} &= 0 \quad\Rightarrow\quad \frac{\mathrm{d}}{\mathrm{d}t}(m\dot{x}) + \frac{\partial V}{\partial x} = 0 \quad\Rightarrow\quad m\ddot{x} = -\frac{\partial V}{\partial x}; \\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{y}} - \frac{\partial L}{\partial y} &= 0 \quad\Rightarrow\quad \frac{\mathrm{d}}{\mathrm{d}t}(m\dot{y}) + \frac{\partial V}{\partial y} = 0 \quad\Rightarrow\quad m\ddot{y} = -\frac{\partial V}{\partial y}; \\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{z}} - \frac{\partial L}{\partial z} &= 0 \quad\Rightarrow\quad \frac{\mathrm{d}}{\mathrm{d}t}(m\dot{z}) + \frac{\partial V}{\partial z} = 0 \quad\Rightarrow\quad m\ddot{z} = -\frac{\partial V}{\partial z}. \end{aligned}$$

We can rewrite all three equations of motion into standard form

$$m\ddot{\mathbf{x}} = \mathbf{F} ; \quad \mathbf{F} \equiv -\nabla V .$$ ◗

◖ **Example 1.7: Plane pendulum**. A plane pendulum has a single degree of freedom. We choose the angle φ as the independent variable.:



$$x(t) = l\sin\varphi(t)\,; \quad y(t) = -l\cos\varphi(t)\,,$$
$$\dot{x} = l\dot{\varphi}\cos\varphi\,; \quad \dot{y} = l\dot{\varphi}\sin\varphi\,,$$
$$T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) = \frac{1}{2}ml^2\dot{\varphi}^2\,,$$
$$V = mgy = -mgl\cos\varphi\,,$$
$$L = T - V = \frac{1}{2}ml^2\dot{\varphi}^2 + mgl\cos\varphi\,.$$

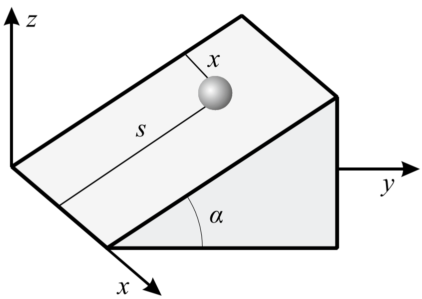
The corresponding Lagrange equation is

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{\varphi}} - \frac{\partial L}{\partial\varphi} = 0 \qquad \Rightarrow \qquad \ddot{\varphi} + \frac{g}{l}\sin\varphi = 0\,.$$

For small angles, sin $\varphi \approx \varphi$, and the equation reduces to the well-known equation for a mathematical pendulum

$$\ddot{\varphi} + \frac{g}{l}\varphi = 0\,.$$ ◗

◖ **Example 1.8: Movement on an inclined plane**. Movement on an inclined plane has two degrees of freedom. As generalized coordinates, we will choose the distances x and s from the edges of the inclined plane. Using the standard procedure, we obtain



$$x(t) = x(t)\,,$$
$$y(t) = s(t)\cos\alpha\,,$$
$$z(t) = s(t)\sin\alpha\,,$$
$$T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) = \frac{1}{2}m(\dot{x}^2 + \dot{s}^2)\,,$$
$$V = mgz = mgs\sin\alpha\,,$$
$$L(s,\dot{x},\dot{s}) = T - V = \frac{m}{2}(\dot{x}^2 + \dot{s}^2) - mgs\sin\alpha$$

and the equations of motion are

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{x}} - \frac{\partial L}{\partial x} = 0\,,$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{s}} - \frac{\partial L}{\partial s} = 0\,.$$

After substituting for L, we obtain the final form of the equations for motion on an inclined plane:

$$\ddot{x} = 0\,,$$
$$\ddot{s} = -g\sin\alpha\,.$$

◗

1.1.6 More Examples

◖ **Example 1.9: LC circuit**. We will choose the charge $Q(t)$ drained from the capacitor bank as the generalized coordinate. The generalized velocity corresponds to the electric current $I = \mathrm{d}Q/\mathrm{d}t$.

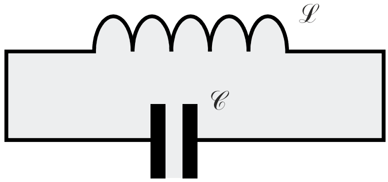


Fig. 1.9: LC circuit

If we denote the inductance by $\mathscr{L}$ and the capacitance by $\mathscr{C}$, then the Lagrangian

$$L(Q,\dot{Q}) = \frac{1}{2}\mathscr{L}\dot{Q}^2 - \frac{Q^2}{2\mathscr{C}}$$

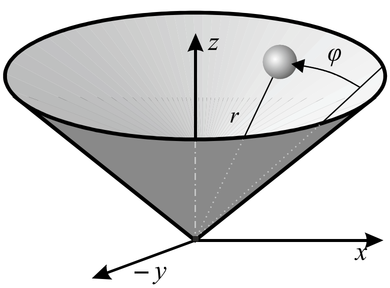
yields the correct LC circuit equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{Q}} - \frac{\partial L}{\partial Q} = 0 \qquad \Rightarrow \qquad \ddot{Q} + \frac{1}{\mathscr{L}\mathscr{C}}Q = 0\,.$$

Note that the first term in the Lagrangian is the energy stored in the coil's magnetic field, and the second term is the energy of the capacitor bank. Once again, the Lagrangian has a structure similar to that in mechanics. In this example, the energy in the coil plays the role of kinetic energy, while the energy stored in the capacitor plays the role of potential energy.

◗

◖ **Example 1.10: Point mass on a conical surface**. The motion has two degrees of freedom. As generalized coordinates, we will choose the distance r of the particle from the vertex of the cone and the polar angle φ. We will thus use two of the spherical coordinates; the third – the deviation θ_0 from the z-axis – is constant on the conical surface. Using (3.30) or (3.33), we can easily derive



$$x(t) = r(t)\cos\varphi(t)\sin\theta_0\,,$$
$$y(t) = r(t)\sin\varphi(t)\sin\theta_0\,,$$
$$z(t) = r(t)\cos\theta_0\,;$$
$$T(r,\dot{r},\dot{\varphi}) = \frac{m}{2}(\dot{r}^2 + r^2\sin^2\theta_0\,\dot{\varphi}^2)\,;$$
$$V(r) = mgz = mgr\cos\theta_0\,;$$
$$L(r,\dot{r},\dot{\varphi}) = \frac{m}{2}(\dot{r}^2 + r^2\dot{\varphi}^2\sin^2\theta_0) - mgr\cos\theta_0\,.$$

The corresponding Lagrange equations are

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{r}}-\frac{\partial L}{\partial r}=0\,; \qquad \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\varphi}}-\frac{\partial L}{\partial \varphi}=0\,.$$

After substituting, we have:

$$m\ddot{r}=mr\sin^2\theta_0\dot{\varphi}^2-mg\cos\theta_0\,,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(mr^2\dot{\varphi}\sin^2\theta_0)=0\,.$$

Note that the equation for r on the right-hand side involves the sum of the centrifugal force and the corresponding component of the gravitational force. The equation for the angle φ is nothing other than the law of conservation of angular momentum. ◗

◖ **Example 1.11: Pendulum on a cart.** A horizontally movable pendulum can be implemented, for example, using a cart on a rail. The system has two degrees of freedom. We choose the horizontal position $x(t)$ of the cart and the angle $\varphi(t)$ of the pendulum as the generalized coordinates. We will denote the Cartesian coordinates of the cart by the index a and the Cartesian coordinates of the pendulum by the index b. The rest of the procedure is standard.

$$x_a(t)=x(t)\,; \qquad x_b(t)=x(t)+l\sin\varphi(t)\,,$$
$$y_a(t)=0\,; \qquad y_b(t)=-\,l\cos\varphi(t)\,;$$

$$L(\varphi,\dot{x},\dot{\varphi})=\frac{1}{2}M_a\left(\dot{x}_a^2+\dot{y}_a^2\right)+\frac{1}{2}M_b\left(\dot{x}_b^2+\dot{y}_b^2\right)-M_a g y_a-M_b g y_b=$$
$$=\frac{1}{2}M_a\dot{x}^2+\frac{1}{2}M_b\left(\dot{x}^2+l^2\dot{\varphi}^2+2l\dot{x}\dot{\varphi}\cos\varphi\right)+M_b gl\cos\varphi\,.$$

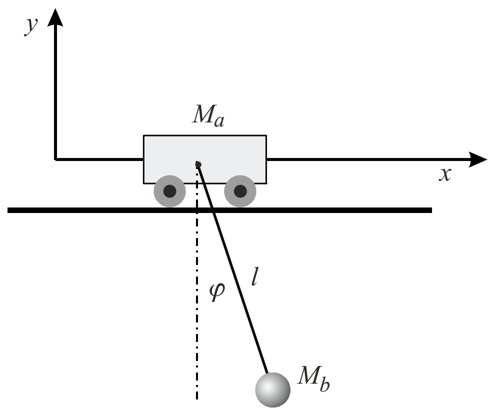


Fig. 1.11: Pendulum on a cart (The example comes from Lev Davidovich Landau)

◗

1.2 Conservation Laws in Nature

1.2.1 Emmy Noether Theorem

The discovery of any quantity that remains unchanged (is conserved) as a system evolves over time is of great importance in physics. In mechanics, we refer to these quantities as integrals of motion. Let us recall some conservation laws: conservation of momentum, angular momentum, energy, etc.; in quantum theory conservation of electric charge, spin, isospin, baryon number, parity, etc.

It is necessary to clarify the nature of these conservation laws and under what conditions they hold. Emmy Noether succeeded in doing this theoretically in 1916:

> Every symmetry in nature is associated with a conserved physical quantity. This quantity is defined by the symmetry in question and is conserved only as long as the underlying symmetry holds.

When observing the phenomena around us, it is therefore very important to look for various types of symmetry. Let's now look at some examples of symmetry:

1) We have built a mechanical device on our desk. We will start the device and observe its behavior. If we conduct the same experiment on the same desk in the next room, the result will be the same. However, if we perform the same experiment on a desk in a room one floor above, the outcome may be different because the Earth's gravitational field has a different strength at that desk. This physical situation is *symmetric with respect to horizontal translation* but is not symmetric with respect to vertical translation.

2) A constant current flows through the conductor. A time-invariant (stationary) magnetic field has formed around the conductor. We will release an electron into this field and observe its trajectory. If we release the electron one minute later (the electron's initial velocity and position must be the same), the resulting trajectory will be identical. Here we are talking about *symmetry with respect to a time translation*. If the current were not constant, this symmetry would be broken; the magnetic field would vary at different times, and the electron trajectories would differ.

3) In the strong interaction (holds the atomic nucleus together), neutrons and protons behave identically, whereas in the electromagnetic interaction they behave differently (proton is charged). The exchange of neutron and proton is a symmetric operation in the strong interaction but an asymmetric one in the electromagnetic one.

4) *Examples of other symmetries*: rotational symmetry; mirror symmetry (reversal of left and right); the results of experiments are the same in all coordinate systems moving uniformly in a straight line relative to one another (Lorentz symmetry).

In theoretical mechanics, we will explore the laws of conservation of momentum, angular momentum, and energy, as well as the symmetries associated with these conservation laws. In quantum theory, we will explore some of the other important symmetries that lead to the conservation of electric charge, spin, isospin, parity, quark color and flavor, and other quantum numbers.

1.2.2 Conservation of Momentum

Let's imagine that the Lagrangian does not depend on some generalized coordinate, specifically q_k :

$$L = L(t, q_1, \dots, q_{k-1}, q_{k+1}, \dots, q_f, \dot{q}_1, \dots, \dot{q}_f) \qquad \Leftrightarrow \qquad \frac{\partial L}{\partial q_k} = 0\,. \tag{1.11}$$

A generalized coordinate that does not appear in the Lagrangian is called a *cyclic* coordinate. In this case, neither the equations of motion nor the outcome of the experiment depend on q_k. The situation is symmetric with respect to a spatial translation in the generalized coordinate q_k (translational symmetry; see the first example of symmetries). From the equation of motion for this coordinate q_k, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} = 0 \qquad \Rightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} = 0 \qquad \Rightarrow \qquad \frac{\partial L}{\partial \dot{q}_k} = \text{const}\,.$$

So we found the relevant conserved quantity.

Definition of generalized momentum

We will call the *generalized momentum* corresponding to the generalized coordinate q_k

► $$p_k \equiv \frac{\partial L}{\partial \dot{q}_k}\,, \qquad k = 1, \dots, f\,. \tag{1.12}$$

This quantity is conserved if the generalized coordinate q_k is cyclic (does not appear in Lagrangian), i.e., the physical situation is symmetric with respect to a spatial translation in the generalized coordinate q_k. Let us now determine the generalized momenta for Examples 1.6 through 1.11 from Sections 1.1.5 and 1.1.6.

◖ **Example 1.12: A point mass in a potential field *V* (completion)**

$$p_x \equiv \frac{\partial L}{\partial \dot{x}} = m\dot{x}\,; \qquad p_y \equiv \frac{\partial L}{\partial \dot{y}} = m\dot{y}\,; \qquad p_z \equiv \frac{\partial L}{\partial \dot{z}} = m\dot{z}\,.$$

The conservation or non-conservation of momentum will depend on the shape of the potential energy $V(x, y, z)$.

Example 1.13: Plane pendulum (completion)

$$p_\varphi \equiv \frac{\partial L}{\partial \dot{\varphi}} = m l^2 \dot{\varphi}\,.$$

The physical situation is not symmetric with respect to a rotation by an angle $\delta\varphi$ (the gravitational field changes), which is why the coordinate φ appears in L and this generalized momentum is not conserved.

Note: In classical mechanics, the generalized momentum with respect to an angular variable is called angular momentum.

Example 1.14: Movement on an inclined plane (completion)

$$p_x \equiv \frac{\partial L}{\partial \dot{x}} = m\dot{x}\,; \qquad p_s \equiv \frac{\partial L}{\partial \dot{s}} = m\dot{s}\,.$$

The situation is symmetric with respect to a translation in the x-direction; the x-coordinate is cyclic and the momentum p_x is conserved. When translated in the s-direction, the gravitational field changes; L depends on s and the momentum p_s is not conserved.

Example 1.15: LC circuit (completion)

$$p_Q \equiv \frac{\partial L}{\partial \dot{Q}} = \mathscr{L}\dot{Q}\,.$$

The generalized momentum p_Q (magnetic flux) is not conserved; Q is not a cyclic coordinate.

Example 1.16: Point mass on a conical surface (completion)

$$p_r \equiv \frac{\partial L}{\partial \dot{r}} = m\dot{r}\,; \qquad p_\varphi \equiv \frac{\partial L}{\partial \dot{\varphi}} = mr^2 \sin^2\theta_0\;\dot{\varphi}\,.$$

Radial angular momentum p_r is not conserved (the gravitational field changes when there is a displacement in r), while angular momentum p_φ is conserved – the situation is symmetric with respect to a rotation by an angle φ, which is a cyclic coordinate.

Example 1.17: Pendulum on a cart (completion)

$$p_x \equiv \frac{\partial L}{\partial \dot{x}} = (M_a + M_b)\,\dot{x} + M_b l\dot{\varphi}\cos\varphi\,; \qquad p_\varphi \equiv \frac{\partial L}{\partial \dot{\varphi}} = M_b\, l^2\,\dot{\varphi} + M_b\dot{x}l\cos\varphi\,.$$

The momentum p_x is conserved, but the angular momentum p_φ is not conserved

◗

1.2.3 Conservation of Energy

Let the Lagrangian do not depend explicitly on time (it suffices that some of the infinitely many equivalent expressions of the Lagrangian do not depend on time), i.e.,

$$L = L(q_1,\ldots,q_f,\dot{q}_1,\ldots,\dot{q}_f) \qquad \Leftrightarrow \qquad \frac{\partial L}{\partial t} = 0\,. \tag{1.13}$$

This corresponds to the situation symmetric with respect to a time translation. Let's find the total time derivative of the Lagrangian:

$$\frac{\mathrm{d}L}{\mathrm{d}t} = \frac{\partial L}{\partial t} + \frac{\partial L}{\partial q_k}\dot{q}_k + \frac{\partial L}{\partial \dot{q}_k}\frac{\mathrm{d}}{\mathrm{d}t}(\dot{q}_k)\,.$$

Given the assumption that the first term on the right-hand side is zero, we express $\partial L/\partial q_k$ in terms of Lagrange equation (1.7) and obtain

$$\frac{\mathrm{d}L}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_k}\right)\dot{q}_k+\frac{\partial L}{\partial \dot{q}_k}\frac{\mathrm{d}}{\mathrm{d}t}(\dot{q}_k)\,.$$

We simplify the terms on the right-hand side using the rule for the derivative of the product of two functions

$$\frac{\mathrm{d}L}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_k}\dot{q}_k\right)$$

and after rearranging the equation, we find that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_k}\dot{q}_k-L\right)=0 \qquad \Rightarrow \qquad \frac{\partial L}{\partial \dot{q}_k}\dot{q}_k-L=\mathrm{const}\,.$$

So once again, we have found a conserved quantity.

Definition of generalized energy

We call the generalized energy

$$► \qquad E\equiv\frac{\partial L}{\partial \dot{q}_k}\dot{q}_k-L\,. \tag{1.14}$$

This quantity is conserved if the Lagrangian does not depend explicitly on time, i.e., if the physical situation is symmetric with respect to time translation.

In Examples 1.6 through 1.11, energy is conserved, the Lagrangian functions do not depend explicitly on time, and all situations are symmetric with respect to time translation. We obtain following formulas:

$$E_{1.6} = \frac{\partial L}{\partial \dot{x}}\dot{x}+\frac{\partial L}{\partial \dot{y}}\dot{y}+\frac{\partial L}{\partial \dot{z}}\dot{z}-L = \frac{1}{2}m(\dot{x}^2+\dot{y}^2+\dot{z}^2)+V(x,y,z)\,,$$

$$E_{1.7} = \frac{\partial L}{\partial \dot{\varphi}}\dot{\varphi}-L = \frac{1}{2}ml^2\dot{\varphi}^2-mgl\cos\varphi\,,$$

$$E_{1.8} = \frac{\partial L}{\partial \dot{x}}\dot{x}+\frac{\partial L}{\partial \dot{s}}\dot{s}-L = \frac{1}{2}m(\dot{x}^2+\dot{s}^2)+mgs\sin\alpha\,,$$

$$E_{1.9} = \frac{\partial L}{\partial \dot{Q}}\dot{Q}-L = \frac{1}{2}\mathscr{L}\dot{Q}^2+\frac{Q^2}{2\mathscr{C}}\,,$$

$$E_{1.10} = \frac{\partial L}{\partial \dot{r}}\dot{r}+\frac{\partial L}{\partial \dot{\varphi}}\dot{\varphi}-L = \frac{1}{2}m(\dot{r}^2+r^2\sin^2\theta_0\dot{\varphi}^2)+mgr\cos\theta_0\,,$$

$$E_{1.11} = \frac{\partial L}{\partial \dot{x}}\dot{x}+\frac{\partial L}{\partial \dot{\varphi}}\dot{\varphi}-L = \frac{1}{2}M_a\dot{x}^2+\frac{1}{2}M_b(\dot{x}^2+l^2\dot{\varphi}^2+2l\dot{x}\dot{\varphi}\cos\varphi)-M_bgl\cos\varphi\,.$$

Note that in all of these simple examples we have

$$E=T+V\,. \tag{1.15}$$

However, this relation holds only for specific forms of the Lagrangian. In the general case, *neither the Lagrangian nor the energy can be divided into kinetic and potential components*. Nevertheless, the energy is still always defined by the relation (1.14).

◖ **Example 1.18: Lack of energy conservation**. Let's conclude with an example where energy is not conserved. Consider a pendulum whose suspension cable is slowly wound up by an auxiliary motor at the point of attachment (a crane with a suspended load). The length of the suspension cable decreases over time

$$l = l_0 - ct\,,$$

c is the winding speed. Lagrange function of the pendulum

$$L = \frac{1}{2} m(l_0 - ct)^2 \dot{\varphi}^2 + mg(l_0 - ct)\cos\varphi$$

now explicitly depends on time, and energy is not conserved. Let's swing the pendulum and observe its oscillations. Let's do the same thing a minute later. The experiment will turn out differently because the suspension has shortened somewhat in the meantime. The physical situation is not symmetric with respect to the time translation. The reason for the non-conservation of energy is obvious here – the additional motor, which is not included in our system. ◗

We can see, then, that the fundamental laws of conservation in mechanics are a direct consequence of the properties of the space and time around us. If space is homogeneous (the same at all points), momentum is conserved; if space is isotropic (the same in all directions), angular momentum is conserved; if space remains unchanged over time, energy is conserved.

spatial homogeneity	→	momentum conservation
spatial isotropy	→	angular momentum conservation
invariance over time	→	energy conservation

◖ **Example 1.19: Brachistochrone (completion)**

We now have sufficient mathematical knowledge to solve the brachistochrone problem from the introduction of the Section 1.1.2. The task was to find the curve between two points along which a body can travel in the shortest time by sliding freely from point A to point B, where the difference in height between the two points is H. The problem led to finding the minimum of the functional (1.1).

$$T = \int_{x_A}^{x_B} \sqrt{\frac{1 + y'^2}{2g(H - y)}}\,\mathrm{d}x\,.$$

The independent variable in this problem is not time t, but the spatial coordinate x. The Euler-Lagrange equations will therefore take the form:

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial F}{\partial y'} - \frac{\partial F}{\partial y} = 0\,; \qquad F = \sqrt{\frac{1 + y'^2}{2g(H - y)}}\,.$$

A direct solution would be highly disadvantageous. Since the independent variable x is not represented in the functional, the "energy" must be conserved

$$E \equiv \frac{\partial F}{\partial y'} y' - F \qquad \Rightarrow \qquad \frac{y'}{\sqrt{2g(H-y)}} \frac{1}{\sqrt{1+y'^2}} y' - \sqrt{\frac{1+y'^2}{2g(H-y)}} = E_0 \,.$$

This is the first integral of the Euler-Lagrange equations and thus a first-order differential equation. Note that in this case, "energy" cannot be divided into a "kinetic" part involving derivatives of the unknown function and a "potential" part without derivatives. After a simple rearrangement, we have

$$E_0 \sqrt{2g(H-y)} \sqrt{1+y'^2} = -1 \,.$$

Let's raise the expression to the second power

$$2E_0^2 g(H-y)(1+y'^2) = 1 \qquad \Rightarrow$$

$$H - y = \frac{K}{1+y'^2}\,; \qquad K \equiv \frac{1}{2E_0^2 g}\,.$$

The simplest integration is parametric, i.e., the substitution $y' = \text{tg}\,\varphi$. The parametric solution for y is then

$$H - y = \frac{K}{1+\text{tg}^2\,\varphi} \qquad \Rightarrow \qquad y = H - K\cos^2\varphi \,. \tag{1.16}$$

It remains to find a solution for x from the defining relation for the substitution; $\text{d}y/\text{d}\varphi$ we express from (1.16):

$$y' = \text{tg}\,\varphi \qquad \Rightarrow \qquad \frac{\text{d}y}{\text{d}\varphi}\frac{\text{d}\varphi}{\text{d}x} = \frac{\sin\varphi}{\cos\varphi} \qquad \Rightarrow \qquad 2K\sin\varphi\cos\varphi\frac{\text{d}\varphi}{\text{d}x} = \frac{\sin\varphi}{\cos\varphi}\,.$$

By separating, we have

$$\text{d}x = 2K\cos^2\varphi\,\text{d}\varphi \,,$$

and after integration

$$x = K\varphi + K(\sin 2\varphi)/2 + L \,. \tag{1.17}$$

The integration constants K and L in equations (1.16) and (1.17) can be determined from the requirement that the solution must pass through the points (0, H) and (l, 0). For our purposes, a general solution that is part of the cycloid is sufficient:

$$x = K\varphi + K(\sin 2\varphi)/2 + L\,;$$
$$y = H - K\cos^2\varphi \,.$$

◗

1.3 Hamilton canonical equations

In this chapter, we will explore another form of the equations of motion – Hamilton equations. Unlike Lagrange equations (second-order differential equations), Hamilton equations are first-order equations, but there are twice as many of them.

1) A large number of numerical methods have been developed for solving first-order differential equations, and thus Hamilton equations are generally more suitable for numerical solutions than Lagrange equations.
2) Using Hamilton equations, it is easy to express the time evolution of any dynamic variable, i.e., not just the chosen generalized coordinates.
3) Hamilton equations can be rewritten in a very simple form using so-called Poisson brackets, which, from a mathematical point of view, are a Lie algebra. The properties of a Lie algebra are determined independently of the objects that form it. Therefore, it will be possible to easily transfer this structure to quantum mechanics.

1.3.1 Hamilton Equations

Using the definition of generalized momentum (1.12), we can rewrite the Lagrange equations (1.7) in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k}-\frac{\partial L}{\partial q_k}=0 \quad \wedge \quad p_k \equiv \frac{\partial L}{\partial \dot{q}_k} \qquad \Rightarrow \qquad \dot{p}_k \equiv \frac{\partial L}{\partial q_k}, \tag{1.18}$$

which strongly resembles Newton equations in Cartesian coordinates. Let us now find the differential of energy using its defining equation (1.14)

$$\begin{aligned} E &= p_k \dot{q}_k - L(t,q,\dot{q}) \qquad \Rightarrow \\ \mathrm{d}E &= \dot{q}_k\,\mathrm{d}p_k + p_k\,\mathrm{d}\dot{q}_k - \frac{\partial L}{\partial t}\mathrm{d}t - \frac{\partial L}{\partial q_k}\mathrm{d}q_k - \frac{\partial L}{\partial \dot{q}_k}\mathrm{d}\dot{q}_k\,. \end{aligned}$$

In the last but one term, we express $\partial L/\partial q_k$ from the equation of motion (1.18); in the final term, we use the definition of generalized momentum:

$$\mathrm{d}E = \dot{q}_k\,\mathrm{d}p_k + p_k\,\mathrm{d}\dot{q}_k - \frac{\partial L}{\partial t}\mathrm{d}t - \dot{p}_k\,\mathrm{d}q_k - p_k\,\mathrm{d}\dot{q}_k\,.$$

The terms with generalized velocity differentials are subtracted, leaving

$$\mathrm{d}E = -\frac{\partial L}{\partial t}\mathrm{d}t - \dot{p}_k\,\mathrm{d}q_k + \dot{q}_k\,\mathrm{d}p_k\,. \tag{1.19}$$

We will denote the function whose differential we have just found as

$$E = H(t,\mathbf{q},\mathbf{p}). \tag{1.20}$$

The coefficients in the differential (1.19) must be the corresponding partial derivatives of the function *H*:

$$-\frac{\partial L}{\partial t}=\frac{\partial H}{\partial t}; \qquad -\dot{p}_k=\frac{\partial H}{\partial q_k}; \qquad \dot{q}_k=\frac{\partial H}{\partial p_k}. \tag{1.21}$$

The function H is called the Hamiltonian. The Hamiltonian is the energy expressed in terms of the variables t, q_k, p_k. In (1.19), the velocity differentials have been subtracted, so it is always possible to find a transformation

$$t,\mathbf{q},\dot{\mathbf{q}} \qquad \to \qquad t,\mathbf{q},\mathbf{p}, \tag{1.22}$$

so that the energy is a function of generalized coordinates and generalized momenta. This transformation is called the *Legendre dual transformation*. The last two equations in relation (1.21) are Hamilton canonical equations (*canon* = law, set of rules):

► $$\dot{q}_k=\frac{\partial H}{\partial p_k}; \qquad \dot{p}_k=-\frac{\partial H}{\partial q_k}. \tag{1.23}$$

When solving the problem using Hamilton equations

1) We will determine the generalized momentum and generalized energy from the Lagrangian.
2) We eliminate generalized velocities from the generalized energy. We express them in terms of generalized momenta, i.e., we perform the so-called *Legendre dual transformation*.
3) We will write down Hamilton equations.
4) We solve them simultaneously for both positions and momenta.

Hamilton equations are equations used to determine the time evolution of the variables $q_k(t)$ and $p_k(t)$. They are first-order differential equations, but there are twice as many of them as there are second-order Lagrange equations. We must supplement the system of Hamilton equations with initial conditions

$$q_k(t_0)=q_{k0}; \qquad p_k(t_0)=p_{k0}. \tag{1.24}$$

◖ **Example 1.20: Plane motion of a planet (2D problem)**

Let the mass of the planet be *m* and the mass of the Sun be *M*. We assume $M \gg m$; that is, the Sun does not move. The motion has two degrees of freedom; as generalized coordinates, we choose polar coordinates $q_1=r(t)$; $q_2=\varphi(t)$, i.e., the distance of the planet from the Sun and the angle of the line connecting the planet and the Sun measured from the chosen direction. From (3.33) and the law of gravitation, we know that

$$T=\frac{1}{2}m(\dot{r}^2+r^2\dot{\varphi}^2), \qquad V=-G\frac{mM}{r}, \quad \text{i.e.}$$

$$L(r,\dot{r}\dot{\varphi})=T-V=\frac{1}{2}m(\dot{r}^2+r^2\dot{\varphi}^2)+G\frac{mM}{r}. \tag{1.25}$$

If we were solving a problem involving Lagrange equations, we would have

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot r}-\frac{\partial L}{\partial r}=0 \qquad \Rightarrow \qquad \ddot r - r\dot\varphi^2 + G\frac{M}{r^2}=0\,,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot \varphi}-\frac{\partial L}{\partial \varphi}=0 \qquad \Rightarrow \qquad r^2\ddot\varphi + 2r\dot r\dot\varphi = 0\,.$$

The equations of motion do not depend on the mass m of the planet under consideration. This is typical of gravity: bodies move along the same trajectories in a given gravitational field. Therefore, it is possible to describe gravity using curved space-time. Let us now determine the generalized momenta and the generalized energy of the system:

$$p_r \equiv \frac{\partial L}{\partial \dot r} = m\dot r\,; \qquad p_\varphi \equiv \frac{\partial L}{\partial \dot\varphi} = mr^2\dot\varphi\,,$$

$$E = \frac{\partial L}{\partial \dot r}\dot r + \frac{\partial L}{\partial \dot\varphi}\dot\varphi - L \;=\; \frac{1}{2}m\dot r^2 + \frac{1}{2}mr^2\dot\varphi^2 - G\frac{mM}{r} \;=\; T_r + T_\varphi + V\,.$$

The angular momentum p_φ (where φ is the cyclic coordinate) and the generalized energy E are conserved. The energy breaks down into three components: radial kinetic energy, angular energy (associated with the planet's orbit), and potential energy. We express the generalized velocities in terms of the generalized momenta

$$\dot r = \frac{p_r}{m}\,; \qquad \dot\varphi = \frac{p_\varphi}{mr^2}$$

and substitute them into the generalized energy (we perform Legendre dual transformation). This gives us the Hamiltonian

$$H(r,\varphi,p_r,p_\varphi) = \frac{p_r^2}{2m} + \frac{p_\varphi^2}{2mr^2} - G\frac{mM}{r}\,.$$

Hamilton canonical equations are

$$\dot r = \frac{\partial H}{\partial p_r} = \frac{p_r}{m}\,, \qquad \dot p_r = -\frac{\partial H}{\partial r} = +\frac{p_\varphi^2}{mr^3} - G\frac{mM}{r^2}\,,$$

$$\dot\varphi = \frac{\partial H}{\partial p_\varphi} = \frac{p_\varphi}{mr^2}\,, \qquad \dot p_\varphi = -\frac{\partial H}{\partial \varphi} = 0\,.$$

These equations must be supplemented with initial conditions $r(t_0)$, $\varphi(t_0)$, $p_r(t_0)$, $p_\varphi(t_0)$. This is a system of four differential equations for functions $r(t)$, $\varphi(t)$, $p_r(t)$, $p_\varphi(t)$.

◗

Definition of phase space

We refer to *phase space* as a $2f$-dimensional space into which we map the values of generalized coordinates and generalized momenta. A point in phase space represents the state of the system. The time evolution **q**(t), **p**(t) of the system's state is represented in phase space as a *phase trajectory*. Configuration space is a subspace of phase space. In the following section, we will examine the phase trajectory of a harmonic oscillator.

1.3.2 Harmonic Oscillator

The harmonic oscillator is one of the most important physical systems. As a first approximation, it can be used to describe the behavior of a particle in a potential well, and it appears in both quantum theory and quantum field theory. Any field (such as an electromagnetic field) can always be represented as a system of harmonic oscillators. Therefore, we will examine the harmonic oscillator in greater detail.

Let's consider a particle in a potential energy field with a minimum at the point x_0 and a minimum value of $V_0 = V(x_0)$. Let's perform a second-order Taylor expansion of $V(x)$ in the vicinity of the minimum:

$$V(x) = V(x_0) + V'(x_0)\cdot(x - x_0) + \frac{1}{2}V''(x_0)\cdot(x-x_0)^2 + \cdots \quad .$$

At the minimum, $V'(x_0) = 0$, and therefore

$$\blacktriangleright \quad V(x) \approx V(x_0) + \frac{1}{2}V''(x_0)(x-x_0)^2 \approx V_0 + \frac{1}{2}k(x-x_0)^2 , \qquad k \equiv V''(x_0) . \tag{1.26}$$

We have replaced the potential energy with a parabolic function – see Fig. 1.12.

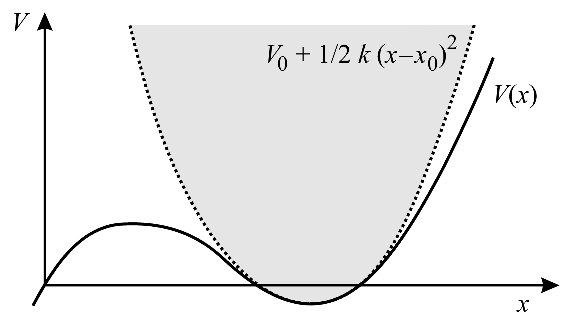


Fig. 1.12: Harmonic oscillator

A *harmonic oscillator* is a system in which the potential energy is parabolic (1.26). For example, a body suspended on a spring in a gravitational field satisfies this relationship quite closely. The quantity $k \equiv V''(x_0)$ is called the *stiffness of the oscillator*.

Let us choose a coordinate system such that the potential energy minimum is at the origin ($x_0 = 0$), and let $V(x_0) = 0$ (we can change the potential energy by an additive constant; the force $F = -\,\mathrm{d}V/\mathrm{d}x$ remains unchanged); with this choice, the potential energy function is $V(x) = 1/2\ kx^2$. Let us first solve problem using Lagrange equations:

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2 \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} = 0 \quad \Rightarrow \quad \ddot{x} + \frac{k}{m}x = 0 . \tag{1.27}$$

The general solution to this equation is

$$x(t) = c_1 \cos\omega t + c_2 \sin\omega t , \qquad \omega \equiv \sqrt{\frac{k}{m}} . \tag{1.28}$$

The solution is derived from the following initial conditions:

$$x(0) = A\,; \qquad \dot{x}(0) = 0 \qquad \Rightarrow \qquad x(t) = A\cos\omega t\,. \tag{1.29}$$

Near the minimum of potential energy, a particle performs oscillatory motion with an angular frequency $\omega = (k/m)^{1/2}$. The angular frequency ω is more commonly used as a parameter of the oscillator than its stiffness k. The Lagrangian is then

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}m\omega^2 x^2\,. \tag{1.30}$$

Let's now solve the problem using Hamilton equations; first, we'll find the generalized momentum and energy:

$$p = \frac{\partial L}{\partial \dot{x}} = m\dot{x} \qquad \Rightarrow \qquad \dot{x} = \frac{p}{m}\,,$$
$$E = \frac{\partial L}{\partial \dot{x}}\dot{x} - L = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}m\omega^2 x^2\,.$$

After eliminating velocity from the energy E, we obtain the Hamiltonian

► $$H(x,p) = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2 \tag{1.31}$$

and Hamilton equations

$$\dot{x} = \frac{\partial H}{\partial p} = \frac{p}{m}\,, \qquad \dot{p} = -\frac{\partial H}{\partial x} = -m\omega^2 x\,.$$

Solving this system with the same initial conditions yields the following results

$$\begin{aligned} x(t) &= A\cos\omega t\,, \\ p(t) &= -mA\omega\sin\omega t\,. \end{aligned} \tag{1.32}$$

Note that $p = m\dot{x}$. If we eliminate time from (1.32) (leaving only trigonometric functions on the right-hand sides, squaring the equation, and summing), we obtain the equation for the trajectory in phase variables x and p:

► $$\left(\frac{x}{A}\right)^2 + \left(\frac{p}{mA\omega}\right)^2 = 1\,. \tag{1.33}$$

The phase trajectory of a harmonic oscillator is an ellipse. We can imagine the motion as the coordinates of a point moving along the ellipse (its projection onto both axes).

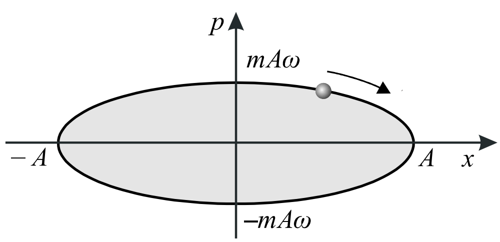


Fig. 1.13: Phase portrait of the harmonic oscillator

Finally, we will determine the classical probability density of a particle's position between the endpoints $-A$ and A. The probability that a particle is located within a distance Δx of the point x is given by:

$$\Delta P \cong \frac{2\Delta t}{T} = \frac{2\Delta x/v(x)}{2\pi/\omega} = \frac{\omega}{\pi v(x)}\Delta x\,.$$

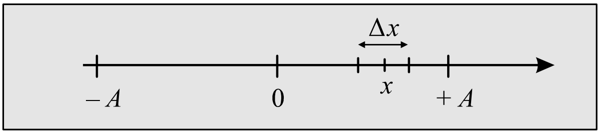


Fig. 1.14: Derivation of the probability of the oscillator's occurrence

We have designated the period as T; the time the particle spends in the vicinity of point x is $2\Delta t$. The particle passes through this region twice during a period of T (back and forth), which is why $2\Delta t$ appears in the numerator. The probability density is

$$w(x) = \frac{\mathrm{d}P}{\mathrm{d}x} = \frac{\omega}{\pi v(x)}\,. \tag{1.34}$$

We determine the dependence $v(x)$ from the law of conservation of energy

$$\frac{1}{2}mv^2 + \frac{1}{2}m\omega^2 x^2 = \frac{1}{2}m\omega^2 A^2 \qquad \Rightarrow \qquad v(x) = \omega\sqrt{A^2 - x^2}\,. \tag{1.35}$$

The final relationship has the form

► $$w(x) = \frac{1}{\pi\sqrt{A^2 - x^2}}\,. \tag{1.36}$$

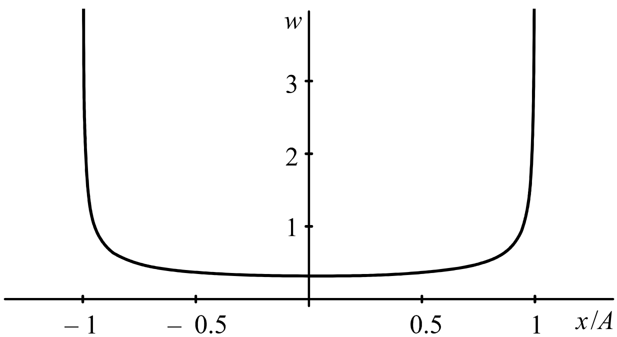


Fig. 1.15: The probability density of the oscillator's occurrence

The probability density of a particle's occurrence is highest at the turning points $-A$ and A, and lowest at the minimum of the potential energy. In quantum theory, we observe a modification of this behavior for particles of the microscopic world (see Chapter 2.3). Note also that the total probability of a particle occurring in the region $(-A, A)$ is equal to one, regardless of the fact that the probability density diverges at the endpoints:

$$\int_{-A}^{+A} w(x)\,\mathrm{d}x = \int_{-A}^{+A} \frac{1}{\pi\sqrt{A^2 - x^2}}\mathrm{d}x = \frac{1}{\pi}\left[\arcsin\left(\frac{x}{A}\right)\right]_{-A}^{+A} = 1\,. \tag{1.37}$$

1.3.3 Poisson Formulation of Hamilton Equations

Before reading this section, the reader should be familiar with the definition and properties of a Lie algebra (see Section 3.3.5). Consider a general dynamical variable $A(\mathbf{q},\mathbf{p})$, which is a function of generalized coordinates and generalized momenta (coordinates, momentum, potential energy, the product of potential and kinetic energy…). Its evolution over time is given by the relation

$$\dot{A}=\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{\partial A}{\partial q_k}\dot{q}_k+\frac{\partial A}{\partial p_k}\dot{p}_k = \frac{\partial A}{\partial q_k}\cdot\frac{\partial H}{\partial p_k}-\frac{\partial A}{\partial p_k}\cdot\frac{\partial H}{\partial q_k}, \tag{1.38}$$

where we expressed the time derivatives of the phase variables **q** and **p** in terms of Hamilton equations.

Definition of Poisson brackets

Let $f(\mathbf{q},\mathbf{p})$ and $g(\mathbf{q},\mathbf{p})$ be two functions of the phase variables **q** and **p**. The function

► $$\{f,g\}\equiv\frac{\partial f}{\partial q_k}\cdot\frac{\partial g}{\partial p_k}-\frac{\partial f}{\partial p_k}\cdot\frac{\partial g}{\partial q_k} \tag{1.39}$$

we call the Poisson bracket of the functions f and g. The time evolution (1.38) of a general dynamical variable is, by definition (1.39), given by the Poisson bracket of the corresponding dynamical variable and the Hamiltonian:

► $$\dot{A}=\{A,H\}\,. \tag{1.40}$$

Note: For $A = A(t,\mathbf{q},\mathbf{p})$, we have $\mathrm{d}A/\mathrm{d}t = \partial A/\partial t + \{A, H\}$. In many cases, systems in which the dynamic variables do not depend explicitly on time are satisfactory.

Properties of Poisson brackets:

$$\begin{aligned}
&1)\ \{f,g\}=-\{g,f\}\,,\\
&2)\ \{f+g,h\}=\{f,h\}+\{g,h\};\quad \{\alpha f,h\}=\alpha\{f,h\};\quad \alpha\in R\,,\\
&3)\ \{f,\{g,h\}\}+\{g,\{h,f\}\}+\{h,\{f,g\}\}=0\,,\\
&4)\ \{fg,h\}=f\{g,h\}+\{f,h\}g\,,\\
&5)\ \{f,gh\}=g\{f,h\}+\{f,g\}h \qquad \text{(follows from the previous).}
\end{aligned} \tag{1.41}$$

The proof of these relationships is trivial and follows directly from the definition of the Poisson bracket (1.39). The *Poisson brackets form a Lie algebra* on the space of functions (see Section 3.3.5). It is very important to know the Poisson brackets between generalized coordinates and momenta:

$$\{q_i,q_j\}=\frac{\partial q_i}{\partial q_k}\cdot\frac{\partial q_j}{\partial p_k}-\frac{\partial q_i}{\partial p_k}\cdot\frac{\partial q_j}{\partial q_k}=\delta_{ik}\cdot 0-0\cdot\delta_{jk}=0\,,$$

$$\{p_i,p_j\}=\frac{\partial p_i}{\partial q_k}\cdot\frac{\partial p_j}{\partial p_k}-\frac{\partial p_i}{\partial p_k}\cdot\frac{\partial p_j}{\partial q_k}=0\cdot\delta_{jk}-\delta_{ik}\cdot 0=0\,,$$

$$\{q_i, p_j\} = \frac{\partial q_i}{\partial q_k}\cdot\frac{\partial p_j}{\partial p_k} - \frac{\partial q_i}{\partial p_k}\cdot\frac{\partial p_j}{\partial q_k} = \delta_{ik}\cdot\delta_{jk} = \delta_{ij}\ .$$

The Poisson bracket is nonzero only for the generalized coordinate and the corresponding momentum; in that case, it is equal to one.

► $$\{q_i, q_j\} = \{p_i, p_j\} = 0\ ; \qquad \{q_i, p_j\} = \delta_{ij}\ . \tag{1.42}$$

These relations determine the entire Lie algebra of Poisson brackets. If we know their properties (1.41) and relations (1.42), we can solve problems in mechanics without needing definition (1.39).

◖ Example 1.21: Harmonic oscillator

$$T = \frac{p^2}{2m}, \qquad V = \frac{1}{2}m\omega^2 x^2\ ; \qquad H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2\ ;$$

$$\dot{x} = \{x, H\} = \frac{\partial x}{\partial x}\cdot\frac{\partial H}{\partial p} - \frac{\partial x}{\partial p}\cdot\frac{\partial H}{\partial x} = \frac{\partial H}{\partial p} = \frac{p}{m},$$

$$\dot{p} = \{p, H\} = \frac{\partial p}{\partial x}\cdot\frac{\partial H}{\partial p} - \frac{\partial p}{\partial p}\cdot\frac{\partial H}{\partial x} = -\frac{\partial H}{\partial x} = -m\omega^2 x\ .$$

We can also easily determine the time evolution of any dynamic variable, such as potential energy:

$$\dot{V} = \{V, H\} = \frac{\partial V}{\partial x}\cdot\frac{\partial H}{\partial p} - \frac{\partial V}{\partial p}\cdot\frac{\partial H}{\partial x} = \omega^2 xp\ .$$

However, we can also determine the time evolution from the properties of the Lie algebra of Poisson brackets (1.41) and (1.42) without knowing their definition. Let's illustrate this using the example of generalized momentum:

$$\dot{x} = \{x, H\} = \frac{\partial x}{\partial x}\cdot\frac{\partial H}{\partial p} - \frac{\partial x}{\partial p}\cdot\frac{\partial H}{\partial x} = \frac{\partial H}{\partial p} = \frac{p}{m}\quad ,$$

$$\dot{p} = \{p, H\} = \left\{p, \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2\right\} \overset{(1.41.2)}{=} \frac{1}{2m}\{p, p^2\} + \frac{1}{2}m\omega^2\{p, x^2\} =$$

$$\overset{(1.41.4)}{=} \frac{1}{2m}\big(p\{p,p\} + \{p,p\}p\big) + \frac{1}{2}m\omega^2\big(x\{p,x\} + \{p,x\}x\big) =$$

$$= \frac{p}{m}\{p,p\} + m\omega^2 x\{p,x\} =$$

$$\overset{(1.41.1)}{=} \frac{p}{m}\{p,p\} - m\omega^2 x\{x,p\} \overset{(1.42)}{=} -m\omega^2 x\quad .$$

We would proceed in the same manner with other dynamic variables. In quantum theory, this structure remains the same; only the objects we work with will be different.

◗

1.3.4 Numerical Solution of Hamilton Equations

Explicit solutions can be found only in exceptional cases. As a rule, we must rely on numerical solutions to the problem. In the text so far, we have learned to formulate the problem using a system of differential equations supplemented with appropriate initial conditions. Most mathematical software (e.g., "Mathematica," "Reduce," "Maple," etc.) can solve a problem formulated in this way numerically and sometimes even analytically. For more inquisitive students who would like to simulate the evolution of the system on a computer themselves, we present here at least one numerical method suitable for finding a numerical solution. We have chosen the 4th-order Runge-Kutta method, which is easy to implement and yet sufficiently accurate.

Let $\boldsymbol{\xi} = (\mathbf{q}, \mathbf{p})$ denote the set of generalized coordinates and momenta. Let the set of functions $\xi_k\ (t)$; $k = 1$, ... $2f$ satisfy the system of equations

$$\dot{\xi}_k = f_k(t, \xi)\,.$$

We divide the time axis into segments with an interval of Δt. Suppose we know the solution at some time t (for example, at t_0 – the initial condition). Then we determine

$$\begin{aligned}
K_{1,k} &= f_k(t, \xi_1, \cdots, \xi_{2f})\,,\\
K_{2,k} &= f_k\left(t + \frac{1}{2}\Delta t, \xi_1(t) + \frac{1}{2}K_{1,1}\,\Delta t, \cdots, \xi_{2f}(t) + \frac{1}{2}K_{1,2f}\,\Delta t\right),\\
K_{3,k} &= f_k\left(t + \frac{1}{2}\Delta t, \xi_1(t) + \frac{1}{2}K_{2,1}\,\Delta t, \cdots, \xi_{2f}(t) + \frac{1}{2}K_{2,2f}\,\Delta t\right),\\
K_{4,k} &= f_k\left(t + \Delta t, \xi_1(t) + K_{3,1}\,\Delta t, \cdots, \xi_{2f}(t) + K_{3,2f}\,\Delta t\right)
\end{aligned}$$

and we obtain an approximate solution at time $t + \Delta t$ from the following equations

$$\xi_k(t + \Delta t) \cong \xi_k(t) + \frac{1}{6}(K_{1,k} + 2K_{2,k} + 2K_{3,k} + K_{4,k}) \cdot \Delta t \quad ; \quad k = 1, \ldots, 2f\,.$$

This gives us the solution at time $t + \Delta t$, and we can repeat the process. Questions regarding the accuracy of the calculation, convergence, and a number of other methods can be found in the literature specializing in this topic. Some additional methods for solving ordinary differential equations are also presented in the companion textbook [2].

1.4 Some Problems in Classical Mechanics

1.4.1 Charged Particle in Electromagnetic Field

The motion of particles in an electromagnetic field is discussed in detail in the companion textbook [2]. For the sake of completeness, we will now present the basic equations for the non-relativistic case in a given electromagnetic field. We can describe electric and magnetic fields either in terms of electric field intensity **E** and magnetic flux density **B**, or using the four-potential (ϕ, **A**). The conversion relations are

► $$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \frac{\partial \phi}{\partial \mathbf{x}}, \tag{1.43}$$

► $$\mathbf{B} = \operatorname{rot} \mathbf{A}. \tag{1.44}$$

We assume that $\phi(t, \mathbf{x})$ and $\mathbf{A}(t, \mathbf{x})$ are given functions. The problem of the motion of a charged particle in a conservative electrostatic field is described by the Lagrangian function of the form $L = T - V$, i.e.,

$$L = \frac{1}{2} m \mathbf{v}^2 - Q\phi. \tag{1.45}$$

If a magnetic field is present, the system is no longer conservative (there is no potential energy), and the Lagrangian takes the form

► $$L = \frac{1}{2} m \mathbf{v}^2 - Q\phi + Q\mathbf{A} \cdot \mathbf{v}. \tag{1.46}$$

The first term is the kinetic energy of a free particle; the remaining two terms represent the particle's interaction with the electric and magnetic fields. The derivation can be found either in Section 1.6.3, Equation (1.260), or in companion textbook [2]. The second and third terms are, in fact, the scalar product of the four-potential of the field and the four-vector of the charge flux caused by the particle's motion. This ensures that it is a scalar quantity. Here, we will assume that we have "guessed" the correct Lagrangian. But then we must prove that the equations of motion derived from it are consistent with nature. We will show that the corresponding Lagrangian equations are identical to the well-known Lorentz equations of motion. In components, we have

$$L = \frac{1}{2} m v_j v_j - Q\phi(t, \mathbf{x}) + QA_j(t, \mathbf{x}) v_j \,;$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial v_i} - \frac{\partial L}{\partial x_i} = 0,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(m v_i + QA_i) + Q\frac{\partial \phi}{\partial x_i} - Q\frac{\partial A_j}{\partial x_i} v_j = 0,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(mv_i) + Q\frac{\partial A_i}{\partial t} + Q\frac{\partial A_i}{\partial x_j}\frac{\mathrm{d}x_j}{\mathrm{d}t} + Q\frac{\partial \phi}{\partial x_i} - Q\frac{\partial A_j}{\partial x_i}v_j = 0,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(mv_i) = Q\left[-\frac{\partial A_i}{\partial t} - \frac{\partial \phi}{\partial x_i} + v_j\left(\frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j}\right)\right],$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(m\,\mathbf{v}) = Q\left[-\frac{\partial \mathbf{A}}{\partial t} - \frac{\partial \phi}{\partial \mathbf{x}} + \mathbf{v}\times\mathrm{rot}\,\mathbf{A}\right],$$

► $$\frac{\mathrm{d}}{\mathrm{d}t}(m\,\mathbf{v}) = Q[\,\mathbf{E} + \mathbf{v}\times\mathbf{B}\,], \tag{1.47}$$

which is the well-known Lorentz equation of motion. The term $\mathbf{v}\times\mathrm{rot}\,\mathbf{A}$ can be expressed in standard form using the Levi-Civita tensor (see 3.3.4 Vector Identities). We will now derive the momentum, energy, and Hamiltonian of the particle using the standard procedure:

► $$\mathbf{p} \equiv \frac{\partial L}{\partial \mathbf{v}} = m\,\mathbf{v} + Q\,\mathbf{A} \tag{1.48}$$

► $$\mathscr{E} = \frac{\partial L}{\partial \mathbf{v}}\cdot\mathbf{v} - L = \frac{1}{2}m\,\mathbf{v}^2 + Q\phi \tag{1.49}$$

► $$H = \frac{(\mathbf{p} - Q\,\mathbf{A})^2}{2m} + Q\phi \tag{1.50}$$

Note 1: In this chapter, we denote energy by $\mathscr{E}$ to distinguish it from the electric field intensity **E**.

Note 2: Note that $\mathscr{E} \neq T + V$. Energy does not depend on the potential **A**. This is because the magnetic field does not change energy, but only the direction of velocity. Momentum is also no longer equal to its mechanical counterpart: $\mathbf{p} \neq m\mathbf{v}$.

◖ Example 1.22: Constant homogeneous electric field

Suppose that a particle is exposed to a constant, uniform electric field **E**. The equation of motion for the particle can be written as

$$m\ddot{\mathbf{r}} = Q\mathbf{E}\,. \tag{1.51}$$

Through direct integration, we will gradually achieve

$$\mathbf{v}(t) = \frac{Q}{m}t\,\mathbf{E} + \mathbf{v}_0\,; \tag{1.52}$$

$$\mathbf{r}(t) = \frac{Q}{2m}t^2\,\mathbf{E} + \mathbf{v}_0 t + \mathbf{r}_0\,. \tag{1.53}$$

At first glance, it is clear that the velocity increases beyond all limits as time increases. For higher velocities, it is necessary to use the relativistic Lagrangian. The calculation is given in companion textbook [2]. Acceleration occurs only in the direction of the field;

perpendicular to the field, the particle moves freely. If the field points along the x-axis, i.e., $\mathbf{E} = (E, 0, 0)$, and the particle is at the origin with an initial velocity $\mathbf{v}_0 = (0, v_0, 0)$, i.e., perpendicular to the field, the solution found has the form

$$x = \frac{QE}{2m} t^2 , \qquad y = v_0 t , \qquad z = 0 \ . \tag{1.54}$$

If we eliminate time from the equation, we obtain the equation of a parabola:

$$x = \frac{QE}{2m v_0^2} y^2 . \tag{1.55}$$

In the presence of a homogeneous electric field, a charged particle moves along a parabolic path. In the direction of the field, the motion is uniformly accelerated. If we perform a relativistic calculation (see [2]), the actual curve will be a hyperbolic cosine. ◗

◖ **Example 1.23: Constant homogeneous magnetic field**

Let us now consider the second simplest situation – the motion of a charged particle in a homogeneous magnetic field. For the sake of clarity, we will assume that the magnetic field points along the z-axis, i.e., $\mathbf{B} = (0, 0, B)$, and that the particle is initially at the origin of the coordinate system and has an initial velocity in the direction of the y-axis, i.e., $\mathbf{x}(t_0) = (0, 0, 0)$, $\mathbf{v}(t_0) = (0, v_0, 0)$.

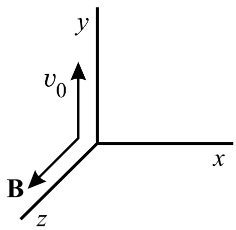


Fig. 1.16: The direction of individual vectors

We can split the motion equation $m\mathrm{d}\mathbf{v}/\mathrm{d}t = Q\, \mathbf{v}\times\mathbf{B}$ into its components:

$$m\ddot{x} = QB\,\dot{y} ,$$

$$m\ddot{y} = -QB\,\dot{x} , \tag{1.56}$$

$$m\ddot{z} = 0 .$$

Solving the third equation is straightforward; given our initial conditions, it is zero, meaning that the motion will occur only in the (x, y) plane. We will solve the system of the first two equations in the complex plane. We will treat the first equation as the real part and the second as the imaginary part:

$$m\ddot{x} + \mathrm{i}\, m\ddot{y} = QB\,\dot{y} - \mathrm{i}\, QB\,\dot{x} .$$

This operation is reversible; at any time, we can separate the real and imaginary parts and recover the original equations. After a simple rearrangement and denoting the combination QB/m as ω_c (we will later determine the meaning of this quantity), we obtain

$$\ddot{x} + \mathrm{i}\,\ddot{y} = -\mathrm{i}\,\omega_\mathrm{c}\,(\dot{x} + \mathrm{i}\,\dot{y})\,; \qquad \omega_\mathrm{c} \equiv \frac{QB}{m}\,. \tag{1.57}$$

Introducing the complex position $\xi \equiv x + \mathrm{i}\,y$, the equation takes a simple form

$$\ddot{\xi} + \mathrm{i}\,\omega_\mathrm{c}\,\dot{\xi} = 0\,; \qquad \xi \equiv x + \mathrm{i}\,y\,. \tag{1.58}$$

Of course, we can always return to the original variables x and y. We will seek a solution to this linear equation without the right-hand side in the exponential form $\exp(\lambda t)$. After substituting, we obtain the characteristic equation

$$\lambda^2 + \mathrm{i}\,\omega_\mathrm{c}\lambda = 0\,; \qquad \Rightarrow \qquad \lambda_1 = 0\,; \quad \lambda_2 = -\mathrm{i}\,\omega_\mathrm{c}\,. \tag{1.59}$$

The general solution is a linear combination of the two modes found:

$$\begin{aligned} \xi(t) &= c_1 + c_2\,\mathrm{e}^{-\mathrm{i}\,\omega_\mathrm{c} t}\,; \\ \dot{\xi}(t) &= -\mathrm{i}\,c_2\omega_\mathrm{c}\,\mathrm{e}^{-\mathrm{i}\,\omega_\mathrm{c} t}\,. \end{aligned} \tag{1.60}$$

We can easily find the integration constants from the initial conditions

$$\begin{aligned} \xi(0) &= x(0) + \mathrm{i}\,y(0) = 0\,; \\ \dot{\xi}(0) &= \dot{x}(0) + \mathrm{i}\,\dot{y}(0) = \mathrm{i}\,v_0\,. \end{aligned} \tag{1.61}$$

If we substitute these initial conditions into the equations (1.60), we get

$$\begin{aligned} c_1 + c_2 &= 0\,; & & & c_1 &= +v_0/\omega_\mathrm{c}\,; \\ -\mathrm{i}\,c_2\omega_\mathrm{c} &= \mathrm{i}\,v_0\,, & &\Rightarrow & c_2 &= -v_0/\omega_\mathrm{c}\,. \end{aligned} \tag{1.62}$$

The resulting solution therefore takes the form

► $$\xi(t) = R_\mathrm{L} - R_\mathrm{L}\,\mathrm{e}^{-\mathrm{i}\,\omega_\mathrm{c} t}\,; \qquad R_\mathrm{L} \equiv v_0/\omega_\mathrm{c} = m v_0/QB\,. \tag{1.63}$$

Separating the real and imaginary parts, we obtain coordinates of the moving particle

► $$\begin{aligned} x(t) &= R_\mathrm{L} - R_\mathrm{L}\cos\omega_\mathrm{c} t\,, \\ y(t) &= R_\mathrm{L}\sin\omega_\mathrm{c} t\,. \end{aligned} \tag{1.64}$$

We can find the equation of the trajectory by eliminating time from (1.64):

► $$(x - R_\mathrm{L})^2 + y^2 = R_\mathrm{L}^2\,. \tag{1.65}$$

We see that the motion occurs along a circle with radius $|R_\mathrm{L}|$, centered at $S = [R_\mathrm{L}, 0]$, and with an angular frequency of ω_c. We call the quantity R_L the Larmor (gyration) radius and the quantity ω_c the cyclotron (gyration) frequency. Depending on the particle's charge, the Larmor radius can have either a positive or negative value; similarly, the cyclotron frequency can have both signs (a negative value indicates counterclockwise rotation).

The magnetic field does not affect the particle's motion along the field. Perpendicular to the field, the Lorentz force acts, causing the particle's trajectory to curve into a circle. At a non-zero initial velocity $v_z(0)$, the particle's motion consists of uniform linear motion along the field and Larmor rotation, resulting in helical motion.

The electric field itself, on the other hand, has no effect on the particle's motion across the field (in the non-relativistic case) or only a very slight effect (in the relativistic case). Acceleration occurs in the direction of the field.

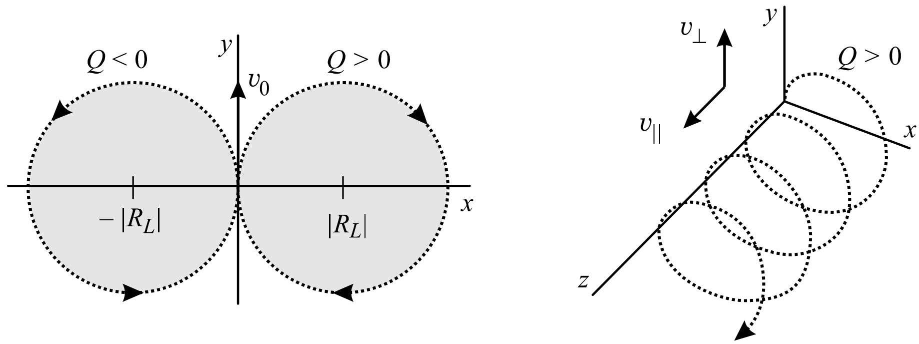


Fig. 1.17. The motion of a charged particle in a homogeneous magnetic field

More complex cases involving the motion of charged particles in electric and magnetic fields are discussed (based on Hamilton equations) in the companion textbook [2].

1.4.2 Motion in a Rotating Reference Frame

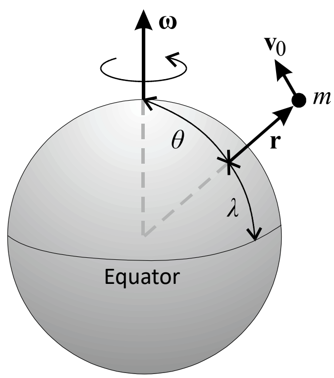


Fig. 1.18: Rotating reference frame

To find the equation of motion in a non-inertial rotating reference frame, we need to know certain vector identities:

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = (\mathbf{b}\times\mathbf{c})\cdot\mathbf{a} = (\mathbf{c}\times\mathbf{a})\cdot\mathbf{b}\,, \tag{1.66}$$

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \mathbf{b}(\mathbf{a}\cdot\mathbf{c}) - \mathbf{c}(\mathbf{a}\cdot\mathbf{b})\,, \tag{1.67}$$

$$(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d}) = (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}) - (\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})\,. \tag{1.68}$$

These identities can be easily derived using the definition of the vector product via the Levi-Civita tensor (see Sections 3.3.3 and 3.3.4). The first identity shows that the individual factors in the product $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$ can be swapped cyclically. The second identity is the well-known "bac–cab" rule. The proof of the third identity can be carried out (just as with the first two) simply by expanding the left-hand side from the definition of the vector product via the Levi-Civita tensor.

Let us assume that we are observing motions on a rotating Earth. One coordinate system is fixed in space (inertial), and the other rotates along with the Earth (non-inertial). Both systems have their origin at the center of the Earth, and the position vector of the observed body is the same in both systems (with different coordinates), i.e.,

$$\mathbf{r}_{\text{in}} = \mathbf{r}_{\text{rot}} = \mathbf{r} . \tag{1.69}$$

The velocities of a body of mass m (see Fig. 1.18) will differ in both reference frames by a velocity v_0 caused by the Earth's rotation. This velocity is proportional to the angular velocity, the distance from the center of the Earth, and the sine of the polar angle θ (the velocity is zero at the pole and maximum at the equator), i.e.,

$$v_0 = \omega r \sin\theta . \tag{1.70}$$

The direction of this velocity is perpendicular to both the vectors $\boldsymbol{\omega}$ and $\mathbf{r}$, therefore:

$$\mathbf{v}_0 = \boldsymbol{\omega}\times\mathbf{r} . \tag{1.71}$$

There is a simple relationship between the velocity in an inertial and a rotating frame:

$$\mathbf{v}_{\text{in}} = \mathbf{v}_{\text{rot}} + \boldsymbol{\omega}\times\mathbf{r} . \tag{1.72}$$

That is all we need to know to construct the Lagrangian. In an inertial frame of reference, the Lagrangian will be equal to

$$L = \frac{1}{2} m\mathbf{v}_{\text{in}}^2 - V(\mathbf{r}_{\text{in}}) . \tag{1.73}$$

We substitute the velocity from (1.72) and the position from (1.69) into the equation

$$L = \frac{1}{2} m\left(\mathbf{v}_{\text{rot}} + \boldsymbol{\omega}\times\mathbf{r}\right)^2 - V(\mathbf{r}) . \tag{1.74}$$

We denote the velocity in the rotating frame of reference that interests us as $\mathbf{v}$. The resulting Lagrangian for motion in a non-inertial rotating frame of reference is

► $$L = \frac{1}{2} m(\mathbf{v} + \boldsymbol{\omega}\times\mathbf{r})^2 - V(\mathbf{r}) . \tag{1.75}$$

To find the momentum and energy and to set up the equation of motion, we need the partial derivatives $\partial L/\partial\mathbf{v}$ and $\partial L/\partial\mathbf{r}$. Let us expand the first term in the Lagrangian:

$$L = \frac{1}{2} m(\mathbf{v} + \boldsymbol{\omega}\times\mathbf{r})\cdot(\mathbf{v} + \boldsymbol{\omega}\times\mathbf{r}) - V(\mathbf{r}) \qquad \Rightarrow$$

$$L = \frac{1}{2} m\,\mathbf{v}^2 + m(\boldsymbol{\omega}\times\mathbf{r})\cdot\mathbf{v} + \frac{1}{2} m(\boldsymbol{\omega}\times\mathbf{r})^2 - V(\mathbf{r}) \tag{1.76}$$

From this form of the Lagrangian, we determine the derivative with respect to $\mathbf{v}$:

$$\frac{\partial L}{\partial\mathbf{v}} = m\,\mathbf{v} + m\,\boldsymbol{\omega}\times\mathbf{r} . \tag{1.77}$$

To find the derivative with respect to the spatial variables, we must modify the Lagrangian (1.76) slightly. In the second term on the right-hand side, we rearrange the individual terms according to relation (1.66) so that **r**, with respect to which we wish to differentiate, is outside the vector product. In the third term, we modify the scalar product (**ω**×**r**)·(**ω**×**r**) according to Crammer relation (1.68)

$$L = \frac{1}{2} m\,\mathbf{v}^2 + m(\mathbf{v}\times\boldsymbol{\omega})\cdot\mathbf{r} + \frac{1}{2} m\,\omega^2\mathbf{r}^2 - \frac{1}{2} m\,(\boldsymbol{\omega}\cdot\mathbf{r})^2 - V(\mathbf{r}) \tag{1.78}$$

Now we can easily find the derivative of the Lagrangian with respect to the variable **r**:

$$\frac{\partial L}{\partial \mathbf{r}} = m(\mathbf{v}\times\boldsymbol{\omega}) + m\,\omega^2\mathbf{r} - m\,\boldsymbol{\omega}(\boldsymbol{\omega}\cdot\mathbf{r}) - \frac{\partial V}{\partial \mathbf{r}} \tag{1.79}$$

We simplify the second and third rhs terms using identity (1.67) to obtain the final form:

$$\frac{\partial L}{\partial \mathbf{r}} = m(\mathbf{v}\times\boldsymbol{\omega}) + m\,\boldsymbol{\omega}\times(\mathbf{r}\times\boldsymbol{\omega}) - \frac{\partial V}{\partial \mathbf{r}} \tag{1.80}$$

Eqs. (1.77) and (1.80) are the derivatives of the Lagrangian with respect to velocity and radius vector. Now we can determine the momentum and energy of the moving object:

► $$\mathbf{p} \equiv \frac{\partial L}{\partial \mathbf{v}} = m\mathbf{v} + m(\boldsymbol{\omega}\times\mathbf{r})\,, \tag{1.81}$$

► $$E \equiv \frac{\partial L}{\partial \mathbf{v}}\cdot\mathbf{v} - L = \frac{1}{2} m\mathbf{v}^2 - \frac{1}{2} m(\boldsymbol{\omega}\times\mathbf{r})^2 + V(\mathbf{r})\,. \tag{1.82}$$

Momentum consists of both classical mechanical momentum $m\mathbf{v}$ and a non-inertial rotational component $m\boldsymbol{\omega}\times\mathbf{r}$. Energy is the sum of kinetic energy, rotational energy, and potential energy. From the perspective of an observer in an inertial frame, rotational energy is negative. To predict the motion of bodies, it is most important to know the equation of motion, which we can now easily derive:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \mathbf{v}} - \frac{\partial L}{\partial \mathbf{r}} = 0 \qquad \Rightarrow$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(m\mathbf{v} + m(\boldsymbol{\omega}\times\mathbf{r})\big) - \left(m\,(\mathbf{v}\times\boldsymbol{\omega}) + m\,\boldsymbol{\omega}\times(\mathbf{r}\times\boldsymbol{\omega}) - \frac{\partial V}{\partial\,\mathbf{r}} \right) = 0 \qquad \Rightarrow$$

► $$\frac{\mathrm{d}m\mathbf{v}}{\mathrm{d}t} = \mathbf{F} + 2m\,(\mathbf{v}\times\boldsymbol{\omega}) + m\,\boldsymbol{\omega}\times(\mathbf{r}\times\boldsymbol{\omega})\,. \tag{1.83}$$

Result (1.83) is the equation of motion for a body moving in a rotating reference frame. On the left is the time derivative of mechanical momentum; on the right are: potential force, Coriolis force, and centrifugal force. The Coriolis force is partly responsible for the direction of rotation of a vortex in a funnel (different on each hemisphere), for the tilting of a pendulum's plane, and for other phenomena. On the Sun, the Coriolis force is very important in the formation of helical fluctuations in the velocity field, which ultimately contribute to the reversal of the solar magnetic dipole. The Coriolis force does not act on bodies moving parallel to the Earth's axis of rotation. The centrifugal force is zero at the poles and maximum at the equator, where it reaches a value of $mr\omega^2$.

◖ Example 1.24: Falling stone

Imagine dropping a stone from a height of 65 meters (equivalent to the Petřín Lookout Tower). How far will the stone deviate from the vertical after hitting the ground due to the Coriolis force? In which direction will the stone deviate from the vertical as it falls? Calculate for Prague (latitude $\lambda = 50°$) .

Solution: We will solve an equation of motion with gravitational and Coriolis forces

$$\frac{\mathrm{d}m\mathbf{v}}{\mathrm{d}t} = m\mathbf{g} + 2m(\mathbf{v}\times\boldsymbol{\omega}) .$$

The second term on the rhs is significantly smaller than the first and can be regarded as a small disturbance. The solution can be found using an iterative method, i.e.

$$\frac{\mathrm{d}m\mathbf{v}^{(k+1)}}{\mathrm{d}t} = m\mathbf{g} + 2m(\mathbf{v}^{(k)}\times\boldsymbol{\omega}) .$$

Let's choose some solution $\mathbf{v}^{(0)}$, substitute it into the right-hand side, and compute $\mathbf{v}^{(1)}$. Then we substitute $\mathbf{v}^{(1)}$ into the right-hand side and compute $\mathbf{v}^{(2)}$, and so on. For the initial zero solution, we have the sequence

$$\begin{aligned} \mathbf{v}^{(0)} &= 0 ; \\ \mathbf{v}^{(1)} &= \mathbf{g}\,t ; \\ \mathbf{v}^{(2)} &= \mathbf{g}\,t + (\mathbf{g}\times\boldsymbol{\omega})\,t^2 . \end{aligned}$$

The first iterative solution $\mathbf{v}^{(1)}$ represents free fall unaffected by the Coriolis force. The second iterative solution $\mathbf{v}^{(2)}$ incorporates the effect of the Coriolis force. Since we consider the Coriolis force to be a small disturbance, we can stop the iteration after the second term. During integration, we used a zero initial velocity, i.e., simple free fall. If $\mathbf{v}_0$ is non-zero, $\mathbf{v}^{(1)} = \mathbf{v}_0 + \mathrm{g}t$. For free fall, the solution found without an initial velocity is sufficient. Integrating the iterative solution gives us the position of the stone:

$$\mathbf{r}(t) \doteq \mathbf{r}^{(2)}(t) = \mathbf{r}_0 + \mathbf{g}\frac{t^2}{2} + (\mathbf{g}\times\boldsymbol{\omega})\frac{t^3}{3} . \qquad (1.84)$$

Let us now choose a specific coordinate system on the Earth's surface (we will shift the solution from the center of the Earth to the surface; the position appears only in the first term on the right-hand side). The z-axis will point vertically, the x-axis will be oriented eastward, and the y-axis northward. This is a right-handed coordinate system, which ensures the correct signs of the vector products. In Figure 1.18, θ is the polar angle and $\lambda = 90° - \theta$ is the latitude. The individual vectors in solution (1.84) will be:

$$\begin{aligned} \mathbf{r}^{(0)} &= (0,0,H) ; \\ \mathbf{g} &= (0,0,-g) ; \\ \boldsymbol{\omega} &= (0,\ \omega\cos\lambda,\ \omega\sin\lambda) . \end{aligned} \qquad (1.85)$$

The free fall of a stone is thus described by the following equations after expressing the vector product

$$x = \frac{g\omega t^3}{3}\cos\lambda;$$
$$y = 0; \qquad (1.86)$$
$$z = H - \frac{gt^2}{2}.$$

As it falls, the stone will be deflected by the Coriolis force in the direction of the x-axis, i.e., to the east. Now we need to determine the magnitude of the deflection. If we set $z = 0$ in the third equation (1.86), we can find the time it takes for the stone to hit the ground. We substitute this into the equation for x to obtain the resulting distance

$$\Delta x = x_{\text{fin}} - x_0 = \frac{g\omega}{3}\left(\frac{2H}{g}\right)^{3/2}\cos\lambda \approx 7\ \text{mm}. \qquad (1.87)$$

◗

1.4.3 Two-Body Problem, Kepler Problem

Let's consider two bodies that interact with each other via a force (gravitational, electrostatic, or some other force acting along the line connecting the two bodies). Their equations of motion are:

$$m_1\ddot{\mathbf{r}}_1 = \mathbf{F}_{12};$$
$$m_2\ddot{\mathbf{r}}_2 = \mathbf{F}_{21}. \qquad (1.88)$$

According to the law of action and reaction:

$$\mathbf{F}_{12} = -\mathbf{F}_{21} \qquad \Rightarrow \qquad \mathbf{F}_{12} + \mathbf{F}_{21} = 0. \qquad (1.89)$$

From the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, we switch to another set of six coordinates – the position of the center of mass and the relative position of the second body with respect to the first:

►
$$\mathbf{r}_T \equiv \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2};$$
$$\mathbf{r} \equiv \mathbf{r}_2 - \mathbf{r}_1. \qquad (1.90)$$

From Eqs. (1.88), we can easily derive the equations of motion in the new variables:

$$\ddot{\mathbf{r}}_T = \frac{\mathbf{F}_{12} + \mathbf{F}_{21}}{m_1 + m_2} = 0;$$

$$\ddot{\mathbf{r}} = \ddot{\mathbf{r}}_2 - \ddot{\mathbf{r}}_1 = \frac{\mathbf{F}_{21}}{m_2} - \frac{\mathbf{F}_{12}}{m_1} = \left(\frac{1}{m_2} + \frac{1}{m_1}\right)\mathbf{F}_{21}.$$

In the new variables, the equations of motion therefore take a very simple form

$$\ddot{\mathbf{r}}_\mathrm{T} = 0\,;$$
$$\mu\ddot{\mathbf{r}} = \mathbf{F}_{21}\,, \tag{1.91}$$

where μ is the so-called reduced mass given by the equation

► $$\mu \equiv \frac{m_1 m_2}{m_1 + m_2}\,. \tag{1.92}$$

From the first equation (1.91), it can be seen that the center of mass of the system can move only at a constant velocity. From the second equation, it is clear that the motion of two bodies can be treated as the motion of a single body with reduced mass μ, which moves relative to the first of the two bodies. If the mass of the first body is significantly greater than the mass of the second body, the reduced mass is approximately equal to the mass of the smaller of the two bodies. In this approximation, the larger body does not move at all. An example of this is the motion of planets around the Sun. The orbiting planets have a minimal effect on the Sun's motion.

It is worth noting that the motion of two bodies with forces along the line connecting them is always planar. This follows from the law of conservation of angular momentum:

$$\frac{\mathrm{d}\mathbf{b}}{\mathrm{d}t} = 0 \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r}\times m\mathbf{v}) = 0 \quad \Rightarrow \quad \mathbf{v}\times m\mathbf{v} + \mathbf{r}\times m\mathbf{a} = 0 \quad \Rightarrow$$
$$\mathbf{r}\times m\mathbf{a} = 0 \quad \Rightarrow \quad \mathbf{a}\,||\,\mathbf{r}\,. \tag{1.93}$$

The acceleration with which bodies act on each other lies along the line connecting them. If the motion occurs in a plane, the acceleration can never act outside that plane; therefore, the motion will take place only within that plane.

Kepler problem

Let us now consider the motion of a planet of mass m around the Sun of mass M ($m \ll M$). If the masses of the two bodies were comparable, we could easily reduce the problem to the relative motion of a body of reduced mass around the other body. Our problem is planar. Therefore, we will use polar coordinates, in which the Lagrangian function has the form (1.25)

$$L(r,\dot{r},\dot{\varphi}) = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\varphi}^2) + G\frac{mM}{r}\,. \tag{1.94}$$

In the system, angular momentum and energy will be conserved:

$$b \equiv \frac{\partial L}{\partial \dot{\varphi}} = mr^2\dot{\varphi}\,, \tag{1.95}$$

$$E \equiv \frac{\partial L}{\partial \dot{r}}\dot{r} + \frac{\partial L}{\partial \dot{\varphi}}\dot{\varphi} - L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\varphi}^2) - G\frac{mM}{r} = \frac{1}{2}m\dot{r}^2 + \frac{b^2}{2mr^2} - G\frac{mM}{r}\,. \tag{1.96}$$

In the law of conservation of energy, we expressed the time derivative of the angle φ from the law of conservation of angular momentum. Instead of solving the second-order equations of motion, we can integrate the first-order conservation laws. From both conservation laws, we calculate the time derivatives of the generalized coordinates:

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{b}{mr^2}, \tag{1.97}$$

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \sqrt{\frac{2}{m}\left(E - \frac{b^2}{2mr^2} + G\frac{mM}{r}\right)}. \tag{1.98}$$

This is a system of differential equations for the variables $r(t)$ and $\varphi(t)$. In Kepler problem, we are not interested in the time evolution of the coordinates, but only in the shape of the planet's orbit $r(\varphi)$. Therefore, we divide the second equation by the first:

$$\frac{\mathrm{d}\,r}{\mathrm{d}\varphi} = \frac{\sqrt{\frac{2}{m}\left(E - \frac{b^2}{2mr^2} + G\frac{mM}{r}\right)}}{b/mr^2}. \tag{1.99}$$

This equation for $r(\varphi)$ can be solved directly by separation followed by integration. However, let's take a slightly different approach. We'll square the equation and rewrite:

$$\frac{b^2}{2mr^4}(r')^2 = E - \frac{b^2}{2mr^2} + \frac{GmM}{r}$$

The prime denotes the derivative with respect to φ. Let's choose the substitution

$$r = \frac{1}{\xi}; \qquad r' = -\frac{1}{\xi^2}\xi', \tag{1.100}$$

after which we will receive

$$\frac{b^2}{2Gm^2M}(\xi')^2 + \frac{b^2}{2Gm^2M}\xi^2 - \xi = \frac{E}{GmM}. \tag{1.101}$$

If we differentiate the equation once more, we get

$$\frac{b^2}{Gm^2M}\xi''\xi' + \frac{b^2}{Gm^2M}\xi\xi' - \xi' = 0 \qquad \Rightarrow$$

$$\frac{b^2}{Gm^2M}(\xi'' + \xi) = 1 \qquad \Rightarrow$$

$$\xi'' + \xi = 1/p; \qquad p \equiv \frac{b^2}{Gm^2M}. \tag{1.102}$$

This is a linear second-order equation with a right-hand side. The particular solution is $1/p$; the homogeneous one $\cos(\varphi-\varphi_0)$. The general solution is therefore

$$\xi = C\cos(\varphi - \varphi_0) + \frac{1}{p}. \tag{1.103}$$

The solution has two integration constants, φ_0 and C. The constant φ_0 determines the origin of the polar angle. By rotating the coordinate system, we can choose $\varphi_0 = 0$.

The constant C ensures, that the solution found satisfies the original equation (1.101), i.e., the equation before differentiation (which is not an equivalent transformation):

$$C = \sqrt{\frac{2E}{GmMp} + \frac{1}{p^2}} \ . \tag{1.104}$$

We substitute C into the solution (1.103) and then return to the original variable r:

$$r = \frac{p}{1 + \left(\sqrt{\frac{2Ep}{GmM} + 1} \right) \cos\varphi} \ .$$

The resulting solution therefore has the form

$$\blacktriangleright \qquad r = \frac{p}{1 + \varepsilon \cos\varphi}; \qquad \varepsilon \equiv \sqrt{\frac{2Ep}{GmM} + 1}\, ; \qquad p \equiv \frac{b^2}{Gm^2 M} \ . \tag{1.105}$$

This is the equation of a conic section (see Section 3.10.1) with numerical eccentricity ε. In the gravitational field of a central body, other bodies will therefore move along conic sections. For $E < 0$ is $\varepsilon < 1$ and the motion follows an ellipse. Negative energy implies a gravitational bound of the two bodies. This case applies to the planets of the Solar System. For $E = 0$ is $\varepsilon = 1$ and the motion is parabolic. For $E > 0$ is $\varepsilon > 1$ and the body moves along a hyperbola. In all cases, we would determine the minimum and maximum orbital distances by finding the extrema of the function (1.105).

Effective potential

The energy of a moving body is given from the Lagrange function by equation (1.96)

$$E \equiv \frac{1}{2} m\dot{r}^2 + \frac{1}{2} mr^2 \dot{\varphi}^2 - G\frac{mM}{r} \ . \tag{1.106}$$

Energy consists of radial kinetic energy, the angular component of kinetic energy, and potential energy. Expressions involving generalized velocities contribute to the kinetic energy, while the expression involving only position contributes to the potential energy. However, if we express the second term using the law of conservation of angular momentum (1.95), we obtain

$$E = \frac{1}{2} m\dot{r}^2 + \frac{b^2}{2mr^2} - G\frac{mM}{r} \ . \tag{1.107}$$

The second term now depends only on position, so we can assign it to the potential. Whether we interpret the term as kinetic or potential is therefore relative and depends on our point of view. Let’s introduce the so-called effective potential:

$$\blacktriangleright \qquad \begin{aligned} E &= \frac{1}{2} m\dot{r}^2 + V_{\mathrm{eff}}(r)\, ; \\ V_{\mathrm{eff}}(r) &\equiv \frac{b^2}{2mr^2} - G\frac{mM}{r} \ . \end{aligned} \tag{1.108}$$

From the first equation, we can easily determine the radial velocity of the object

$$\dot{r} = \sqrt{\frac{2}{m}\left(E - V_{\text{eff}}(r)\right)} \,. \tag{1.109}$$

It is clear that movement can only occur in those areas of effective potential where the following applies

► $$E \geq V_{\text{eff}}(r) \,. \tag{1.110}$$

The graph of the effective potential is shown in Fig. 1.19, where it can be seen that for $E > 0$, the motion is unbounded ($r \in \langle r_{\min}, \infty)$), and the motion follows a hyperbola. Conversely, for $E < 0$, the motion is bounded ($r \in \langle r_{\min}, r_{\max} \rangle$), and the motion follows an ellipse. The limiting cases are $E = 0$ (motion along a parabola) and $E = E_{\min}$ (motion along a circle $r = r_0$).

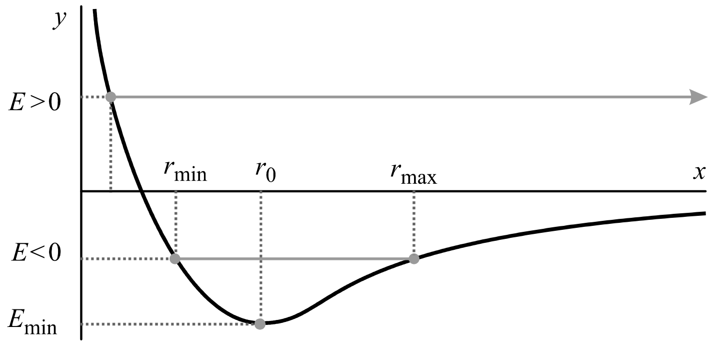


Fig. 1.19: Effective potential

◖ **Example 1.25: Earth as an oscillator**

The Earth's motion around the Sun can be understood as motion in an effective potential near a minimum. Such motion is approximately harmonic (the Earth's radial distance from the Sun fluctuates periodically). The potential can be approximated by a parabolic function near the minimum. Assume that the Earth's angular momentum is $b = 2{,}7\times10^{40}$ kg m^{2}s^{-1}. Determine the minimum of the effective potential and the period

Solution: Using the standard procedure (see Section 1.3.2), we determine the minimum effective potential (1.108) and the stiffness of the oscillations. From the stiffness, we can then easily find the period of the motion:

$$r_0 = \frac{b^2}{Gm^2M} \approx 150\times10^6 \text{ km};$$

$$k = V''_{\text{eff}}(r_0) = \frac{G^4 m^7 M^4}{b^6};$$

$$T = \frac{2\pi}{\omega} = \frac{2\pi}{\sqrt{k/m}} = \frac{2\pi}{\sqrt{G^4 m^6 M^4 / b^6}} = \frac{2\pi b^3}{G^2 m^3 M^2} \approx 365 \text{ dní.}$$

◗

Kepler laws

> 1) The planets move around the Sun in elliptical orbits, with the Sun at one of the foci.
> 2) The areal velocity of a planet's radius vector is constant.
> 3) The ratio of the square of a planet's orbital period to the cube of its semi-major axis is constant.

Kepler third law can be expressed by a simple equation:

► $$\frac{T^2}{a^3} = \frac{4\pi^2}{G(m+M)} . \tag{1.111}$$

We use m to denote the mass of the planet and M to denote the mass of the Sun.

Re 1). Motion along an ellipse follows immediately from equation (1.105).

Re 2). This is a simple consequence of the law of conservation of angular momentum. We can express the change in area during the planet's motion using the vector product

$$\mathrm{d}\mathbf{S} = \frac{1}{2}\mathbf{r}\times\mathrm{d}\mathbf{r} .$$

For the areal velocity, we then have

$$\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}t} = \frac{1}{2}\mathbf{r}\times\mathbf{v} = \frac{1}{2m}\mathbf{r}\times m\mathbf{v} = \frac{\mathbf{b}}{2m} . \tag{1.112}$$

The law of conservation of surface velocity is therefore simply another way of expressing the law of conservation of a planet's angular momentum.

Re 3). In relative coordinates, according to equation (1.91) , the following is valid

$$\mu\ddot{\mathbf{r}} = \mathbf{F} \quad\Rightarrow\quad \frac{mM}{m+M}\ddot{\mathbf{r}} = G\frac{mM}{r^2}\frac{\mathbf{r}}{r} \quad\Rightarrow\quad m\ddot{\mathbf{r}} = G\frac{m(m+M)}{r^2}\frac{\mathbf{r}}{r} .$$

In relative coordinates the sum of the masses of both bodies appears in place of the Sun's mass; otherwise, the equation of motion is identical to the equation for a central gravitational field. Let us now integrate equation (1.112) for the areal velocity

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{b}{2m} \quad\Rightarrow\quad S = \frac{b}{2m}T \quad\Rightarrow\quad \pi a_0 b_0 = \frac{b}{2m}T \quad\Rightarrow$$

$$2\pi a_0 b_0 m = bT . \tag{1.113}$$

We used equation (3.524) for the area of an ellipse, $S = \pi a_0 b_0$, which is given in Section 3.10.1 (we added the subscript 0 to avoid confusion with the angular momentum). We will attempt to convert all quantities to the characteristics of an elliptical orbit. We will substitute the angular momentum from equation (1.105); however, unlike in the central field, the mass of the Sun will appear here in the sum with the mass of the planet, i.e.

$$p = \frac{b^2}{Gm^2(m+M)} . \tag{1.114}$$

Substituting b from equation (1.114) into equation (1.113) gives

$$2\pi a_0 b_0 m = \sqrt{pGm^2(m+M)}\, T \qquad \Rightarrow \qquad 4\pi^2 a_0^2 b_0^2 = Gp(m+M)T^2 .$$

We will now convert all the parameters of the ellipse into the semi-major axis and eccentricity. We will use equations (3.153), i.e., $b_0{}^2 = a_0{}^2(1-\varepsilon^2)$; $p = a_0\,(1-\varepsilon^2)$. The result is the desired equation

$$\frac{T^2}{a_0^3} = \frac{4\pi^2}{G(m+M)} .$$

1.4.4 Lagrange Points

The three-body problem, in which three bodies interact gravitationally, cannot be solved analytically in its full scope. In the case of the so-called *restricted circular three-body problem* (see assumptions below), five equilibrium points can be found in the vicinity of two mutually orbiting bodies. If we place a small test body at these points, the gravitational forces from both bodies will be in equilibrium with the centrifugal force of motion at that location. These points were first discovered by the French-Italian mathematician Joseph Louis Lagrange, which is why they bear his name [9]–[10].

Assumptions

1) This is *restrictive problem*, i.e., two bodies are large and one is small (the test body). This test body does not affect the gravitational potential at that location.
2) This is a circular problem, i.e., two large bodies orbit each other in such a way that their distance remains constant. The relative motion of one body with respect to the other follows a circular path. Examples of such systems include the Earth-Moon, Earth-Sun, and Jupiter-Sun pairs.

Problem Statement

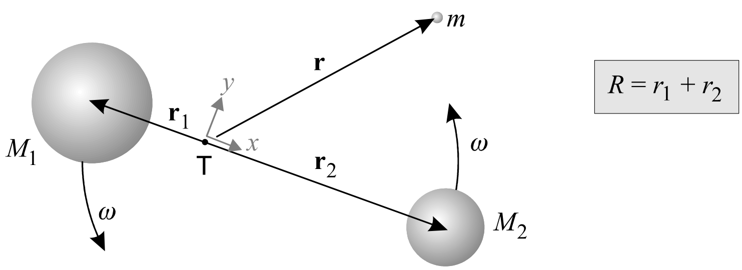


Fig. 1.20: Center of mass is denoted by T, and the distance between the two large bodies is *R*.

We assume that two massive bodies with masses M_1 and M_2 orbit around a common center of mass T in such a way that their mutual distance R remains constant, i.e., one body orbits the other in a circle (this corresponds to a minimum on the effective potential graph). We are looking for positions of a test body of small mass m ($m \ll M_1, M_2$) in which all forces (gravitational, centrifugal) are zero and the body is at rest (thus the Coriolis force is also zero). We will formulate the problem in a coordinate system that has its origin at the center of mass and rotates along with both bodies at an angular velocity given by Kepler third law (1.111)

$$\omega^2 = \frac{G(M_1 + M_2)}{R^3}; \qquad R = r_1 + r_2 . \tag{1.115}$$

The angular velocity, as a vector, points perpendicular to the plane of motion of both bodies (e.g., along the z-axis). The line connecting the two bodies forms one of the coordinate axes of our system (e.g., the x-axis). This is a non-inertial coordinate system in which the motion of the test body is given by the Lagrangian function (1.75):

$$L = \frac{1}{2} m(\mathbf{v} + \boldsymbol{\omega} \times \mathbf{r})^2 + G\frac{mM_1}{|\mathbf{r} - \mathbf{r}_1|} + G\frac{mM_2}{|\mathbf{r} - \mathbf{r}_2|} \tag{1.116}$$

The Lagrangian can be expressed in terms of effective potential energy as

$$L = T - V_{\text{eff}} ; \tag{1.117}$$

$$T \equiv \frac{1}{2} m\mathbf{v}^2 , \tag{1.118}$$

$$V_{\text{eff}} \equiv -m\mathbf{v} \cdot (\boldsymbol{\omega} \times \mathbf{r}) - \frac{1}{2} m(\boldsymbol{\omega} \times \mathbf{r})^2 - G\frac{mM_1}{|\mathbf{r} - \mathbf{r}_1|} - G\frac{mM_2}{|\mathbf{r} - \mathbf{r}_2|} . \tag{1.119}$$

The force acting on our test body (i.e., the equation of motion) takes the form (1.83), which we derived for a non-inertial coordinate system:

$$\blacktriangleright \quad \frac{\mathrm{d}m\mathbf{v}}{\mathrm{d}t} = 2m(\mathbf{v} \times \boldsymbol{\omega}) + m\,\boldsymbol{\omega} \times (\mathbf{r} \times \boldsymbol{\omega}) - G\frac{mM_1}{|\mathbf{r} - \mathbf{r}_1|^3}(\mathbf{r} - \mathbf{r}_1) - G\frac{mM_2}{|\mathbf{r} - \mathbf{r}_2|^3}(\mathbf{r} - \mathbf{r}_2) . \tag{1.120}$$

On the right are the individual forces (Coriolis, centrifugal, and gravitational). We choose the coordinate system such that the x-axis lies on the line connecting the two bodies, the y-axis is perpendicular to the x-axis (in the plane of motion), and the z-axis is perpendicular to the plane of motion. Searching equilibrium points can proceed in two ways. The first option is to find the extrema of the effective potential (1.119). The first term will be zero, since in equilibrium the test body does not move ($\mathbf{v} = 0$). This approach is suitable for graphical solutions – the equipotentials of the function V_{eff} can be plotted using appropriate software (such as Mathematica).

The second option is to solve the condition for zero force (1.120). At equilibrium points, the test body does not move; its velocity relative to our coordinate system is zero, and therefore the first term (the Coriolis force) is also zero. Thus, it is merely a balance between gravitational and centrifugal forces. However, the Coriolis force has a significant influence on the stability of the found Lagrange points.

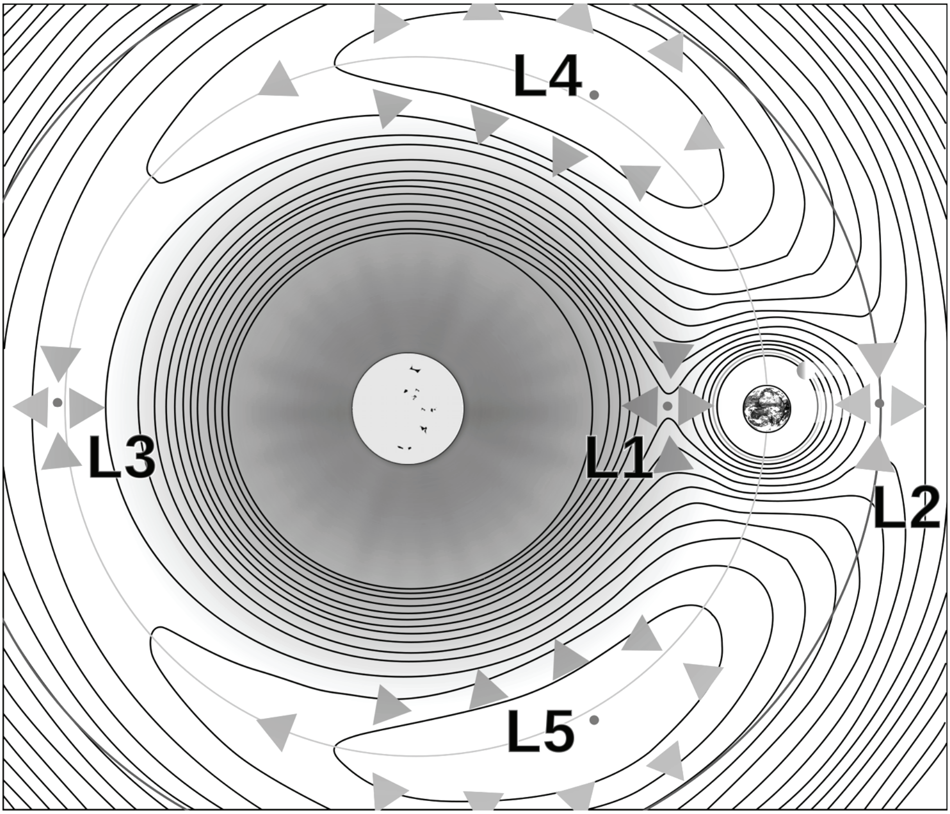


Fig. 1.21. Equipotential lines of the effective potential of the Sun-Earth system. The arrows indicate the direction "from the hill to the valley". NASA/WMAP.

Lagrange points

We find the Lagrange points from the equation

$$m\,\boldsymbol{\omega}\times(\mathbf{r}\times\boldsymbol{\omega})-G\frac{mM_1}{\left|\mathbf{r}-\mathbf{r}_1\right|^3}\left(\mathbf{r}-\mathbf{r}_1\right)-G\frac{mM_2}{\left|\mathbf{r}-\mathbf{r}_2\right|^3}\left(\mathbf{r}-\mathbf{r}_2\right)=0\,, \qquad (1.121)$$

which expresses the balance between centrifugal and gravitational forces. The position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are known in our coordinate system; we substitute them into (1.121):

$$\mathbf{r}_1=\left(-\mu_2 R,0,0\right); \qquad \mathbf{r}_2=\left(+\mu_1 R,0,0\right);$$

$$\boldsymbol{\omega}=(0,0,\omega)\,; \qquad \omega^2=\frac{GM}{R^3}\,; \qquad (1.122)$$

$$\mathbf{r}=(x,y,0)\,, \qquad \overline{\mathbf{r}}=(\overline{x},\overline{y},0)\equiv(x/R,y/R,0)\,,$$

where we denoted the variables

$$\mu_1 \equiv \frac{M_1}{M}; \qquad \mu_2 \equiv \frac{M_2}{M}; \qquad M \equiv M_1 + M_2 ;$$
$$\bar{x} \equiv \frac{x}{R}; \qquad \bar{y} \equiv \frac{y}{R} \tag{1.123}$$

The constants μ_1 and μ_2 are relative masses (they express what fraction of the total mass is accounted for by the first and second bodies, respectively). The variables $\bar{x}$, $\bar{y}$ are dimensionless coordinates of the Lagrange point in units of the distance between the two bodies. The result is a system of two equations for the two unknown coordinates of the Lagrange point:

$$\bar{x} - \frac{\mu_1(\bar{x}+\mu_2)}{\left[(\bar{x}+\mu_2)^2+\bar{y}^2\right]^{3/2}} - \frac{\mu_2(\bar{x}-\mu_1)}{\left[(\bar{x}-\mu_1)^2+\bar{y}^2\right]^{3/2}} = 0 ;$$
$$\bar{y} - \frac{\mu_1\bar{y}}{\left[(\bar{x}+\mu_2)^2+\bar{y}^2\right]^{3/2}} - \frac{\mu_2\bar{y}}{\left[(\bar{x}-\mu_1)^2+\bar{y}^2\right]^{3/2}} = 0 . \tag{1.124}$$

One solution follows immediately from the second equation, namely $\bar{y} = 0$. These are Lagrange points located on the x-axis, i.e., on the line passing through both bodies. Substituting $\bar{y} = 0$ into the first equation, we have

$$\bar{x} - \frac{\mu_1 \operatorname{sgn}(\bar{x}+\mu_2)}{(\bar{x}+\mu_2)^2} - \frac{\mu_2 \operatorname{sgn}(\bar{x}-\mu_1)}{(\bar{x}-\mu_1)^2} = 0 . \tag{1.125}$$

This is a fifth-degree equation with three real solutions (Lagrange points L_1, L_2, L_3). In the equation, we will retain the smaller of the two parameters, for example, the mass μ_2:

$$\bar{x}(\bar{x}+\mu_2)^2(\bar{x}-1+\mu_2)^2 \pm (1-\mu_2)(\bar{x}-1+\mu_2)^2 \pm \mu_2(\bar{x}+\mu_2)^2 = 0 . \tag{1.126}$$

The possible combinations of signs for the second and third terms are

$$+\,-\ (L_1);\ +\,+\ (L_2);\ -\,-\ (L_3) .$$

The solution can be found either numerically or by expanding to the lowest order in terms of μ_2. For example, for the Sun-Earth system, $\mu_2 \approx 3\times10^{-6}$, and expanding in terms of this parameter is valid. The resulting positions are

$$\blacktriangleright\ L_1 = R\left[1-\left(\frac{\mu_2}{3}\right)^{1/3}, 0\right]; \quad L_2 = R\left[1+\left(\frac{\mu_2}{3}\right)^{1/3}, 0\right]; \quad L_3 = -R\left[1+\frac{5\mu_2}{12}, 0\right]. \tag{1.127}$$

For the Sun-Earth system, the distance R is equal to one astronomical unit, i.e., approximately 150×10^6 km. The Lagrange points L_1 and L_2 are located approximately 1.5×10^6 km from Earth.

To find the solution for $\bar{y} \neq 0$, we go back to equations (1.124). We can take advantage of the symmetry of the problem with respect to the x-axis. In this case, the solution can be found analytically:

$$L_4 = R\left[\frac{\mu_1 - \mu_2}{2}, +\frac{\sqrt{3}}{2}\right]; \qquad L_5 = R\left[\frac{\mu_1 - \mu_2}{2}, -\frac{\sqrt{3}}{2}\right]. \tag{1.128}$$

Lagrange points L_1, L_2. Both of these points lie on a straight line passing through both bodies. In the case of the Sun-Earth system, point L_1 lies 1.5×10^6 km from Earth toward the Sun, and point L_2 lies at the same distance away from the Sun. Point L_1 is highly advantageous for observing the Sun. The ISEE-3 probe was first placed there in 1978. Later, point L_1 became home to one of the most famous solar probes, SOHO. At point L_1, the Sun's orbital speed is greater than at the Earth's location. However, the Earth's gravitational pull slows the probe at L_1 to the same angular velocity. Conversely, at the L_2 point, which is farther from the Sun, probes for deep-space research are located – in the past, WMAP, Planck; now, James Webb Space Telescope. The probes would orbit the Sun more slowly here, but Earth's gravitational pull gives them the correct speed. In terms of stability, both L_1 and L_2 are Lagrange points; when deflected in one direction, a restoring force acts on the probe; when deflected in the other direction, the probes begin to move away exponentially with a characteristic constant $\tau \approx 23$ days, which means that the deviation grows according to the function $\exp(t/\tau)$. Probes at these points must always perform orbit corrections every few months.

Lagrange point L_3. For the Sun-Earth system, this point lies on the opposite side of the Sun, slightly beyond Earth's orbit. Any object at this point is permanently invisible to us because it lies behind the Sun's disk. This led to speculation that a planet permanently invisible to us might be located at the L_3 point, which was named Planet X. The Lagrange point L_3 is again a saddle point; it is unstable with a characteristic period of $\tau \approx 150$ years. Any long-term existence of a planet at this location is therefore completely ruled out.

Lagrange points L_4, L_5. These points do not lie on the line connecting the two bodies, but form equilateral triangles with them. The distances of points L_4 and L_5 from both bodies are equal to R, i.e., the distances between the bodies themselves. Both points lie at the maxima of the effective potential, and therefore, at first glance, they might appear to be unstable. However, at any deviation, the Coriolis force begins to act on the test body, causing it to return. The result is an orbit around points L_4 and L_5. For the Sun–Earth system, the orbital period is 89 days. Bodies located at the Lagrange points L_4 and L_5 are called Trojans. The best-known are the Trojans of the Sun–Jupiter system.

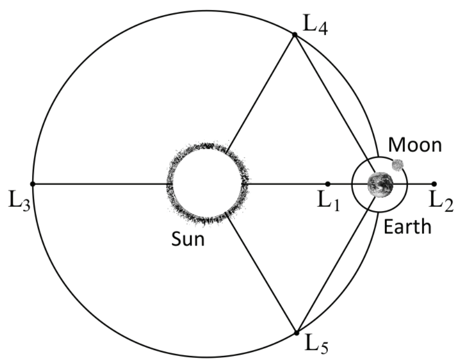


Fig. 1.22: Lagrange points of the Sun-Earth system

1.4.5 Energy Dissipation

Lagrange equations, in the form in which we derived them, apply to non-dissipative systems – that is, systems in which no portion of the energy is irreversibly converted into heat. This is suitable for all fundamental interactions in nature. However, for more technical applications, it is sometimes necessary to include heat losses in the calculation. It is sufficient to have electrical resistance present in an electrical circuit or for a moving body to experience friction with the surrounding air. In these cases, the equations of motion always contain terms proportional to the generalized velocity. The equation of motion for a flying stone serves as an example:

$$\frac{\mathrm{d}\,m\mathbf{v}}{\mathrm{d}t} = m\mathbf{g} - \alpha\mathbf{v}\,. \tag{1.129}$$

The first term on the right is the gravitational force, and the second is the drag force. The drag force is proportional to velocity (or, in some cases, to the square of velocity) and acts in the opposite direction to the motion. Another example is a simple circuit in which all three basic components – a capacitor, an inductor, and a resistor – are connected in series. The sum of the voltages across all components must equal zero, that is

$$\mathscr{L}\frac{\mathrm{d}I}{\mathrm{d}t} + \frac{Q}{\mathscr{C}} + \mathscr{R}I = 0\,. \tag{1.130}$$

The generalized variable (see Section 1.1.6, LC circuit) is the charge. The "motion" equation takes the form

$$\mathscr{L}\ddot{Q} + \mathscr{R}\dot{Q} + \frac{Q}{\mathscr{C}} = 0\,. \tag{1.131}$$

The second term is again proportional to the generalized velocity and represents energy dissipation in the system. If we wish to generalize the Lagrange equations to include energy dissipation, they will take the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} + \alpha_{kl}\,\dot{q}_l = 0\,. \tag{1.132}$$

In the last term, we used Einstein summation convention. This leads us to reformulate Lagrange equations into the form

►
$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} = -\frac{\partial R}{\partial \dot{q}_k}\,;\qquad R = \frac{1}{2}\alpha_{kl}\dot{q}_k\dot{q}_l\,. \tag{1.133}$$

The derivatives of the quadratic function on the right-hand side give rise to dissipative terms that are linear in the generalized velocities. This function is called the *Rayleigh dissipative function*. In the case of dissipation, to correctly formulate the problem, we must "guess" two functions: the Lagrangian and the Rayleigh dissipation function. If both functions have the correct form, we obtain equations that are consistent with natural phenomena.

◖ Example 1.26: RLC circuit

Equation (1.131) is obtained from Lagrange equations (1.133) if we choose

$$L(Q,\dot{Q}) = \frac{1}{2}\mathscr{L}\dot{Q}^2 - \frac{Q^2}{2\mathscr{C}}; \qquad R(\dot{Q}) = \frac{1}{2}\mathscr{R}\dot{Q}^2 . \tag{1.134}$$

The Lagrangian has a similar form as in mechanics (its terms resemble the difference between kinetic and potential energy). The Rayleigh dissipation function is proportional to the energy loss from the system per unit time, since $\mathrm{d}E/\mathrm{d}t = UI = \mathscr{R}I^2 = \mathscr{R}\,(\mathrm{d}Q/\mathrm{d}t)^2$.

◗

In the case of dissipative processes, we define momentum and energy in the same way as usual, i.e.

$$p_k \equiv \frac{\partial L}{\partial \dot{q}_k}; \tag{1.135}$$

$$E \equiv \frac{\partial L}{\partial \dot{q}_k}\dot{q}_k - L . \tag{1.136}$$

However, in the case of dissipation, energy will not be conserved, even if the Lagrangian does not depend explicitly on time. Let us find the change in energy with respect to time in this case:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_k}\dot{q}_k - L\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_k}\right)\dot{q}_k + \frac{\partial L}{\partial \dot{q}_k}\ddot{q}_k - \frac{\partial L}{\partial q_k}\dot{q}_k - \frac{\partial L}{\partial \dot{q}_k}\ddot{q}_k =$$

$$= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_k}\right)\dot{q}_k - \frac{\partial L}{\partial q_k}\dot{q}_k = \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_k}\right) - \frac{\partial L}{\partial q_k}\right]\dot{q}_k = -\frac{\partial R}{\partial \dot{q}_k}\dot{q}_k = -2R$$

In the final step, we used Euler theorem on the derivative of homogeneous functions, which for a quadratic function $f(x)$ takes the form $x\mathrm{d}f/\mathrm{d}x = 2f$ (the proof is trivial; it is simply the derivative of a square). The result obtained

$$\frac{\mathrm{d}E}{\mathrm{d}t} = -2R \tag{1.137}$$

clearly links Rayleigh dissipation function to the energy loss from the system per unit time. Equation (1.137) can be written as the law of conservation of energy as follows:

$$E + \int_0^t 2R\,\mathrm{d}t = \text{const} . \tag{1.138}$$

If we perform a transformation from the variables $(q, \mathrm{d}q/\mathrm{d}t)$ to the variables (q, p) in generalized energy, we obtain the Hamiltonian $H(q, p)$. Let us now show what the Hamilton equations will look like in the case of energy dissipation. It clearly holds that

$$\mathrm{d}L = \frac{\partial L}{\partial q_k}\mathrm{d}q_k + \frac{\partial L}{\partial \dot{q}_k}\mathrm{d}\dot{q}_k \qquad \Rightarrow$$

$$\mathrm{d}L = \left(\dot{p}_k + \frac{\partial R}{\partial \dot{q}_k} \right) \mathrm{d}q_k + p_k \mathrm{d}\dot{q}_k \qquad \Rightarrow$$

$$\mathrm{d}L = \left(\dot{p}_k + \frac{\partial R}{\partial \dot{q}_k} \right) \mathrm{d}q_k + \mathrm{d}(p_k \dot{q}_k) - \dot{q}_k \mathrm{d}p_k \qquad \Rightarrow$$

$$\mathrm{d}\left(p_k \dot{q}_k - L \right) = -\left(\dot{p}_k + \frac{\partial R}{\partial \dot{q}_k} \right) \mathrm{d}q_k + \dot{q}_k \mathrm{d}p_k \qquad \Rightarrow$$

$$\mathrm{d}H = -\left(\dot{p}_k + \frac{\partial R}{\partial \dot{q}_k} \right) \mathrm{d}q_k + \dot{q}_k \mathrm{d}p_k \ .$$

In deriving the first relation, we used Lagrange equations in the form (1.133) and the definition of momentum (1.135). Since the Hamiltonian H is a function only of generalized coordinates and momenta, the coefficients of the corresponding differentials must be equal to the corresponding partial derivatives, i.e.,

$$-\left(\dot{p}_k + \frac{\partial R}{\partial \dot{q}_k} \right) = \frac{\partial H}{\partial q_k} ; \qquad \dot{q}_k = \frac{\partial H}{\partial p_k} .$$

From here, we can easily derive Hamilton equations for the case of energy dissipation:

► $$\begin{aligned} \dot{q}_k &= \frac{\partial H}{\partial p_k} ; \\ \dot{p}_k &= -\frac{\partial H}{\partial q_k} - \frac{\partial R}{\partial \dot{q}_k} . \end{aligned} \qquad (1.139)$$

We must express the term $\partial R/\partial \dot{q}_k$ in terms of generalized positions and momenta, i.e., as a function of (q, p).

1.4.6 Inverse Problem

A very interesting task is to find the Lagrangian for a given problem. If we know absolutely nothing, we try to estimate the Lagrangian as some combination of scalars in the theory. If we obtain equations of motion that agree with the experiment, our efforts have been crowned with success. For a given problem, there exists an infinite number of Lagrangian functions, which may differ by the total time derivative of any function. In other words, if we add (or subtract) $\mathrm{d}f/\mathrm{d}t$ to (or from) the found Lagrangian function, where $f = f(t, q, \mathrm{d}q/\mathrm{d}t)$, we obtain the same Lagrangian equations. This is used to subsequently simplify the found Lagrangian function into the simplest possible form.

However, we often do not start from scratch when looking for a Lagrangian function. If we know the equations of motion, we try to find a suitable Lagrangian function from them. Let us denote the left-hand sides of the equations of motion by ε_k:

$$\varepsilon_k \equiv \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} = 0 \qquad (1.140)$$

The inverse problem refers to finding the Lagrangian from known differential equations of the form $\varepsilon_k = 0$. This problem does not always have a solution; that is, for some systems of equations, the Lagrangian does not exist and a variational formulation is not possible. In 1969, the Italian theoretical physicist Enzo Tonti derived sufficient conditions for the existence of the Lagrangian [8]. If the left-hand sides of the differential equations satisfy the conditions

$$\frac{\partial \varepsilon_k}{\partial \ddot{q}_l} - \frac{\partial \varepsilon_l}{\partial \ddot{q}_k} = 0 \,;$$

► $$\frac{\partial \varepsilon_k}{\partial \dot{q}_l} + \frac{\partial \varepsilon_l}{\partial \dot{q}_k} = 2\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \varepsilon_l}{\partial \ddot{q}_k} \,; \tag{1.141}$$

$$\frac{\partial \varepsilon_k}{\partial q_l} - \frac{\partial \varepsilon_l}{\partial q_k} = -\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial \varepsilon_l}{\partial \dot{q}_k}\right] + \frac{\mathrm{d}^2}{\mathrm{d}t^2}\left[\frac{\partial \varepsilon_l}{\partial \ddot{q}_k}\right]$$

For every k, l, the equations $\varepsilon_k = 0$ are variational, i.e., there exists a Lagrangian L and the following holds

► $$L = -q_k \int_0^1 \varepsilon_k(t, \tau q, \tau \dot{q}, \tau \ddot{q})\, \mathrm{d}\tau \,. \tag{1.142}$$

In equation (1.142), the summation convention applies. The minus sign is included here solely to ensure that the equations of motion take the form given in (1.140). If the sign were reversed, we would obtain the equations $-\varepsilon_k = 0$. If Tonti variational conditions (1.141) are not satisfied, it is possible to seek functions f_k such that a Lagrangian function exists for the equations

$$\tilde{\varepsilon}_k \equiv f_k \varepsilon_k = 0 \,. \tag{1.143}$$

In this context, k is not included in the sum; that is, we multiply each equation separately by some function f_k. If the invariance of the equations is due to terms that are linear in the generalized velocities $\mathrm{d}q_k/\mathrm{d}t$, we divide the equations into two parts such that

$$\varepsilon_k = \varepsilon_{0k}(q, \dot{q}, \ddot{q}) + \alpha_{kl}\dot{q}_k \,; \qquad \alpha_{kl} = \alpha_{lk} \,. \tag{1.144}$$

The dissipative processes occurring in the system can be described by the Rayleigh function

$$R = \frac{1}{2}\alpha_{kl}\dot{q}_k \dot{q}_l \,. \tag{1.145}$$

An inverse variational problem can be solved by finding the Lagrangian for the equations $\varepsilon_{0k} = 0$; the system then satisfies the Lagrangian equations of the form (1.133) with the Rayleigh dissipation function

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} = -\frac{\partial R}{\partial \dot{q}_k} \,. \tag{1.146}$$

◖ Example 1.27: Movement of the crossbar

A metal bar of mass *m* is placed across two horizontal electrodes. We apply a voltage U_0 from a battery to the electrodes. An electric current begins to flow through the bar, creating a magnetic field that starts to move the bar. We choose the position of the bar $x(t)$ and the charge $Q(t)$ that has flowed through the circuit since the beginning as the generalized variables.

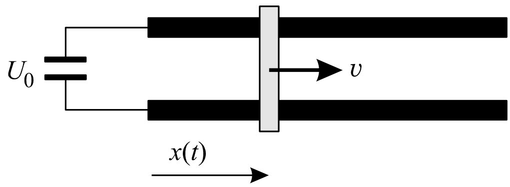


Fig. 1.23: Movement of the crossbar

Let's assume that (for example, from an experiment) we know the equations describing the problem:

$$m\ddot{x} = \frac{1}{2}\mathscr{L}_1\dot{Q}^2 - \beta\dot{x}\,; \tag{1.147}$$

$$(\mathscr{L}_0 + \mathscr{L}_1 x)\ddot{Q} + \mathscr{R}\dot{Q} + \frac{Q}{\mathscr{C}} + \mathscr{L}_1\dot{x}\dot{Q} = U_0\,. \tag{1.148}$$

From an electrical standpoint, this is a circuit consisting of a capacitor bank with capacitance $\mathscr{C}$, resistance $\mathscr{R}$, and variable inductance $\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_1 x$. The last term on the left side of the equation (1.148) describes the interconnection between the mechanical and electrical parts. From a mechanical standpoint, the crossbar is accelerated by the force $\mathscr{L}_1(\mathrm{d}Q/\mathrm{d}t)^2/2$ and decelerated by the frictional force $-\beta\,\mathrm{d}x/\mathrm{d}t$. Let us denote

$$\varepsilon_x \equiv m\ddot{x} - \frac{1}{2}\mathscr{L}_1\dot{Q}^2 + \beta\dot{x}\,; \tag{1.149}$$

$$\varepsilon_Q \equiv (\mathscr{L}_0 + \mathscr{L}_1 x)\ddot{Q} + \mathscr{R}\dot{Q} + \frac{Q}{\mathscr{C}} + \mathscr{L}_1\dot{x}\dot{Q} - U_0\,. \tag{1.150}$$

Tonti's variational conditions are satisfied only if $\beta = 0$ and $\mathscr{R} = 0$. This is quite natural, since these are the coefficients of the dissipative terms (circuit resistance and crossbar friction). We therefore introduce the Rayleigh dissipation function

$$R \equiv \frac{1}{2}\mathscr{R}\dot{Q}^2 \;+\; \frac{1}{2}\beta\dot{x}^2 \tag{1.151}$$

and we will seek the Lagrangian function only for the left-hand sides of the equations without dissipative terms

$$\varepsilon_{0x} \equiv m\ddot{x} - \frac{1}{2}\mathscr{L}_1\dot{Q}^2\,; \tag{1.152}$$

$$\varepsilon_{0Q} \equiv (\mathscr{L}_0 + \mathscr{L}_1 x)\ddot{Q} + \frac{Q}{\mathscr{C}} + \mathscr{L}_1\dot{x}\dot{Q} - U_0\,. \tag{1.153}$$

These equations already satisfy Tonti's conditions, and we can calculate the Lagrangian using the following formula

$$L = -x\int_0^1 \varepsilon_{0x}(\tau x, \tau Q, \tau \dot{x}, \tau \dot{Q}, \tau \ddot{x}, \tau \ddot{Q})\,\mathrm{d}\tau - Q\int_0^1 \varepsilon_{0Q}(\tau x, \tau Q, \tau \dot{x}, \tau \dot{Q}, \tau \ddot{x}, \tau \ddot{Q})\,\mathrm{d}\tau\,. \quad (1.154)$$

After substitution and a simple integration, we obtain the Lagrangian

$$L = -x\left(\frac{m_0\ddot{x}}{2} - \frac{\mathscr{L}_1\dot{Q}^2}{6}\right) - Q\left(\frac{\mathscr{L}_0\ddot{Q}}{2} + \frac{\mathscr{L}_1 x\ddot{Q}}{3} + \frac{\mathscr{L}_1\dot{x}\dot{Q}}{3} + \frac{Q}{2\mathscr{C}} - U_0\right). \quad (1.155)$$

Although the Lagrangian (1.155), together with the Rayleigh function (1.151), yields the correct initial equations, the Lagrangian is quite complicated. We will take advantage of the fact that the equations remain unchanged if we add (or subtract) the total derivative of any function with respect to time to (or from) the Lagrange function. We will modify the terms containing second derivatives as follows:

$$x\ddot{x} = \frac{\mathrm{d}}{\mathrm{d}t}(x\dot{x}) - \dot{x}^2\ ;$$

$$Q\ddot{Q} = \frac{\mathrm{d}}{\mathrm{d}t}(Q\dot{Q}) - \dot{Q}^2\ ;$$

$$Qx\ddot{Q} = \frac{\mathrm{d}}{\mathrm{d}t}(Qx\dot{Q}) - \dot{Q}\frac{\mathrm{d}}{\mathrm{d}t}(Qx)\,.$$

We substitute these expressions into the Lagrangian and omit the full derivatives, which do not affect the equations of motion. The result is now much clearer:

$$L = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}(\mathscr{L}_0 + \mathscr{L}_1 x)\dot{Q}^2 - \frac{Q^2}{2\mathscr{C}} + QU_0\,. \quad (1.156)$$

The interpretation of this Lagrangian is clear. The first term is the particle's kinetic energy, the second is the energy associated with inductance, the third is the energy associated with the system's capacitance (which represents potential energy, hence the negative sign), and the last term relates to the definition of the variable Q (the total charge that has flowed through the circuit since the beginning). If we chose the charge on the capacitor bank as the generalized variable, this term would be zero. The Lagrangian (1.156) together with the Rayleigh function (1.151) provide a natural solution to our problem. A more detailed solution for the case where the crossbar is formed by plasma in a rail accelerator can be found in [13].

◗

1.4.7 Adiabatic Invariants

Periodic motion

Let's imagine a system that moves periodically. Examples include a spring, a pendulum, the Earth orbiting the Sun, or an electron bouncing between magnetic mirrors. In phase space (where the axes represent generalized coordinates and momenta), such motion forms a closed curve along which the system moves over and over again. Let us assume that the motion is periodic in a certain generalized variable q, to which the canonically conjugate generalized momentum p corresponds.

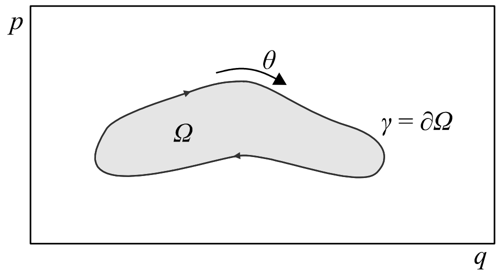


Fig. 1.24: Periodic motion in phase space

A phase trajectory encloses a region Ω in the (q, p) plane, of which it is the boundary. In mathematics, we denote the boundary of a region symbolically as follows: $\gamma = \partial\Omega$ (read as "the curve γ is the boundary of the set Ω"). We denote the area enclosed by the phase trajectory as J; its numerical value can be easily determined as an integral

► $$J = \oint_\gamma p\, \mathrm{d}q\,. \tag{1.157}$$

If the phase trajectory were oriented in the opposite direction from that shown in the figure, we would obtain an area enclosed by a phase trajectory with a negative sign. In periodic motion, the area remains constant. Similarly, the energy is conserved

$$H = p\dot{q} - L\,. \tag{1.158}$$

The Lagrangian can be expressed as

$$L = p\dot{q}\,. \tag{1.159}$$

In the expression, we have omitted the movement constant H, which does not alter the equations of motion. The action integral calculated over the period is

$$\begin{aligned} S &= \int_t^{t+T} L\,\mathrm{d}t = \int_t^{t+T} p\dot{q}\,\mathrm{d}t = \\ &= \int_t^{t+T} p\frac{\mathrm{d}q}{\mathrm{d}t}\mathrm{d}t = \oint_\gamma p\,\mathrm{d}q = J \end{aligned}\,. \tag{1.160}$$

We can thus interpret the area J enclosed by the phase trajectory as the work done by the system over one period. The quantities J and H remain constant during periodic motion. If we observe a system with higher energy (a pendulum with greater amplitude), both J and H will increase. The two quantities are dependent on each other,

$$J = J(H); \qquad H = H(J) . \tag{1.161}$$

We have already identified two meanings of the quantity J: it is the area enclosed by the phase trajectory and the action integral over a period. However, the quantity J has another important meaning. We can interpret it as a kind of generalized momentum of the system. We denote the corresponding canonically conjugate coordinate by θ. The new coordinate θ must satisfy Hamilton equation

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{\partial H}{\partial J} . \tag{1.162}$$

Since $\partial H/\partial J$ is a constant for a system with a given energy (it can be shown that it is equal to the reciprocal of the period of motion), the new generalized variable θ increases linearly with time. It can thus symbolize, for example, the increasing angle as the system orbits along the phase trajectory. We can easily parameterize the phase trajectory (the boundary of the set Ω) using this variable. For the time derivatives of any quantity on the boundary $\gamma = \partial\Omega$, we can therefore write (θ is the only parameter that uniquely determines the position on the boundary)

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial \theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{\partial f}{\partial \theta}\frac{\partial H}{\partial J} . \tag{1.163}$$

We can see that the new canonically coupled variables J and θ are extremely useful. The momentum J is an integral of motion that describes the area enclosed by a phase trajectory, and the coordinate θ allows us to uniquely parameterize the boundary of this area (the phase trajectory) and convert the total time derivatives on the boundary into partial derivatives.

Adiabatic approximation

Let us now assume that some parameter λ changes slowly during periodic motion. This could be the length of a pendulum's string or the magnetic field surrounding an electron orbiting in a circle around magnetic field lines. Over the course of one period, the change is insignificant, i.e., the following relationship holds

► $$\frac{\mathrm{d}\lambda}{\mathrm{d}t} \ll \frac{\lambda}{T} ; \tag{1.164}$$

However, over many time periods, the change in conditions can be significant. We refer to such changes as *adiabatic changes*. In this case, energy is not conserved. The change in energy is proportional to the change in the parameter λ. Both changing quantities can be combined into a new variable that remains constant even during adiabatic changes; that is, there exists a quantity $A(H, \lambda)$ such that

$$A(H,\lambda) = \text{const} . \tag{1.165}$$

We call such a quantity an *adiabatic invariant*. Let us now prove that the area of the phase trajectory is an adiabatic invariant, i.e., the generalized momentum J given by the relation (1.157). Unlike energy, this quantity is conserved during adiabatic changes.

Let us therefore find the time derivative of J, using the parameterization of the curve by the canonical conjugate variable θ for the generalized momentum J:

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\oint_\gamma p\mathrm{d}q = \frac{\mathrm{d}}{\mathrm{d}t}\int_0^{2\pi} p\frac{\partial q}{\partial \theta}\mathrm{d}\theta =$$

$$= \int_0^{2\pi}\left(\frac{\mathrm{d}p}{\mathrm{d}t}\frac{\partial q}{\partial \theta} + p\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial q}{\partial \theta}\right)\mathrm{d}\theta .$$

We substitute the time derivatives at the boundary of the region according to equation (1.163), i.e.

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \int_0^{2\pi}\frac{\partial H}{\partial J}\left(\frac{\partial p}{\partial \theta}\frac{\partial q}{\partial \theta} + p\frac{\partial}{\partial \theta}\frac{\partial q}{\partial \theta}\right)\mathrm{d}\theta .$$

In the second term, we will perform integration per partes

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \int_0^{2\pi}\frac{\partial H}{\partial J}\left(\frac{\partial p}{\partial \theta}\frac{\partial q}{\partial \theta} - \frac{\partial p}{\partial \theta}\frac{\partial q}{\partial \theta}\right)\mathrm{d}\theta + \left[p\frac{\partial q}{\partial \theta}\right]_0^{2\pi} = 0 .$$

The first term on the right-hand side is clearly zero; we can consider the second term to be zero if the adiabatic approximation holds. We have thus found an adiabatic invariant for which, under the adiabatic approximation, the following holds:

► $$J = \oint p\mathrm{d}q = \text{const} . \qquad (1.166)$$

◖ **Example 1.28: Harmonic oscillator**

Suppose that, for some reason, the frequency of a harmonic oscillator changes slowly in an adiabatic manner (for example, by changing the length of the suspension of a mathematical pendulum or the stiffness of the spring on which the body swings). The phase trajectory of the oscillator is an ellipse defined by the equation

$$\frac{1}{2}m\omega^2 x^2 + \frac{p^2}{2m} = E . \qquad (1.167)$$

We can determine the semi-axes of our ellipse from its segmented form.

$$\left(\frac{x}{a}\right)^2 + \left(\frac{p}{b}\right)^2 = 1 ; \qquad a = \sqrt{\frac{2E}{m\omega^2}} ; \qquad b = \sqrt{2mE} . \qquad (1.168)$$

We can now easily define our adiabatic invariant as the area of an ellipse with semi-axes a and b:

$$J = \oint p\mathrm{d}x = \pi ab = \frac{2\pi E}{\omega} = \text{const}. \qquad (1.169)$$

Although both energy and frequency change slowly over time, their ratio remains constant even after many cycles. The concept of the adiabatic pendulum was proposed by Albert Einstein as early as 1911. ◗

◖ **Example 1.29: Variable magnetic field**

Suppose that an electron undergoes Larmor precession (see Example 1.23) in a slowly varying magnetic field. We choose the angle φ as the coordinate (determines the position on the circle). The relevant generalized momentum is the angular momentum:

$$J = \int_0^{2\pi} p_\varphi \mathrm{d}\varphi = \int_0^{2\pi} m_\mathrm{e} v R_L \mathrm{d}\varphi = 2\pi m_\mathrm{e} v R_\mathrm{L} .$$

Substituting the Larmor radius from equation (1.63), we have

$$J = 2\pi \frac{m_\mathrm{e}^2 v^2}{QB} \approx \frac{v^2}{B} .$$

As the particle moves, the magnetic field changes slowly. As a result, the particle's velocity changes in such a way that the ratio v^2/B remains constant. This is known as the first adiabatic invariant in a magnetic field. Typically, a combination with slightly different constants is chosen (though these constants are not important).

$$J_1 = \frac{m_\mathrm{e} v^2}{2B} = \mathrm{const} . \tag{1.170}$$

For more details on adiabatic invariants for a charged particle moving in a magnetic field, see companion textbook [2]. A more general overview is provided in [11]. ◗

1.4.8 Canonical Transformations

We do not always succeed in choosing the optimal generalized coordinates. From any set of generalized coordinates and momenta (q, p), we can always transition to another system of new coordinates and momenta (Q, P). If we require that the new variables also be canonically conjugate with one another, i.e., that

► $$\begin{aligned} \{p_k, q_l\} &= \delta_{kl} ; \\ \{P_k, Q_l\} &= \delta_{kl} . \end{aligned} \tag{1.171}$$

We are talking about the so-called *canonical transformation*. If we want to automatically ensure that the new set of variables (Q, P) is canonical, we can use a simple mechanism of *generating functions* to generate it, which we will now describe. Suppose we want to transition from

► $$\begin{aligned} q_1, q_2, \ldots, q_f, p_1, p_2, \ldots, p_f \quad &\to \quad Q_1, Q_2, \ldots, Q_f, P_1, P_2, \ldots, P_f , \\ H(t, q, p) \quad &\to \quad \bar{H}(t, Q, P) , \end{aligned} \tag{1.172}$$

where we have denoted $\bar{H}$ as the Hamiltonian in the new variables. Of course, we want the Hamilton equations to hold for both the new and old coordinates, i.e., we want Hamilton principle – from which they were derived – to hold:

$$\delta \int_{t_A}^{t_B} L \mathrm{d}t = \delta \int_{t_A}^{t_B} (p_k \dot{q}_k - H) \, \mathrm{d}t = 0 ; \qquad \wedge \qquad \delta \int_{t_A}^{t_B} L \mathrm{d}t = \delta \int_{t_A}^{t_B} (P_k \dot{Q}_k - \bar{H}) \, \mathrm{d}t = 0 .$$

For both equations to hold simultaneously, the integrands may differ by the total derivative of any function of the old and new variables $\mathcal{V}(t, q, Q)$:

$$p_k \dot{q}_k - H = P_k \dot{Q}_k - \bar{H} + \frac{\mathrm{d}\mathcal{V}}{\mathrm{d}t}; \qquad \mathcal{V} = \mathcal{V}(t, q, Q). \tag{1.173}$$

This is because the changes in both the old and new coordinates are zero at the ends of the trajectory:

$$\delta \int_{t_A}^{t_B} \frac{\mathrm{d}\mathcal{V}}{\mathrm{d}t} \mathrm{d}t = [\delta \mathcal{V}]_{t_A}^{t_B} = \left[\frac{\partial \mathcal{V}}{\partial q_k} \delta q_k + \frac{\partial \mathcal{V}}{\partial Q_k} \delta Q_k \right]_{t_A}^{t_B} = 0.$$

Let us now take the derivative of the function $\mathcal{V}(t, q, Q)$ in equation (1.173)

$$p_k \dot{q}_k - H = P_k \dot{Q}_k - \bar{H} + \frac{\partial \mathcal{V}}{\partial t} + \frac{\partial \mathcal{V}}{\partial q_k} \dot{q}_k + \frac{\partial \mathcal{V}}{\partial Q_k} \dot{Q}_k \qquad \Rightarrow$$

$$\left(p_k - \frac{\partial \mathcal{V}}{\partial q_k} \right) \dot{q}_k + \left(\bar{H} - H - \frac{\partial \mathcal{V}}{\partial t} \right) - \left(P_k + \frac{\partial \mathcal{V}}{\partial Q_k} \right) \dot{Q}_k = 0.$$

We can satisfy this equality with simple requirements:

► $$p_k = \frac{\partial \mathcal{V}}{\partial q_k}; \qquad P_k = -\frac{\partial \mathcal{V}}{\partial Q_k}; \qquad \bar{H} = H + \frac{\partial \mathcal{V}}{\partial t}. \tag{1.174}$$

The function $\mathcal{V}$ is called the *generating function*. It is an arbitrary function of time, the old variables q, and the new variables Q, i.e., $\mathcal{V}(t, q, Q)$. If relations (1.174) hold, we calculate the new momenta according to this formula, Hamilton equations will hold in the new variables, and furthermore, the new variables created in this way will be canonically coupled, i.e., $\{Q_k, P_l\} = \delta_{kl}$.

Hamilton-Jacobi equation

Using a generating function, we can create various transformations to new variables. A generating function is, in fact, any function of the old and new coordinates. In principle, it could also be a function of old coordinates and new momenta, old momenta and new coordinates, or old momenta and new momenta. The differences between coordinates and momenta are blurred in Hamilton theory. The transformation relations (1.174) would be only slightly different.

Naturally, the question arises as to how to choose the generating function so that the solution in the new variables is as simple as possible. We can even require that the solution in the new coordinates and momenta is constant, i.e., that the right-hand side of Hamilton equations is zero:

$$Q_k = \text{const} \quad \Rightarrow \quad \dot{Q}_k = +\frac{\partial \bar{H}}{\partial P_k} = 0; \qquad P_k = \text{const} \quad \Rightarrow \quad \dot{P}_k = -\frac{\partial \bar{H}}{\partial Q_k} = 0. \tag{1.175}$$

In this case, the new generalized coordinates and momenta will be cyclic, i.e., they will not appear in the Hamiltonian. We can require that the Hamiltonian in the new variables be exactly zero. We denote the generating function that leads to constant generalized coordinates and momenta by S and call it the *principal Hamiltonian*. The transformation equations (1.174) will now be:

$$p_k = \frac{\partial S}{\partial q_k};$$

► $$P_k = -\frac{\partial S}{\partial Q_k}; \qquad (1.176)$$

$$0 = H(t, q_k, p_k) + \frac{\partial S}{\partial t}.$$

We substitute the first equation into the Hamiltonian in the third equation. In the second equation, we use the fact that the new coordinates $Q_k = \alpha_k$ and momenta $P_k = \beta_k$ are constants:

► $$\frac{\partial S}{\partial t} + H\left(t, q_k, \frac{\partial S}{\partial q_k}\right) = 0 \qquad \Rightarrow \qquad S(t, q_k, \alpha_k) \qquad (1.177)$$

► $$\beta_k = -\frac{\partial S(t, q_k, \alpha_k)}{\partial \alpha_k} \qquad \Rightarrow \qquad q_k = q_k(t, \alpha_k, \beta_k); \qquad (1.178)$$

Equation (1.177) is known as the Hamilton–Jacobi equation [12]. The generating function S, which leads to constant generalized coordinates and momenta, has the form of an action integral:

$$S = S(t, q_k, \alpha_k) \quad \Rightarrow \quad \frac{\mathrm{d}S}{\mathrm{d}t} = \frac{\partial S}{\partial t} + \frac{\partial S}{\partial q_k}\dot{q}_k = -H + p_k \dot{q}_k = L \quad \Rightarrow \quad S = \int L \,\mathrm{d}t .$$

The procedure is as follows:

1) Using arbitrary generalized coordinates and momenta, we construct the Hamiltonian (i.e., the energy expressed in terms of generalized coordinates and momenta).
2) In this Hamiltonian, we replace all occurrences of the momentum p_k with the expression $\partial S/\partial q_k$ and construct the Hamilton-Jacobi equation (1.177) for the generating function S.
3) We are solving the Hamilton-Jacobi equation. The solution is a generating function $S(t, q_k, \alpha_k)$, where α_k are integration constants. These constants serve as new generalized coordinates.
4) The transformation (1.178), given by S, leads to new momenta that are also constant; therefore, we denote them by β_k. These transformation relations contain new generalized coordinates α_k (constants), new generalized momenta β_k (constants), and time. Therefore, we can express the original coordinates q_k as functions of time and the found constants. This brings us back to solving the problem in the original coordinates.

◖ **Example 1.30: Free fall**

Solving for free fall from a height of h using the Hamilton-Jacobi equation is as absurd as hunting sparrows with an artillery cannon. However, as an exercise to familiarize yourself with the Hamilton-Jacobi equation, this approach is useful. Let's assume that the y-axis points upward. The Lagrangian and the Hamiltonian will have the simple form (Step 1)

$$L = \frac{1}{2}mv^2 - mgy\,; \qquad H = \frac{p^2}{2m} + mgy\,. \tag{1.179}$$

Now we will set up the Hamilton-Jacobi equation (Step 2):

$$\frac{\partial S}{\partial t} + \frac{1}{2m}\left(\frac{\partial S}{\partial y}\right)^2 + mgy = 0\,. \tag{1.180}$$

This is a partial differential equation that we will solve by separation: $S(t, y) = S_1(t) + S_2(y)$. Since time appears only in the first term, $S_1(t)$ must be linear in time, i.e., for example, $S_1 = -\alpha t$. Substituting this into the Hamilton-Jacobi equation, we obtain

$$-\alpha + \frac{1}{2m}\left(\frac{\mathrm{d}S_2}{\mathrm{d}y}\right)^2 + mgy = 0 \qquad \Rightarrow \qquad S_2(y) = \int \sqrt{2m(\alpha - mgy)}\,\mathrm{d}y \qquad \Rightarrow$$

$$S(t, y, \alpha) = -\alpha t - \frac{(2m\alpha - 2m^2 gy)^{3/2}}{3m^2 g}\,. \tag{1.181}$$

So we have found S, which is a function of time, the old generalized variable y, and the new generalized variable $Q = \alpha$, which is constant. This is the result of the third step. Now we will perform the final step—the transformation to the new momentum $P = \beta$, which is also constant:

$$\beta = -\frac{\partial S(t, y, \alpha)}{\partial \alpha} \qquad \Rightarrow \qquad \beta = t + \sqrt{\frac{2\alpha}{mg^2} - \frac{2y}{g}} \qquad \Rightarrow \qquad y = \frac{\alpha}{mg} - \frac{1}{2}g(t-\beta)^2\,.$$

From the initial conditions $y(t_0) = h$ a $y'(t_0) = 0$, we obtain $\alpha/mg = h$, $\beta = t_0$, that is

$$y = h - \frac{1}{2}g(t - t_0)^2\,. \tag{1.182}$$

◗

1.5 Nonlinear Dynamical Systems

Hamilton equations describing mechanical systems lead to a system of first-order differential equations for the variables **q** and **p**. Let $\boldsymbol{\xi} = (\mathbf{q}, \mathbf{p})$ denote the set of phase variables of the system. The differential equations derived from Hamilton equations then take the form:

$$\begin{aligned} \dot{\xi}_1 &= f_1(t, \xi_1, \xi_2, \ldots, \xi_N)\,; \\ \dot{\xi}_2 &= f_2(t, \xi_1, \xi_2, \ldots, \xi_N)\,; \\ &\vdots \\ \dot{\xi}_N &= f_N(t, \xi_1, \xi_2, \ldots, \xi_N)\,, \end{aligned} \tag{1.183}$$

in other words

$$\dot{\xi}_k = f_k(t, \boldsymbol{\xi})\,, \qquad k = 1, \ldots, N\,. \tag{1.184}$$

The number of equations N need not necessarily be even (coordinates and their corresponding momenta); we omit the equations for conserved variables from the system and do not solve them. Time is usually not explicitly included on the right-hand sides—such systems of equations are called *autonomous*. In the following text, we will deal only with autonomous systems of equations

$$\dot{\xi}_k = f_k(\boldsymbol{\xi})\,, \qquad k = 1, \ldots, N\,. \tag{1.185}$$

The simplest case is that of linear equations of the form

$$\begin{aligned} \dot{\xi}_1 &= a_{11}\xi_1 + \cdots, a_{1N}\xi_N\,, \\ &\vdots \\ \dot{\xi}_N &= a_{N1}\xi_1 + \cdots, a_{NN}\xi_N\,, \end{aligned} \tag{1.186}$$

in other words

$$\dot{\boldsymbol{\xi}} = \mathbf{A}\boldsymbol{\xi}\,. \tag{1.187}$$

Solution is simple. We find the eigenvalues and eigenvectors of matrix $\mathbf{A}$:

$$\mathbf{A}\boldsymbol{\eta}^{(l)} = \lambda_l\,\boldsymbol{\eta}^{(l)}\,. \tag{1.188}$$

Rearranging equation (1.188) gives $(\mathbf{A} - \lambda\mathbf{1})\boldsymbol{\eta} = 0$. This equation will have a non-trivial solution only if

$$\det(\mathbf{A} - \lambda\,\mathbf{1}) = 0\,, \tag{1.189}$$

which is the equation for the eigenvalues λ. From (1.188) we can then calculate the eigenvectors. The solution to the system of linear differential equations is every expression

$$\boldsymbol{\xi} = \boldsymbol{\eta}\,\exp(\lambda t)\,,$$

because

$$\frac{\mathrm{d}\boldsymbol{\xi}}{\mathrm{d}t} = \lambda \boldsymbol{\eta} \exp(\lambda t) = \mathbf{A}\boldsymbol{\eta} \exp(\lambda t) = \mathbf{A}\boldsymbol{\xi} .$$

The solution is a linear combination of the solutions for the individual eigenvalues:

$$\boldsymbol{\xi}(t) = c_1 \boldsymbol{\eta}^{(1)} \mathrm{e}^{\lambda_1 t} + c_2 \boldsymbol{\eta}^{(2)} \mathrm{e}^{\lambda_2 t} + \cdots . \tag{1.190}$$

When dealing with oscillations, the eigenvalues λ are complex ($\lambda_k = \delta + \mathrm{i}\,\omega_k$). The individual terms in the sum (1.190) are known as the eigenmodes of the oscillations. The number of eigenfrequencies is less than or equal to the order of matrix **A**.

> **Note:** The result (1.190) holds only if the eigenvalues of matrix **A**, determined from equation (1.189), are all distinct. If any eigenvalue is a k-fold root of equation (1.189), then the corresponding coefficient of the linear combination (1.190) will be a polynomial of degree $k - 1$.

◖ **Example 1.31: Harmonic oscillator**

Hamilton equations for a harmonic oscillator are given by

$$\dot{x} = \frac{p}{m} ;$$
$$\dot{p} = -m\omega^2 x .$$

Leaving aside the irrelevant constants, we need to solve a system of equations

$$\begin{aligned} \dot{\xi}_1 &= \xi_2 \\ \dot{\xi}_2 &= -\xi_1 \end{aligned} \qquad \Rightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix},$$

where $\xi_1 = x$ and $\xi_2 = p$. From the equation for eigenvalues (1.189), we can easily determine the eigenvalues $\lambda_{1,2} = \pm\mathrm{i}$, and from the equation for eigenvectors (1.188) , the corresponding eigenvectors

$$\boldsymbol{\eta}^{(1)} = c \cdot \begin{pmatrix} 1 \\ +\mathrm{i} \end{pmatrix};$$
$$\boldsymbol{\eta}^{(2)} = c \cdot \begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix}.$$

The general solution to the system is therefore

$$\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = c_1 \cdot \begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix} \mathrm{e}^{\mathrm{i}t} + c_2 \cdot \begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix} \mathrm{e}^{-\mathrm{i}t} ,$$

which, for the initial conditions $x(0) = \xi_1(0) = A;\ p(0) = \xi_2(0) = 0,$ yields the known solution

$$x = \xi_1 = A\cos t ;$$
$$p = \xi_2 = -A\sin t .$$

◗

1.5.1 Stability Matrix and Phase Portrait of the System

If the system of differential equations is nonlinear, the solution may be much more complicated than the result in (1.190).

Equilibrium points of the solution

These are points in phase space from which the system does not evolve spontaneously. They are defined by the relation $\xi_k/\mathrm{d}t = 0$. We find them by setting the right-hand sides of the system of differential equations (1.183) equal to zero:

► $$f_k(\boldsymbol{\xi}) = 0\,; \qquad \text{Equations for stationary points.} \tag{1.191}$$

Note: If we “place! the system exactly at a equilibrium point (i.e., set it up with such initial conditions), it will remain at that point in phase space once and for all.

Solution stability

We will examine whether equilibrium points are stable with respect to small disturbances (perturbations). We can imagine slightly displacing a system placed at a stationary point and observing whether it spontaneously returns to the stationary point (stable point) or moves away from it (unstable point). Let us therefore seek a solution to the system of equations (1.183) in the form

$$\xi_k = \xi_k^{(\mathrm{S})} + \delta\xi_k\,, \qquad k = 1,\ldots,N\,,$$

where $\boldsymbol{\xi}^{(\mathrm{S})}$ is a equilibrium point satisfying $f_k(\boldsymbol{\xi}^{(\mathrm{S})}) = 0$; $\delta\boldsymbol{\xi}$ is a small first-order perturbation. We substitute this form into the original system of equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\xi_k^{(\mathrm{S})} + \delta\xi_k\right) = f_k(\boldsymbol{\xi}^{(\mathrm{S})} + \delta\boldsymbol{\xi})$$

and we will perform a first-order Taylor expansion of the right-hand side

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\xi_k^{(\mathrm{S})}\right) + \frac{\mathrm{d}}{\mathrm{d}t}\left(\delta\xi_k\right) = f_k(\boldsymbol{\xi}^{(\mathrm{S})}) + \left.\frac{\partial f_k}{\partial \xi_l}\right|_{\boldsymbol{\xi}^{(S)}} \cdot \delta\xi_l\,.$$

Given the stationarity of $\boldsymbol{\xi}^{(\mathrm{S})}$, the first terms on both sides cancel out, and we can write

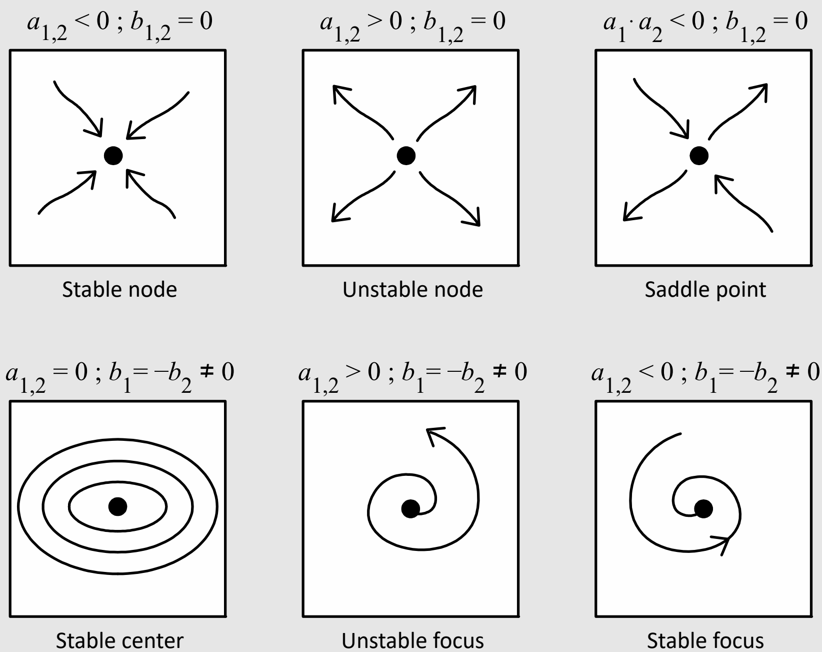
► $$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \delta\xi_1 \\ \vdots \\ \delta\xi_N \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix} \cdot \begin{pmatrix} \delta\xi_1 \\ \vdots \\ \delta\xi_N \end{pmatrix}, \tag{1.192}$$

where we have denoted

► $$a_{kl} \equiv \left. \frac{\partial f_k}{\partial \xi_l} \right|_{\boldsymbol{\xi}^{(S)}} \tag{1.193}$$

is the so-called *stability matrix*. It consists of the partial derivatives of the right-hand sides of equations (1.183) with respect to the individual variables at the stationary point under consideration. The system of equations (1.192) for small perturbations $\delta\boldsymbol{\xi}$ is linearized, and therefore we can find its solution using the eigenvalues and eigenvectors of matrix $\mathbf{A}$. If $\mathrm{Re}(\lambda) < 0$, the given mode $\exp(\lambda t)$ will be damped, and the solution is stable in the corresponding eigenvector direction. If $\mathrm{Re}(\lambda) > 0$, the mode is unstable in the given direction. If $\lambda = \pm \mathrm{i}\, b$, a small perturbation will cause the system to oscillate in the vicinity of the stationary point.

Note 1: For a system of two differential equations, the 2×2 stability matrix will have two eigenvalues, $\lambda_1 = a_1 + \mathrm{i}\, b_1$ and $\lambda_2 = a_2 + \mathrm{i}\, b_2$, and the following situations are possible:·



Note 2: Based on our knowledge of the equilibrium points, eigenvalues, and eigenvectors of the stability matrix, we are usually able to estimate the phase portrait of the system. Examples are provided below.

◖ Example 1.32: Nonlinear oscillator

Let's consider the following system of equations

$$\dot{\xi}_1 = \xi_2$$
$$\dot{\xi}_2 = -\xi_1 + \varepsilon \xi_1^2 \,.$$

In contrast to a standard harmonic oscillator, there is an additional nonlinear term with coefficient ε. First, we determine the equilibrium points A and B from the condition that the right-hand sides are zero:

$$\begin{array}{l} \xi_2 = 0 \\ -\xi_1 + \varepsilon \xi_1^2 = 0 \end{array} \quad \Rightarrow \quad \begin{array}{l} A = [0,0] \\ B = [1/\varepsilon, 0] \end{array}$$

and plot them in phase space. Then we find the stability matrix in general form:

$$\mathbf{A} = \begin{pmatrix} \dfrac{\partial f_1}{\partial \xi_1} & \dfrac{\partial f_1}{\partial \xi_2} \\ \dfrac{\partial f_2}{\partial \xi_1} & \dfrac{\partial f_2}{\partial \xi_2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1+2\varepsilon\xi_1 & 0 \end{pmatrix}.$$

We will determine this matrix at the equilibrium points A and B. We will calculate the eigenvalues and eigenvectors from equations (1.189) and (1.188):

$$A: \quad \mathbf{A} = \begin{pmatrix} 0 & +1 \\ -1 & 0 \end{pmatrix} \quad \Rightarrow \quad \lambda_{1,2} = \pm \mathrm{i} \quad \Rightarrow \quad \text{perturbation } \mathrm{e}^{\pm \mathrm{i} t}$$

$$B: \quad \mathbf{A} = \begin{pmatrix} 0 & +1 \\ +1 & 0 \end{pmatrix} \quad \Rightarrow \quad \begin{cases} \lambda_1 = +1, \quad \boldsymbol{\eta}_1 = c\begin{pmatrix} +1 \\ +1 \end{pmatrix}, \quad \text{perturbation } \mathrm{e}^{+t} \\ \lambda_2 = -1, \quad \boldsymbol{\eta}_2 = c\begin{pmatrix} +1 \\ -1 \end{pmatrix}, \quad \text{perturbation } \mathrm{e}^{-t}\,. \end{cases}$$

We plot the identified stability modes and their corresponding eigenvectors in phase space:

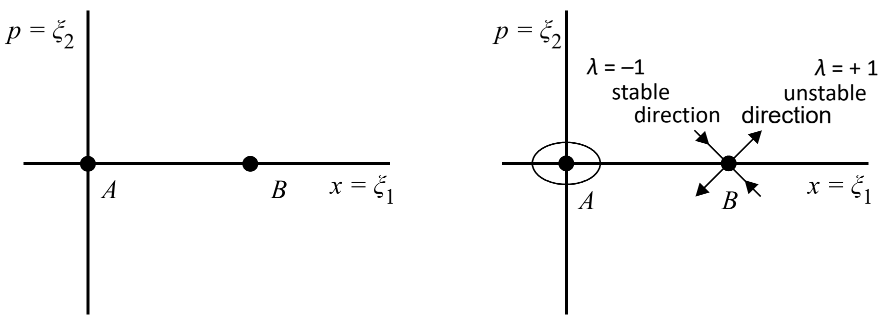


Fig. 1.26: Step-by-step creation of a phase portrait of a nonlinear oscillator

From the equilibrium points, their stability types, and eigenvectors, it is usually possible to estimate the entire phase portrait of the system:

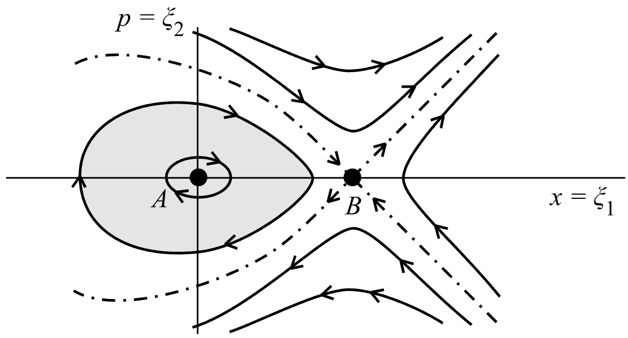


Fig. 1.27: Phase portrait of a nonlinear oscillator

◗

◖ Example 1.33: Particle in a periodic potential

A typical example of motion in a periodic potential is the motion of a charged particle in a crystal lattice. Suppose the particle is in a potential energy field given by the equation

$$V(x) = -V_0 \cos\frac{2\pi x}{a}.$$

The negative sign in the equation ensures that the potential has a minimum at the origin; this negative sign is not essential for solving the problem. The period of the potential is a and the amplitude is V_0. It is clear that particles with total energy $E < V_0$ can be trapped in the potential energy minima (oscillating), and particles with energy $E > V_0$ can move freely. The corresponding Hamilton equations will be:

$$H = \frac{p^2}{2m} - V_0 \cos\frac{2\pi x}{a} \qquad \Rightarrow \qquad \begin{aligned} \dot{x} &= \{x, H\} = \frac{\partial H}{\partial p} = \frac{p}{m}; \\ \dot{p} &= \{p, H\} = -\frac{\partial H}{\partial x} = -\frac{2\pi V_0}{a}\sin\frac{2\pi x}{a}. \end{aligned}$$

As in the first example, we will ignore the insignificant constants (which are determined by the choice of units and coordinates) and solve a system of equations of the form

$$\dot{\xi}_1 = \xi_2,$$
$$\dot{\xi}_2 = -\sin\xi_1.$$

In the vicinity of the origin, if we replace the "sinus" function with its argument, this equation would describe a harmonic oscillator (the potential energy reaches its minimum at the origin). In general, this equation is nonlinear due to the "sinus" function. We will proceed as in the previous example. We will determine the steady-state points, find the stability matrix at those points, determine the eigenvalues and eigenvectors, and reconstruct the phase portrait of the system:

Equilibrium points:

$$\begin{matrix} \xi_2 = 0 \\ \sin\xi_1 = 0 \end{matrix} \quad \Rightarrow \quad A_k = [k\pi, 0]; \qquad k = 0, \pm 1, \pm 2, \cdots$$

Stability matrix:

$$\mathbf{A} = \begin{pmatrix} \dfrac{\partial f_1}{\partial \xi_1} & \dfrac{\partial f_1}{\partial \xi_2} \\ \dfrac{\partial f_2}{\partial \xi_1} & \dfrac{\partial f_2}{\partial \xi_2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\cos\xi_1 & 0 \end{pmatrix}.$$

For even values of k:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \Rightarrow \quad \lambda_{1,2} = \pm \mathrm{i} \quad \Rightarrow \quad \text{porucha } \mathrm{e}^{\pm \mathrm{i} t}.$$

For odd values of k:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \Rightarrow \quad \begin{cases} \lambda_1 = +1, \quad \boldsymbol{\eta}_1 = c\begin{pmatrix} +1 \\ +1 \end{pmatrix}, \quad \text{porucha } \mathrm{e}^{+t}, \\ \lambda_2 = -1, \quad \boldsymbol{\eta}_2 = c\begin{pmatrix} +1 \\ -1 \end{pmatrix}, \quad \text{porucha } \mathrm{e}^{-t}. \end{cases}$$

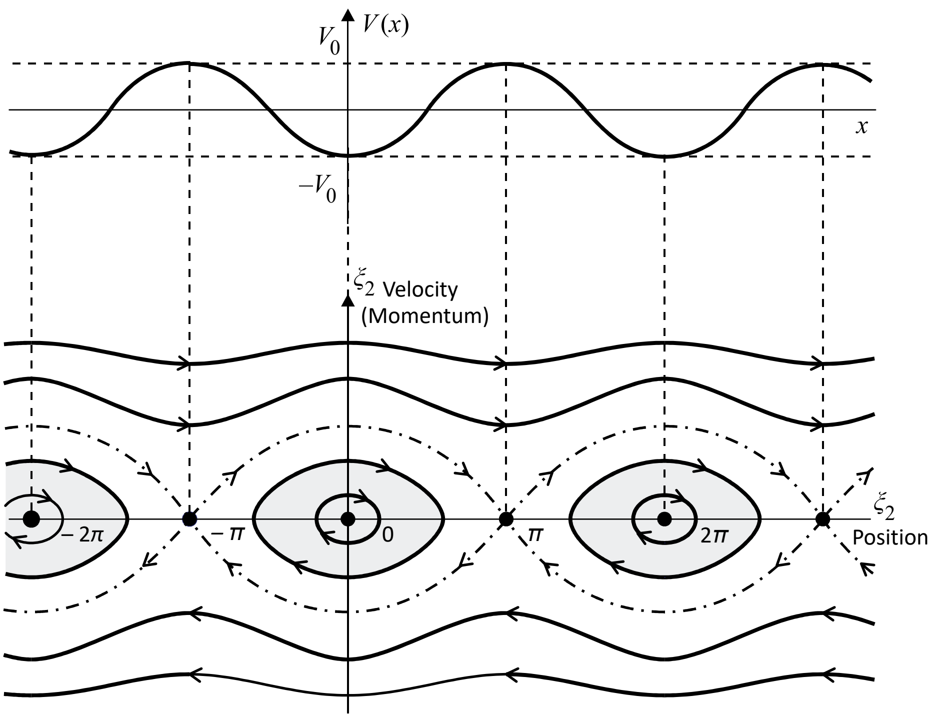


Fig. 1.28: Phase portrait of a particle in a periodic potential

Low-energy particles oscillate at the potential energy minima (they are trapped). Higher-energy particles move either in the positive direction of the *x*-axis (the top two trajectories) or in the negative direction of the *x*-axis (the bottom two trajectories). The higher the particle's velocity, the less its motion is affected by the periodic potential. The curve separating the different types of trajectories (shown as a dashed line in the previous examples) is called the *separatris*. ◗

1.5.2 Potential Method

The stability problem can be solved in ways other than by calculating the stability matrix. In some cases, we can find the so-called potential of the system. This is a function $\phi\,(\xi_1, \ldots \xi_N)$ at whose maxima the system is unstable (analogous to a ball at the top of a hill) and at whose minima the system is stable (analogous to a ball in a valley) If we know the potential $\phi\,(\xi_1, \ldots \xi_N)$, we can visualize this function as the elevation ϕ of the terrain above the space $\phi\,(\xi_1, \ldots \xi_N)$. Hills, valleys, saddles, and other features of this terrain correspond to the same types of stability that a ball placed at a given point on the terrain would have in a gravitational field.

In the one-dimensional case, we have a single differential equation

$$\dot{\xi} = f(\xi)\,. \tag{1.194}$$

Using the procedure from the previous chapter, we would first determine the steady-state points from the equation $f(\xi) = 0$, then the single-element stability matrix $a = \mathrm{d}f/\mathrm{d}\xi$ and its value at the found steady-state points. For $a > 0$, the system is unstable, and for $a < 0$, the system is stable (the disturbance is given by the exponential function e^{at}).

Definition of potential

We call the quantity in equation (1.194) the potential

► $$\phi(\xi) \equiv -\int f(\xi)\,\mathrm{d}\xi\,. \tag{1.195}$$

From the definition itself, we can easily show that the following holds

ϕ has extreme	$\Rightarrow$	$\mathrm{d}\phi/\mathrm{d}\xi = 0$	$\Rightarrow$	$f(\xi) = 0$	$\Rightarrow$	equilibrium point
ϕ has maximum	$\Rightarrow$	$\mathrm{d}^2\phi/\mathrm{d}\xi^2 < 0$	$\Rightarrow$	$a = \mathrm{d}f/\mathrm{d}\xi > 0$	$\Rightarrow$	unstable,
ϕ has minimum	$\Rightarrow$	$\mathrm{d}^2\phi/\mathrm{d}\xi^2 > 0$	$\Rightarrow$	$a = \mathrm{d}f/\mathrm{d}\xi < 0$	$\Rightarrow$	stable

In the multidimensional case, the procedure for system (1.183) is similar. We define the differential form

$$\mathrm{d}\phi \equiv -f_1(\boldsymbol{\xi})\,\mathrm{d}\xi_1 - f_2(\boldsymbol{\xi})\,\mathrm{d}\xi_2 - \cdots - f_N(\boldsymbol{\xi})\,\mathrm{d}\xi_N \tag{1.196}$$

and we seek a potential ϕ such that $f_k = -\partial\phi/\partial\xi_k$. Equation (1.196) is then the total differential of the function ϕ. If the differential form (1.196) is not integrable, we can seek an integration factor $\mu(\boldsymbol{\xi})$ such that the form is integrable

$$\mathrm{d}\phi \equiv -f_1\,\mu\,\mathrm{d}\xi_1 - f_2\,\mu\,\mathrm{d}\xi_2 - \cdots - f_N\mu\,\mathrm{d}\xi_N\,.$$

For $N = 1, 2$, an integration factor always exists. From the form of the resulting function ϕ, we can easily determine the stability of the system. For comparison, the following example is solved using both the stability matrix and the potential method.

◖ Example 1.34: Mexican hat potential

Let us now consider the properties of the following differential equation with the variable parameter ε. The goal is to find the equilibrium points and the general properties of the solutions:

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = \varepsilon\xi - \delta\xi^3 ; \qquad \delta > 0 ; \qquad \xi \in R . \tag{1.197}$$

The analysis using the stability matrix separately for negative and positive ε gives:

$\varepsilon < 0$: equilibrium point A: $\xi_S = 0$;

stability matrix: $a = \varepsilon - 3\delta\xi_S^2 = \varepsilon < 0 \quad \Rightarrow$

point A je stable

$\varepsilon > 0$: equilibrium points A, B, C: $\xi_S = 0$; $\xi_S = \pm\sqrt{\varepsilon/\delta}$

stability matrix:

$$a = \varepsilon - 3\delta\xi_S^2 = \begin{cases} \varepsilon & \text{for point } A \quad \Rightarrow \quad A \text{ is unstable} \\ -2\varepsilon & \text{for points } B, C \quad \Rightarrow \quad B, C \text{ are stable} \end{cases}$$

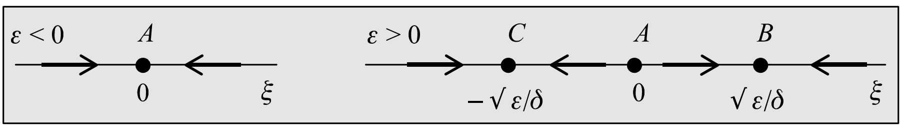


Fig. 1.29: Stability of the equilibrium points

Let's now solve the same problem using the potential method:

$$\phi(\xi) \equiv -\int f(\xi)\,\mathrm{d}\xi = -\varepsilon\frac{\xi^2}{2} + \delta\frac{\xi^4}{4} .$$

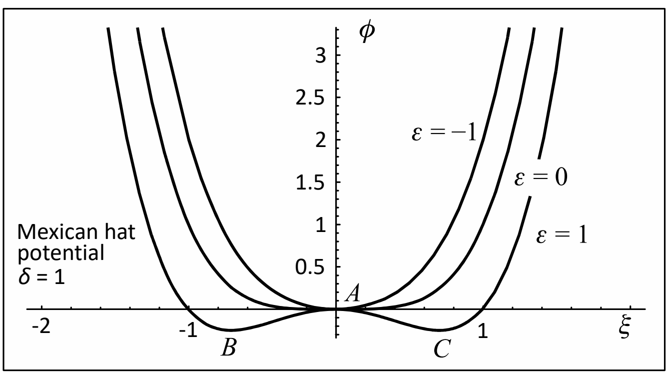


Fig. 1.30: Mexican hat potential

The figure shows the potential curve for $\delta = 1$ and various values of the parameter ε. We see that for $\varepsilon < 0$, ϕ has a single minimum at the origin, where the stable point A is located. For $\varepsilon > 0$, this point becomes a maximum and is unstable. However, two minima appear at the points $\xi = \pm (\varepsilon/\delta)^{1/2}$, at which the system is stable. Due to the characteristic shape of the function ϕ for $\varepsilon > 0$, this function is called the "Mexican hat potential." ◗

1.5.3 Bifurcation

We call a *bifurcation* a sudden change in the phase portrait of a system that occurs when a control parameter of the governing equations is changed continuously. In Example 1.34 from the previous chapter, the phase portrait looks different for $\varepsilon < 0$ and for $\varepsilon > 0$. When ε is changed slowly, the phase portrait of the system changes slowly. The exception is the point $\varepsilon = 0$. The phase portraits for $\varepsilon < 0$ and $\varepsilon > 0$ are not topologically equivalent (they cannot be mapped onto each other by a continuous mapping).

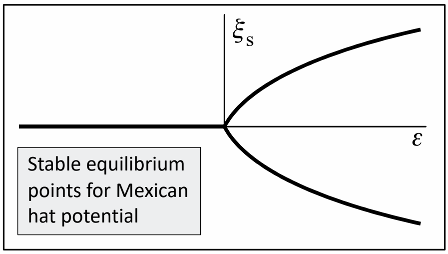


Fig. 1.31: Bifurcation – branching of solutions

A typical phenomenon in bifurcation is the branching of solutions. In Example 1.34, for $\varepsilon < 0$, there is a single stable point $\xi_S = 0$; for $\varepsilon > 0$, there are two stable points $\xi_S = \pm (\varepsilon/\delta)^{1/2}$, and the point $\xi_S = 0$ becomes unstable. Depending on the type of solution branching, bifurcations can be classified as supercritical, subcritical, and transcritical:

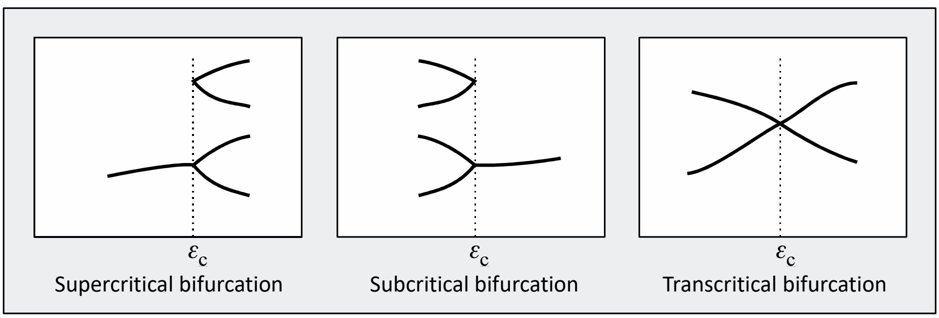


Fig. 1.32. Different types of bifurcations

Continuous phase transitions – a typical bifurcation

The concept of the Mexican hat is used in the theory of continuous phase transitions. Discontinuous phase transitions are changes in a substance in which the internal energy, volume, entropy, etc., change abruptly (melting, solidification, boiling). Continuous phase transitions are changes in a substance in which the first and higher derivatives of the aforementioned quantities change abruptly: specific heat, thermal expansion, elastic modulus, susceptibility, etc. The first comprehensive theory of second-order phase transitions was proposed by Lev Davidovich Landau.

A typical continuous phase transition is the change in the behavior of a ferromagnet at the Curie temperature T_C. For illustrative purposes, let us consider a single infinite series of spins $\sigma_1, \sigma_2, \sigma_3, \ldots$, which can be oriented only up or down (corresponding to values of $\sigma_a = \pm 1$) with a simple interaction energy given by the relation

$$H = -J \sum_{\langle \sigma_a \sigma_b \rangle} \delta_{\sigma_a \sigma_b} \ .$$

Summation is performed over the nearest neighbors. Thus, if two neighboring spins are aligned, they contribute a value of $-J$ to the total energy; if they are misaligned, they contribute nothing at all. At low temperatures ($T < T_C$), spins tend to adopt the state with the lowest possible energy, i.e., they are oriented *mostly* in the same direction. Thus, two typical configurations are possible:

$$\uparrow\uparrow\uparrow\uparrow\downarrow\uparrow\uparrow\uparrow\uparrow\downarrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow \quad \text{or} \quad \downarrow\downarrow\downarrow\downarrow\downarrow\downarrow\uparrow\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow.$$

As the temperature increases, a phase transition occurs at $T = T_C$. At temperatures $T > T_C$, the spins are random $\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow\uparrow\uparrow\downarrow\downarrow\uparrow\uparrow\downarrow\downarrow\uparrow$ and the ferromagnetic properties are lost. If we introduce the *order parameter* (magnetization) as the average value of the spin

$$M \equiv \frac{1}{N} \sum_a \sigma_a \, ,$$

then, in the low-temperature phase as the temperature decreases, $M \to \pm 1$, and in the high-temperature phase as the temperature increases, $M \to 0$. The "Mexican hat" potential and the associated equation (1.197) describe precisely such a phase transition very well. The quantity ξ corresponds to the ordering parameter, i.e., $\xi = M$, and the control parameter corresponds to the quantity $\varepsilon = T_C - T$:

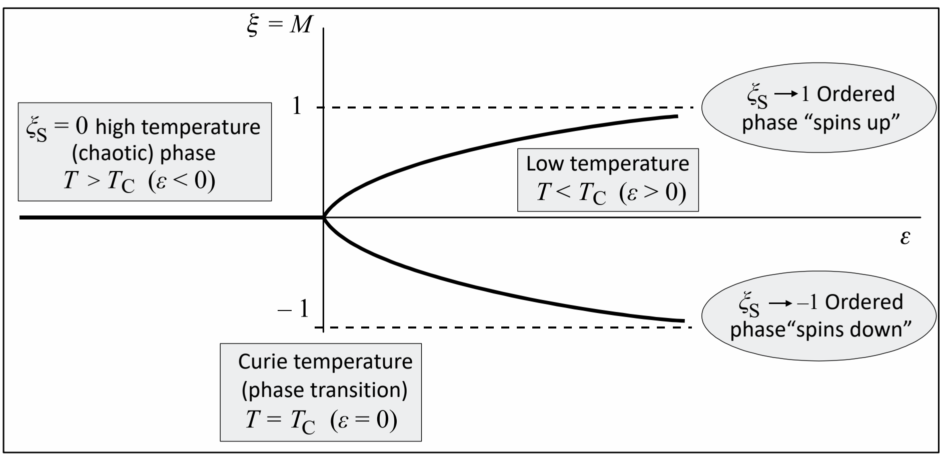


Fig. 1.33: Phase transition in a ferromagnet

Note: Potentials similar to the "Mexican hat" potential are used not only to describe phase transitions, but also, for example, in inflationary models of the early universe and to describe spontaneous symmetry breaking in nature.

◖ **Example 1.35: Hopf bifurcation**

Let us now consider a system of differential equations that describes the motion of a system in polar coordinates

$$\dot{r} = r(\varepsilon + \delta r^2)\,;$$
$$\dot{\varphi} = \omega\,; \qquad \delta > 0, \quad r \geq 0, \quad \varphi \in R\,. \tag{1.198}$$

This is a system of equations describing the motion of the system in polar coordinates. The solution for the angle is immediate: $\varphi(t) = \varphi_0 + \omega t$. Thus, the angle φ represents counterclockwise rotational motion with angular frequency ω. Only a single equation remains for the non-negative radial distance $r(t)$. We can easily find the solutions for stationary points and stability:

$\varepsilon < 0$: equilibrium points A, B: $r_S = 0$; $r_S = \sqrt{|\varepsilon|/\delta}$

stability matrix:

$$a = \varepsilon + 3\delta r_S^2 = \begin{cases} \varepsilon & \text{pro bod } A \;\Rightarrow\; A \text{ is stable} \\ -2\varepsilon & \text{pro bod } B \;\Rightarrow\; B \text{ is unstable} \end{cases}$$

$\varepsilon > 0$: equilibrium point A: $r_S = 0$

stability matrix:

$$a = \varepsilon + 3\delta r_S^2 = \varepsilon > 0 \qquad \Rightarrow \qquad A \text{ je unstable.}$$

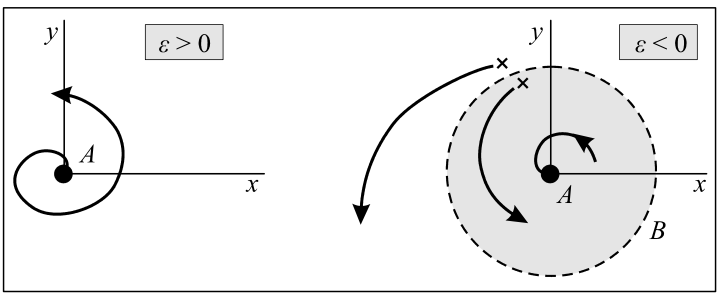


Fig. 1.34: Hopf bifurcation

For $\varepsilon > 0$, the origin is an unstable focus. For $\varepsilon < 0$, the origin is a stable focus, and the "point" B defined by the relation $r_S = \sqrt{|\varepsilon|/\delta}$ is unstable. In fact, B forms a complete set of points – a circle – in the Cartesian coordinate system. Systems with the initial condition $r > r_S$ will spiral away from the center, and systems with $r < r_S$ will spiral toward the center. All trajectories move away from the set B. The figure shows two trajectories with similar initial conditions, whose distance increases exponentially with time. This is the so-called Lyapunov instability, which we will discuss in the next chapter.

◗

1.5.4 Lyapunov Stability, Limit Cycle, Attractor

Let us examine how two trajectories with similar initial conditions, $\boldsymbol{\xi}_0$ and $\boldsymbol{\xi}_0+\boldsymbol{\varepsilon}$, evolve over time:

Fig. 1.35: Lyapunov instability (on the left)

We say that a trajectory is *Lyapunov-unstable* if there exists a trajectory with a nearby initial condition that will diverge exponentially from the trajectory under consideration over time. We say that a trajectory is *Lyapunov-stable* if all trajectories that are nearby at time t_0 will converge to it exponentially. If the distance between the two trajectories changes exponentially over time, then

$$\|\boldsymbol{\xi}(t,\boldsymbol{\xi}_0+\boldsymbol{\varepsilon})-\boldsymbol{\xi}(t,\boldsymbol{\xi}_0)\| \sim \mathrm{e}^{\lambda t}$$

and we can easily determine

$$\lambda = \lim_{\substack{t\to\infty \\ \boldsymbol{\varepsilon}\to 0}} \frac{1}{t} \ln \|\boldsymbol{\xi}_{\boldsymbol{\varepsilon}} - \boldsymbol{\xi}\|.$$

The coefficient λ is called the Lyapunov exponent. If $\lambda > 0$, we refer to a Lyapunov-unstable trajectory. If $\lambda < 0$, it is a Lyapunov-stable trajectory. If $\lambda = 0$, the dependence is non-exponential, such as power-law, and we cannot speak of Lyapunov stability or instability.

In multidimensional problems with $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_N)$, the Lyapunov exponent depends on how the limit $\boldsymbol{\varepsilon} \to 0$ is taken. This yields N first-order Lyapunov coefficients (in the direction of the coordinate axes). However, we can also consider the entire bundle of nearby trajectories from a two- or three-dimensional region (see Fig. 1.36).

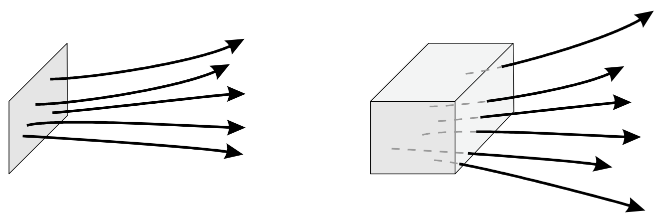


Fig. 1.36: 2D and 3D Lyapunov exponents

We then speak of multidimensional Lyapunov exponents (of the second order, third order, ...). A trajectory is Lyapunov-stable if all Lyapunov coefficients λ are greater than or equal to 0. An example of a Lyapunov-unstable trajectory is the set $r_S = \sqrt{(|\varepsilon|/\delta)}$ for $\varepsilon < 0$ in the last example on the Hopf bifurcation. Trajectories with $r \geq r_S$ are Lyapunov-unstable. Trajectories for which $r < r_S$ are Lyapunov-stable. Another example of Lyapunov instability is the billiards game with obstacles shown in Figure 1.37. Note that two nearby trajectories quickly lose their connection in later times.

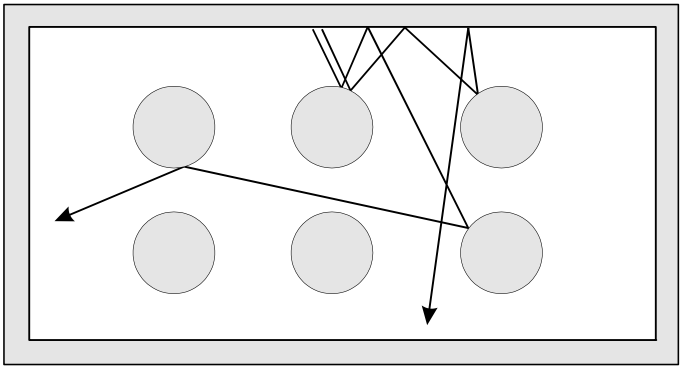


Fig. 1.37: Billiards as an illustration of Lyapunov instability

◖ **Example 1.36: Van der Pol oscillator**

The system is described by differential equations proposed by the Dutch physicist Balthasar van der Pol (1889–1959) to describe the electrical circuit of a vacuum tube:

$$\begin{aligned} \dot{\xi}_1 &= \xi_2 \,, \\ \dot{\xi}_2 &= -\xi_1 + \varepsilon(1-\delta\xi_1^2)\xi_2 \,; \qquad \delta > 0 \,. \end{aligned} \tag{1.199}$$

In this system, trajectories with arbitrary initial conditions converge to a single periodic trajectory, which we call the *limit cycle*. After a sufficiently long time, every trajectory approaches the limit cycle trajectory as closely as one wishes. All trajectories in the vicinity of the limit cycle are Lyapunov-stable. Figure 1.38 shows the phase trajectories for various initial conditions of the van der Pol oscillator with $\delta = 1$ and $\varepsilon = 0.1$:

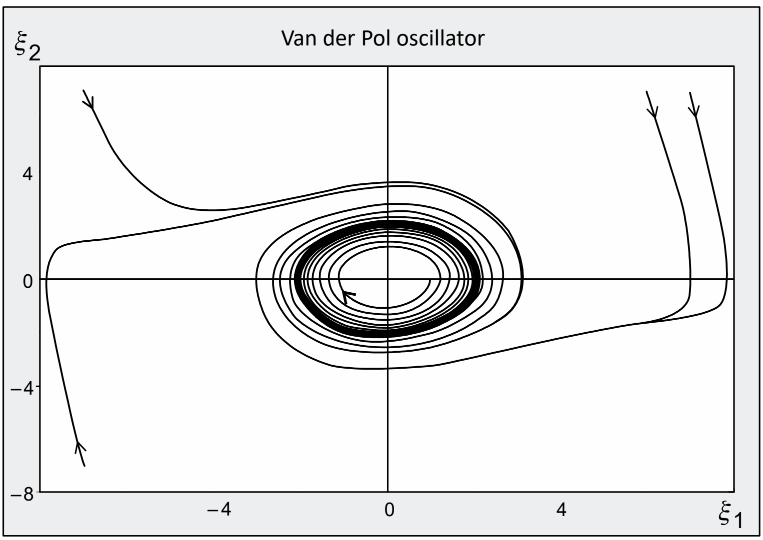


Fig. 1.38: Van der Pol oscillator. The limit cycle is a thick closed curve

◗

Some concepts from the theory of sets

The distance between two points $\rho(\mathbf{A}, \mathbf{B})$:

Let's define the distance between two points as simply as possible, for example as the Cartesian distance given by the Pythagorean theorem

► $$\rho(\mathbf{A},\mathbf{B}) \equiv \sqrt{\sum_{k=1}^{N} (A_k - B_k)^2}\ .$$

There is another formula satisfying the basic requirements for the concept of distance. We can write the distance between two points as $||\mathbf{A}-\mathbf{B}||$, where $||\cdot||$ is defined as

► $$||\mathbf{X}|| \equiv \sqrt{\sum_{k=1}^{N} {X_k}^2}\ .$$

This is the norm (magnitude) of the vector difference **A**–**B**.

Distance between a point and a set $\rho(\mathbf{A},\mathcal{M})$:

We define the distance of a point from a set as the minimum of the distances from all points in the set, including its boundary $(\bar{\mathcal{M}})$:

► $$\rho(\mathbf{A},\mathcal{M}) \equiv \min_{\mathbf{X}\in \bar{\mathcal{M}}} \rho(\mathbf{A},\mathbf{X})\ .$$

The neighborhood of a point $U_\varepsilon(\mathbf{A})$

The neighborhood of a point is a sphere without boundary centered at **A** with radius ε:

► $$U_\varepsilon(\mathbf{A}) \equiv \{\mathbf{X};\ \rho(\mathbf{A},\mathbf{X}) < \varepsilon\}\ .$$

Open set $\mathcal{M}_\mathrm{O}$

An open set is a set such that for every point in it, a neighborhood can be constructed that is entirely contained in the set $\mathcal{M}_\mathrm{O}$:

► $$\text{ke}\ \forall\ \mathbf{X}\in \mathcal{M}_\mathrm{O}\ \exists U_\varepsilon(\mathbf{X}) \subset \mathcal{M}_\mathrm{O}\ .$$

Simply put, open sets do not contain their own boundary.

Closed set $\mathcal{M}_\mathrm{U}$

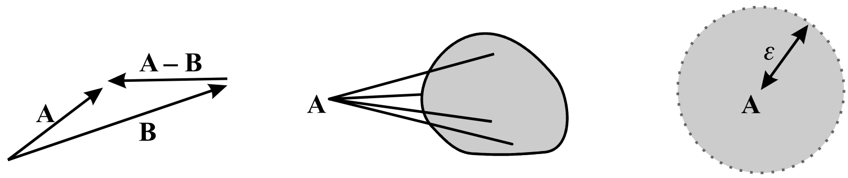
A closed set is a set in which a convergent sequence of points always converges to some element of that set. In other words, it is never the case that we find a sequence that converges but whose limit lies outside the set. Therefore, the following holds:

► $$\mathbf{X}^{(k)} \in \mathcal{M}_\mathrm{U}\ ; \qquad \mathbf{X}^{(k)} \to \mathbf{X} \qquad \Rightarrow \qquad \mathbf{X} \in \mathcal{M}_\mathrm{U}\ ;$$

Simply put, closed sets contain their own boundary.

Note 1: In our case of phase space, the points **A**, **B**, and **X** are always N-tuples $(\xi_1, \dots, \xi_N)$.

Note 2: A closed interval and a circle with a boundary are closed sets; an open interval and a circle without a boundary are open sets; a semi-closed interval is neither an open nor a closed set; the empty set and the entire space R^N are both open and closed sets in the sense of the previous definitions.



$\rho(\mathbf{A},\mathbf{B})$ $\rho(\mathbf{A},\mathcal{M})$ $U_\varepsilon(\mathbf{A})$

Fig. 1.39: Defining distance and neighborhood

Concepts related to solving systems of differential equations

Invariant set $\mathcal{J}$

We say that a set is invariant if, starting from any point in the set $\mathcal{J}$ – taken as the initial condition for the system of differential equations (1.183) – a phase trajectory evolves that lies entirely within the set $\mathcal{J}$. Thus, once the system enters the set $\mathcal{J}$, it will remain there at all subsequent times:

► $$\mathcal{J} = \{\mathbf{X};\quad \mathbf{X}_0 = \boldsymbol{\xi}(t_0) \in \mathcal{J} \quad \Rightarrow \quad \mathbf{X} = \boldsymbol{\xi}(t) \in \mathcal{J} \quad \text{pro } \forall\, t > t_0\}.$$

Dense set $\mathcal{D}$

We say that a set is dense (densely covered) if there is a phase trajectory passing through any arbitrarily small neighborhood of every point in the set $\mathcal{D}$.

Chaotic set $\mathcal{X}$

1) Every trajectory in $\mathcal{X}$ is Lyapunov-unstable
2) There exists a trajectory that densely covers $\mathcal{X}$
3) $\mathcal{X}$ is an invariant set

Atractor $\mathcal{A}$

1) Trajectories from the vicinity of $\mathcal{A}$ are “drawn” toward $\mathcal{A}$, i.e., they approach $\mathcal{A}$ as time increases:

   $$\exists\, U_{\mathcal{A}} \supset \mathcal{A}, \text{ že pro } \forall\, \boldsymbol{\xi}(t_0) \in U_{\mathcal{A}} \text{ platí } \lim_{t\to\infty} \rho(\boldsymbol{\xi}(t), \mathcal{A}) = 0$$

2) Exists a trajectory that densely covers $\mathcal{A}$

3) $\mathcal{A}$ is invariant set

4) $\mathcal{A}$ is closed set

Strange atractor S

A strange attractor is a chaotic attractor, i.e., all trajectories of a strange attractor are Lyapunov-unstable.

Limit cycle C

A limit cycle is a closed phase trajectory that is an attractor.

Note 1: Every equilibrium point is an invariant set. Furthermore, every closed trajectory – such as that of a harmonic oscillator – is an invariant set.

Note 2: Every closed trajectory automatically forms an invariant, closed, densely covered set. Moreover, the limit cycle "attracts" trajectories from its surroundings, i.e., it possesses the first property of an attractor.

Note 3: An example of a chaotic set is the billiard table shown in the Fig. 1.37.

Note 4: A strange attractor can only arise in a problem with dimension $N \geq 3$.

Note 5: If, for a system of two equations, the expression $\partial f_1/\partial\xi_1 + \partial f_2/\partial\xi_2$ does not change sign in a simply connected region, then no closed trajectory exists in that region. This criterion is called the *Bendixson criterion* (named after the Swedish mathematician Ivar Otto Bendixson).

◖ **Example 1.37: Brusselator (2D)**

We will examine a chemical reaction of the type

$$\begin{aligned} A &\xrightarrow{k_1} X \\ B + X &\xrightarrow{k_2} Y + D \\ 2X + Y &\xrightarrow{k_3} 3X \\ X &\xrightarrow{k_4} E \end{aligned}$$

The rates of the individual reactions are denoted by $k_1, \dots, k_4$. We denote the concentrations of the reactants and products by n_A, n_B, n_D, and n_E. The variables will be the concentrations of substances X and Y: $\xi_1 = n_X$, $\xi_2 = n_Y$. From the reaction equations, we will formulate the initial system of differential equations

$$\frac{\mathrm{d}\xi_1}{\mathrm{d}t} = k_1 n_A - k_2 n_B \xi_1 + k_3 \xi_1^2 \xi_2 - k_4 \xi_1 ,$$
$$\frac{\mathrm{d}\xi_2}{\mathrm{d}t} = k_2 n_B \xi_1 - k_3 \xi_1^2 \xi_2 .$$

On the right hand side is the list of mechanisms for the creation and destruction of substances X and Y. If we disregard the insignificant constants, the equations have form

$$\frac{\mathrm{d}\xi_1}{\mathrm{d}t} = \alpha - (\beta + 1)\xi_1 + \xi_1^2 \xi_2 ,$$
$$\frac{\mathrm{d}\xi_2}{\mathrm{d}t} = \beta \xi_1 - \xi_1^2 \xi_2 . \tag{1.200}$$

These equations yield a solution in the form of a limit cycle. For values of $\alpha = 2$ and $\beta = 5{,}9$, and various initial conditions, the phase trajectories are shown in the following figure. After a sufficiently long time, the concentrations ξ_1 and ξ_2 periodically vary (oscillate) around certain mean values. This theoretical model of a chemical reaction was proposed by the Russian-Belgian chemist Ilya Prigogine (1917–2003) at the Free University of Brussels.

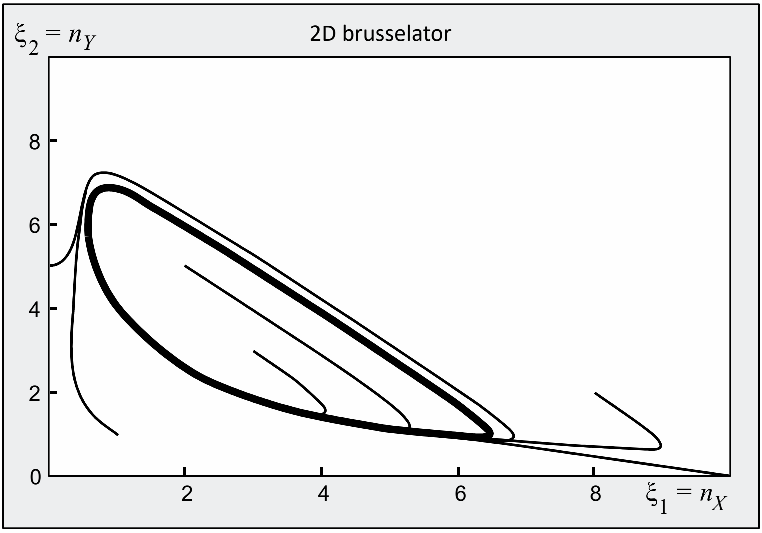


Fig. 1.40: Phase portrait of a 2D brusselator; the limit cycle is shown as a thick line

◗

◖ **Example 1.38: Brusselator (4D)**

Let us assume that the previous reaction takes place simultaneously in two reactors, with substance X being exchanged at a rate of δ_1 and substance Y at a rate of δ_2. We denote the concentrations of substances X and Y in reactors 1 and 2 as follows: $\xi_1 = n_{X1}$, $\xi_2 = n_{Y1}$, $\xi_3 = n_{X2}$, $\xi_4 = n_{Y2}$. The initial equations will be

$$\frac{\mathrm{d}\xi_1}{\mathrm{d}t} = \alpha - (\beta + 1)\xi_1 + \xi_1^2 \xi_2 + \delta_1(\xi_3 - \xi_1) ,$$
$$\frac{\mathrm{d}\xi_2}{\mathrm{d}t} = \beta\xi_1 - \xi_1^2 \xi_2 + \delta_2(\xi_4 - \xi_2) ,$$
$$\frac{\mathrm{d}\xi_3}{\mathrm{d}t} = \alpha - (\beta + 1)\xi_3 + \xi_3^2 \xi_4 + \delta_1(\xi_1 - \xi_3) , \tag{1.201}$$
$$\frac{\mathrm{d}\xi_4}{\mathrm{d}t} = \beta\xi_3 - \xi_3^2 \xi_4 + \delta_2(\xi_2 - \xi_4) .$$

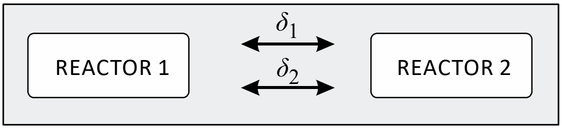


Fig. 1.41: Coupled brusselators

Equations (1.201) form a system of four nonlinear differential equations, whose solutions lead to a strange attractor for certain parameters (the system has a dimension > 3). The following figure shows a portion of the phase trajectory that would densely cover the region of the strange attractor for $\alpha = 2$; $\beta = 5.9$; $\delta_1 = 1.21$ and $\delta_2 = 12.1$.

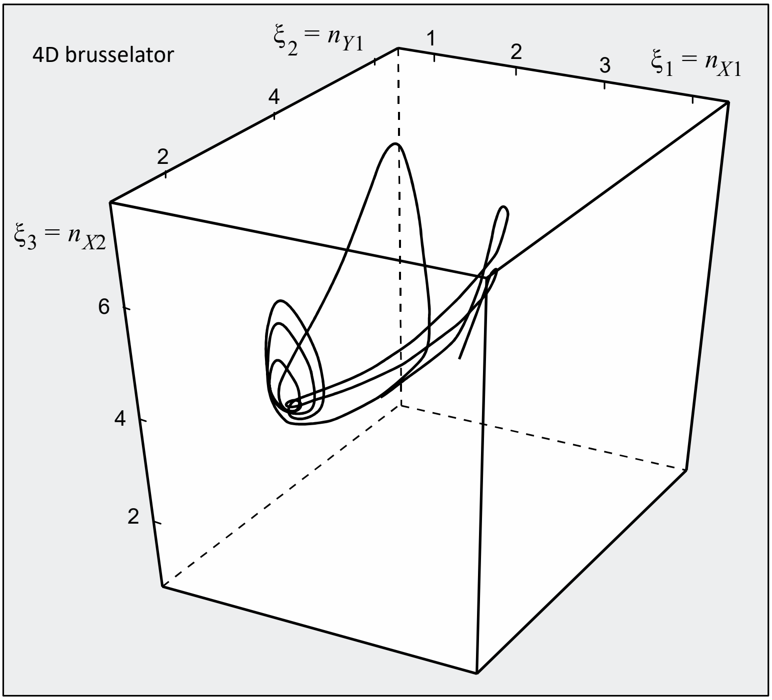


Fig. 1.42: Phase portrait of the 4D brusselator ◗

◖ Example 1.39: Lorenz atractor

This is the best-known example of a strange attractor. The initial set of equations

$$\frac{\mathrm{d}\xi_1}{\mathrm{d}t} = \alpha(\xi_2 - \xi_1),$$
$$\frac{\mathrm{d}\xi_2}{\mathrm{d}t} = -\xi_1 \xi_3 + \beta \xi_1 - \xi_2, \tag{1.202}$$
$$\frac{\mathrm{d}\xi_3}{\mathrm{d}t} = \xi_1 \xi_2 - \gamma \xi_3$$

►

describes fluid between two parallel plates at different temperatures (see [14], [15]). The variables ξ_1, ξ_2, and ξ_3 represent, respectively: the first Fourier component of velocity, and the first and second Fourier components of temperature. The following figure again shows a portion of the phase trajectory that would densely cover the attractor region. The equations were solved numerically for the values $\alpha = 3$; $\beta = 26.5$; $\gamma = 1$.

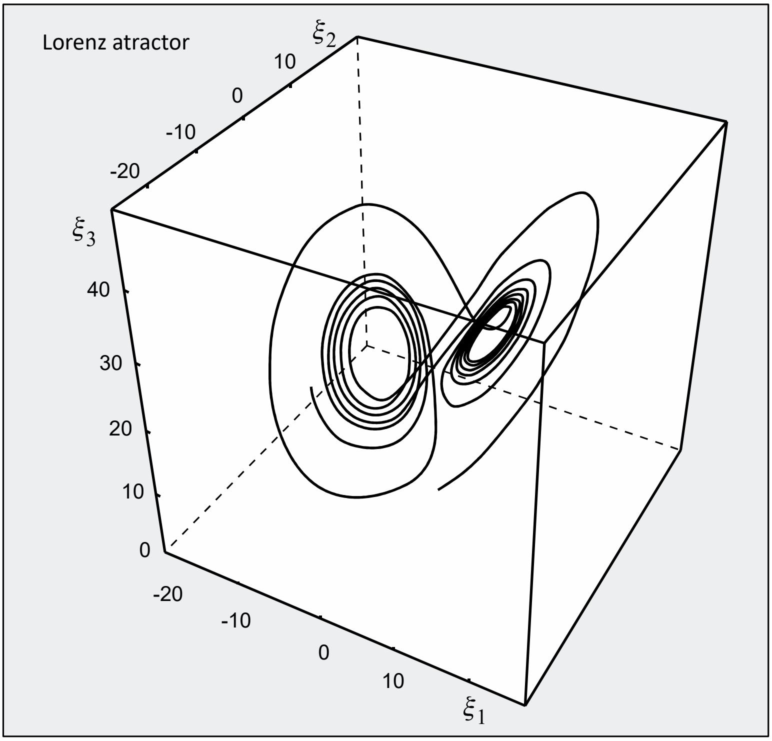


Fig. 1.43: Phase portrait of the Lorenz atractor

1.5.5 Evolutionary Equations

◖ **Example 1.40: Electron-hole plasma in a strong electric field**

In a strong electric field, accelerated electrons and holes cause impact ionization. When an electron collides with a hole, recombination occurs, i.e., the carriers are annihilated. If we denote the electron concentration by $\xi_1 = n_e$ and the hole concentration by $\xi_2 = n_h$, the basic equations for the time evolution of the carrier population will have the form:

$$\frac{d\xi_1}{dt} = \alpha_1 \xi_1 - \beta \xi_1 \xi_2 ,$$
$$\frac{d\xi_2}{dt} = \alpha_2 \xi_2 - \beta \xi_1 \xi_2 . \tag{1.203}$$

The first terms on the rhs describe ionization processes (gain of carriers), while the second terms describe recombination processes (loss of carriers). The following problems have the same form of equations. ◗

◖ **Example 1.41: Predator and Prey System**

We assume that the predator feeds on prey (such as wolves and hares), and that the prey has an ample food supply (for example, by eating grass). If we denote $\xi_1 = n_\mathrm{p}$ as the number of predators in a given area and $\xi_2 = n_\mathrm{k}$ as the number of potential prey, the basic equations describing the temporal evolution of the animal populations will have a form similar to that in the previous example:

$$\begin{aligned} \frac{\mathrm{d}\xi_1}{\mathrm{d}t} &= -\alpha_1 \xi_1 + \beta_1 \xi_1 \xi_2 \,, \\ \frac{\mathrm{d}\xi_2}{\mathrm{d}t} &= +\alpha_2 \xi_2 - \beta_2 \xi_1 \xi_2 \,. \end{aligned} \tag{1.204}$$

The first term in the first equation describes the mortality of predators in the absence of prey ($\xi_2 = 0$). The first term in the second equation describes the reproduction of prey in the absence of predators ($\xi_1 = 0$). The second terms represent the consumption of prey by predators, the so-called “pairwise interaction,” which causes the number of predators to increase and the amount of prey to decrease. ◗

◖ **Example 1.42: Two social groups**

Let us now describe two groups of people with differing views on a particular issue (supporters of two different approaches, theories, opinions, or political parties). If we denote $\xi_1 = n_\mathrm{A}$ as the number of supporters of view A and $\xi_2 = n_\mathrm{B}$ as the number of supporters of view B, the basic equations describing the change over time in the number of supporters will take the form:

$$\begin{aligned} \frac{\mathrm{d}\xi_1}{\mathrm{d}t} &= \alpha_1 \xi_1 + \beta_1 \xi_1 \xi_2 \,, \\ \frac{\mathrm{d}\xi_2}{\mathrm{d}t} &= \alpha_2 \xi_2 - \beta_2 \xi_1 \xi_2 \,. \end{aligned} \tag{1.205}$$

The β coefficients can be either positive or negative; pairwise interactions here refer to encounters between members of different groups, discussions, etc. ◗

◖ **Example 1.43: Chemical reaction**

Let’s consider a chemical reaction of the type

$$\begin{aligned} A + B &\xrightarrow{\;k_1\;} C \,, \\ A + C &\xrightarrow{\;k_2\;} B + D \,. \end{aligned}$$

The equations for the time evolution of the individual concentrations are of the form:

$$\begin{aligned} \frac{\mathrm{d}n_\mathrm{A}}{\mathrm{d}t} &= -k_1 n_\mathrm{A} n_\mathrm{B} - k_2 n_\mathrm{A} n_\mathrm{C} \,, & \frac{\mathrm{d}n_\mathrm{C}}{\mathrm{d}t} &= +k_1 n_\mathrm{A} n_\mathrm{B} - k_2 n_\mathrm{A} n_\mathrm{C} \,, \\ \frac{\mathrm{d}n_\mathrm{B}}{\mathrm{d}t} &= -k_1 n_\mathrm{A} n_\mathrm{B} + k_2 n_\mathrm{A} n_\mathrm{C} \,, & \frac{\mathrm{d}n_\mathrm{D}}{\mathrm{d}t} &= +k_2 n_\mathrm{A} n_\mathrm{C} \,. \end{aligned} \tag{1.206}$$

Substance B acts as a catalyst for the reaction. If A is present in sufficient quantity as a reactant, we can assume n_A = const. and solve only three equations. ◗

All the equations from the previous examples have the same form

$$\blacktriangleright \qquad \frac{\mathrm{d}\xi_k}{\mathrm{d}t} = \alpha_{kj}\,\xi_j + \beta_k^{jl}\xi_j\,\xi_l \tag{1.207}$$

and are called *evolutionary equations*. Note that the double subscripts indicate summation. A characteristic feature is the linear combination of various pairwise interactions. Typical solutions include oscillations and limit cycles; in more than three dimensions, chaotic sets and strange attractors arise. Let us now examine the solution to a system of two equations of the form

$$\blacktriangleright \qquad \begin{aligned} \frac{\mathrm{d}\xi_1}{\mathrm{d}t} &= \alpha_1\,\xi_1 + \beta_1\,\xi_1\,\xi_2\,, \\ \frac{\mathrm{d}\xi_2}{\mathrm{d}t} &= \alpha_2\,\xi_2 + \beta_2\,\xi_1\,\xi_2\,. \end{aligned} \tag{1.208}$$

We use the standard procedure to determine the equilibrium points:

$$\xi^{(1)} = (0,0)\,;$$

$$\xi^{(2)} = \left(-\frac{\alpha_2}{\beta_2}, -\frac{\alpha_1}{\beta_1}\right).$$

Let's not forget the meaning of the individual variables ξ. These are generally the numbers of individuals of a certain type. Therefore, only non-negative values make sense. From the stability matrix, we determine that the first equilibrium point

$\xi^{(1)} = (0,0)$ is for	$\alpha_1\,,\alpha_2 > 0$	unstable
	$\alpha_1\,,\alpha_2 < 0$	stable
	$\alpha_1{\cdot}\alpha_2 < 0$	saddle point

For the second equilibrium point we have

$\xi^{(2)} = (-\alpha_2/\beta_2\,, -\alpha_1/\beta_1)$ is for

	$\alpha_1\,,\alpha_2 > 0$	unstable in one direction
	$\alpha_1\,,\alpha_2 < 0$	stable in one direction
	$\alpha_1{\cdot}\alpha_2 < 0$	oscillation center (different signs for α_1, α_2)

In the latter case, the oscillations occur around a equilibrium point with a frequency of

$$\blacktriangleright \qquad \omega = \sqrt{|\,\alpha_1\cdot\alpha_2\,|}. \tag{1.209}$$

These oscillations represent an *oscillating equilibrium* between individuals of both types; their numbers are maintained within the limits set by the oscillations. A prime example of such a system is the predator-prey system (Example 1.41). In the system of electrons and holes in a strong electric field (Example 1.40), it is not possible to achieve an oscillating equilibrium. The numbers of individuals in two various social groups (Example 1.42) may or may not oscillate, just as the concentrations of substances in chemical reactions (Example 1.43) may or may not oscillate.

Lotka-Volterra equations

If we add the control terms f_k to the right-hand sides of the evolutionary equations, we obtain the so-called *Lotka–Volterra equations*:

$$\blacktriangleright \qquad \frac{\mathrm{d}\xi_k}{\mathrm{d}t} = \alpha_{kj}\,\xi_j \;+\; \beta_k^{jl}\xi_j\xi_l \;+\; f_k\,. \tag{1.210}$$

In a predator-prey system, control terms can describe, for example, the supply of food from an external source or external regulation of the animal population. The values of f_k can be constant or various functions of time (periodic hunting). The range of solution types for Lotka–Volterra systems is already very rich for the two-dimensional case. In different regions of phase space, we find various types of solutions – oscillations, stable and unstable foci, stable regions, unstable regions, and saddles. With periodic control terms, we observe resonance and system excitation. For example, a predator-prey equation with a control term

$$\begin{aligned} \frac{\mathrm{d}\xi_1}{\mathrm{d}t} &= -\,\xi_1 + \xi_1\,\xi_2 + 1/4\,, \\ \frac{\mathrm{d}\xi_2}{\mathrm{d}t} &= +\,\xi_2 - \xi_1\,\xi_2\,. \end{aligned} \tag{1.211}$$

has a solution at point $\xi^{(S)} = (1,\ 3/4)$ in the form of a stable focus.

Logistic equation

The simplest type of evolutionary equation is a differential equation in which the rate of change over time is proportional to the number of individuals:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \alpha n\,. \tag{1.212}$$

Based on the sign of α, the solution is an increasing or decreasing exponential function

$$n(t) = n(0)\,\mathrm{e}^{\alpha t}\,. \tag{1.213}$$

This solution is always only an approximation of reality. Nothing can grow exponentially for a sufficiently long period of time. An exponential solution describes the initial increase of instabilities or various transient solutions. After a certain period of time, pairwise processes take over, leading to the saturation of the solution. This situation is described by the so-called logistic equation

$$\blacktriangleright \qquad \frac{\mathrm{d}n}{\mathrm{d}t} = \alpha n - \beta n^2\,. \tag{1.214}$$

We can easily find the steady-state (steady) solution by setting $\mathrm{d}n/\mathrm{d}t$ equal to zero:

$$n_\mathrm{S} \equiv \lim_{t\to\infty} n(t) = \frac{\alpha}{\beta}\,. \tag{1.215}$$

To solve equation (1.214) , we will use a substitution

$$n(t) = \frac{\alpha}{\beta}\xi^r(t)\,. \tag{1.216}$$

The constant term before the solution represents the saturated value. We will find a power of r that simplifies the differential equation as much as possible. After performing the substitution, we obtain the equation

$$r\frac{\mathrm{d}\xi}{\mathrm{d}t} = \alpha\xi - \alpha\xi^{r+1}\,. \tag{1.217}$$

It is obvious that choosing $r = -1$ transforms this equation into a linear equation with a constant right-hand side, whose solution can be easily found. The analytical solution to the original logistic equation is

► $$n(t) \;=\; n_S \frac{e^{\alpha t}}{e^{\alpha t} + n_S/n_0 - 1} \;=\; \frac{n_S}{1+\left(n_S/n_0 - 1\right)e^{-\alpha t}}\,. \tag{1.218}$$

We can easily verify that the limit of the solution as $t \to \infty$ yields a saturated value (1.215), and the limit without pairwise interaction – i.e., as $\beta \to 0$ ($n_S \to \infty$) – yields an exponential function (1.213). The logistic equation describing restricted exponential growth was first proposed by the Belgian mathematician Pierre François Verhulst (1804–1849) in 1838.

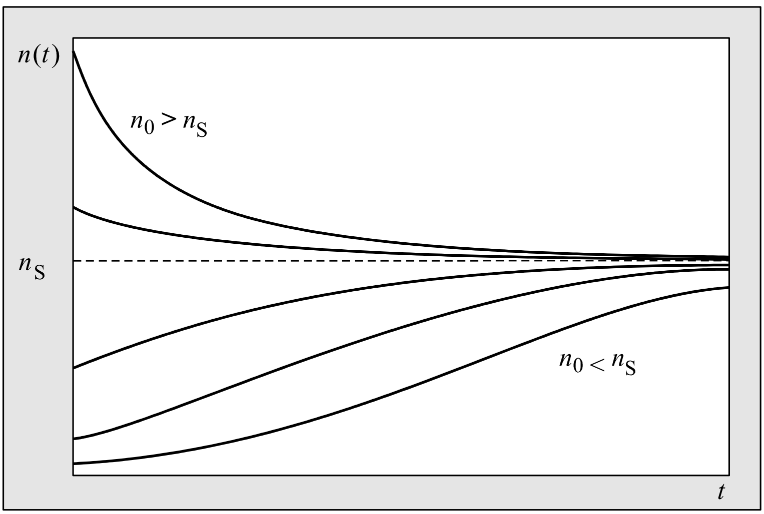


Fig. 1.44: Solution of the logistics equation

1.6 Lagrange Equations for Field Problems

1.6.1 Lagrange Equations, Scalar Fields

To study this chapter, you must be familiar with covariant and contravariant indices. If you are not familiar with this technique, please read Section 3.3.6 Tensors and Metrics, first. In classical mechanics, we sought the dependence of generalized coordinates $q_k(t)$ on time. For field problems, we will seek the spacetime dependence of fields $\varphi_k(t, \mathbf{x})$. Instead of the Lagrangian, we will use the density of the Lagrangian, which depends on time, space, fields, and their derivatives:

$$\mathscr{L} = \mathscr{L}(t, x, y, z, \varphi_1, \ldots \varphi_N, \partial\varphi_1/\partial t, \partial\varphi_1/\partial x, \ldots \partial\varphi_N/\partial z)\,,$$

which we will write in the abbreviated form as

$$\mathscr{L} = \mathscr{L}(x^\mu, \varphi_k, \varphi_{k,\alpha})\,.$$

The following will hold for the integral of the action

$$S = \int_\Omega \mathscr{L}(x^\mu, \varphi_k, \varphi_{k,\alpha})\,\mathrm{d}^3\mathbf{x}\,\mathrm{d}t = \int_\Omega \mathscr{L}(x^\mu, \varphi_k, \varphi_{k,\alpha})\,\mathrm{d}^4 x\,.$$

Just as in the mechanics of rigid bodies, we will seek the necessary conditions for the extremality of the action integral, and the variations of the fields will be defined at the same time (but this time also in the spatial coordinate), which ensures the interchangeability of variations and partial derivatives. At the boundary of the domain $\partial\Omega$, we require that the variations be zero; thus, relations analogous to (1.4), (1.5), and (1.6) must hold here:

$$\begin{aligned} \delta\varphi_k &= \varphi_{k,\mathrm{virt}}(x^\mu) - \varphi_{k,\mathrm{real}}(x^\mu)\,; \\ \delta\varphi_k(\partial\Omega) &= 0\,; \\ \delta\,\partial_\mu\varphi_k &= \partial_\mu\delta\,\varphi_k\,. \end{aligned} \tag{1.219}$$

Let us therefore require that the variation of the action integral be zero:

$$\delta\int_\Omega \mathscr{L}(x^\mu, \varphi_k, \varphi_{k,\alpha})\,\mathrm{d}^4 x = 0\,.$$

Thanks to the interchangeability of variations and derivatives, we can bring the variation into the integral and apply it to all variables (with the exception of x^μ, since this is a variation within the same event):

$$\int_\Omega \left[\frac{\partial\mathscr{L}}{\partial\varphi_k}\delta\varphi_k + \frac{\partial\mathscr{L}}{\partial\varphi_{k,\alpha}}\delta\varphi_{k,\alpha}\right]\mathrm{d}^4 x = 0\,.$$

In the last term, we swap the variation and the derivative: $\delta\varphi_{k,\alpha} = \delta\partial_\alpha\varphi_k = \partial_\alpha\delta\varphi_k$ and perform integration per partes (using Gauss theorem). The integral at the boundary is zero by (1.219), so we have:

$$\int_\Omega \left[\frac{\partial\mathscr{L}}{\partial\varphi_k} - \partial_\alpha \frac{\partial\mathscr{L}}{\partial\varphi_{k,\alpha}}\right]\delta\varphi_k \,\mathrm{d}^4x = 0\,.$$

Since the integral must be zero for any spacetime region Ω, the integrand must also be zero (more precisely, *almost everywhere*, i.e., except on sets of dimension less than 4):

$$\left[\frac{\partial\mathscr{L}}{\partial\varphi_k} - \partial_\alpha \frac{\partial\mathscr{L}}{\partial\varphi_{k,\alpha}}\right]\delta\varphi_k = 0\,. \tag{1.220}$$

If the fields φ_k are independent, the coefficients of the linear combination (1.220) will be zero (the entire expression takes the form $\Sigma c_k\,\delta\varphi_k = 0$), that is

$$\frac{\partial\mathscr{L}}{\partial\varphi_k} - \partial_\alpha \frac{\partial\mathscr{L}}{\partial\varphi_{k,\alpha}} = 0\,.$$

We will rewrite the expression into the standard form of Lagrange equations

► $$\partial_\alpha\left[\frac{\partial\mathscr{L}}{\partial\varphi_{k,\alpha}}\right] - \frac{\partial\mathscr{L}}{\partial\varphi_k} = 0\,; \qquad k = 1,\ldots,N\,. \tag{1.221}$$

Unlike Lagrange equations for point masses and rigid bodies, the first term does not contain only a time derivative, but includes derivatives with respect to all four variables. Lagrange equations, expressed in terms of a single field, take the form:

$$\frac{\partial}{\partial t}\left[\frac{\partial\mathscr{L}}{\partial(\partial\varphi/\partial t)}\right] + \frac{\partial}{\partial x}\left[\frac{\partial\mathscr{L}}{\partial(\partial\varphi/\partial x)}\right] + \frac{\partial}{\partial y}\left[\frac{\partial\mathscr{L}}{\partial(\partial\varphi/\partial y)}\right] + \frac{\partial}{\partial z}\left[\frac{\partial\mathscr{L}}{\partial(\partial\varphi/\partial z)}\right] - \frac{\partial\mathscr{L}}{\partial\varphi} = 0\,. \tag{1.222}$$

Note: The Lagrangian is not uniquely determined. The function $\tilde{\mathscr{L}} \equiv \mathscr{L} + \partial_\mu K^\mu$ yields the same field equations for any four-vector K^μ. This freedom can be used to construct the "most elegant" Lagrangian possible.

◖ Example 1.44: Wave equation

Let us find the Lagrange equations for the Lagrange function of the scalar field φ, which contains only the derivatives of this field according to the following formula:

$$\mathscr{L} = \left(\partial_\mu\varphi\right)\left(\partial^\mu\varphi\right). \tag{1.223}$$

Solution: The Lagrangian is a scalar (this is ensured by having one upper and one lower index; the expression remains unchanged when the basis or coordinate system is changed). If we write out the Lagrangian, we have:

$$\mathscr{L} = -\frac{1}{c^2}\left(\frac{\partial\varphi}{\partial t}\right)^2 + \left(\frac{\partial\varphi}{\partial x}\right)^2 + \left(\frac{\partial\varphi}{\partial y}\right)^2 + \left(\frac{\partial\varphi}{\partial z}\right)^2. \tag{1.224}$$

After performing all the derivatives, Lagrange equation (1.222) gives

$$-\frac{2}{c^2}\frac{\partial^2\varphi}{\partial t^2}+2\frac{\partial^2\varphi}{\partial x^2}+2\frac{\partial^2\varphi}{\partial y^2}+2\frac{\partial^2\varphi}{\partial z^2}=0 \qquad \Rightarrow$$

$$-\frac{1}{c^2}\frac{\partial^2\varphi}{\partial t^2}+\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}+\frac{\partial^2\varphi}{\partial z^2}=0\,.$$

The simplest form of the Lagrangian for a scalar field thus leads to the wave equation. In the Lagrangian (1.223), the coefficient ½ is usually written, i.e.:

► $$\mathscr{L}=\frac{1}{2}\left(\partial_\mu\varphi\right)\left(\partial^\mu\varphi\right) \qquad \Rightarrow \qquad \Box\varphi=0\,. \tag{1.225}$$

There are two reasons: Lagrange equations yield the wave equation (without the need to multiply by a factor of 2), and the entire expression (1.225) is analogous to kinetic energy (half of the square of the derivatives).

Other solution: Let's solve the same example for the Lagrangian (1.223) without decomposing it into components. The left-hand side of the Lagrangian equations is:

$$\partial_\alpha\left[\frac{\partial\mathscr{L}}{\partial\varphi_{,\alpha}}\right]-\frac{\partial\mathscr{L}}{\partial\varphi}=\partial_\alpha\left[\frac{\partial\mathscr{L}}{\partial\varphi_{,\alpha}}\right]=\partial_\alpha\left[\frac{\partial\mathscr{L}}{\partial(\partial_\alpha\varphi)}\right]=$$

$$=\partial_\alpha\frac{\partial}{\partial(\partial_\alpha\varphi)}\left[\frac{1}{2}\left(\partial_\mu\varphi\right)\left(\partial^\mu\varphi\right)\right]=$$

$$=\frac{1}{2}\partial_\alpha\frac{\partial}{\partial(\partial_\alpha\varphi)}\left[g^{\mu\nu}\left(\partial_\mu\varphi\right)\left(\partial_\nu\varphi\right)\right]=$$

$$=\frac{1}{2}g^{\mu\nu}\partial_\alpha\left[\delta^\alpha{}_\mu\left(\partial_\nu\varphi\right)+\left(\partial_\mu\varphi\right)\delta^\alpha{}_\nu\right]=$$

$$=\frac{1}{2}\partial_\alpha\left[\delta^\alpha{}_\mu\left(\partial^\mu\varphi\right)+\left(\partial^\nu\varphi\right)\delta^\alpha{}_\nu\right]=$$

$$=\frac{1}{2}\partial_\alpha\left[\left(\partial^\alpha\varphi\right)+\left(\partial^\alpha\varphi\right)\right]=$$

$$=\partial_\alpha\partial^\alpha\varphi=\Box\varphi\,.$$

The result is once again the wave equation

$$\Box\varphi=0\,.$$

The shortest solution:

$$\partial_\alpha\left[\frac{\partial\mathscr{L}}{\partial\varphi_{,\alpha}}\right]-\frac{\partial\mathscr{L}}{\partial\varphi}=\frac{1}{2}\partial_\alpha\frac{\partial}{\partial\varphi_{,\alpha}}\left(\varphi_{,\beta}\varphi^{,\beta}\right)=\frac{1}{2}\partial_\alpha 2\varphi^{,\alpha}=\partial_\alpha\partial^\alpha\varphi=\Box\varphi\,.$$

◗

Klein–Gordon field

► $$\mathscr{L} = \frac{1}{2}\left(\partial_\mu \varphi\right)\left(\partial^\mu \varphi\right) + \frac{1}{2}\kappa^2\varphi^2 \quad \Rightarrow \quad (\Box - \kappa^2)\varphi = 0\,. \tag{1.226}$$

The Lagrangian of this field is quadratic in both its derivatives and the field itself. The second term would correspond to the potential energy density in classical mechanics ($L = T - V$). The resulting Lagrangian equation is linear and is suitable, for example, for plasma waves or for the quantum description of particles with zero spin.

Mexican hat potential

► $$\mathscr{L} = \frac{1}{2}\left(\partial_\mu \varphi\right)\left(\partial^\mu \varphi\right) + \frac{1}{2}\kappa^2\varphi^2 - \frac{\beta}{4}\varphi^4 \quad \Rightarrow \quad \Box\varphi - \kappa^2\varphi + \beta\varphi^3 = 0\,. \tag{1.227}$$

If we interpret the second term as the potential energy density, we obtain the value

► $$\mathscr{V} = -\frac{1}{2}\kappa^2\varphi^2 + \frac{\beta}{4}\varphi^4\,.$$

We have already encountered a similar function of a real variable in the Section 1.5.2.

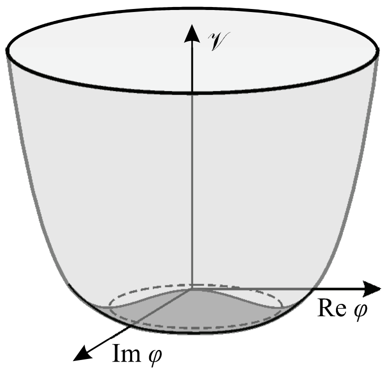


Fig. 1.45: Mexican hat potential

The potential has cylindrical symmetry and an infinite number of minima localized on a circle of radius R. Selecting one of these minima breaks the cylindrical symmetry. The corresponding Lagrangian equation is nonlinear. The nonlinear term can eliminate the spreading of the wave packet (dispersion) caused by the linear term. The equation therefore has some solutions in the form of a soliton – a wave packet with a constant shape.

Sin-Gordon equation

► $$\mathscr{L} = \frac{1}{2}\left(\partial_\mu \varphi\right)\left(\partial^\mu \varphi\right) - \kappa^2\cos\varphi \quad \Rightarrow \quad \Box\varphi - \kappa^2\sin\varphi = 0\,. \tag{1.228}$$

This well-known nonlinear equation again yields soliton solutions. If we expand the trigonometric function into a linear term, we obtain the Klein–Gordon equation; if we expand it into the first two terms, we obtain an equation corresponding to the "Mexican hat" potential.

1.6.2 Canonically Conjugate Fields

Just as we previously introduced the canonically conjugate momentum and then the energy for a given generalized coordinate, we can also define the canonically conjugate energy field and energy density in the continuous case using the following relations

► $$\pi_k(t,\mathbf{x}) \equiv \frac{\partial \mathscr{L}}{\partial \varphi_{k,t}}; \tag{1.229}$$

► $$\mathscr{H}(t,\mathbf{x}) \equiv \sum_k \left(\frac{\partial \mathscr{L}}{\partial \varphi_{k,t}} \varphi_{k,t} \right) - \mathscr{L}. \tag{1.230}$$

Poisson brackets between fields and conjugate fields yield the same result as before

► $$\begin{gathered} \{\varphi_k(t,\mathbf{x}), \varphi_l(t,\mathbf{x}')\} = 0, \\ \{\pi_k(t,\mathbf{x}), \pi_l(t,\mathbf{x}')\} = 0, \\ \{\varphi_k(t,\mathbf{x}), \pi_l(t,\mathbf{x}')\} = \delta_{kl}\delta(\mathbf{x}-\mathbf{x}'). \end{gathered} \tag{1.231}$$

The notation for the equation describing the time evolution of fields φ_k is similar

► $$\dot{\varphi}_k = \{\varphi_k, \mathscr{H}\}. \tag{1.232}$$

In all these expressions, the covariance (the same form in different coordinate systems) of the equations with respect to the Lorentz transformation seems to have been lost; expressions (1.229) through (1.232) do not appear to be relativistic. However, this is only an illusion. If we introduce the energy-momentum tensor via the relation (which is a relativistic generalization of relations 1.226 and 1.227)

$$T^{\alpha}{}_{\beta} \equiv \frac{1}{c}\sum_k \left(\frac{\partial \mathscr{L}}{\partial(\partial_\alpha \varphi_k)} \partial_\beta \varphi_k \right) - \frac{1}{c} g^{\alpha}{}_{\beta} \mathscr{L}, \tag{1.233}$$

the energy density of the field will be given by the component

$$\mathscr{H} = cT^0{}_0 \tag{1.234}$$

and the field momentum density will be proportional to the canonically conjugate fields

$$\mathscr{P}_n = T^0{}_n = \sum_k \pi_k \partial_n \varphi_k \qquad \Rightarrow \qquad \vec{\mathscr{P}} = \sum_k \pi_k \vec{\nabla} \varphi_k. \tag{1.235}$$

The quantities $T^l{}_0$ represent the energy flux and $T^l{}_n$ the momentum flux. From the expressions for the tensor (1.233) and the field equations (1.221), it can be easily shown that the energy-momentum tensor satisfies the continuity equation

$$\partial_\alpha T^{\alpha}{}_{\beta} = 0. \tag{1.236}$$

Other relations can also be easily rewritten in relativistic form.

1.6.3 Maxwell Equations, Electromagnetic Field

We usually describe electromagnetic fields using Maxwell equations

► $$\operatorname{div}\mathbf{B} = 0\,, \tag{1.237}$$

► $$\operatorname{rot}\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}\,, \tag{1.238}$$

► $$\operatorname{div}\mathbf{D} = \rho_Q\,, \tag{1.239}$$

► $$\operatorname{rot}\mathbf{H} = \mathbf{j}_Q + \frac{\partial \mathbf{D}}{\partial t}\,, \tag{1.240}$$

which we will supplement with material relationships

► $$\begin{aligned} \mathbf{D} &= \varepsilon_0 \mathbf{E} + \mathbf{P}\,, \\ \mathbf{B} &= \mu_0 (\mathbf{H} + \mathbf{M})\,, \end{aligned} \tag{1.241}$$

where the vector **P** is the polarization of the medium (electric dipole moment density) and **M** is the magnetization (magnetic dipole moment density).

Field potentials

Equation (1.237) implies the existence of a function $\mathbf{A}(t, \mathbf{x})$ such that

► $$\mathbf{B} = \operatorname{rot}\mathbf{A}\,. \tag{1.242}$$

The equation is then satisfied automatically, since the divergence of the curl of each function is zero. The quantity **A** is called the vector potential. If we substitute expression (1.242) into equation (1.238), we obtain the relation

$$\operatorname{rot}\left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}\right) = 0\,,$$

from which it follows that there exists a function ϕ such that $\mathbf{E} + \partial\mathbf{A}\partial t = -\nabla\phi$. Equation (1.238) is again satisfied automatically, and for the electric field we have the expression

► $$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}\,. \tag{1.243}$$

We call the function ϕ the scalar potential; in the case of stationary fields, (1.243) reduces to the well-known relation

$$\mathbf{E} = -\nabla\phi\,, \tag{1.244}$$

where the minus sign reflects the fact that the force points toward the minimum of the potential energy. We can thus describe the electromagnetic field using just four quantities: the scalar and vector potentials. These four quantities form a relativistic four-vector (see Sec. 3.3.6 Tensors and Metrics; for more details, see the companion textbook [1])

► $$A^{\mu} = \begin{pmatrix} \phi/c \\ \mathbf{A} \end{pmatrix}. \tag{1.245}$$

If we know the four-vector A^{μ}, we can easily determine the electric and magnetic fields from equations (1.243) and (1.242). In a certain sense, the vectors **E** and **B** are preferred over the vectors **D** and **H**, since we can determine them directly from the potentials. The electromagnetic field is a derivative of the potentials; both equations (1.242) and (1.243) can be easily written using the electromagnetic field tensor

► $$F^{\mu\nu} \equiv \partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu} = \begin{pmatrix} 0 & E^x/c & E^y/c & E^z/c \\ -E^x/c & 0 & B^z & -B^y \\ -E^y/c & -B^z & 0 & B^x \\ -E^z/c & B^y & -B^x & 0 \end{pmatrix}. \tag{1.246}$$

It is a second-order antisymmetric tensor that has only six independent components (namely, the electric and magnetic fields). The field components can be easily derived from the corresponding entries of the tensor.

Ambiguity of potentials, calibration invariance

As we have seen earlier, potentials are not uniquely determined; two different potentials may correspond to the same electromagnetic field. If we introduce new, transformed potentials using the so-called *gradient transformation*

► $$\tilde{A}^{\mu} \equiv A^{\mu} + \partial^{\mu} f\,, \tag{1.247}$$

where f is an arbitrary twice-differentiable function, the field remains unchanged:

$$\tilde{F}^{\mu\nu} = \partial^{\mu}\left(A^{\nu} + \partial^{\nu} f\right) - \partial^{\nu}\left(A^{\mu} + \partial^{\mu} f\right) = \partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu} = F^{\mu\nu}\,. \tag{1.248}$$

This arbitrariness in potentials can be advantageously exploited in formulating the simplest possible version of Maxwell equations in terms of potentials.

Maxwell equations expressed in terms of a field tensor

We used Maxwell equations (1.237) and (1.238) to introduce the electromagnetic field potentials. Using the electromagnetic field tensor, we can easily rewrite the remaining two equations containing source terms into the form

► $$F^{\mu\nu}{}_{,\nu} = \mu_0 j^{\mu}\,, \tag{1.249}$$

where the four-vector j^{μ} represents the sources of magnetic fields

$$j^{\mu} \equiv \begin{pmatrix} \rho_Q c \\ \mathbf{j}_Q \end{pmatrix}. \tag{1.250}$$

This form of Maxwell equations is clearly relativistic.

Maxwell equations in potentials

Let's rewrite Maxwell equations in the form (1.249) using potentials:

$$F^{\mu\nu}{}_{,\nu} = \mu_0 j^\mu \;;$$

$$\partial_\nu \left(\partial^\mu A^\nu - \partial^\nu A^\mu\right) = \mu_0 j^\mu \;;$$

$$\partial^\mu \partial_\nu A^\nu - \partial_\nu \partial^\nu A^\mu = \mu_0 j^\mu \;. \tag{1.251}$$

The second term on the left-hand side is D'Alembert wave operator applied to the four-potential; the right-hand side clearly represents the field sources. The only "flaw" in the equations written in potentials is the first term. Here we will take advantage of the considerable arbitrariness in the potentials given by the calibration transformation. Let us assume that the quantity $\partial_\nu A^\nu$ is equal to some function of time and space $F(t, \mathbf{x})$, i.e.

$$\partial_\nu A^\nu = F(t, \mathbf{x}) ,$$

and let us choose, using the gradient transformation (1.247), another four-potential for which we will require that

$$\partial_\nu \tilde{A}^\nu = 0 \qquad \Rightarrow \qquad \partial_\nu A^\nu + \partial_\nu \partial^\nu f = 0 \qquad \Rightarrow \qquad F(t, \mathbf{x}) + \Box f = 0 \,.$$

Such a gradient transformation will always exist. The function f that generates the transformation need only be chosen so that it satisfies the equation

$$\Box f = -F(t, \mathbf{x}) \,. \tag{1.252}$$

In the new potentials, the first term in equation (1.251) is zero, and Maxwell equations take on a simple form

$$\Box \tilde{A}^\mu = -\mu_0 j^\mu \;; \tag{1.253}$$

$$\partial_\mu \tilde{A}^\mu = 0 \,. \tag{1.254}$$

These are ordinary wave equations for the four-potential A^μ, in which the source terms are the components of the four-vector j^μ. The equations are further supplemented by the Lorenz calibration condition (1.254). We have shown that the arbitrariness of the potentials can be used to ensure that the Lorentz calibration condition is satisfied. However, even the requirement that it be satisfied does not uniquely determine the potentials! From equation (1.252) , it is clear that the function f is not defined uniquely and that any solution to the wave equation can be added to it.

$$\Box f_0 = 0 \,. \tag{1.255}$$

Therefore, another gradient transformation is possible

$$\tilde{\tilde{A}}^\mu = \tilde{A}^\mu + \partial^\mu f_0 ,$$

which can be used, for example, to set the scalar potential to zero. We conclude this section by noting that we can always choose potentials such that Maxwell equations take a simple form

►
$$\begin{aligned} \Box A^\mu &= -\mu_0 j^\mu \;; \\ \partial_\mu A^\mu &= 0 \,. \end{aligned} \tag{1.256}$$

Lagrange formulation of Maxwell equations

The density of the Lagrangian function describing the interaction of charged particles with a field is given by

► $$\mathscr{L} = \mathscr{L}_{\text{part}} + \mathscr{L}_{\text{int}} + \mathscr{L}_{\text{field}}\ . \tag{1.257}$$

The first part describes the motion of particles, which we discussed in Section 1.4.1. The interaction term must be some combination of the four-current j^μ (which describes particles) and the four-potential A^μ (which describes the field). Simply taking the simplest scalar variant yields the correct field equations:

$$\mathscr{L}_{\text{int}} = j_\mu A^\mu\ . \tag{1.258}$$

For a point particle, the four-flux is given by the relation

$$\begin{pmatrix} Qc\delta(\mathbf{x}-\mathbf{x}') \\ Q\mathbf{v}\delta(\mathbf{x}-\mathbf{x}') \end{pmatrix}, \tag{1.259}$$

The vector $\mathbf{x}$ is the position vector of the particle, and the vector $\mathbf{x}'$ describes the observer's position. The Lagrangian function of interaction is given by the integral

$$L_{\text{int}} = \int \mathscr{L}_{\text{int}}\, \mathrm{d}^3\mathbf{x}' = \int j_\mu A^\mu\, \mathrm{d}^3\mathbf{x}' = \int \left(-Q\phi + Q\mathbf{A}\cdot\mathbf{v}\right)\delta(\mathbf{x}-\mathbf{x}')\,\mathrm{d}^3\mathbf{x}' \quad \Rightarrow$$

$$L_{\text{int}} = -Q\phi(t,\mathbf{x}) + Q\mathbf{A}(t,\mathbf{x})\cdot\mathbf{v}\ . \tag{1.260}$$

The field term of the Lagrangian must consist of the electromagnetic field tensor; the simplest scalar is the combination $F_{\mu\nu}F^{\mu\nu}$, and the field term of the Lagrangian should be proportional to this expression. We determine the proportionality constant so that we obtain the correct field equations (in this case, Maxwell equations):

► $$\mathscr{L}_{\text{field}} = -\frac{1}{4\mu_0} F_{\mu\nu}F^{\mu\nu} \tag{1.261}$$

Both components of the Lagrangian density for the electromagnetic field are

► $$\mathscr{L}_{\text{elmg}} = \mathscr{L}_{\text{field}} + \mathscr{L}_{\text{int}} = -\frac{1}{4\mu_0} F_{\mu\nu}F^{\mu\nu} + j_\mu A^\mu \tag{1.262}$$

Finally, let's verify that the field Lagrange equations yield Maxwell equations:

$$\partial_\alpha \left[\frac{\partial \mathscr{L}}{\partial A^\beta{}_{,\alpha}}\right] - \frac{\partial \mathscr{L}}{\partial A^\beta} = 0\ ;$$

$$-\frac{1}{4\mu_0}\partial^\alpha \left[\frac{\partial\left(F_{\mu\nu}F^{\mu\nu}\right)}{\partial A^{\beta,\alpha}}\right] - j_\beta = 0\ ;$$

$$-\frac{1}{4}\partial^\alpha \left[2F_{\mu\nu}\frac{\partial F^{\mu\nu}}{\partial A^{\beta,\alpha}}\right] = \mu_0 j_\beta\ ;$$

$$-\frac{1}{2}\partial^\alpha\left[\left(\partial_\mu A_\nu-\partial_\nu A_\mu\right)\frac{\partial}{\partial(\partial^\alpha A^\beta)}\left(\partial^\mu A^\nu-\partial^\nu A^\mu\right)\right]=\mu_0 j_\beta\,;$$

$$-\frac{1}{2}\left[\partial^\alpha\left(\partial_\mu A_\nu-\partial_\nu A_\mu\right)\left(\delta^\mu{}_\alpha\delta^\nu{}_\beta-\delta^\nu{}_\alpha\delta^\mu{}_\beta\right)\right]=\mu_0 j_\beta\,;$$

$$-\frac{1}{2}\partial^\alpha\left[\partial_\alpha A_\beta-\partial_\beta A_\alpha+\partial_\alpha A_\beta-\partial_\beta A_\alpha\right]=\mu_0 j_\beta\,;$$

$$-\partial^\alpha F_{\alpha\beta}=\mu_0 j_\beta\,;$$

$$-F^{\alpha\beta}{}_{,\alpha}=\mu_0 j^\beta\,;$$

$$F^{\beta\alpha}{}_{,\alpha}=\mu_0 j^\beta\,.$$

which are Maxwell equations in the form of (1.249).

Lorentz's equation of motion

Of course, Lorentz equation of motion (1.47) still applies to the motion of particles, and it can be easily written using the electromagnetic field tensor. Let us first introduce the particle's proper time $d\tau$ as the time passing directly at the particle. For the interval at the particle's location, we have

$$\mathrm{d}s^2=-c^2\mathrm{d}t^2+\mathrm{d}\mathbf{x}^2=-c^2\mathrm{d}\tau^2\,. \tag{1.263}$$

There is therefore a relationship between the proper and the laboratory time.

$$\begin{aligned}&-c^2\mathrm{d}\tau^2=-c^2\mathrm{d}t^2+\mathrm{d}\mathbf{x}^2\qquad\Rightarrow\\&\mathrm{d}\tau=\sqrt{1-v^2/c^2}\;\mathrm{d}t=\frac{\mathrm{d}t}{\gamma}\end{aligned}\;. \tag{1.264}$$

Proper time is an invariant that we will use when introducing four-velocity and four-momentum:

► $$U^\alpha\equiv\frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\,; \tag{1.265}$$

► $$P^\alpha\equiv m_0U^\alpha\,. \tag{1.266}$$

The equation of motion for a charged particle is then given by

► $$\frac{\mathrm{d}P^\alpha}{\mathrm{d}\tau}\equiv QF^{\alpha\beta}U_\beta\,. \tag{1.267}$$

By simply substituting the electromagnetic field tensor (1.246) and using the identity $\mathrm{d}/\mathrm{d}\tau=\gamma\,\mathrm{d}/\mathrm{d}t$, we can easily verify that equation (1.267) is an elegant reformulation of the Lorentz equation of motion.

Summary

Let us now list the key Lagrangian functions and action integrals $S = \int L\, \mathrm{d}t = \int \mathscr{L}\, \mathrm{d}^4x$ for electricity and magnetism

	Particle	Interaction	Field
$\mathscr{L}$	–	$j_\mu A^\mu$	$-\dfrac{1}{4\mu_0} F_{\mu\nu} F^{\mu\nu}$
L	$-m_0 c^2 \sqrt{1-\dfrac{\mathbf{v}^2}{c^2}}$	$-Q\phi + Q\mathbf{A}\cdot\mathbf{v}$	–
S	$-m_0 c^2 \int \mathrm{d}\tau$	$\int j_\mu A^\mu\, \mathrm{d}^4x$	$-\dfrac{1}{4\mu_0}\int F_{\mu\nu}F^{\mu\nu}\, \mathrm{d}^4x$

For a material point, the density of the Lagrangian makes no sense, even though it is possible to write it down in principle. In expressing the particle Lagrangian, we used the relation $\mathrm{d}\tau = (1-\mathbf{v}^2/c^2)^{1/2}\mathrm{d}t$. Similarly, the total Lagrangian makes no sense for a field that is spread out in time and space. If we consider only the particle Lagrangian, we obtain the equation of motion for a free particle:

$$\frac{\mathrm{d}P^\alpha}{\mathrm{d}\tau} = 0 \,. \tag{1.268}$$

If we consider the Lagrangian functions for the particle and for the interactions, we obtain the equation of motion

$$\frac{\mathrm{d}P^\alpha}{\mathrm{d}\tau} = QF^{\alpha\beta}U_\beta \,. \tag{1.269}$$

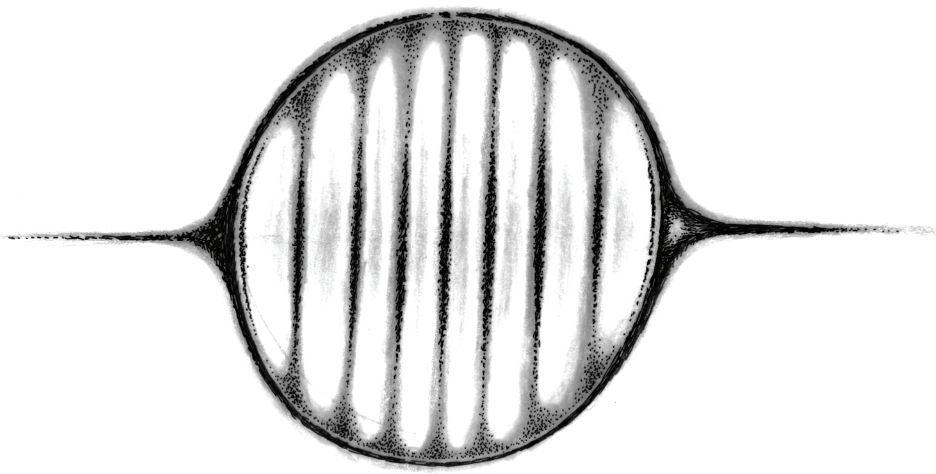
If we consider only the Lagrangian for the fields, we obtain Maxwell equations in a vacuum:

$$F^{\mu\nu}{}_{,\nu} = 0 \,. \tag{1.270}$$

And if we consider the field and interaction parts of the Lagrangian, we obtain Maxwell equations with source terms:

$$F^{\mu\nu}{}_{,\nu} = \mu_0 j^\mu \,. \tag{1.271}$$

2. Quantum Theory



2.1 Introduction

2.1.1 Microworld and Macroworld

Theoretical mechanics is based on how we perceive the world around us on our scale, in what is known as the macroworld. If we attempt to apply the laws of theoretical mechanics to objects of minute dimensions, such as atoms or particles (the so-called *microworld*), the predictions will no longer agree with experimental results. For example, the very act of measurement can influence objects in the microworld. If we want to determine the position of a soccer ball, we capture photons reflected from the ball with our eyes and process the information. If we want to determine the position of an electron, the reflected photon – from which we infer the position – imparts a non-negligible momentum to the electron and changes its state. Perhaps the greatest difference between phenomena in the macroworld and the microworld relates to *commutativity*. In the macroscopic world, we have grown accustomed to the fact that the phenomena we observe are commutative – the order does not matter. It makes no difference whether we first perform measurement A and then measurement B, or vice versa. In short, $AB = BA$. In the microworld, however, this is not the case. The act of measurement affects the state of objects, and it matters which measurement we perform first. This is also the main reason for the failure of classical mechanics in describing the microworld. Classical mechanics is based on commutative mathematical objects. The only non-commutative structure in mechanics is the *Poisson bracket*, and it is only auxiliary.

The first phenomena in the microworld that stood in stark contrast to classical mechanics were discovered in the early 20$^{\text{th}}$ century. Their analysis led to the birth of quantum theory – one of the two most successful theories in human history (the other being Einstein general theory of relativity). The predictions of today's quantum theory agree with experimental results to many significant digits. The fundamental equations and relationships remain consistent with classical mechanics, but they apply to entirely different objects. For example, the Lie algebra of Poisson brackets is applied to operators that represent dynamic variables. Therefore, it is necessary to thoroughly familiarize yourself with Sections 3.4 and 3.3.5 before studying quantum theory.

Let's now look at some of the fundamental differences between the world of the *microworld* and the *macroworld*:

► Discrete values of certain dynamic variables

In some situations, we can measure only certain values. Most often, these are energy or angular momentum. In the macroworld, measured values are always continuous. More precisely, they are so close together that we are unable to distinguish between them.

► Wave-particle duality

Objects in the microworld sometimes behave like waves and sometimes like particles. For example, light manifests itself as particles (photons) in the photoelectric effect and as waves during interference or diffraction. And it's not just light. An electron, which we think of as a particle, behaves as a wave in an electron microscope and exhibits wave-like properties in other situations. The same is true for neutrons and other particles.

► Non-commutativity of the measurement

When measuring the values of two dynamic variables (such as position and velocity), the result may depend on the order in which the measurements are taken. This is because the act of measurement affects the state of the system; after the measurement, the system is generally in a different state than before the measurement.

► Uncertainty relations

In some cases, increasing the measurement accuracy of one dynamic variable will reduce the measurement accuracy of another dynamic variable. These measurements influence each other and are non-commutative. The measurement of a given generalized coordinate and its corresponding generalized momentum will always be affected.

► Indeterminacy of quantum theory

Two experiments conducted under identical conditions may yield different results. When we conduct many experiments, we find that the results are probabilistic in nature. We are therefore only able to predict the probability of observing a particular possible outcome, not which specific outcome will occur.

► Superposition of states

In the macroworld, you can never be in two places at once – for example, in a lecture hall and at a restaurant. In the microworld, it is possible. Objects in the microscopic world can be, and indeed commonly are, in a superposition of two or more states.

2.1.2 Experiments Leading to Quantum Theory

Let's now summarize the experimental facts that led to the birth of quantum theory:

Blackbody radiation

In an absolute black body (such as any star), matter and radiation are in equilibrium at a specific temperature T. If we observe the radiation from an absolute black body, we find that it emits radiation with varying intensities at different frequencies. The experimentally observed curve of energy emitted per unit frequency is shown in the figure. Theoretical calculations of the blackbody radiation curve, performed by Lord Rayleigh, James Jeans, and Wilhelm Wien, led to different dependencies. They either diverged in the infrared (IR) or in the ultraviolet (UV) region of the spectrum.

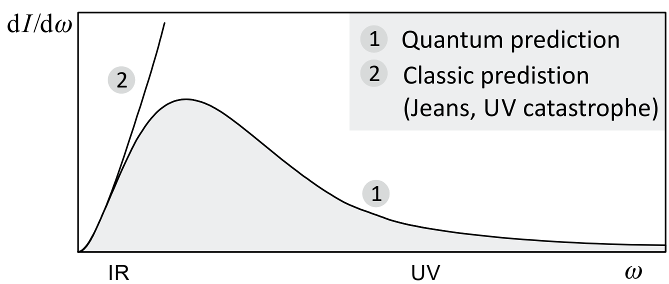


Fig. 2.1: Blackbody radiation

It was not until August 1900 that Max Planck guessed the correct formula by comparing various functions with the measured data. His result was: $\mathrm{d}I/\mathrm{d}\omega \sim \omega^3 \exp[-\,\mathrm{const}\;\omega/T\,]$. Over the next two months, Planck also derived this relationship theoretically, assuming that the energy of light of a certain frequency ω does not vary continuously but is an integer multiple of the fundamental energy quantum

► $$E = \hbar\omega; \qquad \hbar = 1{,}05457\times10^{-34}\ \mathrm{Js}. \tag{2.1}$$

The quantity $\hbar$ is called the reduced Planck constant (sometimes also the Planck-Dirac constant). Planck originally used the assumption of energy quantization to simplify mathematical calculations. It later turned out that the energy of electromagnetic radiation of a certain frequency is indeed quantized, i.e., its observed values are not continuous but change in discrete steps by the fundamental energy quantum $\hbar\omega$.

Note: In his calculations, Planck used ordinary frequency instead of angular frequency, and thus the proportionality constant he introduced had a different value:

$$E = h\nu; \qquad h = 2\pi\hbar = 6.62607\times10^{-34}\ \mathrm{Js}. \tag{2.2}$$

The reduced Planck constant has a real physical significance (it is the elementary quantum of angular momentum); we denote it by the barred letter ħ and will use it throughout this textbook.

Photoelectric effect

When light (electromagnetic radiation) strikes the surface of a metal, an electron may be knocked out of the metal and leave the surface. Electrons are released from the metal at light frequencies higher than the threshold frequency ω_0, which is characteristic of that particular metal. If light with a frequency lower than the threshold is used, no electrons will be emitted, no matter how intense the light is. This experiment contradicts the idea of light as an electromagnetic wave. The photoelectric effect should occur at any frequency, and sufficient energy for emission should be obtainable by increasing the intensity of the incident light.

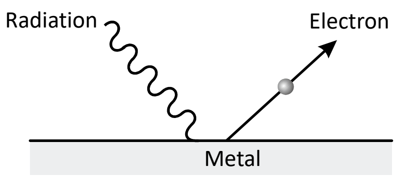


Fig. 2.2: Photoelectric effect

Albert Einstein provided the solution in 1905. Electromagnetic waves behave like particles in the photoelectric effect. Today, we call these particles *photons*. The energy of a single photon of radiation with frequency ω is precisely the energy of a single energy quantum (2.1). The explanation of the photoelectric effect is now very simple. On the surface of a metal, a photon collides with an electron. For a photon to knock out an electron, it must have higher energy than the binding energy of the electron in the metal: $\hbar\omega \geq E_\mathrm{i}$. The threshold frequency is obviously $\omega_0 = E_\mathrm{i}/\hbar$. The total energy balance for the photon and the electron

► $$\hbar\omega = E_\mathrm{i} + \frac{1}{2} m_\mathrm{e} v^2 \qquad (2.3)$$

is known as Einstein equation for the photoelectric effect. The energy of the incident photon is used to eject an electron from the metal and to provide the kinetic energy of the emitted electron. We can therefore consider electromagnetic waves to be a collection of photons. This is why, even in the radiation of an absolutely black body, the energy of radiation at a given frequency changes in discrete steps – each step represents the addition or subtraction of a single photon.

Note: Albert Einstein was awarded the 1921 Nobel Prize in Physics for his explanation of the photoelectric effect. Although his theory of general relativity was a more significant achievement, at the time many physicists regarded it more as a controversial hypothesis than as a new theory of gravitational interaction.

Compton effect

In 1923, Arthur Compton discovered that X-rays reflected from the surface of graphite change their wavelength. According to classical theory, the waves should excite the surface electrons, which would then generate a wave of the same frequency. The explanation turned out to be simple. The photons behave as particles again; they collide with electrons and lose some of their energy during the collision, which is why their wavelength changes. In extremely hot plasma, the opposite phenomenon can occur: here, photons gain energy when they collide with energetic electrons; this is known as the *inverse Compton effect*.

Electron diffraction

The photoelectric effect has shown that waves can behave like particles in certain situations. Conversely, particles sometimes behave like waves. For example, a beam of electrons passing through a single slit or a double slit produces a characteristic diffraction pattern when it strikes a screen. We cannot predict in advance where each electron will strike, but with a large number of electrons, we can determine the probabilities of impact at a specific location on the screen. The resulting diffraction pattern is therefore a typical statistical phenomenon.

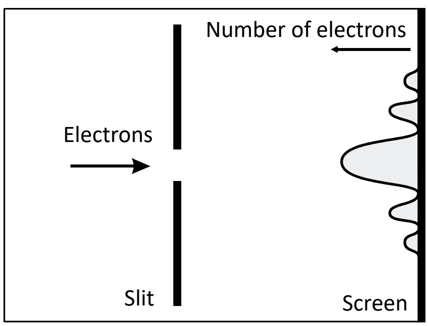


Fig. 2.3: The diffraction of electrons at a slit

Today, the wave properties of electrons are utilized, for example, in electron microscopes. Electrons have a significantly shorter wavelength than visible light, and therefore the resolving power of an electron microscope is considerably higher than that of an optical microscope. The wave properties of electrons were first observed by American physicists Clinton Davisson and Lester Germer in 1927. They were studying the reflection of electrons off the surface of nickel. After annealing the nickel, recrystallization occurred, and the reflected electrons began to exhibit a diffraction pattern on precise, large crystals.

Note: We describe particles using a four-vector of quantities (E, $\mathbf{p}$). The definitions of energy E and momentum $\mathbf{p}$ are related to symmetries under time and space translations (see Noether theorem in Chapter 1.2). We describe waves using a four-vector of quantities (ω, $\mathbf{k}$). The angular frequency ω is defined as the change of the wave phase with respect to time, $\omega = \partial\varphi/\partial t$, and the wave vector $\mathbf{k}$ is the change of the wave phase with respect to spatial coordinates, $\mathbf{k} = \partial\varphi/\partial\mathbf{x}$. For a periodic process with constant period T in time and λ in space (wavelength), we can write $\omega = 2\pi/T$, $k = 2\pi/\lambda$. Louis de Broglie proposed the hypothesis that objects in the microworld behave both as waves and as particles (we refer to this as wave-particle duality). Today, we write the conversion relationship in the form:

► $$E = \hbar\omega, \qquad \mathbf{p} = \hbar\mathbf{k}. \tag{2.4}$$

If we were to use a natural system of units chosen such that $\hbar = 1$, the relations in (2.4) would take an even simpler form. We are often interested in the wavelength of the wave corresponding to a specific particle, such as an electron in an electron microscope. From the second relation in (2.4), we have $mv = 2\pi\hbar/\lambda$, and thus

$$\lambda = \frac{2\pi\hbar}{mv}. \tag{2.5}$$

Existence of the atom

According to the classical planetary model of the atom, negatively charged electrons orbit a positively charged nucleus, just as the planets in the solar system orbit the Sun. The planets are held in their orbits by gravitational force, while for electrons in the electron shell, the centrifugal force is balanced by the attractive Coulomb force.

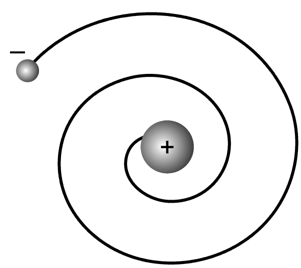


Fig. 2.4: The paradox of the atomic shell

There is, however, a fundamental difference between gravitational and electromagnetic phenomena. Maxwell theory of the electromagnetic field implies that every charged particle moving with acceleration emits electromagnetic waves and thus loses energy. During the circular motion of an electron around the nucleus, the direction of velocity changes; the acceleration dv/dt is non-zero (pointing toward the center of the atom, i.e., centripetal acceleration), and the electron loses energy through radiation. It should therefore move in a spiral and eventually collide with the atomic nucleus. For hydrogen, for example, this process should take only 10^{-11} seconds. According to classical theory, therefore, no atoms should exist after a very short time! The Danish physicist Niels Bohr was the first to point out this paradox.

Niels Bohr developed the so-called *Bohr model of the atom* based on three artificial postulates that he added to classical theory:

1) Electrons move only along so-called *stationary orbits* – that is, orbits in which the corresponding de Broglie wavelength, as given by the equation (2.5), is "wrapped" around the orbit, meaning that the circumference of the orbit is an n-fold multiple of the wavelength. The result is a simple quantization condition

$$2\pi r_n = n\lambda; \qquad \lambda = \frac{2\pi\hbar}{mv_n}. \tag{2.6}$$

The index n numbers the possible energy levels of an electron in an atom (r_n is the possible orbital radius, v_n is the velocity in the n^{th} orbital, and E_n is the corresponding energy) according to the number of wavelengths of the electron in its orbit.

2) An electron does not emit light in a stationary orbit.

3) When an electron jumps between two stationary energy levels, a photon is emitted with an energy equal to the difference in energy between those levels.

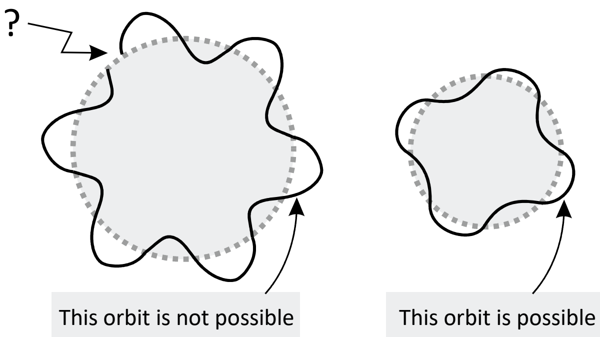


Fig. 2.5: Bohr's atomic model

This simple Bohr model of the atom does not resolve the above paradox; rather, it is a postulate or a statement of experimentally known facts. Furthermore, this model is applicable only to the simplest atoms with a single electron in the atomic shell (H, He^{+}). However, this simple model was the first to correctly determine the energy levels of the electron in the hydrogen atom and explain the spectrum of the hydrogen atom.

Heisenberg uncertainty principle

When measuring the position and momentum of a microscopic object, the measurement uncertainties Δx and Δp will satisfy the relation discovered by Werner Heisenberg

► $$\Delta x \, \Delta p \geq \frac{\hbar}{2}. \tag{2.7}$$

The more precisely we determine an object's position, the less precisely we can determine its corresponding momentum, and vice versa. The very act of measurement affects our object, but the relation holds even if we do not perform the measurement at all. This is a fundamental limit imposed by nature, beyond which we cannot see. This relation holds for any generalized coordinate and its corresponding generalized momentum.

For example, the simple diffraction of light at a slit can be understood as a consequence of the uncertainty principle for photons. The passage of photons through a slit is nothing more than an attempt to determine their position y with a precision of Δy (the width of the slit). The photons that passed through the slit certainly had a y-coordinate equal to the slit's y-coordinate at the moment of passage. If we reduce the slit width by Δy, we increase the precision of the y measurement; however, according to relations (2.7), the uncertainty Δp_y in determining the corresponding component of momentum will increase. The result is the well-known diffraction phenomenon – photons emerge from the slit in various directions with a mean square fluctuation in momentum Δp_y given by Heisenberg uncertainty relations.

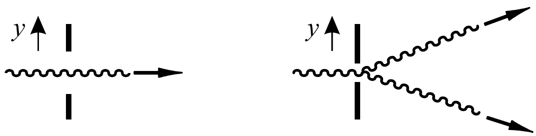


Fig. 2.6: Particle diffraction at a slit and the uncertainty principle

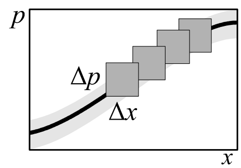
The list of experimental facts we have presented above is by no means exhaustive. However, they have all contributed to the emergence of quantum theory, which describes the world of atoms and elementary particles – a world that is unfamiliar to us.

2.2 Quantum Theory Basic Principles

Classical mechanics has failed to describe the phenomena of the microworld, primarily because it is based on commutative objects. In the microworld, however, phenomena do not commute. The fundamental goal will therefore be to use non-commutative objects (operators) instead of dynamic variables and to find the relationship between operators and actual measured quantities. Mathematical foundations for studying these phenomena can be found in Chapter 3.4, which you should review before continuing.

2.2.1 Basic Axioms and Definitions

I. Redefining the state



In classical mechanics, the state of a particle is determined by its position and momentum. In microworld these quantities cannot be measured simultaneously and measuring one affects the measurement of the other. It is necessary to redefine the concept of state. Phase trajectories can no longer be described by curves in the microworld. We see them with an accuracy given by the uncertainty relations $\Delta x\Delta p \geq \hbar/2$. We can imagine that we see the phase trajectory as a blurred line with a resolution given by a rectangle with an area of $\hbar/2$ (if we track one coordinate and its corresponding momentum). Let us first introduce some concepts.

Compatibility: We say that two dynamic variables are compatible if measuring one quantity does not affect the measurement of the other. An example of compatible variables is the coordinates (x, y); an example of incompatible variables is the coordinates and the corresponding momentum (x, p_x). Compatibility is a symmetric property:

$$(A \text{ comp } B) \quad \Rightarrow \quad (B \text{ comp } A) \ . \tag{2.8}$$

However, compatibility is not a transitive property. From $(x \text{ comp } y) \wedge (y \text{ comp } p_x)$ does not imply that $(x \text{ comp } p_x)$ must hold. In general, we can write

$$(A \text{ comp } B) \wedge (B \text{ comp } C) \quad \not\Rightarrow \quad (A \text{ comp } C) \ . \tag{2.9}$$

Complete set of observables: This is the maximal independent set of mutually compatible dynamic variables. Any additional dynamic variable is no longer compatible with them. In non-relativistic theory, the best-known complete sets of observables are (x, y, z) or (p_x, p_y, p_z). For the central field, the complete set of observables includes energy, square of the angular momentum, and one of its components (E, L^2, L_3). All three coordinates or all three components of momentum can be measured together. It is no longer possible to measure all three components of angular momentum.

State of the system: Let's say that we know the state of the system if we know the measurement result of some complete set of observables. We will therefore define the state as only what can actually be measured simultaneously.

A fundamental feature of the new theory must be non-commuting objects – operators. Instead of the dynamic variables of classical mechanics (coordinates, energy, etc.), we will use operators (the coordinate operator, the energy operator, etc.). The non-commutativity of these operators will express the non-commutativity of the act of measuring in the microworld. The quantities measured by an instrument in the microworld are real numbers, sometimes continuous (particle position), sometimes discrete (for example energy levels of an electron bound in an atom). How can we obtain a set of real numbers of continuous or discrete nature from a dynamic variable operator? Such a set is precisely the spectrum of Hermitian operators (see Section 3.4.5 Spectral Theory). *We will therefore assign Hermitian operators to dynamic variables*.

Every operator acts on elements of some Hilbert space $\mathcal{H}$. We must ask what meaning Hilbert space itself will have in our theory, as well as the vectors on which the operators act. We will see later that the choice of Hilbert space does not matter much. What is essential are the relationships between the dynamic variables, now operators. Quantum mechanics based on the $\mathcal{L}^2$ space of quadratically integrable functions is the well-known Schrödinger wave mechanics. The quantum theory based on the ℓ^2 space of sequences that are square-summable is Heisenberg matrix mechanics. At first glance, the two theories seem different. Nevertheless, the eigenvalues of the operators in both theories are the same, and thus both theories yield the same predictions. Hilbert space, with all its vectors and the operators acting on them, corresponds to the properties of the entire system in classical mechanics. *Instead of a system, we will therefore now speak about the Hilbert space of a system* (for example, the Hilbert space of an electron).

All that remains is to solve the final puzzle – what are the elements of Hilbert space for? As we noted in the introduction, in the microworld, the very act of measurement influences the state of the system. Before measurement, the system is in a different state than after measurement. In quantum theory, the act of measuring a dynamic variable is represented by the Hermitian operator of that variable. By applying this operator to an element of the space, we obtain a different element of that space. And that is exactly what we are looking for. *The elements (vectors) of the space thus represent the state of the system*. The act of measurement corresponds to the application of the relevant operator to the state (element of the space), and the new state is the element that arises from the application of the operator.

An eigenvalue represents a measured value and thus provides information about the state of the system. We know that the multiples of any eigenvector are themselves eigenvectors. Therefore, for a given eigenvalue, there exists a complete eigenvector (direction) in $\mathcal{H}$. Therefore, the state of the system must correspond to an entire beam in $\mathcal{H}$, not just a single vector. This brings us to the three fundamental *axioms* of quantum theory, which describe how classical and quantum concepts correspond to one another:

System	$\rightarrow$	Hilbert space $\mathcal{H}$
State of the system	$\rightarrow$	Ray generated by $\mid\psi>$
Dynamical variable A	$\rightarrow$	Hermitian operator $\hat{\mathbf{A}}$

An important principle is associated with the linearity of the theory being developed:

■ **Principle of superposition:** Let $\varphi\rangle \in \mathcal{H}$ a $|\psi\rangle \in \mathcal{H}$ represent two different states of the system. Then vector $\alpha_1 |\varphi\rangle + \alpha_2|\psi\rangle$ is also a physically realizable state. Without this requirement, it would not be possible to construct a linear theory. Moreover, this is the aforementioned property of the quantum world. A system in the microworld can (unlike in the macroscopic world) be in a superposition of multiple states.

II. Measurement in quantum theory

The act of measuring a dynamic variable A in a given state means applying the operator $\hat{\mathbf{A}}$ to that dynamic variable in the state $|\,\psi\,\rangle$. The operator $\hat{\mathbf{A}}$ and the state $|\,\psi\,\rangle$ must therefore unambiguously determine what can and cannot be measured on the system. The answer to this question is provided by the so-called *interpretation postulates*:

■ **Postulate A:** When measuring the dynamic variable A, we can only measure one of the eigenvalues $\{a_j\}$ of the $\hat{\mathbf{A}}$ operator of this dynamic variable:

$$\hat{\mathbf{A}}|\,j\,\rangle = a_j\ |\,j\,\rangle\,. \tag{2.10}$$

■ **Postulate B:** Observing the dynamic variable A on a system set to the eigenstate $|\,j\,\rangle$ of the operator $\hat{\mathbf{A}}$ will certainly result in the measurement of the eigenvalue a_j.

■ **Postulate C:** If the system is in a general state $|\,\psi\,\rangle \in \mathcal{H}$, repeated measurements of the quantity A yield different results a_j. The average value of these repeated measurements will be equal to

$$\langle\, A\,\rangle = \langle\,\psi\,|\,\hat{\mathbf{A}}\,|\,\psi\,\rangle\,. \tag{2.11}$$

Note 1: We shouldn't think of repeated measurements as constantly repeating the exact same measurements on the same system. In practice, that would be impossible. It is difficult to repeat a measurement on a single electron. We must have a large number of systems in the same state (such as a beam of electrons) and repeat the measurement on many different electrons (systems).

Note 2: The expression for the mean is the simplest possible expression consisting of the operator $\hat{\mathbf{A}}$ and the condition $|\,\psi >$, which yields a real number. It is customary to denote the mean by $\langle\,A\,\rangle$ or $\overline{A}$.

Note 3: We automatically assume that the state vectors are normalized to one. If the state vector is not normalized, we must divide the expression for the mean by the square of the norm of the state vector:

►
$$\langle\, A\,\rangle = \frac{\langle\,\psi\,|\,\hat{\mathbf{A}}\,|\,\psi\,\rangle}{\langle\,\psi\,|\,\psi\,\rangle}\,. \tag{2.12}$$

Note 4: The expression for the mean value written in the space $\mathcal{L}^2(\mathcal{R}^3)$ gives:

$$\langle\, A\,\rangle = \frac{\int \psi^*(\mathbf{x})\,\hat{\mathbf{A}}\,\psi(\mathbf{x})\,\mathrm{d}^3\mathbf{x}}{\int \psi^*(\mathbf{x})\,\psi(\mathbf{x})\,\mathrm{d}^3\mathbf{x}}\,. \tag{2.13}$$

Note 5: If all systems are in their eigenstates $|\,j\,\rangle$ of the operator $\hat{\mathbf{A}}$, the average value given by Postulate C naturally corresponds to the eigenvalue specified by Postulate B, and all measurements yield the same result in this exceptional case (no summation over j is performed):

$$\langle\, A\,\rangle = \frac{\langle\, j\,|\,\hat{\mathbf{A}}\,|\,j\,\rangle}{\langle\, j\,|\,j\,\rangle} = \frac{a_j\langle\, j\,|\,j\,\rangle}{\langle\, j\,|\,j\,\rangle} = a_j\,.$$

III. Statistical Interpretation of the state vector

If we expand the state vector into an orthonormal finite-dimensional basis | n > or an infinite-dimensional basis | x >, the expansion will take a very simple form (see Section 3.4.4 Expanding an Element to the Base)

$$|\psi\rangle = \sum |n\rangle\langle n|\psi\rangle = \sum \psi_n |n\rangle \text{ resp.}$$
$$|\psi\rangle = \int |x\rangle\langle x|\psi\rangle \mathrm{d}x = \int \psi(x)|x\rangle \mathrm{d}x . \tag{2.14}$$

We interpret the expansion coefficients ψ_n and $\psi(x)$, respectively, as the *probability amplitude* that the system is found in the state | n 〉, respectively | x 〉. This interpretation is justified by the fact that these are projections of the state vector onto the corresponding basis element. It immediately follows from the normality of the state that

$$\sum \psi_n^* \psi_n = 1 , \text{ resp.} \qquad \int \psi^*(x)\psi(x)\mathrm{d}x = 1 \tag{2.15}$$

and expressions

► $$w_n = \psi_n^* \psi_n , \text{ resp.} \qquad w(x) = \psi^*(x)\psi(x) \tag{2.16}$$

we therefore interpret as the probability of the state | n 〉 occurring, or the probability density of finding the system in the state | x 〉. The probabilities are automatically normalized to one. The second of the relations (2.14) represents the superposition of the system in several positions.

IV. Correspondence principle

The last of the fundamental principles of quantum theory is the principle of correspondence. It defines which parts of classical mechanics can be used in quantum theory.

■ **Correspondence principle for the basic relations.** The fundamental relationship between dynamic variables in classical mechanics and the corresponding operators in quantum mechanics may differ only in the order of the operators.

■ **Correspondence principle for Poisson brackets.** The structure of Poisson brackets in theoretical mechanics is identical to the structure of commutators in quantum theory:

$$\begin{matrix} A \to \hat{\mathbf{A}} & & \\ B \to \hat{\mathbf{B}} & \Rightarrow & \{A,B\} = C \to [\hat{\mathbf{A}},\hat{\mathbf{B}}] = k\hat{\mathbf{C}} . \\ C \to \hat{\mathbf{C}} & & \end{matrix}$$

The first part of the correspondence principle applies to simple relations between dynamic variables that do not involve derivatives. For example, the definition of the Hamiltonian function in a potential field V

$$H \equiv \frac{p_x^2 + p_y^2 + p_z^2}{2m} + V(x,y,z) \tag{2.17}$$

appears in the definition of the Hamiltonian operator

► $$\hat{\mathbf{H}} \equiv \frac{\hat{\mathbf{P}}_x^2 + \hat{\mathbf{P}}_y^2 + \hat{\mathbf{P}}_z^2}{2m} + V(\hat{\mathbf{X}},\hat{\mathbf{Y}},\hat{\mathbf{Z}}) . \tag{2.18}$$

The expression for potential energy is a typical operator function (see Sections 3.4.2 and 3.4.5). For expressions of the form $A = xp$, the quantum analogue cannot be uniquely determined. It can be either

$$\hat{\mathbf{A}} = \hat{\mathbf{X}}\hat{\mathbf{P}}, \text{ or } \hat{\mathbf{A}} = \hat{\mathbf{P}}\hat{\mathbf{X}}.$$

Operators do not commute, so their order matters. The correct version of the two possible ones must be chosen based on experiment. Similarly, different quantum theories can be derived from different Lagrangian functions of the same system, and the correct version must again be chosen based on how nature actually behaves.

The second part of the correspondence principle concerns Poisson brackets – expressions that, in classical mechanics, contain derivatives. In quantum theory, Poisson brackets correspond to the commutators of dynamical variables. However, one cannot equate the commutation relation with a Poisson bracket. There are two reasons for this:

1) *Dimensional*: Poisson brackets contain derivatives that introduce physical dimensions into expressions, whereas commutators do not. Therefore, it is necessary to use the dimensional conversion factor k.
2) *Fundamental:* In quantum theory, we can assign only Hermitian operators to dynamic variables (they have real eigenvalues, which we interpret as measurable values). If the operators corresponding to A and B are Hermitian, then the operator corresponding to C must also be Hermitian. This can again be ensured by using the constant k.

Let us now determine a condition on the constant k that follows from the requirement that the operators be Hermitian:

$$\begin{aligned} [\hat{\mathbf{A}},\hat{\mathbf{B}}] &= k\hat{\mathbf{C}} \\ \hat{\mathbf{A}}\hat{\mathbf{B}} - \hat{\mathbf{B}}\hat{\mathbf{A}} &= k\hat{\mathbf{C}} \qquad /\,^{\dagger} \\ \hat{\mathbf{B}}^{\dagger}\hat{\mathbf{A}}^{\dagger} - \hat{\mathbf{A}}^{\dagger}\hat{\mathbf{B}}^{\dagger} &= k^{*}\hat{\mathbf{C}}^{\dagger} \qquad /\,\hat{\mathbf{O}}^{\dagger} = \hat{\mathbf{O}} \\ \hat{\mathbf{B}}\hat{\mathbf{A}} - \hat{\mathbf{A}}\hat{\mathbf{B}} &= k^{*}\hat{\mathbf{C}} \\ -[\hat{\mathbf{A}},\hat{\mathbf{B}}] &= k^{*}\hat{\mathbf{C}}. \end{aligned}$$

In the derivation, we used equation (3.267) for the Hermitian conjugate of the product of two operators. If we compare the initial and final expressions, $k^* = -k$ must hold. However, only purely imaginary numbers satisfy this condition. The conversion constant must therefore take the form:

$$k = \mathrm{i}\hbar. \tag{2.19}$$

The constant $\hbar$ is a real number and is the only fundamental constant in quantum theory. This constant appears in all predictions of quantum theory (for example, in the energy spectrum of an electron bound in atoms, in the relations for blackbody radiation, in Heisenberg uncertainty relations, etc.). Its value can be measured experimentally based on these predictions and is equal to:

► $$\hbar = 1{,}054572 \times 10^{-34}\ \mathrm{J\,s}. \tag{2.20}$$

This is known as the reduced Planck constant. The correspondence principle for Poisson brackets can be briefly written as

► $$\{A,B\} \;\rightarrow\; \frac{1}{\mathrm{i}\hbar}[\hat{\mathbf{A}},\hat{\mathbf{B}}]. \tag{2.21}$$

These are the fundamental, irreducible principles upon which the theory is built. We could simply list the axioms, postulates, and principles marked with a square in this chapter. The supplementary texts merely aim to demonstrate that this particular choice of fundamental axioms is natural and will lead to the desired outcome. However, only experiments verifying the predictions derived from these principles can determine the validity of the fundamental principles.

2.2.2 Heisenberg Uncertainty Principle

It is easy to determine whether the measurements of two dynamic variables influence each other or not. All you need to know is the commutator of the operators of these variables. If this commutator is zero, then

$$\hat{\mathbf{A}}\hat{\mathbf{B}} = \hat{\mathbf{B}}\hat{\mathbf{A}} \tag{2.22}$$

and the measurements do not influence one another. The fundamental commutators for coordinates and momenta can be derived from the correspondence principle, while the others follow from the properties of the commutators. The following relation (1.42) holds for the Poisson brackets between coordinates and momenta:

$$\{x_k, x_l\} = \{p_k, p_l\} = 0\,, \qquad \{x_k, p_l\} = \delta_{kl}\,. \tag{2.23}$$

According to the principle of correspondence, the commutative relations are:

► $$\begin{gathered}[\hat{\mathbf{X}}_k,\ \hat{\mathbf{X}}_l\,] = [\,\hat{\mathbf{P}}_k,\ \hat{\mathbf{P}}_l\,] = 0,\\ [\hat{\mathbf{X}}_k,\ \hat{\mathbf{P}}_l\,] = \mathrm{i}\hbar\hat{\mathbf{1}}\delta_{kl}\,.\end{gathered} \tag{2.24}$$

It is clear that all three coordinates or momenta can be measured simultaneously. Furthermore, measurements such as the x coordinate and the p_y momentum do not influence each other. *The only measurements that influence each other* (non-zero commutator) *are the measurement of the generalized coordinate and its momentum*.

Equations (2.24) are the fundamental commutation relations in quantum theory. It would be possible to derive other, more complex commutation relations from Poisson brackets as well. However, it is more advantageous to derive them from the fundamental relations (2.24) and the properties of the Lie algebra. This frees us from classical mechanics, and we do not need to return to it for every commutation relation. Quantum mechanics begins to "take on a life of its own." What it has taken over from classical mechanics are only the relations (2.24) between coordinates and momenta.

The commutator between two components of angular momentum will be:

$$\begin{aligned}
&[\hat{\mathbf{L}}_1,\hat{\mathbf{L}}_2] = [\hat{\mathbf{X}}_2\hat{\mathbf{P}}_3 - \hat{\mathbf{X}}_3\hat{\mathbf{P}}_2, \hat{\mathbf{X}}_3\hat{\mathbf{P}}_1 - \hat{\mathbf{X}}_1\hat{\mathbf{P}}_3] =\\
&= [\hat{\mathbf{X}}_2\hat{\mathbf{P}}_3, \hat{\mathbf{X}}_3\hat{\mathbf{P}}_1] - [\hat{\mathbf{X}}_2\hat{\mathbf{P}}_3, \hat{\mathbf{X}}_1\hat{\mathbf{P}}_3] - [\hat{\mathbf{X}}_3\hat{\mathbf{P}}_2, \hat{\mathbf{X}}_3\hat{\mathbf{P}}_1] + [\hat{\mathbf{X}}_3\hat{\mathbf{P}}_2, \hat{\mathbf{X}}_1\hat{\mathbf{P}}_3] =\\
&= \hat{\mathbf{X}}_2[\hat{\mathbf{P}}_3, \hat{\mathbf{X}}_3\hat{\mathbf{P}}_1] + [\hat{\mathbf{X}}_2, \hat{\mathbf{X}}_3\hat{\mathbf{P}}_1]\hat{\mathbf{P}}_3 - \hat{\mathbf{X}}_2[\hat{\mathbf{P}}_3, \hat{\mathbf{X}}_1\hat{\mathbf{P}}_3] - [\hat{\mathbf{X}}_2, \hat{\mathbf{X}}_1\hat{\mathbf{P}}_3]\hat{\mathbf{P}}_3 -\\
&- \hat{\mathbf{X}}_3[\hat{\mathbf{P}}_2, \hat{\mathbf{X}}_3\hat{\mathbf{P}}_1] - [\hat{\mathbf{X}}_3, \hat{\mathbf{X}}_3\hat{\mathbf{P}}_1]\hat{\mathbf{P}}_2 + \hat{\mathbf{X}}_3[\hat{\mathbf{P}}_2, \hat{\mathbf{X}}_1\hat{\mathbf{P}}_3] + [\hat{\mathbf{X}}_3, \hat{\mathbf{X}}_1\hat{\mathbf{P}}_3]\hat{\mathbf{P}}_2 =\\
&= \hat{\mathbf{X}}_2\hat{\mathbf{X}}_3[\hat{\mathbf{P}}_3, \hat{\mathbf{P}}_1] + \hat{\mathbf{X}}_2[\hat{\mathbf{P}}_3, \hat{\mathbf{X}}_3]\hat{\mathbf{P}}_1 + \hat{\mathbf{X}}_3[\hat{\mathbf{X}}_2, \hat{\mathbf{P}}_1]\hat{\mathbf{P}}_3 + [\hat{\mathbf{X}}_2, \hat{\mathbf{X}}_3]\hat{\mathbf{P}}_1\hat{\mathbf{P}}_3 -
\end{aligned}$$

$$-\hat{\mathbf{X}}_2\hat{\mathbf{X}}_1[\hat{\mathbf{P}}_3,\hat{\mathbf{P}}_3]-\hat{\mathbf{X}}_2[\hat{\mathbf{P}}_3,\hat{\mathbf{X}}_1]\hat{\mathbf{P}}_3-\hat{\mathbf{X}}_1[\hat{\mathbf{X}}_2,\hat{\mathbf{P}}_3]\hat{\mathbf{P}}_3-[\hat{\mathbf{X}}_2,\hat{\mathbf{X}}_1]\hat{\mathbf{P}}_3\hat{\mathbf{P}}_3-$$
$$-\hat{\mathbf{X}}_3\hat{\mathbf{X}}_3[\hat{\mathbf{P}}_2,\hat{\mathbf{P}}_1]-\hat{\mathbf{X}}_3[\hat{\mathbf{P}}_2,\hat{\mathbf{X}}_3]\hat{\mathbf{P}}_1-\hat{\mathbf{X}}_3[\hat{\mathbf{X}}_3,\hat{\mathbf{P}}_1]\hat{\mathbf{P}}_2-[\hat{\mathbf{X}}_3,\hat{\mathbf{X}}_3]\hat{\mathbf{P}}_1\hat{\mathbf{P}}_2+$$
$$+\hat{\mathbf{X}}_3\hat{\mathbf{X}}_1[\hat{\mathbf{P}}_2,\hat{\mathbf{P}}_3]+\hat{\mathbf{X}}_3[\hat{\mathbf{P}}_2,\hat{\mathbf{X}}_1]\hat{\mathbf{P}}_3+\hat{\mathbf{X}}_1[\hat{\mathbf{X}}_3,\hat{\mathbf{P}}_3]\hat{\mathbf{P}}_2+[\hat{\mathbf{X}}_3,\hat{\mathbf{X}}_1]\hat{\mathbf{P}}_3\hat{\mathbf{P}}_2=$$
$$=0-\hat{\mathbf{X}}_2[\hat{\mathbf{X}}_3,\hat{\mathbf{P}}_3]\hat{\mathbf{P}}_1+0+0-0-0-0-0-0-0-0-0+0+0+\hat{\mathbf{X}}_1[\hat{\mathbf{X}}_3,\hat{\mathbf{P}}_3]\hat{\mathbf{P}}_2+0=$$
$$=-\mathrm{i}\hbar\hat{\mathbf{X}}_2\hat{\mathbf{1}}\hat{\mathbf{P}}_1+\mathrm{i}\hbar\hat{\mathbf{X}}_1\hat{\mathbf{1}}\hat{\mathbf{P}}_2=\mathrm{i}\hbar(\hat{\mathbf{X}}_1\hat{\mathbf{P}}_2-\hat{\mathbf{X}}_2\hat{\mathbf{P}}_1)=\mathrm{i}\hbar\hat{\mathbf{L}}_3.$$

It is clear that the procedure is very time-consuming, but straightforward. We gradually "break down" the commutation relation we are looking for according to the rules of Lie algebra into elementary relations between coordinates and momenta. Virtually all symbolically oriented programs or programming languages handle this task for us without any problems and include packages for computing commutation relations.

Commutation relations for the other components of angular momentum we can obtain through a simpler process: by cyclically permuting the coordinate axes (1 → 2 → 3 → 1). The complete commutation relations for angular momentum are then:

► $$[\hat{\mathbf{L}}_1,\hat{\mathbf{L}}_2]=\mathrm{i}\hbar\hat{\mathbf{L}}_3,\qquad[\hat{\mathbf{L}}_2,\hat{\mathbf{L}}_3]=\mathrm{i}\hbar\hat{\mathbf{L}}_1,\qquad[\hat{\mathbf{L}}_3,\hat{\mathbf{L}}_1]=\mathrm{i}\hbar\hat{\mathbf{L}}_2.\tag{2.25}$$

As a result, it is not possible to measure any two components of angular momentum simultaneously. Measuring one component will affect the measurement of any other component. Let's introduce the square of the angular momentum operator

$$\hat{\mathbf{L}}^2\equiv\hat{\mathbf{L}}_1^2+\hat{\mathbf{L}}_2^2+\hat{\mathbf{L}}_3^2\,.\tag{2.26}$$

As before, we will calculate the commutation relations of the square of the angular momentum with the individual components. This time, when "breaking down" the commutation relations, it is sufficient to arrive only at relations (2.25) for the angular momentum. We already know their result. We will do it for the third component:

$$[\hat{\mathbf{L}}^2,\hat{\mathbf{L}}_3]=[\hat{\mathbf{L}}_1^2+\hat{\mathbf{L}}_2^2+\hat{\mathbf{L}}_3^2,\hat{\mathbf{L}}_3]=[\hat{\mathbf{L}}_1^2,\hat{\mathbf{L}}_3]+[\hat{\mathbf{L}}_2^2,\hat{\mathbf{L}}_3]+[\hat{\mathbf{L}}_3^2,\hat{\mathbf{L}}_3]=$$
$$\hat{\mathbf{L}}_1[\hat{\mathbf{L}}_1,\hat{\mathbf{L}}_3]+[\hat{\mathbf{L}}_1,\hat{\mathbf{L}}_3]\hat{\mathbf{L}}_1+\hat{\mathbf{L}}_2[\hat{\mathbf{L}}_2,\hat{\mathbf{L}}_3]+[\hat{\mathbf{L}}_2,\hat{\mathbf{L}}_3]\hat{\mathbf{L}}_2+0+0=$$
$$-\mathrm{i}\hbar\hat{\mathbf{L}}_1\hat{\mathbf{L}}_2-\mathrm{i}\hbar\hat{\mathbf{L}}_2\hat{\mathbf{L}}_1+\mathrm{i}\hbar\hat{\mathbf{L}}_2\hat{\mathbf{L}}_1+\mathrm{i}\hbar\hat{\mathbf{L}}_1\hat{\mathbf{L}}_2=0\,.$$

We get the same result for any component:

► $$[\hat{\mathbf{L}}^2,\hat{\mathbf{L}}_k]=0\,,\qquad k=1,2,3\,.\tag{2.27}$$

It is therefore not possible to measure two different components of angular momentum at the same time. However, it is always possible to measure the square of the angular momentum and one of its components. From the considerations made so far, it is clear that we can simultaneously measure the dynamic variables $\{x, y, z\}$ or $\{p_x, p_y, p_z\}$ or $\{L^2, L_3\}$. In Section 2.5, we will see that in the case of a spherically symmetric potential the complete set of observables consists of the triplet $\{E, L^2, L_3\}$.

We have thus found a simple method for determining which quantities can be measured simultaneously and which cannot. It suffices to find the commutator of the corresponding operators. However, this method only provides a yes/no answer. In cases where dynamic variables cannot be measured simultaneously, we are interested in how much the act of measuring one variable disturbs the act of measuring the other. This question is answered by Heisenberg uncertainty relations, which we will now derive.

Before that, let us present an overview of basic statistical concepts and their operator analogues in quantum theory:

Statistics		**Quantum theory**	
Mean	$\bar{a}$	Mean	$\bar{a} \equiv \langle \psi \mid \hat{\mathbf{A}} \mid \psi \rangle$
Fluctuation	$\Delta a \equiv a - \bar{a}$	Fluctuation operator	$\Delta \hat{\mathbf{A}} \equiv \hat{\mathbf{A}} - \bar{a}\,\hat{\mathbf{1}}$
Mean of fluctuations	$\overline{\Delta a} = 0$	Mean of fluctuations	$\langle \psi \mid \Delta \hat{\mathbf{A}} \mid \psi \rangle = 0$
Variance	$\overline{(\Delta a)^2} = \overline{a^2} - \bar{a}^2$	Variance	$\langle \psi \mid \Delta \hat{\mathbf{A}}^2 \mid \psi \rangle = \langle \psi \mid \hat{\mathbf{A}}^2 \mid \psi \rangle - \langle \psi \mid \hat{\mathbf{A}} \mid \psi \rangle^2$
Standard deviation	$\Delta a_{\mathrm{sd}} = \sqrt{\overline{(\Delta a)^2}}$	Standard deviation	$\Delta a_{\mathrm{sd}} = \sqrt{\langle \psi \mid \Delta \hat{\mathbf{A}}^2 \mid \psi \rangle}$

Try to prove both classical and quantum statistical formulas. In both cases, all you need to do is substitute the values from the relevant definitions. Now we can proceed to derive the uncertainty relations. Suppose we have two incompatible variables:

$$A, B \to \hat{\mathbf{A}}, \hat{\mathbf{B}}\,; \qquad [\hat{\mathbf{A}}, \hat{\mathbf{B}}] = \hat{\mathbf{C}}\,. \tag{2.28}$$

Let's find the product of the standard deviations of the measurements:

$$\left(\Delta a_{\mathrm{sd}}\right)^2 \left(\Delta b_{\mathrm{sd}}\right)^2 = \langle \psi \mid (\Delta\hat{\mathbf{A}})^2 \mid \psi \rangle \langle \psi \mid (\Delta\hat{\mathbf{B}})^2 \mid \psi \rangle \overset{(*1)}{=}$$

$$= \langle \Delta\hat{\mathbf{A}}\psi \mid \Delta\hat{\mathbf{A}}\psi \rangle \langle \Delta\hat{\mathbf{B}}\psi \mid \Delta\hat{\mathbf{B}}\psi \rangle = \| \Delta\hat{\mathbf{A}}\psi \|^2 \cdot \| \Delta\hat{\mathbf{B}}\psi \|^2 \overset{(*2)}{\geq}$$

$$\geq |\langle \Delta\hat{\mathbf{A}}\psi \mid \Delta\hat{\mathbf{B}}\psi \rangle|^2 = |\langle \psi \mid \Delta\hat{\mathbf{A}}\Delta\hat{\mathbf{B}} \mid \psi \rangle|^2 \overset{(*3)}{=}$$

$$= |\langle \psi \mid \frac{1}{2}(\Delta\hat{\mathbf{A}}\Delta\hat{\mathbf{B}} + \Delta\hat{\mathbf{B}}\Delta\hat{\mathbf{A}}) \; + \; \frac{1}{2}(\Delta\hat{\mathbf{A}}\Delta\hat{\mathbf{B}} - \Delta\hat{\mathbf{B}}\Delta\hat{\mathbf{A}}) \mid \psi \rangle|^2 \overset{(*4)}{\geq}$$

$$\geq |\langle \psi \mid \frac{1}{2}(\Delta\hat{\mathbf{A}}\Delta\hat{\mathbf{B}} - \Delta\hat{\mathbf{B}}\Delta\hat{\mathbf{A}}) \mid \psi \rangle|^2 =$$

$$= |\frac{1}{2}\langle \psi \mid [\Delta\hat{\mathbf{A}}, \Delta\hat{\mathbf{B}}] \mid \psi \rangle|^2 \overset{(*5)}{=}$$

$$= |\frac{1}{2}\langle \psi \mid [\hat{\mathbf{A}}, \hat{\mathbf{B}}] \mid \psi \rangle|^2 =$$

$$= |\frac{1}{2}\langle \psi \mid \hat{\mathbf{C}} \mid \psi \rangle|^2 .$$

The following tricks were used in the derivation (marked with an asterisk:

(1) Use of operator hermiticity;
(2) Schwarz's lemma, see Section 3.3.2, Equation (3.165);
(3) Division into a symmetric (*S*) and antisymmetric (*D*) part;
(4) The symmetric part *S* is real (it is the sum of two complex conjugates), while the antisymmetric part *D* is purely imaginary (it is the difference of two complex conjugates); together, they form a complex number for which the following holds $|S+D| = |x+\mathrm{i}y| = (x^2+y^2)^{1/2} \geq |y| = |D|$;
(5) The unit operator in the definition of $\Delta\hat{\mathbf{A}}$ commutes with everything.

After taking the square root of the last expression, we obtain the Heisenberg relations:

► $$\Delta a_{\mathrm{SD}}\,\Delta b_{\mathrm{SD}} \geq \frac{1}{2}|\langle \psi|\hat{\mathbf{C}}|\psi\rangle|. \tag{2.29}$$

If we know the result of the commutation relation between the operators conjugate with two dynamic variables, we can use Heisenberg relations to determine the extent to which one measurement influences the other. It depends on the state in which the system is prepared. Only if both dynamic variables are in the *generalized coordinate –momentum* relationship, the mutual influence does not depend on the state of the system:

$$[\hat{\mathbf{X}},\hat{\mathbf{P}}] = \mathrm{i}\hbar\hat{\mathbf{1}} \quad \Rightarrow \quad \Delta x\,\Delta p \geq \frac{\hbar}{2}|\langle \psi|\hat{\mathbf{1}}|\psi\rangle| = \frac{\hbar}{2}|\langle \psi|\psi\rangle| = \frac{\hbar}{2}.$$

This is the most well-known form of uncertainty relations:

► $$\Delta x\,\Delta p \geq \frac{\hbar}{2}. \tag{2.30}$$

Heisenberg uncertainty principle is a very important feature of quantum theory in the microworld. We already mentioned the example of a slit through which a beam of light passes. If the slit is wide enough, there is no significant diffraction of light. We know the momentum of the photons in the plane of the slit almost exactly (it is approximately zero; the photons travel perpendicular to this plane), but we do not know the exact point at which the photon passed through the slit. If we narrow the slit, we do obtain more precise information about the photon’s position; however, light diffraction occurs, and we lose information about the photon’s momentum in the plane of the slit. Measuring the position degrades information about the momentum.

Another example is an electron in an atomic shell. It is localized within a certain small region (on the order of 10^{-10} m; of course, we do not know its exact position), and this must correspond to a certain momentum (or velocity) in accordance with Heisenberg relations. For example, an electron cannot “come to a stop” in an atomic shell. A similar situation occurs in a crystalline substance at very low temperatures. As the temperature of the substance decreases, the chaotic motion of the particles decreases; however, this motion can never completely stop. If all motion of the substance ceased at absolute zero, the ions would be located precisely at the vertices of the crystal lattice (we would know their exact positions) and would not move (we would know their exact momentum, which would be zero). Knowing both quantities contradicts Heisenberg uncertainty principle. Even at absolute zero, a crystal will undergo so-called zero vibrations. Absolute zero is not a state of matter with zero particle motion, but a state with the minimum possible motion allowed by the laws of quantum mechanics. For the same reason, the lowest energy state of a harmonic oscillator is not zero, but a harmonic oscillator always performs at least so-called zero-point vibrations (see Chapter 2.3).

As we have seen, the uncertainty relation (2.30) holds for any generalized coordinate and its corresponding momentum. It is therefore not possible, for example, to simultaneously measure the angle of a pendulum and its corresponding angular momentum. In general, for any two canonically coupled variables (q, p), the following holds

$$\blacktriangleright \qquad \Delta q\,\Delta p \geq \frac{\hbar}{2}\,. \tag{2.31}$$

Lagrange formulation of the electromagnetic field (see Chapter 1.6) shows that electromagnetic field potentials can be understood as continuous generalized coordinates and that there is a corresponding canonically conjugate field momentum. Heisenberg uncertainty relations apply to both quantities, so it is impossible for both the field and its momentum to be zero simultaneously in a vacuum. A vacuum is therefore not a space without fields, but one with a minimal amount of various fluctuations of all possible fields permitted by the laws of quantum theory. The situation is very similar to the definition of absolute zero. Thanks to quantum theory, a true vacuum can never be empty; it is a non-trivial dynamic system full of field fluctuations, which manifest as the temporary creation of particle-antiparticle pairs decaying again after a fairly short time.

The uncertainty relations (2.30) apply to all three components of the position vector. In the context of relativity, however, the spatial components are part of the four-vector $(ct, \mathbf{x})$, just as the momentum components are part of the four-vector $(E/c, \mathbf{p})$. The speed of light only ensures that all four components have the same dimension. The uncertainty relation can also be written for the time component of four-vectors:

$$\blacktriangleright \qquad \Delta E\,\Delta t \geq \frac{\hbar}{2}\,. \tag{2.32}$$

The particle-antiparticle pair with total energy ΔE can arise in a vacuum, as if from nothing, provided that it annihilates within a time shorter than $\Delta t = \hbar/(2\Delta E)$. Although such a process violates the law of conservation of energy (it is a sort of "borrowing" of energy from the vacuum "on credit"), it will be undetectable to us because it violates the uncertainty principle (2.32). These processes do indeed occur in a vacuum; the constant creation and annihilation of electron-positron pairs results, for example, in observable vacuum polarization or the Lamb shift of spectral lines. There is speculation that the accelerated expansion of the universe, discovered by Adam Riess and Saul Perlmutter in 1998, could originate from the non-trivial behavior of the quantum vacuum.

We can also apply the relationship (2.32) when describing the formation of spectral lines in the atomic shell. If an electron's transition between two energy levels takes a finite time Δt, the resulting photon will not have a precisely defined energy; the minimum energy uncertainty will be $\Delta E \sim \hbar/(2\Delta t)$, and the spectral line will never be perfectly sharp – it will always have a finite width given by the uncertainty relations (this is a statistical phenomenon caused by the many electron transitions responsible for the formation of the spectral line).

The uncertainty principle (2.29) also applies to variables that are not canonically coupled. The extent to which one measurement influences the other depends on the state of the system. For example, the kinetic and potential energy operators generally do not commute with each other. As a result, it is not possible to determine both kinetic and potential energy precisely at the same time, and a particle can (unlike in classical physics) "swing" over a potential barrier. This is the so-called *tunneling effect*, which is another interesting consequence of quantum mechanics and whose essence lies in Heisenberg uncertainty relations.

◖ Example 2.1: Gaussian wave packet

Let us now show that if a particle has a wave function in the form of a Gaussian distribution, then $\Delta x \Delta p = \hbar/2$ holds for it; that is, in Heisenberg uncertainty relations (2.30), the equality holds, and the maximum possible information can be obtained through the act of measurement. All the integrals needed to calculate the individual steps of this example can be found in Section 3.10.4.

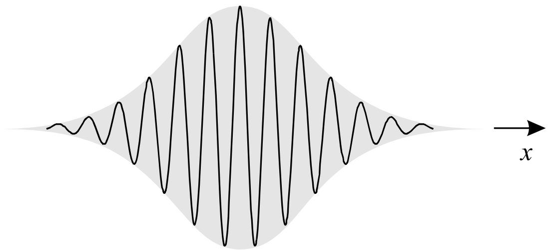


Fig. 2.8: Gaussian wave packet

Suppose that the packet-shaped wave function has a form given by the function $\cos(k_0 x)$ and the envelope is shaped by a Gaussian exponential function:

► $$\psi(x) \approx \cos(k_0 x)\, \mathrm{e}^{-\alpha x^2} . \tag{2.33}$$

The wave vector of the wave is k_0; this corresponds, in terms of wave-particle duality, to the momentum of the described object, $\hbar k_0$. In exponential notation, including normalization to unity, the wave function and the probability will have the form (3.576):

$$\begin{aligned} \psi(x) &= (2\alpha/\pi)^{1/4}\, \mathrm{e}^{\mathrm{i}k_0 x}\, \mathrm{e}^{-\alpha x^2} ; \\ w(x) \equiv \psi^* \psi &= (2\alpha/\pi)^{1/2}\, \mathrm{e}^{-2\alpha x^2} . \end{aligned} \tag{2.34}$$

If we include the time evolution, our wave function would also have a time component. Let's now calculate the mean position, the mean squared position, and the standard deviation. The calculations are straightforward using the relations from Section 3.10.4:

► $$\begin{aligned} \langle x \rangle &= \int_{-\infty}^{+\infty} x\, w(x)\, \mathrm{d}x = 0 ; \\ \langle x^2 \rangle &= \int_{-\infty}^{+\infty} x^2\, w(x)\, \mathrm{d}x = 1/(4\alpha) ; \\ \Delta x &\equiv \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = 1/\sqrt{4\alpha}. \end{aligned} \tag{2.35}$$

We can decompose our wave function into individual Fourier components using the Fourier transform (3.495) and (3.496)

$$\begin{aligned} \psi(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathscr{A}(k)\, \mathrm{e}^{\mathrm{i}kx}\, \mathrm{d}k; \\ \mathscr{A}(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \psi(x)\, \mathrm{e}^{-\mathrm{i}kx}\, \mathrm{d}x. \end{aligned} \tag{2.36}$$

We will perform the calculation by converting the exponent to a square.

$$\mathscr{A}(k)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\psi(x)\,\mathrm{e}^{-\mathrm{i}kx}\;\mathrm{d}x=$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\left(\frac{2\alpha}{\pi}\right)^{1/4}\mathrm{e}^{\mathrm{i}k_0x}\;\mathrm{e}^{-\alpha x^2}\;\mathrm{e}^{-\mathrm{i}kx}\;\mathrm{d}x=$$

$$=\frac{1}{\sqrt{2\pi}}\left(\frac{2\alpha}{\pi}\right)^{1/4}\int_{-\infty}^{+\infty}\mathrm{e}^{-\alpha[x^2-\mathrm{i}(k-k_0)x/\alpha]}\;\mathrm{d}x=$$

$$=\frac{1}{\sqrt{2\pi}}\left(\frac{2\alpha}{\pi}\right)^{1/4}\mathrm{e}^{-(k-k_0)^2/(4\alpha)}\int_{-\infty}^{+\infty}\mathrm{e}^{-\alpha[x-\mathrm{i}(k-k_0)x/(2\alpha)]^2}\;\mathrm{d}x=$$

$$=\frac{1}{\sqrt{2\pi}}\left(\frac{2\alpha}{\pi}\right)^{1/4}\mathrm{e}^{-(k-k_0)^2/(4\alpha)}\int_{-\infty}^{+\infty}\mathrm{e}^{-\alpha[x-x_0]^2}\;\mathrm{d}x=$$

$$=\frac{1}{(2\pi\alpha)^{1/4}}\mathrm{e}^{-(k-k_0)^2/(4\alpha)}\;.$$

For the amplitude $\mathscr{A}(k)$, we therefore have the final relationship

► $$\mathscr{A}(k)=\frac{1}{(2\pi\alpha)^{1/4}}\mathrm{e}^{-(k-k_0)^2/(4\alpha)}\;. \tag{2.37}$$

The amplitude $\mathscr{A}(k)$ can be understood as a wave function in k-space (or in momentum space, since $p = \hbar k$), and its square is the probability density in k-space:

$$w_p(k)=\mathscr{A}^*\mathscr{A}=\frac{1}{(2\pi\alpha)^{1/2}}\mathrm{e}^{-(k-k_0)^2/(2\alpha)}\;. \tag{2.38}$$

We can verify that the probability is normalized, i.e., $\int w_p(k)\,\mathrm{d}k = 1$. Next, we will determine the mean values of the momentum and the square of the momentum of the object:

$$\langle\, p\,\rangle=\int_{-\infty}^{+\infty}\hbar k\,w_p(k)\,\mathrm{d}k=\hbar k_0\;;$$

► $$\langle\, p^2\,\rangle=\int_{-\infty}^{+\infty}(\hbar k)^2\,w_p(k)\,\mathrm{d}k=\hbar^2\alpha+\hbar^2k_0^2\;; \tag{2.39}$$

$$\Delta p\equiv\sqrt{\langle\, p^2\,\rangle-\langle\, p\,\rangle^2}=\hbar\sqrt{\alpha}\;.$$

We now have all we need to complete the left-hand side of Heisenberg relation:

► $$\Delta x\Delta p=\frac{1}{\sqrt{4\alpha}}\hbar\sqrt{\alpha}=\frac{\hbar}{2}\,. \tag{2.40}$$

The Gaussian wave packet thus minimizes Heisenberg uncertainty relations: the minimum of Heisenberg uncertainty relations is achieved for the Gaussian wave packet.

◗

2.2.3 Energy Eigenstates, Schrödinger Equation

In the previous chapter, we learned how to determine which dynamic variables can be measured simultaneously and which cannot. Using Heisenberg uncertainty relations, we can also qualitatively describe the extent to which one measurement disturbs another. Now we will turn to the second fundamental task of quantum mechanics: finding the energy operator's spectrum – the energy values that can be measured in the system. We can formulate this task, for example, as follows:

$$\hat{\mathbf{H}}\,|\,n> = E_n\,|\,n>, \tag{2.41}$$

$$\hat{\mathbf{H}} \equiv \frac{\hat{\vec{\mathbf{P}}}^2}{2m} + V(\hat{\vec{\mathbf{X}}}) = \frac{\hat{\mathbf{P}}_x^2 + \hat{\mathbf{P}}_y^2 + \hat{\mathbf{P}}_z^2}{2m} + V(\hat{\mathbf{X}}, \hat{\mathbf{Y}}, \hat{\mathbf{Z}}), \tag{2.42}$$

$$\left[\hat{\mathbf{X}}_k, \hat{\mathbf{X}}_l\right] = \left[\hat{\mathbf{P}}_k, \hat{\mathbf{P}}_l\right] = 0, \qquad \left[\hat{\mathbf{X}}_k, \hat{\mathbf{P}}_l\right] = \mathrm{i}\hbar\,\hat{\mathbf{1}}\,\delta_{kl}\,. \tag{2.43}$$

We will look for the eigenvalues of the energy operator (Hamiltonian) from eq. (2.41). This equation for the eigenvalues of the Hamiltonian is called *Schrödinger equation*. The energy operator (Hamiltonian) is given by equation (2.42). The basic operators that make up the Hamiltonian operator are subject to the commutation relations (2.24) and (2.43), respectively. It does not matter much which Hilbert space we choose. In the next chapter, we will see the solution of the harmonic oscillator in various spaces $\mathcal{H}$; we will always obtain the same spectrum of the Hamiltonian operator. In a given space, the most important thing is to choose Hermitian operators for the generalized coordinates and momenta such that they satisfy the commutation relations (2.43).

Let us now consider the formulation of Schrödinger equation in the space $\mathcal{L}^2(\mathcal{R}^3)$ of functions that are square-integrable over the entire space $\mathcal{R}^3$. One of the simplest operators on this space is the coordinate multiplication operator. We identify this operator with the coordinate operator:

$$\hat{\mathbf{X}} = x\,; \qquad \hat{\mathbf{Y}} = y\,; \qquad \hat{\mathbf{Z}} = z\,. \tag{2.44}$$

Now we need to find Hermitian operators for momentum that satisfy the commutation relations (2.43). In one dimension, the derivative operator and the coordinate multiplication operator satisfy the commutation relation (see Example 3.31 in Sec. 3.4.2)

$$[\hat{\mathbf{D}}, \hat{\mathbf{X}}] = \hat{\mathbf{1}}, \text{ resp} \qquad [\hat{\mathbf{X}}, \hat{\mathbf{D}}] = -\hat{\mathbf{1}} \tag{2.45}$$

It is clear that the relation (2.43) satisfies in one dimension the operator

► $$\hat{\mathbf{P}} \equiv -\mathrm{i}\hbar\,\mathrm{d}/\mathrm{d}x\,. \tag{2.46}$$

If we choose the coordinate operator to be simply the product of the coordinates, then the operator (2.46) corresponds to the momentum operator. The derivative operator itself is not Hermitian, but the derivative operator multiplied by a purely imaginary constant is Hermitian (see Example 3.33 in Sec. 3.4.2). In three dimensions, we proceed in exactly the same way. If we require that the coordinate operators have the simple form (2.44) while the relations (2.43) hold, the momentum operator must have the form

$$\hat{\mathbf{P}}_x = -\mathrm{i}\hbar\,\partial/\partial x;$$
► $$\hat{\mathbf{P}}_y = -\mathrm{i}\hbar\,\partial/\partial y; \tag{2.47}$$
$$\hat{\mathbf{P}}_z = -\mathrm{i}\hbar\,\partial/\partial z,$$

or in vector form

► $$\hat{\vec{\mathbf{P}}} = -\mathrm{i}\hbar\nabla\,. \tag{2.48}$$

Schrödinger equation (2.41), together with the energy operator (2.42), the choice of space $\mathcal{H} = \mathcal{L}^2(\mathcal{R}^3)$, and the operators (2.44) and (2.47), then leads to the famous Schrödinger equation in the so-called *x*-representation (the coordinate operator is represented by multiplication by the coordinates):

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{x})\right]\psi_n(\mathbf{x}) \;=\; E_n\,\psi_n(\mathbf{x})\,. \tag{2.49}$$

Solving Schrödinger's equation for a specific potential V yields the spectrum of the energy operator $\{E_n\}$, which is the set of possible measurable energy values for that potential.

Note 1: A solution to the equation (2.49) can be found for every value of energy. However, this solution does not always belong to the space $\mathcal{L}^2(\mathcal{R}^3)$. It is therefore always necessary to select only those solutions from the set of possible solutions that are integrable with respect to the square, i.e., that decreases sufficiently rapidly to infinity to ensure integrability.

Note 2: There is a simple way to estimate the type of spectrum for a given potential. If, in classical mechanics, a particle can move away to infinity, the spectrum of the energy operator is continuous. If it cannot move away to infinity in any direction, the spectrum of the energy operator is discrete.

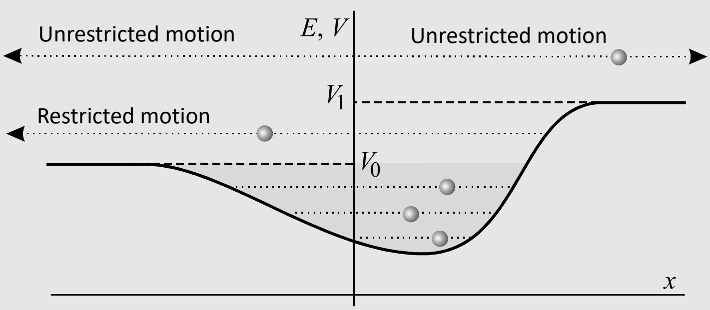


Fig. 2.9: Movements in potential energy with a minimum

In classical mechanics, a particle can move wherever its total energy is greater than its potential energy. This follows from the relation $E = mv^2 + V(x)$. This is essentially the condition that the kinetic energy must be non-negative. In the situation depicted, the energy spectrum is discrete for $E < V_0$ and continuous for $E > V_0$.

Examples of potentials

In the case of a one-dimensional potential, the position of the object x can take on any real value, i.e., $x \in (-\infty, +\infty)$.

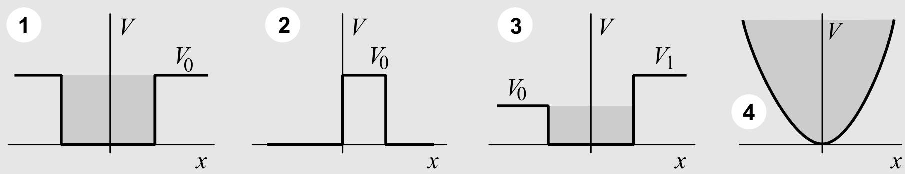


Fig. 2.10: Examples of one-dimensional potentials

1. Symmetric square well. A particle in a potential well has a discrete energy spectrum for $E < V_0$, where, in the classical case, it cannot escape to infinity. This region is shown in gray. For $E > V_0$, the particle in this potential has a continuous energy spectrum. For an infinite potential well ($V_0 \to \infty$), the spectrum is discrete.

2. Barrier. In the classical case, a particle can always move away to infinity. In quantum theory, this corresponds to a completely continuous energy spectrum.

3. Asymmetric square well. The particle has a discrete energy spectrum for $E < V_0$, where, in the classical case, it cannot move away to infinity. The energy spectrum is continuous for $E > V_0$, where, in the classical case, it can move away either to one side ($E > V_0$ and simultaneously $E < V_1$) or to both sides ($E > V_1$).

4. Harmonic oscillator. A harmonic oscillator has a parabolic potential energy curve. In the classical case, the particle oscillates and can never move away to infinity. This corresponds to a discrete energy spectrum in quantum theory.

In the following examples, we consider a spherically symmetric field in which the potential energy is a function only of the radial coordinate, i.e., $V = V(r)$. We impose the constraint $r \geq 0$ on the radial coordinate itself, meaning that a particle cannot have a negative radial coordinate.

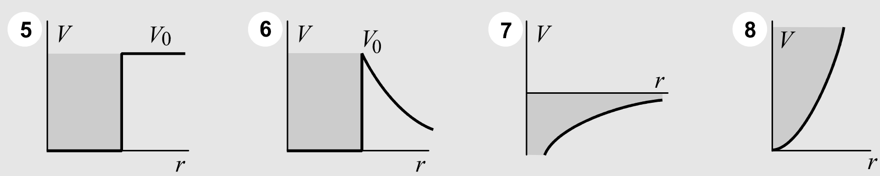


Fig. 2.11: Examples of three-dimensional potentials

5. Spherical well. A particle in a spherical well has a discrete energy spectrum for $E < V_0$ and a continuous energy spectrum for $E > V_0$. A neutron in an atomic nucleus, for example, moves in a similar potential.

6. Coulomb barrier. The potential profile is a combination of the potentials of a spherical well and the Coulombic repulsion. A proton moves in a similar potential within an atomic nucleus. The particle has a discrete energy spectrum for $E < V_0$ and a continuous energy spectrum for $E > V_0$. There is a nonzero probability to penetrate the potential barrier. We call this phenomenon the tunneling effect. It is caused by the fact that the kinetic and potential energy operators do not commute with each other.

7. Coulomb's attractive potential. A particle has a discrete energy spectrum for $E < 0$ and a continuous one for $E > 0$. An electron in the atomic shell moves in a similar potential. States with negative energy are bound states; states with positive energy are free, i.e., the electron is not bound to the atomic nucleus.

8. Spherical harmonic oscillator. In this potential, the particle has only discrete energy states. When the system is displaced in any direction, it returns to the origin according to the equation $V(r) = 1/2\ kr^2$.

2.2.4 Various Interpretations of Quantum Theory

Quantum theory is the first theory in which the experiment itself is an integral part. There is no doubt that an experiment conducted in the microworld affects objects and alters their state. During the experiment, only a single result is selected from possible measurement outcomes. The way this occurs is more of a philosophical than a physical question, and different physicists hold varying views on the course of the measurement process itself. Quantum theory is an elegant mathematical construct whose results are in exceptionally precise agreement with the experiments conducted. However, in many parts of quantum theory, our imagination and so-called "common sense" completely fail us. We can no longer equate objects of the microworld with familiar balls; they possess different, richer properties. An example is intrinsic angular momentum, or spin, whose existence follows from Lorentz symmetry (two experiments conducted in two inertial coordinate systems will yield the same results). It is almost impossible to visualize spin. We can calculate how it combines with a particle's angular momentum; we know its manifestations, how it causes the splitting of spectral lines in a magnetic field, or how it is responsible for molecular bonding. We can imagine that an electron orbits the nucleus (angular momentum) and also rotates around its own axis (spin), but this is merely a mental image that is very far from reality. An electron neither orbits the nucleus nor rotates around any axis. Sometimes, however, even a flawed conception is better than none at all. In quantum theory, there are many things that are unimaginable and incomprehensible to our imperfect senses. In the following text, we will explore some views on the role of quantum theory in describing the real world.

Copenhagen Interpretation

The Copenhagen interpretation was developed in Copenhagen between 1924 and 1927. Today, it is the most widely accepted view of quantum theory. The main authors of the Copenhagen interpretation are Niels Bohr and Werner Heisenberg, who served as Bohr's assistant from 1926. Quantum theory is unable to predict the exact result of a measurement, but only the values that can be measured and the probability with which this will occur. This probability is equal to the square of the wave function (2.16). Before the measurement, the system is in a superposition of states. This superposition is all that can be known about the system. During the measurement, the system transitions to one of the states of this superposition. The act of measurement behaves like a projection operator, which selects one specific projection, one specific state. If we describe the system using a wave function, we say that *collapse of the wave function* occurs.

Before the measurement, the wave function is spread out in space and provides a probability density $\psi^*\psi$ for the occurrence of a particle (object) at points in space. During the measurement, this probability must change dramatically, since after the measurement the particle is localized at a specific point $\mathbf{x}_0$, where the detector found it. The abrupt change occurs immediately before the measurement in the entire wave function. The collapse of the wave function occurs simultaneously throughout all of space and is a non-local process. It is precisely the non-locality of the act of measurement that has been the target of significant criticism from proponents of local theories in which the change is influenced only by an infinitely small neighborhood of that point.

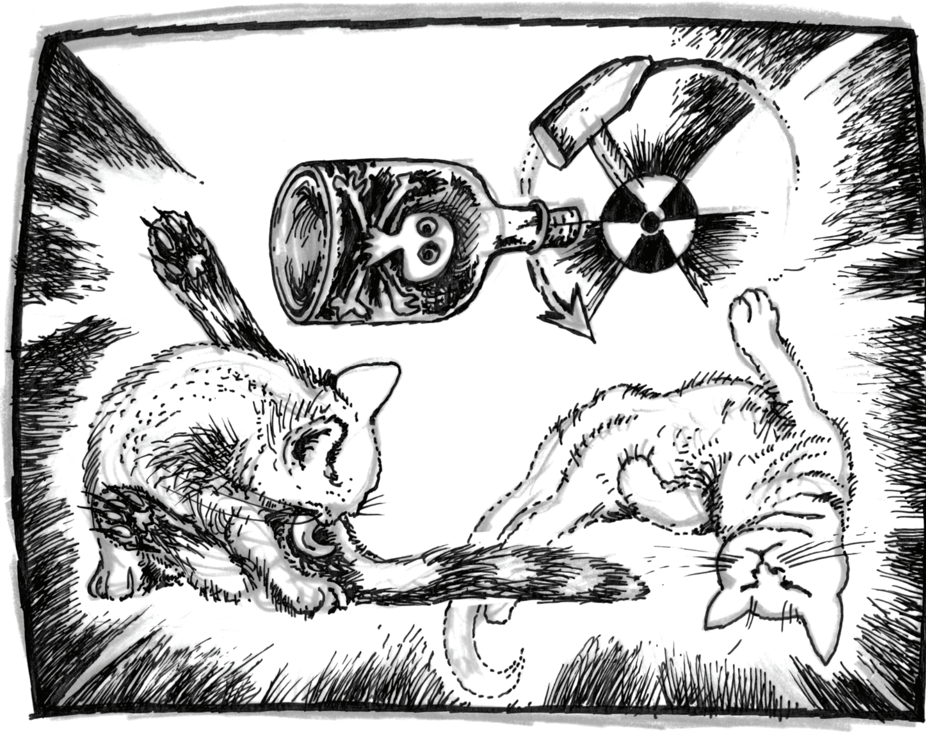


Fig. 2.12: Schrödinger's cat paradox

The Copenhagen interpretation can only be applied to objects in the microworld. When applied to macroscopic objects, it yields clearly nonsensical results; the best-known example is the so-called Schrödinger cat paradox – a thought experiment proposed by Schrödinger. The cat is enclosed in an opaque box containing a vial of lethal poison. A hammer that will shatter the vial is triggered by random radioactive decay. After a certain period of time, there is a fifty-percent chance that the vial has shattered and the cat has been killed. However, until we open the box and verify the cat's actual state, the cat is, from the perspective of quantum theory, in a superposition of both possible states: $|\mathit{dead\ cat}\rangle$ and $|\mathit{alive\ cat}\rangle$. Only the act of measurement causes the collapse of the cat's wave function, and for us, it will definitively be either alive or dead. From this thought experiment, it is clear that we cannot apply quantum theory to a macroscopic object. But this inevitably raises the question: Where is the boundary between the quantum world, where quantum laws apply, and the macroworld, where they clearly do not?

Classical interpretation

The author of the classical interpretation is Albert Einstein, who never accepted the statistical (Copenhagen) interpretation of quantum theory. He expressed this in his famous statement: "*God does not play dice*." According to Einstein, the fact that a system is in a superposition of states prior to measurement is a consequence of our lack of knowledge of all the system's microscopic parameters. During measurement, we then choose one of the possibilities only seemingly. It would be unambiguously determined if we had all the information about the object. Theories of this type are referred to as *theories with hidden parameters*. Today, this interpretation is ruled out based on the confirmation of the invalidity of the so-called Bell inequalities in quantum systems (see Section 2.9.4). Einstein and others were also troubled by non-locality and the impossibility of simultaneously measuring certain quantities. In 1935, Einstein, together with Podolsky and Rosen, formulated a thought experiment (see Section 2.9.3) intended to demonstrate an internal contradiction within quantum theory. Today's perspective sees no such contradiction here.

Many-Worlds Interpretation

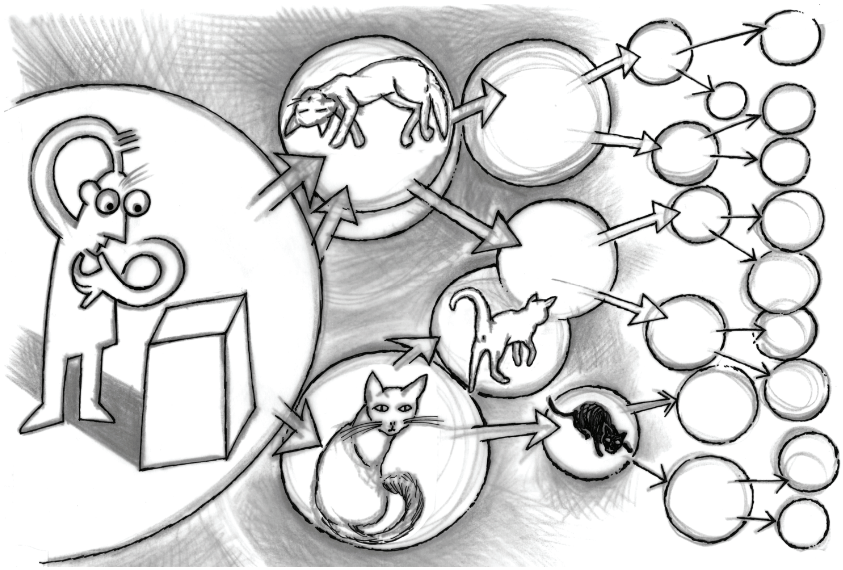


Fig. 2.13: Many-Worlds Interpretation of the Schrödinger cat paradox

The many-worlds interpretation of quantum theory was introduced by American physicist Hugh Everett in 1957. American theorist Bryce DeWitt was a major proponent and advocate of the many-worlds interpretation in the 1960s and 1970s. The core idea is that, upon measurement, one of the superposition states is realized in our world. In other parallel worlds (universes), the other possibilities are realized. Everything that can happen does happen, but in different universes that coexist in parallel. The act of measurement is thus understood as a branching of the object's worldline. In the Schrödinger cat paradox, the cat is indeed in a superposition of two states $|\,dead\,\rangle$ and $|\,alive\,\rangle$ until a measurement is performed. If we find that it is, for example, $|\,alive\,\rangle$, this state will be realized in our universe. In some other universe, the state $|\,dead\,\rangle$ will be realized. It is unclear whether this interpretation yields any new measurable facts.

Holographic interpretation

The American-British theoretical physicist David Bohm (1917–1992) made a significant contribution to our understanding of the nonlocality of quantum theory. We perceive objects such as subatomic particles as separate from one another because we see only a part of their reality. The interconnectedness of the whole has nothing to do with the particle's location in space and time as we perceive it. By measuring one part of the whole, we can obtain information about another part that appears to us to be spatially distant. Every particle is part of this indivisible whole, which had a single wave function at the Big Bang. Today, the whole is contained within each of its parts. Nonlocality is therefore inherent to quantum theory. In this interpretation, it is often expressed as a collective potential that all particles influence and to which each of them responds. Everything that exists in physical reality is stored in smaller parts, and the universe itself is a reflection of this foundation, which we can call a *hologram* (it is possible to reconstruct a larger whole from a smaller one; the larger whole is no longer reality, but merely a kind of image of reality that we perceive with our senses). This interpretation offers a new perspective on the non-locality of quantum theory and could play a role in understanding the structure of quantum theory in the future.

Consciousness-based interpretation

In 1932, the Hungarian mathematician John von Neumann (who, among other things, authored the first proposal for the architecture of today's computers) put forward an intriguing interpretation of quantum theory. Von Neumann posited that the collapse of the wave function into a specific state during the act of measurement is caused by the observer's consciousness. The observer is a natural part of the quantum world, and his consciousness can influence the outcome of the experiment. The Hungarian-American theorist Eugene Wigner also became a proponent of this idea. In general, this interpretation was not accepted among physicists. Again, it is debatable whether it actually yields any new, experimentally verifiable facts.

Boundaries of the Quantum World

An electron is undoubtedly a particle of the quantum world, with all its strange properties – sometimes it behaves like a particle, other times like a wave; it can exist in multiple states at once, and so on. If we place two slits of appropriate width and distance in its path, it will not pass through just one of them, as a particle of the macroscopic world would. It will take advantage of the superposition of states and pass through both slits at once. The two states will then interfere, so in a sense, the electron actually interferes with itself. An interference pattern will appear on the screen after many electrons strike it. If you throw ordinary pebbles or marbles at the double slit, only two maxima will appear on the screen (one opposite each slit). A pebble cannot be in a superposition of states and cannot interfere with itself.

If we emit a sufficient number of electrons, their points of impact will not be aligned with the slits, as with classical particles, but will form fringes similar to the interference pattern observed in waves. These fringes do not disappear even when the electron current is so sparse that there is at most a single electron in the detector at any given moment. On the contrary, the fringes disappear if there is a fundamental possibility of detecting the electron's position.

Where is the boundary between these two worlds? Up to what scale do particles behave quantum mechanically, interfere with themselves, and appear (to us) as strange objects of the microworld? And when does the classical behavior we know so well take over? Experiments conducted by Anton Zeilinger and his colleagues at the University of Vienna around 2005 showed that no sharp boundary exists. The scientists conducted experiments with giant molecules that passed through a Talbot interferometer (which has a large number of slits). The largest molecule, $C_{60}F_{48}$, contained 1,632 nucleons and was approximately spherical in shape. However, the Vienna team also attempted to use molecules of other shapes, such as the flat porphyrin molecule.

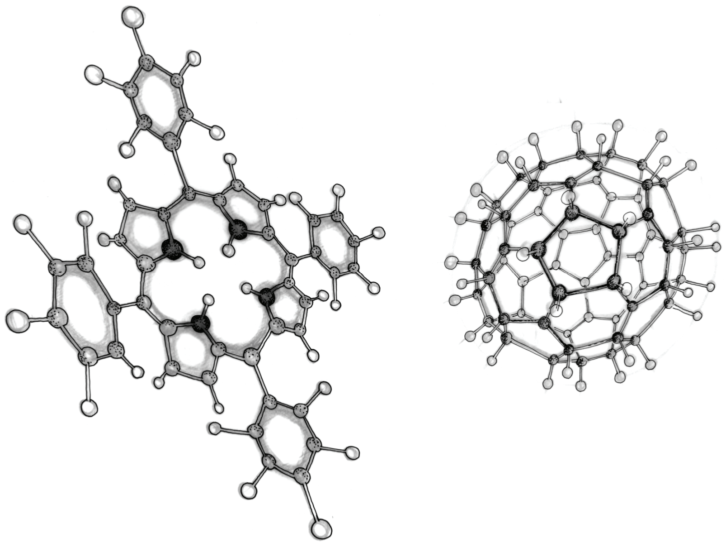


Fig. 2.14 Porphyrin molecules (left) and fullerene $C_{60}F_{48}$ (right) used in multi-slit experiments in Vienna

It turned out that this molecule was capable of interfering with itself only if it did not interact with its surroundings in any way and it was fundamentally impossible to determine its position (through which slit it passed). The interference pattern appeared as the pressure in the apparatus decreased (it was not possible to detect the position of the giant molecules from the reflection of atmospheric atoms off these molecules). The interference pattern also appeared as the temperature dropped. At low temperatures, the molecule no longer emitted any photons from which its location could be determined. The conclusion is simple. Objects behave quantum mechanically if they cannot interact with their surroundings, and classically if they do interact with their surroundings. In such a case, we say that they have entangled states with their surroundings, which means that the wave function of the object together with its surroundings cannot be separated into a simple product in which one part depends only on the object's variables and the other only on the variables describing the surroundings.

So if you want your friend to behave in a quantum way – to become a wave and interfere with himself – you have to cool him down to nearly absolute zero (so he no longer emits any photons) and place him in a vacuum where no surrounding gas molecules will interact with him. At that point, you won't have any interaction with your friend, and you won't know where he is. He will begin to behave like a quantum object, interfering with himself and perhaps even passing through multiple slits at once. At least, current experiments with giant molecules suggest this. Of course, there may be some other fundamental boundary between the quantum world and the macroworld that we have not yet discovered.

Readers can find more detailed information on the various interpretations of quantum theory in the publication [31].

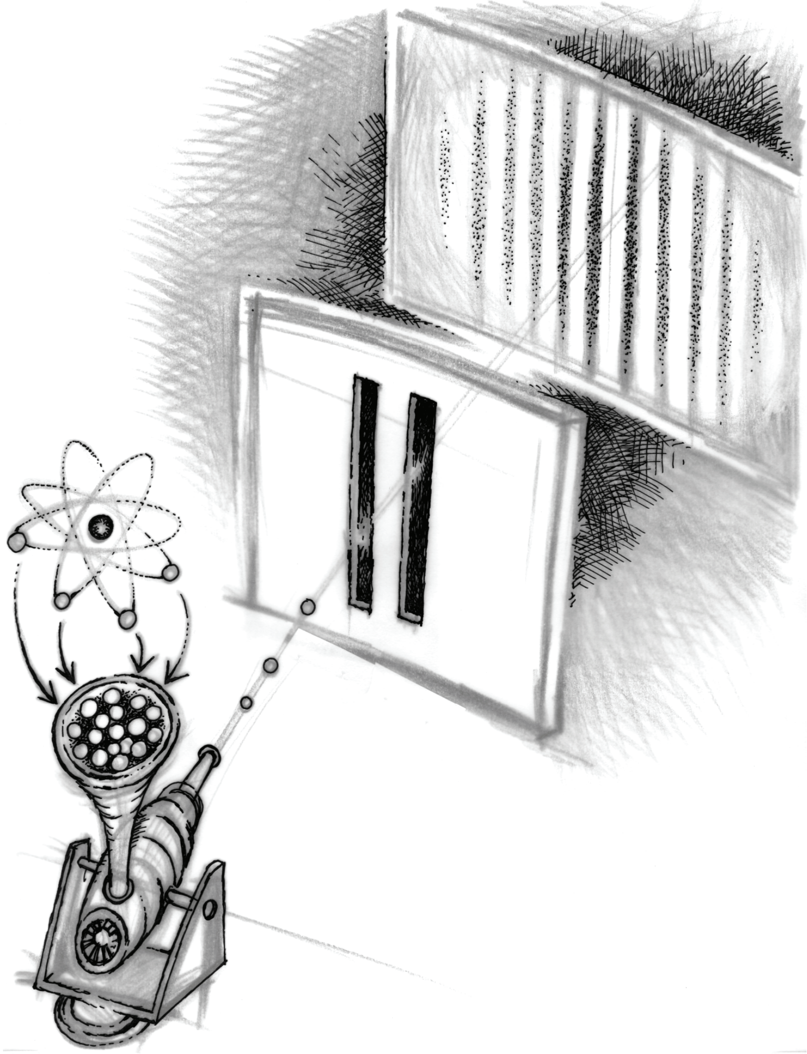


Fig. 2.15: The principle of the double-slit experiment

2.3 Harmonic Oscillator

Using the example of a harmonic oscillator, whose classical solution we learned in Section 1.3.2, we will demonstrate typical methods for finding the eigenvalues of the energy operator. Our task is

$$\hat{\mathbf{H}}\,|\,n\,\rangle = E_n\,|\,n\,\rangle\,,$$

$$► \qquad \hat{\mathbf{H}} \equiv \frac{\hat{\mathbf{P}}^2}{2m} + \frac{1}{2}m\omega^2\hat{\mathbf{X}}^2\,, \tag{2.50}$$

$$\left[\,\hat{\mathbf{X}}, \hat{\mathbf{P}}\,\right] = \mathrm{i}\hbar\,\hat{\mathbf{1}}\,.$$

This is a problem concerning the eigenvalues of the Hamiltonian operator with a specific potential energy profile and given fundamental commutation relations between the position and momentum operators.

In Section 2.3.1, we will solve the problem within the framework of classical Schrödinger wave mechanics. We choose the space $\mathcal{L}^2(\mathcal{R})$ as the Hilbert space; the choice of operators (2.44) and (2.47) leads to the Schrödinger equation (2.49) in one dimension. The solution to this equation is obtained by expanding it into infinite series, which must be "truncated" so that the solution belongs to the space $\mathcal{L}^2(\mathcal{R})$, i.e., is square-integrable. From this, we obtain the spectrum of the energy operator.

In Section 2.3.2, we will demonstrate a solution to problem (2.50) without specifying a representation. We will not choose any particular form of the Hilbert space at all. We will find the solution solely from the formulation of problem (2.50). We will thus see that the specific choice of Hilbert space is not essential. In this approach, we will introduce creation and annihilation operators, which, by their action, shift the energy by one level up or down. These operators are very useful in quantum theory, and therefore we will become familiar with them now in a simple example of harmonic oscillations.

In Section 2.3.3, we will demonstrate a solution to problem (2.50) in the ℓ^2 space of infinite sequences that are square-summable (within the framework of so-called Heisenberg matrix mechanics). Here, the operators will be infinite matrices. You may find it difficult to find the eigenvalues of infinite matrices. However, the problem is not that complicated. If we choose the eigenvectors of the relevant operator as the basis vectors, the matrix corresponding to this operator will be diagonal. The eigenvalues of diagonal matrices are very easy to find—they are simply the elements on the diagonal.

Thus, you will see three different ways of solving the same problem. In quantum theory, what matters is the internal structure of the theory, not the specific representation in which we perform the calculation.

2.3.1 Solution Using Wave Mechanics (Schrödinger)

The Hamiltonian of a one-dimensional harmonic oscillator is given by the sum of the kinetic and potential energies (1.31)

$$H(x,p) = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2\,. \tag{2.51}$$

The Hamilton operator in the space $\mathcal{L}^2(-\infty, +\infty)$ is then given by a simple relation

► $$\hat{\mathsf{H}} = -\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{2}m\omega^2 x^2. \tag{2.52}$$

The corresponding Schrödinger equation for the wave eigenfunction $\psi(x)$ in the space $\mathcal{L}^2(-\infty, +\infty)$ has a simple form

$$\left(-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{2}m\omega^2 x^2\right)\psi(x) = E\,\psi(x). \tag{2.53}$$

This is an ordinary second-order linear differential equation with a nonlinear coefficient for the zero-order derivative. The standard form of this equation (with a unit coefficient for the highest-order derivative) is:

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} + \left(\frac{2mE}{\hbar^2} - \frac{m^2\omega^2}{\hbar^2}x^2\right)\psi = 0. \tag{2.54}$$

We will solve the equation in four steps:

1. Substitution in the independent (inner) variable

We will choose a substitution that makes the equation dimensionless. Let's rearrange the coefficients so that they are symmetric about the variable x

$$\frac{\mathrm{d}^2\psi}{\frac{m\omega}{\hbar}\mathrm{d}x^2} - \frac{m\omega}{\hbar}x^2\,\psi + \frac{2E}{\hbar\omega}\psi\ = 0 \tag{2.55}$$

and let's perform the substitution

$$\xi \equiv \sqrt{\frac{m\omega}{\hbar}}\,x\,, \tag{2.56}$$

after which Schrödinger's equation takes on a dimensionless form

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}\xi^2} - \xi^2\psi + \lambda\psi = 0\,, \qquad \lambda \equiv \frac{2E}{\hbar\omega}. \tag{2.57}$$

2. Substitution in the dependent (outer) variable

In the dependent variable, we will choose a substitution that takes into account the behavior of the wave function as ξ approaches ±∞. For large values of ξ, we can neglect the last term in equation (2.57) compared to the last but one. Approximately, we have

$$\xi \to \pm\infty \qquad \Rightarrow \qquad \frac{\mathrm{d}^2\psi}{\mathrm{d}\xi^2} - \xi^2\psi \approx 0 \qquad \Rightarrow \qquad \psi \approx \mathrm{e}^{\pm\xi^2/2}.$$

(Simply substitute the solution into the original equation and neglect terms with lower powers of ξ.) The positive solution found is evidently not from the $\mathcal{L}^2$ space; the integral of the square over the entire space would be infinite. The wave function must therefore behave like the function $\exp[-\xi^2]$ for large ξ. This leads us to a substitution for the dependent variable

$$\psi(\xi) = \mathrm{e}^{-\xi^2/2}\, u(\xi)\,, \tag{2.58}$$

which gives us the equation

$$u'' - 2\xi u' + (\lambda - 1)\, u = 0\,. \tag{2.59}$$

The derivatives are automatically understood with respect to the new variable ξ. In principle, from a mathematical standpoint, it would be correct to say, "In equation (2.54), we perform substitutions (2.56) and (2.58), and the resulting equation is (2.59)." In points 1 and 2, we just showed what motivates us to make these substitutions, because the procedure is similar for other potential energy profiles.

3. Expanding the solution into a power series

We will look for a solution to the equation (2.59) in the form of a power series

$$u(\xi) = \sum_{k=0}^{\infty} c_k \xi^k .$$

It is easy to find the first and second derivatives

$$u'(\xi) = \sum_{k=0}^{\infty} k\, c_k\, \xi^{k-1}; \qquad u''(\xi) = \sum_{k=0}^{\infty} k(k-1)\, c_k\, \xi^{k-2}.$$

We substitute the expressions for u and its derivatives into the equation (2.59):

$$\sum_{k=0}^{\infty} k(k-1)\, c_k\, \xi^{k-2} - \sum_{k=0}^{\infty} 2k\, c_k\, \xi^k + (\lambda-1) \sum_{k=0}^{\infty} c_k\, \xi^k = 0\,.$$

We will rearrange the terms so that the powers of the variable ξ are the same (in the first term, we set $k - 2 = l$):

$$\sum_{l=-2}^{\infty} (l+1)(l+2)\, c_{l+2}\, \xi^l - \sum_{l=0}^{\infty} 2l\, c_l\, \xi^l + (\lambda-1) \sum_{l=0}^{\infty} c_l\, \xi^l = 0\,.$$

The first two terms of the first sum are zero, so we can set the lower value to $l = 0$:

$$\sum_{l=0}^{\infty} \left[(l+1)(l+2)\, c_{l+2} - (2l+1-\lambda)\, c_l \right] \xi^l = 0\,.$$

For a polynomial expression to be exactly zero for every value of ξ, all coefficients – i.e., the terms in square brackets – must be zero. This gives us a recursive relation for the coefficients c_l of our series:

► $$c_{l+2} = \frac{(2l+1-\lambda)}{(l+1)(l+2)} c_l . \tag{2.60}$$

If we know the coefficients c_0 and c_1, we will know the entire solution, because we can calculate it from the recursive relation

$$c_0 \quad \Rightarrow \quad c_2, c_4, c_6, \dots$$
$$c_1 \quad \Rightarrow \quad c_3, c_5, c_7, \dots .$$

The coefficients c_0 and c_1 thus act as the two integration constants in the solution to the second-order differential equation (2.59). The even terms of the series are calculated using c_0, and the odd terms using c_1.

4. Truncation of the series

The solution found is in the form of an infinite power series. Although it satisfies the original equation, it is not in the $\mathcal{L}^2$ space. For the solution to be in $\mathcal{L}^2$ (i.e., integrable with respect to the square), the series must be finite, i.e., polynomial. In practice, this means that the coefficients of the series must be zero for some $l = n$. In the recursive relation (2.60), the numerator will be zero for this $l = n$, and all derived coefficients c_l with $l \geq n$ will be zero. We see that it will not be possible to "truncate" both even and odd terms of the series in this way. Therefore, only even ($c_0 \neq 0$, $c_1 = 0$) or only odd solutions ($c_0 = 0$, $c_1 \neq 0$) are possible, representing an even or odd polynomial of degree n. The truncation condition (zero numerator) in (2.60) is $2n + 1 - \lambda = 0$, and from this, after expressing λ, follows the energy spectrum of the harmonic oscillator:

► $$E_n = (n + 1/2)\hbar\omega . \qquad (2.61)$$

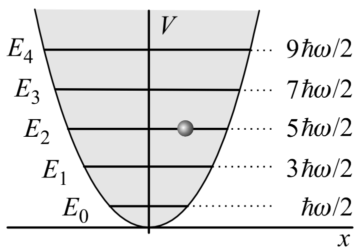


Fig. 2.16: Spectrum of a harmonic oscillator

Note 1: Keep in mind that the energy E (the eigenvalue of the operator $\hat{\mathbf{H}}$) is embedded in the dimensionless constant (eigenvalue) λ throughout the calculation.

Note 2: Schrödinger equation itself has a solution for every energy value. However, these solutions are not square-integrable; only the selection of integrable functions (truncation of the series) leads to a discrete spectrum of the energy operator (only for certain selected energy values do the solutions decrease sufficiently rapidly in the $\pm\infty$ region to be square-integrable). This situation is typical for continuous potential energy functions with a minimum.

Note 3: The ground-state energy $E_0 = \hbar\omega/2$ is nonzero! Even at absolute zero, a harmonic oscillator is not at rest and undergoes zero-point oscillations (such as the oscillations of a crystal lattice). At absolute zero, matter is in a state of lowest possible energy, but not at rest. This is due to the uncertainty principle: we cannot know both the position (zero) and the momentum (also zero) simultaneously.

Note 4: The energy spectrum is equidistant; the difference between any two adjacent energy levels is $\Delta E = E_{n+1} - E_n = \hbar\omega$: This is precisely the well-known Planck relation from the early 20th century. The energy of any oscillation cannot vary continuously, but only in discrete steps (energy quanta)

$$\Delta E = \hbar\omega . \tag{2.62}$$

Note 5: This is also where one of the first opportunities for experimentally determining Planck constant through the measurement of energy quanta arises (for example, in the photoelectric effect: the ejection of electrons from a metal surface by means of energy quanta of electromagnetic radiation – photons). Until then, Planck constant had been the only undetermined parameter of the fundamental quantum postulates. Planck constant, of course, also appears in other equations.

Note 6: The polynomial solutions we found for the function u are called Hermite polynomials and are denoted by $H_n(\xi)$. For a given n, we first determine the dimensionless eigenvalue λ_n

$$\lambda_n \equiv \frac{2E_n}{\hbar\omega} = \frac{2(n+1/2)\,\hbar\omega}{\hbar\omega} = 2n+1$$

and then use c_0 or c_1 (depending on whether the polynomial is even or odd) to determine the other expansion coefficients from the recursive formula (2.60). For $c_0 \neq 0$, $c_1 = 0$ or $c_0 = 0$, $c_1 \neq 0$, the polynomials found are called Hermite polynomials. The first few Hermite polynomials are:

$$H_0(\xi) = 1, \qquad H_3(\xi) = \xi - 2/3\,\xi^3,$$
$$H_1(\xi) = \xi, \qquad H_4(\xi) = 1 - 4\xi^2 + 4/3\,\xi^4,$$
$$H_2(\xi) = 1 - 2\xi^2, \qquad H_5(\xi) = \xi - 4/3\,\xi^3 + 4/15\,\xi^5 \ldots$$

We have set the coefficients c_0 and c_1 equal to one. The degree of the polynomial, n, also indicates the number of zeros of the polynomial (the number of intersections with the ξ-axis).

Note 7: Hermite polynomials can be easily computed in unnormalized form using a recursive formula

$$H_{n+1}(\xi) = 2\xi\,H_n(\xi) - 2n\,H_{n-1}(\xi) .$$

For the first polynomials, we have:

$$H_0(\xi) = 1, \qquad H_3(\xi) = 8\xi^3 - 12\xi,$$
$$H_1(\xi) = 2\xi, \qquad H_4(\xi) = 16\xi^4 - 48\,\xi^2 + 12,$$
$$H_2(\xi) = 4\xi^2 - 2, \qquad H_5(\xi) = 32\xi^5 - 160\xi^3 + 120\xi \ldots$$

The normalization coefficients of the wave function $H_n(\xi)\exp[-\xi^2/2]$ are given by the relation

$$\alpha_n = \frac{1}{\sqrt{\pi^{1/2}\; n!\, 2^n}} .$$

Note 8: The complete solution to the spectral problem is

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega,$$
$$|\,n\,\rangle = \psi_n(\xi) = \alpha_n H_n(\xi)\,\mathrm{e}^{-\xi^2/2}\,; \quad n = 0,1,2,\dots \tag{2.63}$$

The eigenfunctions $\psi_n(\xi)$ form a natural complete orthonormal basis on the Hilbert space $\mathcal{L}^2(-\infty, +\infty)$, which tends to zero "sufficiently rapidly" as $\xi \to \pm\infty$.

Note 9: The probability density that a particle oscillating with energy E_n (an oscillator in the state $|\,n\,\rangle$) is located at position x (or the dimensionless position ξ) is given by the expression $w_n = \psi_n{}^*\psi_n$. This is plotted in the figure for the first few states. The probability has an oscillatory character, and there is a small non-zero probability of the oscillator occurring even beyond the classical turning points. This picture arises for low-temperature systems and is completely different from the classical solution. For large energies (high n), the curve should approach the classical probability of the oscillator occurring (1.36). However, we see that while the oscillations are very dense, there are a significant number of points where the quantum probability is zero. Yet we do not measure anything like this in macroscopic systems. Why? This is due to the resolving power of macroscopic instruments. No instrument can measure position with sufficient precision to detect individual probability minima in high-energy states. In reality, the instrument determines the position with finite precision, within which a number of minima fits, and records only the average value of the probability density. And that is precisely the classical curve, which is shown in the figure as the gray region.

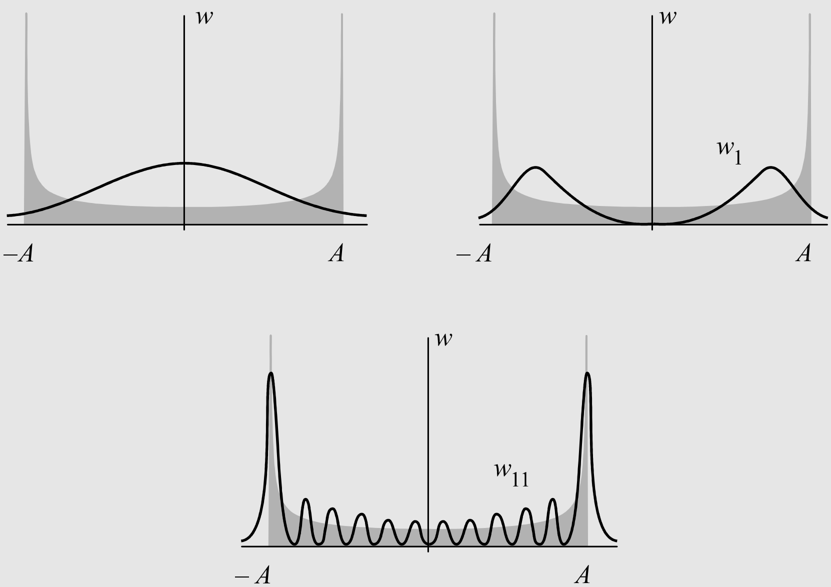


Fig. 2.17: The quantum probability of the oscillator's state. Note that the probability is highest at the edges (except in the ground state). See also Figure 2.18.

2.3.2 Solution Without Representation (Dirac)

We will now solve problem (2.50) in general. First, we rewrite the Hamiltonian operator in dimensionless form:

$$\hat{\mathbf{H}}\left(\hat{\mathbf{X}},\hat{\mathbf{P}}\right) \equiv \frac{1}{2}m\omega^2\hat{\mathbf{X}}^2 + \frac{\hat{\mathbf{P}}^2}{2m} \quad \rightarrow \quad \frac{\hat{\mathbf{H}}}{\hbar\omega} \equiv \frac{m\omega}{2\hbar}\hat{\mathbf{X}}^2 + \frac{1}{2m\hbar\omega}\hat{\mathbf{P}}^2 . \qquad (2.64)$$

It is not necessary to convert to a dimensionless form; all further considerations could also be carried out using the dimensional Hamiltonian, and all the following relations would differ by the constant $\hbar\omega$ by which we divided the Hamiltonian. The results from the dimensionless Hamiltonian are more intuitive. For commuting quantities, the sum of squares can be "de-squared" using the relation $a^2+b^2=(a+\mathrm{i}b)(a-\mathrm{i}b)$. For non-commuting objects, the situation is not so simple. Let us introduce the operators:

►

$$\hat{\mathbf{a}} \equiv \sqrt{\frac{m\omega}{2\hbar}}\,\hat{\mathbf{X}} + \mathrm{i}\sqrt{\frac{1}{2m\hbar\omega}}\,\hat{\mathbf{P}}\,;$$

$$\hat{\mathbf{a}}^\dagger \equiv \sqrt{\frac{m\omega}{2\hbar}}\,\hat{\mathbf{X}} - \mathrm{i}\sqrt{\frac{1}{2m\hbar\omega}}\,\hat{\mathbf{P}}\,. \qquad (2.65)$$

Both of these operators are very important in quantum theory. They are called annihilation and creation operators (we will see the meaning later). Creation and annihilation operators, as some of the few in quantum theory, are not Hermitian and therefore do not act equally on both parts of a scalar product. The creation operator is the Hermitian conjugate of the annihilation operator. Certain important relations hold for them, e.g.:

►

$$\begin{aligned}
&(1) \quad \hat{\mathbf{a}}^\dagger\hat{\mathbf{a}} = \frac{\hat{\mathbf{H}}}{\hbar\omega} - \frac{1}{2},\\
&(2) \quad \hat{\mathbf{a}}\,\hat{\mathbf{a}}^\dagger = \frac{\hat{\mathbf{H}}}{\hbar\omega} + \frac{1}{2},\\
&(3) \quad \hat{\mathbf{X}} = \sqrt{\frac{\hbar}{2m\omega}}\left(\hat{\mathbf{a}}^\dagger + \hat{\mathbf{a}}\right),\\
&(4) \quad \hat{\mathbf{P}} = \mathrm{i}\sqrt{\frac{m\hbar\omega}{2}}\left(\hat{\mathbf{a}}^\dagger - \hat{\mathbf{a}}\right),\\
&(5) \quad \left[\hat{\mathbf{H}},\hat{\mathbf{a}}\right] = -\hbar\omega\hat{\mathbf{a}},\\
&(6) \quad \left[\hat{\mathbf{H}},\hat{\mathbf{a}}^\dagger\right] = +\hbar\omega\hat{\mathbf{a}}^\dagger,\\
&(7) \quad \left[\hat{\mathbf{a}},\hat{\mathbf{a}}^\dagger\right] = \hat{\mathbf{1}}.
\end{aligned} \qquad (2.66)$$

The proof of all the relations is trivial. It suffices to substitute the definitions of the operators $\hat{\mathbf{a}}^\dagger$, $\hat{\mathbf{a}}$ (2.65) and apply the basic commutation relations

$$[\hat{\mathbf{X}},\hat{\mathbf{P}}] = \mathrm{i}\hbar\,\hat{\mathbf{1}}.$$

Relations (1) and (2) are generalizations of the relationship

$$a^2 + b^2 = (a + \mathrm{i}b)(a - \mathrm{i}b)$$

for non-commuting objects and represent the formal square root of the Hamiltonian. The creation and annihilation operators are linear combinations of the position operator and the momentum operator. Conversely, it is therefore possible to express the position and momentum operators as linear combinations of the creation and annihilation operators – see relations (3) and (4). If we know the creation and annihilation operators, we can also reconstruct the Hamiltonian from relations (1) through (4). The commutation relations (5) through (7) express the fundamental properties of the creation and annihilation operators: We will see that relation (5) means that the annihilation operator shifts the system's states down by an energy level of $\hbar\omega$, and relation (6) means that the creation operator shifts the state up by an energy level of $\hbar\omega$.

In the following theorem, we will prove that the operator $\hat{\mathbf{a}}^\dagger$ is a creation operator, i.e., it shifts energy states up by one unit (creates an energy quantum).

The theorem on the creation operator

When the creation operator acts on the current energy state, we transition to the next energy state. Let $\hat{\mathbf{H}}\,|\,n\,\rangle = E_n\,|\,n\,\rangle$; then $\hat{\mathbf{a}}^\dagger\,|\,n\,\rangle \sim |\,n+1\,\rangle$.

Proof:

$$\hat{\mathbf{H}}\hat{\mathbf{a}}^\dagger\,|\,n\,\rangle \overset{(6)}{=} (\hat{\mathbf{a}}^\dagger\,\hat{\mathbf{H}} + \hbar\omega\hat{\mathbf{a}}^\dagger)|\,n\,\rangle = (\hat{\mathbf{a}}^\dagger E_n + \hbar\omega\hat{\mathbf{a}}^\dagger)|\,n\,\rangle = (E_n + \hbar\omega)\,\hat{\mathbf{a}}^\dagger\,|\,n\,\rangle$$

$$\Downarrow$$

$$\hat{\mathbf{H}}\hat{\mathbf{a}}^\dagger\,|\,n\,\rangle = (E_n + \hbar\omega)\,\hat{\mathbf{a}}^\dagger\,|\,n\,\rangle$$

$$\Downarrow$$

$$\hat{\mathbf{a}}^\dagger\,|\,n > \sim |\,n+1 > . \qquad \blacksquare$$

By a similar line of reasoning, we can show from relation (5) in set (2.66) that the annihilation operator satisfies $\hat{\mathbf{a}}\,|\,n\,\rangle \sim |\,n-1\,\rangle$. If we introduce normalization constants (we require that all states be normalized to one, i.e., form an orthonormal basis of the energy operator's eigenvectors), we can simply write the shifts in the energy spectrum caused by the creation and annihilation operators as the following equalities

$$\begin{aligned} \hat{\mathbf{a}}^\dagger\,|\,n\,\rangle &= \alpha_n^+\,|\,n+1\,\rangle\,, \\ \hat{\mathbf{a}}\,|\,n\,\rangle &= \alpha_n^-\,|\,n-1\,\rangle\,. \end{aligned} \qquad (2.67)$$

We will determine the normalization constants α later. For now, we will focus on finding the spectrum of the Hamiltonian operator for a harmonic oscillator without specifying the choice of the corresponding Hilbert space.

The Hamiltonian operator is the sum of the squares of two Hermitian operators and is therefore positive definite, i.e., its eigenvalues are non-negative. The creation and annihilation operators shift the energy spectrum by a constant value (quantum). There must therefore exist a state with the lowest possible non-negative energy. We call this state the ground state and denote it by $|\,\mathrm{GS}\,\rangle$. If we act on the ground state with the annihilation operator, we must obtain the zero vector $|\,\mathbf{0}\,\rangle$ with zero magnitude, which does not form a beam and is not a physical state, because in the ground state there is nothing left to annihilate (we are in the state with the lowest possible energy). Therefore:

$$\hat{\mathsf{H}}\,|\,\mathrm{GS}\,\rangle = E_0\,|\,\mathrm{GS}\,\rangle\,; \qquad \hat{\mathsf{a}}\,|\,\mathrm{GS}\,\rangle = |\,\mathbf{0}\,\rangle\,.$$

Let's find the square of the last relation (the scalar product of an element with itself):

$$\langle\,\mathrm{GS}\,|\,\hat{\mathsf{a}}^\dagger\,\hat{\mathsf{a}}\,|\,\mathrm{GS}\,\rangle = 0 \quad \overset{(2.66.1)}{\Rightarrow} \quad \langle\,\mathrm{GS}\,|\,\frac{\hat{\mathsf{H}}}{\hbar\omega} - \frac{\hat{\mathsf{1}}}{2}\,|\,\mathrm{GS}\,\rangle = 0 \quad \Rightarrow$$

$$\frac{1}{\hbar\omega}\langle\,\mathrm{GS}\,|\,\hat{\mathsf{H}}\,|\,\mathrm{GS}\,\rangle - \frac{1}{2}\langle\,\mathrm{GS}\,|\,\hat{\mathsf{1}}\,|\,\mathrm{GS}\,\rangle = 0 \quad \Rightarrow \quad \left(\frac{E_0}{\hbar\omega} - \frac{1}{2}\right)\langle\,\mathrm{GS}\,|\,\mathrm{GS}\,\rangle = 0 \quad \Rightarrow$$

$$\frac{E_0}{\hbar\omega} - \frac{1}{2} = 0 \quad \Rightarrow \quad E_0 = \frac{\hbar\omega}{2}\,.$$

If we know the ground state energy, we can obtain other energy values by applying the creation operator, which shifts the energy by a constant $\hbar\omega$; it is therefore clear that

$$\begin{aligned} E_1 &= E_0 + \hbar\omega = \frac{3}{2}\hbar\omega\,, \\ E_2 &= E_0 + 2\hbar\omega = \frac{5}{2}\hbar\omega\,, \\ &\vdots \\ E_n &= E_0 + n\hbar\omega = \left(n + \frac{1}{2}\right)\hbar\omega\,; \qquad n = 0,1,2,\ldots \end{aligned}$$

We derived the spectrum of the oscillator only from the properties of the Hamiltonian operator. Nowhere did we choose a specific representation or a specific Hilbert space. The creation and annihilation operators, which we encountered here for the first time, are of considerable importance in quantum field theory, where we use similar operators to create and annihilate particles present in the system. Here, for the harmonic oscillator, we merely create or annihilate an energy quantum, thereby moving up or down one energy level. To complete our derivation, we will finally determine the normalization constants in expression (2.67). Let us start from the basic relations for both operators

$$\begin{aligned} \hat{\mathsf{a}}^\dagger\,|\,n\,\rangle &= \alpha_n^+\,|\,n+1\,\rangle\,, \\ \hat{\mathsf{a}}\,|\,n\,\rangle &= \alpha_n^-\,|\,n-1\,\rangle\,. \end{aligned}$$

First, we'll square both relations

$$\begin{aligned} \langle\,n\,|\,\hat{\mathsf{a}}\,\hat{\mathsf{a}}^\dagger\,|\,n\,\rangle &= \left|\alpha_n^+\right|^2\,\langle\,n+1\,|\,n+1\,\rangle\,, \\ \langle\,n\,|\,\hat{\mathsf{a}}^\dagger\hat{\mathsf{a}}\,|\,n\,\rangle &= \left|\alpha_n^-\right|^2\,\langle\,n-1\,|\,n-1\,\rangle\,. \end{aligned}$$

We express the products of the operators $\hat{\mathsf{a}}$, $\hat{\mathsf{a}}^\dagger$ using relations (1) and (2) of set (2.66):

$$\begin{aligned} \langle\,n\,|\,\frac{\hat{\mathsf{H}}}{\hbar\omega} + \frac{1}{2}\,|\,n\,\rangle &= \left|\,\alpha_n^+\right|^2\,\langle\,n+1\,|\,n+1\,\rangle, \\ \langle\,n\,|\,\frac{\hat{\mathsf{H}}}{\hbar\omega} - \frac{1}{2}\,|\,n\,\rangle &= \left|\,\alpha_n^-\right|^2\,\langle\,n-1\,|\,n-1\,\rangle. \end{aligned}$$

Applying the Hamilton operator from the left, we obtain

$$\left(\frac{E_n}{\hbar\omega} + \frac{1}{2}\right)\langle\, n \,|\, n\,\rangle = \left|\alpha_n^+\right|^2 \;\langle\, n+1 \,|\, n+1\,\rangle\,,$$
$$\left(\frac{E_n}{\hbar\omega} - \frac{1}{2}\right)\langle\, n \,|\, n\,\rangle = \left|\alpha_n^-\right|^2 \;\langle\, n-1 \,|\, n-1\,\rangle\,.$$

Eigenvectors of the energy operator are normalized to one, therefore we have:

$$\left|\,\alpha_n^+\right|^2 = \left(\frac{E_n}{\hbar\omega} + \frac{1}{2}\right) = \left(\frac{(n+1/2)\,\hbar\omega}{\hbar\omega} + \frac{1}{2}\right) = n+1\,,$$

$$\left|\,\alpha_n^-\right|^2 = \left(\frac{E_n}{\hbar\omega} - \frac{1}{2}\right) = \left(\frac{(n+1/2)\,\hbar\omega}{\hbar\omega} - \frac{1}{2}\right) = n\,.$$

The phase factor is not significant when square-rooting a complex number (it does not affect the unit magnitude of the vector). The resulting action of the creation and annihilation operators (2.67), including the normalization constant, is therefore:

► $$\hat{\mathbf{a}}^\dagger\,|\,n\,\rangle = \sqrt{n+1}\,|\,n+1\,\rangle\,, \qquad \hat{\mathbf{a}}\,|\,n\,\rangle = \sqrt{n}\,|\,n-1\,\rangle\,. \tag{2.68}$$

This result is easy to remember: The square root always contains the order number of the higher energy state on either side of the equation. There is another operator with interesting properties:

► $$\hat{\mathbf{N}} \equiv \hat{\mathbf{a}}^\dagger\,\hat{\mathbf{a}}\,. \tag{2.69}$$

Let's apply this operator to the state $|\,n\,\rangle$; using relations (2.68). We obtain

$$\hat{\mathbf{N}}\,|\,n\,\rangle = \hat{\mathbf{a}}^\dagger\,\hat{\mathbf{a}}\,|\,n\,\rangle = \sqrt{n}\,\hat{\mathbf{a}}^\dagger\,|\,n-1\,\rangle = \sqrt{n}\,\sqrt{n}\,|\,n\,\rangle = n\,|\,n\,\rangle\,.$$

The eigenvalue of this operator is the number of quanta present in a given energy state. In quantum field theory, this operator corresponds to the *particle number operator*, and the following relation holds for it

► $$\hat{\mathbf{N}}\,|\,n\,\rangle = n\,|\,n\,\rangle\,. \tag{2.70}$$

2.3.3 Solution Using Matrix Mechanics (Heisenberg)

Let us revisit problem (2.50) of the harmonic oscillator in the space of infinite square-summable sequences (ℓ^2). In the space of n-tuples, the operators are $n\times n$ square matrices. In the space of infinite sequences ($n\to\infty$), the operators will be matrices of infinite dimension. The task is therefore to find infinite-dimensional matrices **X**, **P**, and **H**, that satisfy (2.50). We do not have to search for these matrices "from scratch." With what we know about creation and annihilation operators, we can easily construct them. We find them in the energy representation – that is, we determine the matrix elements of the position, momentum, and energy operators in a basis formed from the eigenvectors of the Hamiltonian operator. We can construct all three operators using creation and annihilation operators according to relation (2.66). And we also know the action of creation and annihilation operators on the chosen basis – see relation (2.68). The construction of the elements of the corresponding matrices is therefore more or less a trivial matter:

$$X_{kl} = \langle k | \hat{\mathbf{X}} | l \rangle = \sqrt{\frac{\hbar}{2m\omega}} \langle k | \left(\hat{\mathbf{a}}^\dagger + \hat{\mathbf{a}} \right) | l \rangle =$$

$$= \sqrt{\frac{\hbar}{2m\omega}} \left(\sqrt{l+1} \langle k | l+1 \rangle + \sqrt{l} \langle k | l-1 \rangle \right) =$$

$$= \sqrt{\frac{\hbar}{2m\omega}} \left(\sqrt{l+1}\, \delta_{k,\, l+1} + \sqrt{l}\, \delta_{k,\, l-1} \right), \qquad k,l = 0,1,2\ldots.$$

We expressed the position operator in terms of (2.66) using the creation and annihilation operators, and then applied them as in (2.68). Similarly, we have for the momentum

$$P_{kl} = \langle k | \hat{\mathbf{P}} | l \rangle = \mathrm{i} \sqrt{\frac{m\hbar\omega}{2}} \langle k | \left(\hat{\mathbf{a}}^\dagger - \hat{\mathbf{a}} \right) | l \rangle =$$

$$= \mathrm{i} \sqrt{\frac{m\hbar\omega}{2}} \left(\sqrt{l+1} \langle k | l+1 \rangle - \sqrt{l} \langle k | l-1 \rangle \right) =$$

$$= \mathrm{i} \sqrt{\frac{m\hbar\omega}{2}} \left(\sqrt{l+1}\, \delta_{k,\, l+1} - \sqrt{l}\, \delta_{k,\, l-1} \right), \qquad k,l = 0,1,2\ldots$$

Finally, we find the matrix of the Hamiltonian operator. This matrix is the only one that must be diagonal, since it is a basis of the eigenstates of the energy operator:

$$H_{kl} = \langle k | \hat{\mathbf{H}} | l \rangle = \hbar\omega \langle k | \left(\hat{\mathbf{a}}^\dagger \hat{\mathbf{a}} + \frac{1}{2} \right) | l \rangle = \cdots$$

$$= \left(l + \frac{1}{2} \right) \hbar\omega\, \delta_{kl}, \qquad k,l = 0,1,2\ldots.$$

Let's now write down the matrices we've found:

$$\mathbf{X} = \sqrt{\frac{\hbar}{2m\omega}} \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \cdots \\ \sqrt{1} & 0 & \sqrt{2} & 0 & \\ 0 & \sqrt{2} & 0 & \sqrt{3} & \\ 0 & 0 & \sqrt{3} & 0 & \\ \vdots & & & & \ddots \end{pmatrix},$$

$$\mathbf{P} = \mathrm{i} \sqrt{\frac{m\hbar\omega}{2}} \begin{pmatrix} 0 & -\sqrt{1} & 0 & 0 & \cdots \\ \sqrt{1} & 0 & -\sqrt{2} & 0 & \\ 0 & \sqrt{2} & 0 & -\sqrt{3} & \\ 0 & 0 & \sqrt{3} & 0 & \\ \vdots & & & & \ddots \end{pmatrix},$$

$$\mathbf{H} = \begin{pmatrix} \frac{\hbar\omega}{2} & 0 & 0 & \cdots \\ 0 & \frac{3\hbar\omega}{2} & 0 & \\ 0 & 0 & \frac{5\hbar\omega}{2} & \\ \vdots & & & \ddots \end{pmatrix}.$$

Verify that indeed $[\mathbf{X}, \mathbf{P}] = \mathrm{i}\hbar\,\mathbf{1}$ and that $\mathbf{H} = \mathbf{P}^2/2m + m\omega^2\mathbf{X}^2/2$ holds, as required by problem (2.50). Given the matrices **X** and **P**, we could already determine the Hamilton matrix directly from this relation. The last thing remaining is to find the eigenvalues of the matrix **H**. This task is extremely simple. For a diagonal matrix, the eigenvalues are precisely the elements on the diagonal. The calculation is simple:

$$\mathbf{H}\,|\psi\rangle = E\,|\psi\rangle \quad \Rightarrow \quad (\mathbf{H} - \mathbf{1}E)\,|\psi\rangle = 0 \quad \Rightarrow \quad \det\,(\mathbf{H} - \mathbf{1}E) = 0 \quad \Rightarrow$$

$$\left(\frac{\hbar\omega}{2} - E\right)\cdot\left(\frac{3\hbar\omega}{2} - E\right)\cdot\left(\frac{5\hbar\omega}{2} - E\right) \cdots = 0 \quad \Rightarrow$$

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega \quad , \qquad n = 0, 1, 2, \ldots$$

So once again, we have a relationship (2.61) for the spectrum of a harmonic oscillator.

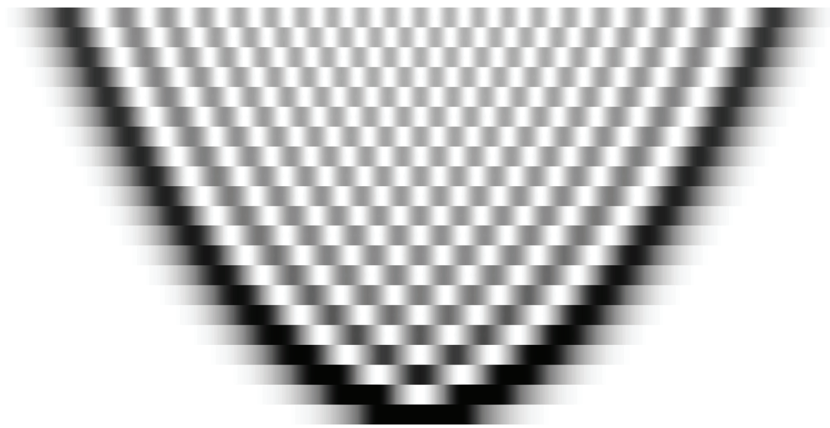


Fig. 2.18: Probability density of the oscillator from the ground state to the 20th state. The lower band corresponds to the ground state, and the upper band to the 20th state. White indicates the minimum probability, and black the maximum probability of occurrence. At the turnpoints (on the parabola), the probability of the particle's occurrence is clearly highest. Source: Wikipedia.

2.4 Simple One-dimensional Systems

In a real-world situation, a particle occasionally finds itself at a potential minimum. For the general form of the potential, an analytical solution is not possible, so various approximations come into play. Near the minimum, the potential can be replaced by a parabolic function, allowing us to use the known solution for a harmonic oscillator. Another option is to replace the potential profile with a rectangular well. The Schrödinger equation in $\mathcal{L}^2$ space leads to a differential equation with constant coefficients, whose solution is exceptionally simple. It is more complicated to connect the solutions in different regions, which, in the case of a finite-height potential well, leads to a rather elegant graphical solution for the energy levels. If the well is narrow and deep, we can replace it in a first approximation with an infinite potential well, and then the entire solution is extremely simple.

2.4.1 Infinite Rectangular Well

Let us consider the motion of a particle in the potential of an infinite rectangular well

$$V(x) = \begin{cases} 0; & x \in (0, L)\,, \\ \infty; & x \notin (0, L)\,. \end{cases} \tag{2.71}$$

This is, of course, a physical idealization that does not exist in the real world. We divide the potential into three regions, as shown in the figure:

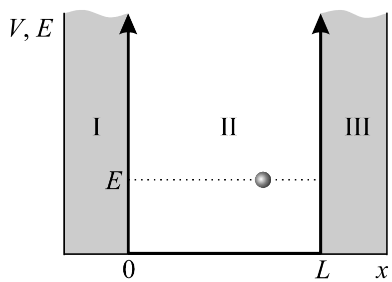


Fig. 2.19: Infinite rectangular well

In regions I and III, the potential is infinite, and the only possible solution to the time-independent Schrödinger equation is $\psi = 0$. From a physical standpoint, this means that the probability of finding a particle outside the potential well is zero. If the potential well were finite (i.e., the potential outside the well were finite), the wave function ψ would be nonzero in the immediate vicinity of the well boundary. The particle would have a small but nonzero probability of existing even beyond the well boundary. In region II, the Schrödinger equation (2.49) takes the form

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} = E\psi, \tag{2.72}$$

which can be rewritten as the standard equation of oscillations in the variable x

$$\frac{d^2\psi}{dx^2} + k^2\psi = 0\,; \qquad k^2 \equiv \frac{2mE}{\hbar^2}\,, \tag{2.73}$$

the solution to which is

$$\psi = A\cos kx + B\sin kx\,. \tag{2.74}$$

The wave function must be continuous at both boundaries of the well; otherwise, the first derivative of the wave function would be the derivative of a step function, i.e., a distribution, and the second derivative would even behave like the derivative of a distribution, which contradicts the original equation (2.72). Therefore, the boundary conditions $\psi(0) = 0$, $\psi(L) = 0$ must hold, from which it follows that

$$A = 0\,; \qquad k = n\frac{\pi}{L}\,; \qquad n = 1, 2, 3\ldots \tag{2.75}$$

Quantization is precisely the result of applying the boundary condition. The value of k in which energy is "hidden" according to equation (2.73) cannot take on arbitrary values. Condition (2.75) is therefore nothing other than the quantization of energy:

► $$E_n = \frac{\pi^2\hbar^2}{2mL^2}n^2\,; \qquad n = 1, 2, 3\ldots \tag{2.76}$$

The ground state energy is non-zero, just as in a harmonic oscillator. The reason for this is once again Heisenberg uncertainty principle – a particle moving within a potential well is localized within a finite region and therefore cannot have zero momentum or energy. For $n = 0$, the solution would be zero, which would contradict the fact that a particle is present in the potential well. Unlike the harmonic oscillator, the energy spectrum is not equidistant, and as n increases, the difference between two adjacent energy levels grows. The wave function itself has the form

$$\psi_n(x) = B\sin k_n x = B\sin\left(n\frac{\pi}{L}x\right). \tag{2.77}$$

The solution consists of entire rays in the Hilbert space $\mathcal{L}^2$. From these, we select unit vectors satisfying

$$\langle\psi_n|\psi_n\rangle = 1 \qquad \Rightarrow \qquad \int_0^L B^2\sin^2(n\pi x/L)\,dx = 1 \qquad \Rightarrow \qquad B = \sqrt{\frac{2}{L}}\,.$$

After taking the square root, the normalizing constant could also have a negative or complex value. Since the task was to select one of the unit vectors of the beam, we can use any solution. The resulting solution in region II is therefore

► $$|n\rangle = \psi_n(x) = \sqrt{\frac{2}{L}}\sin\left(n\frac{\pi}{L}x\right); \qquad n = 1, 2, 3\ldots \tag{2.78}$$

The probability of a particle being in the well (in region II) is then

► $$w_n(x) = \psi_n^*\psi_n = \frac{2}{L}\sin^2\left(n\frac{\pi}{L}x\right); \qquad n = 1, 2, 3\ldots \tag{2.79}$$

Figure 2.20 shows the results of the first three solutions. It is clear that within the infinite rectangular well, there are regions where the probability of finding a particle is zero (as in the harmonic oscillator). Outside the well, the particle does not appear at all.

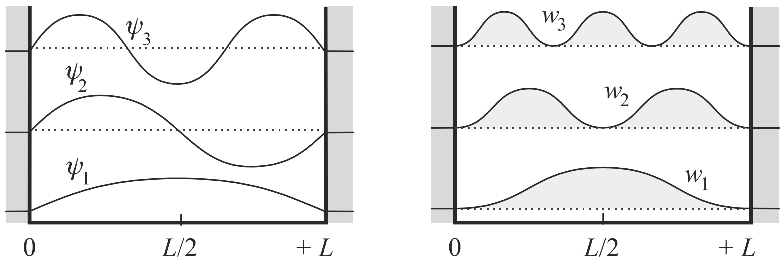


Fig. 2.20: Wave functions (left) and probabilities (right) of the first three states

If we average the position using the probabilities w_n, we can easily see that the average position of the particle is exactly in the center of the potential well, regardless of which energy state the particle is in:

$$\blacktriangleright \qquad \langle x \rangle = \langle n | x | n \rangle = \int_0^L \psi_n^* x \psi_n \, \mathrm{d}x = \int_0^L \frac{2x}{L} \sin^2\left(n \frac{\pi}{L} x \right) \mathrm{d}x = \frac{L}{2}. \qquad (2.80)$$

◖ Example 2.2: Electron in the well

An electron in a semiconductor is located in an electric field whose potential takes the form of a deep well 1 nm wide. Determine the first three energy levels. Then solve the problem for a well 1 mm wide. Find the energy difference between the third and second energy levels.

Solution: We determine the values using the equation (2.76):

width	E_1	E_2	E_3	ΔE_{23}
1 nm	6×10^{-20} J = 0,38 eV	1,5 eV	3,4 eV	1,9 eV
1 mm	6×10^{-32} J = 0,38 peV	1,5 peV	3,4 peV	1,9 peV

For a well 1 nm wide, quantization is significant, and the energy levels are comparable to the energy of an electron in the atomic shell. For a well 1 mm wide, the energies are many orders of magnitude lower, and the differences between energy levels are not observable by conventional means. ◗

2.4.2 Finite Rectangular Well

Another approximation of a potential minimum is a finite well. This is a useful approximation, for example, for neutron-proton interactions in the atomic nucleus, although in this case we should strictly speak solve a three-dimensional finite well. The width of the well in this case is approximately 10^{-15} m, which is the range of the strong interaction. The energy spectrum cannot be determined analytically. Fortunately, there is a simple geometric method that leads to the spectrum. We will assume that the well is symmetrical with respect to the origin of the coordinate system, which will simplify the calculation, which again breaks into three regions. Both the first and second derivatives of the wave function must be continuous at the interface; otherwise, the second derivative in the Schrödinger equation would be an unacceptable distribution.

We will therefore solve Schrödinger equation with a finite-potential well:

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2}+V(x)\psi=E\psi\,,\qquad V(x)=\begin{cases}0; & x\in(-L/2,L/2)\,,\\ V_0; & x\notin(-L/2,L/2)\,.\end{cases}\tag{2.81}$$

In regions I, II, and III, the potential is constant and the solution is simple. In regions I and III, the solution is a combination of increasing and decreasing exponential functions; in each region, only one of these functions can be integrated with a square, and the other must be excluded (by setting its constant to 0). In region II, the solution consists of cosine and sine functions:

$$\begin{aligned}&\psi_{\mathrm{I}}(x)=C\,\mathrm{e}^{hx}\,; && x\in\mathrm{I}\,, && h^2\equiv\frac{2m(V_0-E)}{\hbar^2}\,;\\ &\psi_{\mathrm{II}}(x)=A\cos kx+B\sin kx\,; && x\in\mathrm{II}\,, && \\ &\psi_{\mathrm{III}}(x)=D\,\mathrm{e}^{-hx}\,; && x\in\mathrm{III}\,, && k^2\equiv\frac{2mE}{\hbar^2}\,.\end{aligned}\tag{2.82}$$

The continuity conditions on the left and right sides of the well lead to the equations

$$\begin{aligned}\psi_{\mathrm{II}}(-L/2)&=\psi_{\mathrm{I}}(-L/2)\,;\\ \psi_{\mathrm{II}}(+L/2)&=\psi_{\mathrm{III}}(+L/2)\,;\\ \psi'_{\mathrm{II}}(-L/2)&=\psi'_{\mathrm{I}}(-L/2)\,;\\ \psi'_{\mathrm{II}}(+L/2)&=\psi'_{\mathrm{III}}(+L/2)\,.\end{aligned}\tag{2.83}$$

After substituting, we obtain the relations between the constants A, B, C, and D:

$$\begin{aligned}A\cos(kL/2)-B\sin(kL/2)&=C\,\mathrm{e}^{-hL/2}\,;\\ A\cos(kL/2)+B\sin(kL/2)&=D\,\mathrm{e}^{-hL/2}\,;\\ +Ak\sin(kL/2)+Bk\cos(kL/2)&=Ch\,\mathrm{e}^{-hL/2}\,;\\ -Ak\sin(kL/2)+Bk\cos(kL/2)&=-Dh\,\mathrm{e}^{-hL/2}\,.\end{aligned}\tag{2.84}$$

We'll add and subtract the first pair of equations. We'll do the same with the second pair

$$\begin{aligned}2A\cos(kL/2)&=(D+C)\,\mathrm{e}^{-hL/2}\,;\\ 2B\sin(kL/2)&=(D-C)\,\mathrm{e}^{-hL/2}\,;\\ 2Bk\cos(kL/2)&=(C-D)h\,\mathrm{e}^{-hL/2}\,;\\ 2Ak\sin(kL/2)&=(C+D)h\,\mathrm{e}^{-hL/2}\,.\end{aligned}\tag{2.85}$$

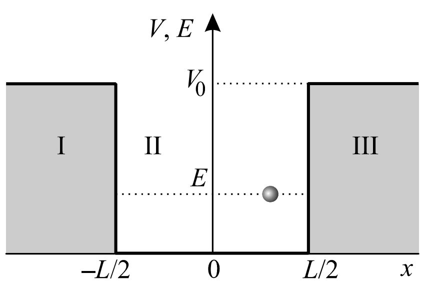


Fig. 2.21: Finite rectangular well

If the constant A is nonzero, we can perform the calculation using the first and last equations; if B is nonzero, we can use the second and third equations:

$$\begin{aligned} A \neq 0 \quad &\Rightarrow \quad k\,\mathrm{tg}\,(kL/2) = h\,, \\ B \neq 0 \quad &\Rightarrow \quad k\,\mathrm{ctg}\,(kL/2) = -h\,. \end{aligned} \tag{2.86}$$

If both constants were nonzero, we would immediately encounter a contradiction. If both constants were zero, we would have only a zero solution for the wave function. Therefore, only two options are possible

1. $A \neq 0$, $B = 0$. According to (2.82), this is an even solution.
2. $A = 0$, $B \neq 0$. According to (2.82), this is an odd solution.

In both cases, it is now easy to calculate the constants C and D from (2.85). We determine the final constant A (in case 1) or B (in case 2) from the normalization condition $\langle \psi | \psi \rangle = 1$. However, we are more interested in the energy spectrum of the particle in the finite well. This is determined by the conditions (2.86), which are transcendental equations. The variables k and h contain the energy. The first condition is for even solutions, the second for odd ones. Let us introduce new variables

$$\xi \equiv kL/2\,; \qquad \eta \equiv hL/2\,. \tag{2.87}$$

Both variables are dimensionless, and the spectral conditions change to

► $$\begin{aligned} \eta &= \xi\,\mathrm{tg}\,\xi\,; \quad &\text{even solutions,} \\ \eta &= -\xi\,\mathrm{ctg}\,\xi\,; \quad &\text{odd solutions.} \end{aligned} \tag{2.88}$$

We can easily plot both relations on a graph in the (ξ, η) plane, i.e., we will plot the values of ξ on the horizontal axis and the calculated values η on the vertical axis. For new variables (ξ, η), there is an interesting property that follows from definitions (2.82):

$$\xi^2 + \eta^2 = \frac{(k^2 + h^2)L^2}{4} = \frac{mV_0 L^2}{2\hbar^2}\,.$$

It is the equation of a circle whose radius is determined by the parameters of the well:

► $$\xi^2 + \eta^2 = R^2\,; \qquad R \equiv \sqrt{\frac{mV_0 L^2}{2\hbar^2}}\,. \tag{2.89}$$

The intersection of this circle with the curves (2.88) provide the energy spectrum

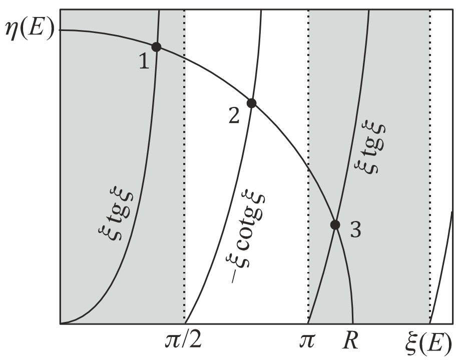


Fig. 2.22: Graphical determination of energy levels using the circle and the curves intersections

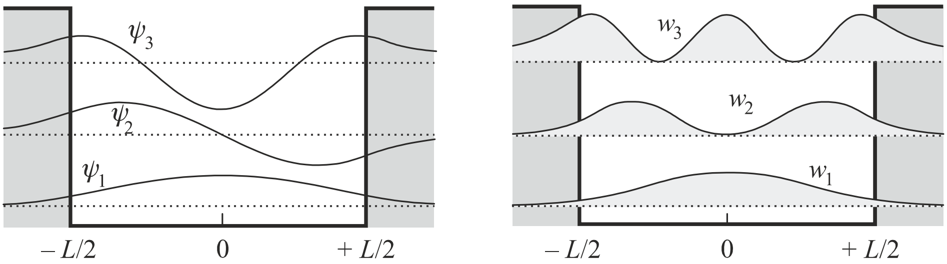


Fig. 2.23: Wave functions (left) and probabilities (right) of the first three states

Unlike in an infinite well, the particle penetrates even into classically forbidden regions, i.e., the region $x < -L/2$ and the region $x > +L/2$. The figure shows that the probability density is nonzero in these forbidden regions. The wave functions alternate between even and odd functions. The probability densities are even functions for all states. The mean position of the particle is therefore 0, i.e., in the center of the well.

Note: For $E < V_0$, i.e., in the region where, in classical mechanics, a particle cannot escape to infinity, the quantum spectrum of the energy operator is discrete. There is always at least one bound state, even in the shallowest potential well. If the particle has $E > V_0$, i.e., it is above the potential well, the solution in all three regions will be linear combinations of sines and cosines, and we will have 6 constants, 4 binding conditions, and one normalization condition. We will not encounter any constraints of the type , and the spectrum will be continuous.

2.4.3 Barrier, Tunneling Effect, and Scattering

Suppose a particle with mass m and energy E strikes a potential barrier (see Figure 2.24) of height V_0 from the left. The particle's energy is less than the height of the potential barrier, so in the classical case the particle could not pass through the barrier because it does not have enough energy to "swing over" the barrier. In quantum mechanics, this is possible. The potential profile has a simple form

$$V(x) = \begin{cases} V_0\,; & x \in (0, L)\,, \\ 0\,; & x \notin (0, L)\,. \end{cases} \tag{2.90}$$

As in previous cases, we divide the solution to Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}x^2} + V(x)\psi = E\psi \tag{2.91}$$

into three regions in which the potential takes constant value. At the barrier, the solution for a flying particle no longer splits into a set of even and odd solutions (as was the case in the symmetric well), so it no longer makes sense to write the solution as a superposition of an odd sine and an even cosine and to keep the origin of the coordinates at the center of the barrier. Instead, we will use a superposition of oscillating exponential functions, which are generally easier to handle (for example, their derivatives are simpler). The solution to the Schrödinger equation in each region will have the form:

$$\begin{aligned}
\psi_{\mathrm{I}}(x) &= A_{\mathrm{I}}\,\mathrm{e}^{\mathrm{i}kx} + B_{\mathrm{I}}\,\mathrm{e}^{-\mathrm{i}kx}\,; && x \in \mathrm{I}, && h^2 \equiv \frac{2m(V_0 - E)}{\hbar^2}\,; \\
\psi_{\mathrm{II}}(x) &= A_{\mathrm{II}}\,\mathrm{e}^{hx} + B_{\mathrm{II}}\,\mathrm{e}^{-hx}\,; && x \in \mathrm{II}, && \\
\psi_{\mathrm{III}}(x) &= A_{\mathrm{III}}\,\mathrm{e}^{\mathrm{i}kx} + B_{\mathrm{III}}\,\mathrm{e}^{-\mathrm{i}kx}\,; && x \in \mathrm{III}, && k^2 \equiv \frac{2mE}{\hbar^2}\,.
\end{aligned} \tag{2.92}$$

Given the probabilistic nature of quantum mechanics, we must repeat the experiment many times, which means that we have prepared a large number of particles in the same state (with the same energy and momentum), which we repeatedly send from the left toward the barrier. The wave function $\psi(x)$ represents the probability amplitude of particle occurrence, and its square $w(x) = \psi^*\psi$ represents the probability density of particle occurrence. In classical wave theory (see, for example, the accompanying textbook [1], plane waves propagating in the (+) and (−) directions along the x-axis have the form

$$\begin{aligned}
\varphi_+(t,x) &= A\,\mathrm{e}^{\mathrm{i}[kx-\omega t]}\,; \\
\varphi_-(t,x) &= A\,\mathrm{e}^{\mathrm{i}[-kx-\omega t]}\,.
\end{aligned} \tag{2.93}$$

From this, we can easily see that the solution in regions I and III is a superposition of waves propagating to the right and to the left. Since no particles are arriving at the barrier from the right, the following must hold

$$B_{\mathrm{III}} = 0\,. \tag{2.94}$$

On the left side (before the barrier), the superposition remains. The wave propagating to the right corresponds to the particles we send toward the barrier, while the left wave is caused by the reflected particles. The continuity conditions for the wave function and its first derivative on the left and right sides of the barrier lead to the equations

$$\begin{aligned}
\psi_{\mathrm{I}}(0) &= \psi_{\mathrm{II}}(0)\,; && \psi'_{\mathrm{I}}(0) = \psi'_{\mathrm{II}}(0)\,; \\
\psi_{\mathrm{II}}(L) &= \psi_{\mathrm{III}}(L)\,; && \psi'_{\mathrm{II}}(L) = \psi'_{\mathrm{III}}(L)\,.
\end{aligned} \tag{2.95}$$

After substituting, we obtain the relationships between the constants A and B:

$$\begin{aligned}
A_{\mathrm{I}} + B_{\mathrm{I}} &= A_{\mathrm{II}} + B_{\mathrm{II}}\,; \\
A_{\mathrm{II}}\,\mathrm{e}^{hL} + B_{\mathrm{II}}\,\mathrm{e}^{-hL} &= A_{\mathrm{III}}\,\mathrm{e}^{\mathrm{i}kL}\,; \\
\mathrm{i}kA_{\mathrm{I}} - \mathrm{i}kB_{\mathrm{I}} &= hA_{\mathrm{II}} - hB_{\mathrm{II}}\,; \\
A_{\mathrm{II}}h\,\mathrm{e}^{hL} - B_{\mathrm{II}}h\,\mathrm{e}^{-hL} &= \mathrm{i}kA_{\mathrm{III}}\,\mathrm{e}^{\mathrm{i}kL}\,.
\end{aligned} \tag{2.96}$$

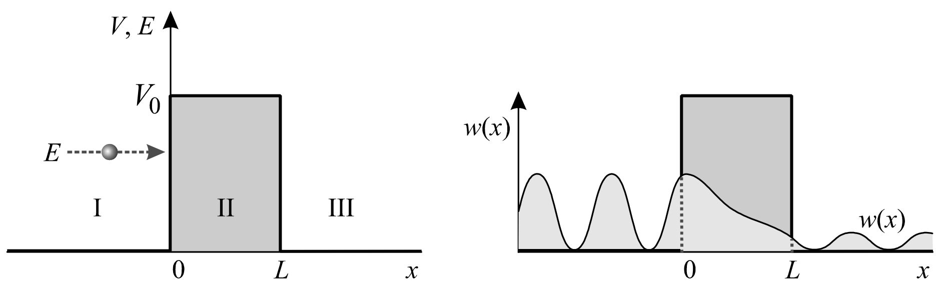


Fig. 2.24: Potential profile (left) and probability of occurrence (right) for a particle traveling toward a one-dimensional rectangular potential barrier

Since we have four conditions for five constants, everything is in order, and one constant remains free for normalizing the found wave function. Another option is to take an unnormalized wave function and set the amplitude of the incident waves equal to

$$A_{\mathrm{I}} = 1 . \tag{2.97}$$

In this case, calculation of the transmission coefficient is easy. It is the ratio of the number of particles that pass through to the number of particles that strike the barrier. We calculate it as the ratio of the probability densities of the transmitted and incident waves:

► $$T = \frac{A_{\mathrm{III}}^* A_{\mathrm{III}}}{A_{\mathrm{I}}^* A_{\mathrm{I}}} = A_{\mathrm{III}}^* A_{\mathrm{III}} . \tag{2.98}$$

Similarly, we can introduce a reflection coefficient

► $$R = \frac{B_{\mathrm{I}}^* B_{\mathrm{I}}}{A_{\mathrm{I}}^* A_{\mathrm{I}}} = B_{\mathrm{I}}^* B_{\mathrm{I}} . \tag{2.99}$$

We rewrite the system (2.96) in matrix form; setting $A_{\mathrm{I}} = 1$ yields the right-hand side of the system:

$$\begin{pmatrix} 1 & -1 & 1 & 0 \\ \mathrm{e}^{hL} & 0 & \mathrm{e}^{-hL} & -\mathrm{e}^{\mathrm{i}kL} \\ h & \mathrm{i}k & -h & 0 \\ h\,\mathrm{e}^{hL} & 0 & -h\,\mathrm{e}^{-hL} & -\mathrm{i}k\,\mathrm{e}^{\mathrm{i}kL} \end{pmatrix} \cdot \begin{pmatrix} A_{\mathrm{II}} \\ B_{\mathrm{I}} \\ B_{\mathrm{II}} \\ A_{\mathrm{III}} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \mathrm{i}k \\ 0 \end{pmatrix} \tag{2.100}$$

Since we want to determine the transmission coefficients of the barrier, we need only determine the constant A_{III} and eliminate the rest. We can transform the matrix into a triangular matrix or use the subdeterminant method, the inverse matrix method, etc. The calculation is straightforward. The result is:

► $$T = \frac{1}{1 + \dfrac{V_0^2 \,\mathrm{sh}^2(hL)}{4E(V_0 - E)}} . \tag{2.101}$$

The reflection coefficient R can be derived from the law of conservation of particles:

► $$R + T = 1 . \tag{2.102}$$

Quantum theory allows a particle to pass through a barrier even if the particle has less energy than the potential energy at the top of the barrier. This is due to the non-commutativity of potential and kinetic energy. Heisenberg uncertainty principle then implies that we cannot measure both values precisely at the same time. The uncertainty in kinetic and potential energy allows a particle to occasionally pass through a barrier. We call this phenomenon the tunneling effect. The probability of this phenomenon decreases exponentially with the thickness of the barrier. A typical example is two regions of metal separated by a thin insulator through which electrons can tunnel. Another example is alpha particles tunneling from an atomic nucleus through a Coulomb barrier (see potential 6 in Fig. 2.11) during radioactive alpha decay. The tunnel diode, the scanning tunneling microscope, and other devices are based on the tunneling effect.

Scanning Tunneling Microscope

An STM (*Scanning Tunneling Microscope*) allows the surface of a solid to be imaged at atomic resolution. The surface is literally scanned by a piezoelectrically deflected tungsten tip. Between the surface and the tip is a non-conductive gap through which electrons tunnel. The resulting tunneling current of electrons is highly sensitive to the distance between the tip and surface irregularities. By measuring this current, we can detect any surface irregularities. In the direction of the surface, the resolution of an STM microscope is on the order of 10^{-10} m; however, in the direction perpendicular to the surface, the resolution is significantly better due to the very strong nonlinear dependence of the current magnitude on the distance from the surface. Ideally, there is a single atom at the tip of the tungsten probe, depending on how well the tip is etched. It is the sharpest tip we can produce and is also used as a cold cathode in scanning electron microscopes. An STM not only makes it possible to visualize the position of atoms on the surface of a crystal lattice, but also to move them from one place to another when the chemical bond with the surface is overcome by an applied electrical voltage and the atom is transferred by the microscope's tip.

The STM microscope was developed by Gerd Binnig and Heinrich Rohrer at IBM's laboratories. They were awarded the 1986 Nobel Prize in Physics for their discovery.

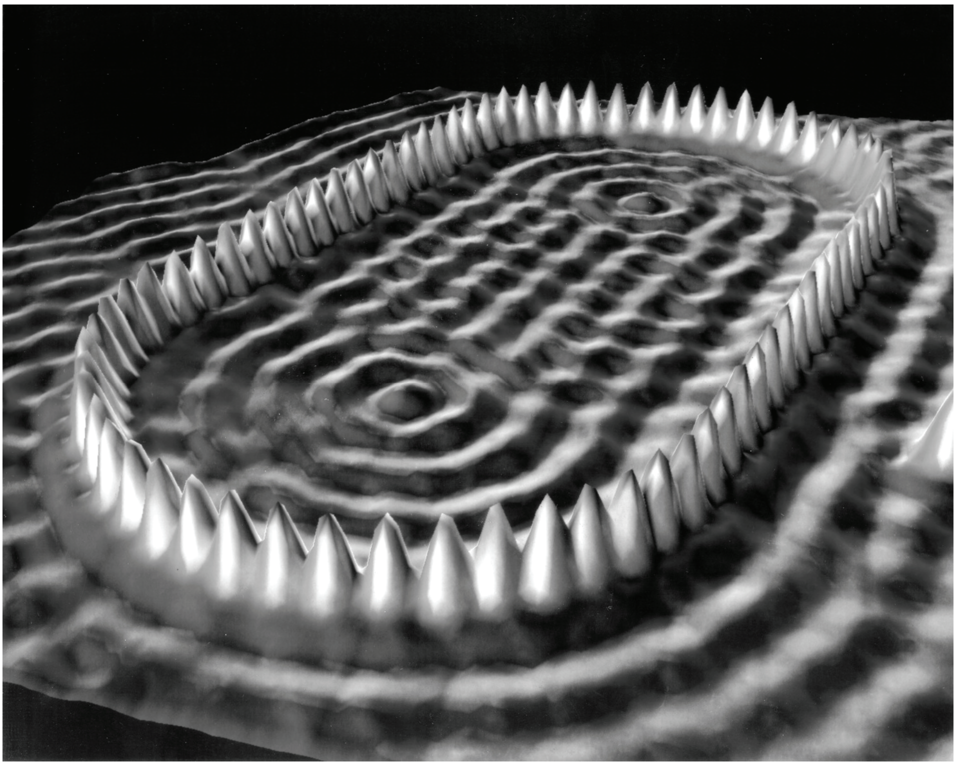


Fig. 2.25: A fence of iron atoms was created on a copper surface using a scanning tunneling microscope. The resulting surface is scanned by the STM, and a topographic map of the surface – the one you are looking at – is generated. Inside the fence, standing density waves of trapped electrons are visible. The tunneling effect has enabled humanity to examine materials at the atomic level and manipulate individual atoms. Crommie, Lutz & Eigler, IBM.

Scattering

A barrier is a simple example of a one-dimensional localized potential. It is non-zero small region of space, into which particles can approach from two directions (from the right and from the left). The incoming particles are described by wave functions $\psi_I^{\pm}$. The particles interact with the region of non-zero potential (scatter off it) and fly out of the region again in two directions (to the right and to the left). They are described by wave functions $\psi_S^{\pm}$. The index S stands for scattered. Both the incoming and outgoing (scattered) wave functions are solutions to the Schrödinger equation with zero potential, meaning they are plane waves. After a certain time, the flow of particles into and out of the potential region stabilizes and any bound states within the potential region are filled:

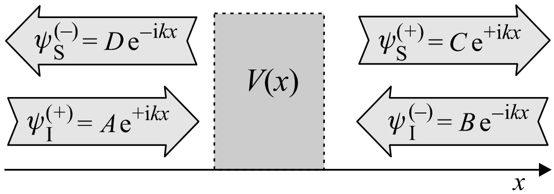


Fig. 2.26: Scattering at a 1D localized potential

The behavior of particles on a localized 1D potential is described by the scattering matrix S, defined by the relation

$$\begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix}. \tag{2.103}$$

The scattering matrix transforms the incoming particle beams into scattered ones. It is a unitary matrix, as can be easily seen from the law of conservation of probability current density (velocity × density) in the steady state:

$$\begin{aligned} j_{\text{left}} &= j_{\text{right}} \qquad \Rightarrow \\ \upsilon_{\text{left}}\left(A^* A - D^* D\right) &= \upsilon_{\text{right}}\left(C^* C - B^* B\right) \qquad \Rightarrow \\ \frac{\hbar k}{m}\left(A^* A - D^* D\right) &= \frac{\hbar k}{m}\left(C^* C - B^* B\right) \qquad \Rightarrow \\ C^* C + D^* D &= A^* A + B^* B\,. \end{aligned}$$

The matrix S therefore does not change the magnitude of the vector. The vector has the same magnitude before and after the matrix acts on it, which is the definition of unitary:

$$\langle\, \psi_S \,|\, \psi_S \,\rangle = \langle\, \psi_I \,|\, \psi_I \,\rangle\,;$$

$$\langle\, S\psi_I \,|\, S\psi_I \,\rangle = \langle\, \psi_I \,|\, \psi_I \,\rangle\,.$$

The eigenvalues of a unitary matrix lie on the unit circle in the complex plane, and we can write them as

$$\lambda_{1,2} = e^{i\delta_{1,2}}\,; \qquad \delta_{1,2} = f_{1,2}(k)\,. \tag{2.104}$$

The scattering at a localized one-dimensional potential is therefore characterized by two angles $\delta_{1,2}(k)$, which are the phases of the eigenvectors of the scattering matrix.

2.4.4 Periodic Potential and Band Spectrum

The motion of particles in a non-localized potential is also very common, for example in the periodic potential of a crystal lattice, which satisfies the basic condition

$$V(x+a) = V(x)\,, \tag{2.105}$$

where a is the period of the potential. The basic properties of a periodic potential can be understood by solving case of an infinite sequence of alternating wells and barriers. This potential is called the Kronig–Penney model. It was first used by the German-American physicist Ralph Kronig and the English mathematician William Penney. Let us assume that the height of the periodic barriers is V_0, their width is L, and periodicity is a.

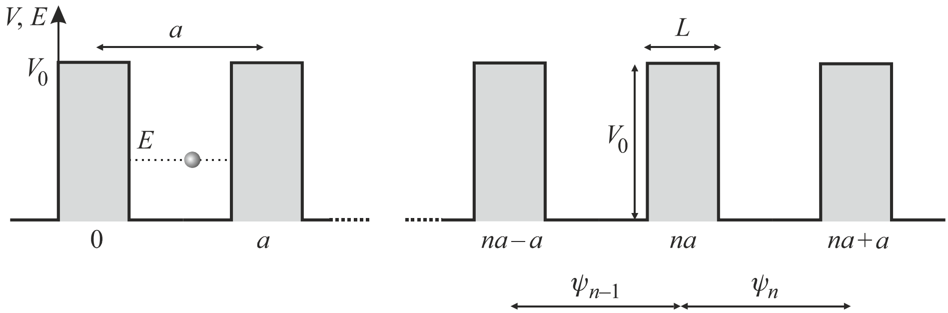


Fig. 2.27: The Kronig–Penney model of a crystal with a periodic potential

Kronig–Penney model

The qualitative nature of the spectrum does not depend on the width of the individual barriers. We will deform them so that their area remains constant, i.e., we will take the limit as $L \to 0$ and $V_0 \to \infty$, such that the product LV_0 remains unchanged. This gives us a potential composed of an infinite series of Dirac delta functions, and a single wave function attachment suffices at each barrier. We will therefore have a potential

$$V(x) = LV_0 \sum_{n=-\infty}^{n=+\infty} \delta(x-na)\,, \tag{2.106}$$

for which we can find a solution on the interval (na, $na + a$), where the potential is zero:

$$\psi_n(x) = A_n \cos[k(x-na)] + B_n \sin[k(x-na)]\,;$$

$$\psi_n'(x) = -A_n k \sin[k(x-na)] + B_n k \cos[k(x-na)]\,; \tag{2.107}$$

$$k \equiv 2mE/\hbar^2\,; \quad n = 0, \pm 1, \pm 2 \ldots$$

The argument of the solution is properly shifted to the local origin at the location of each Dirac delta, so that the sine starts from 0 and the cosine from 1 at each well. When joining the wave functions, we will use three conditions. The wave function is continuous, the first derivative will have a jump (due to Dirac delta), and the periodicity of the potential will lead to the periodicity of the probability density. Let us write out these three conditions for joining at the n^{th} Dirac delta (or at the n^{th} infinitesimal barrier):

1. Continuity of the wave function

At every barrier, we must assume that the wave function is continuous. If it were discontinuous, the first derivative would yield a distribution, and the second derivative contained in Schrödinger equation would be a derivative of that distribution, which could not be compensated for by any other term. Therefore, the following must hold:

$$\psi_{n-1}(na) = \psi_n(na)\,. \tag{2.108}$$

From this, we obtain the first of the three conditions mentioned above:

► $$A_{n-1}\cos(ka) + B_{n-1}\sin(ka) = A_n\,. \tag{2.109}$$

2. Jump in the first derivative

Let's write down Schrödinger's equation for our situation

$$-\frac{\hbar^2}{2m}\psi'' + V(x)\psi = E\psi\,. \tag{2.110}$$

We will integrate the equation in the ε-neighborhood of the n^{th} Dirac delta function:

$$-\frac{\hbar^2}{2m}\int\limits_{na-\varepsilon}^{na+\varepsilon}\psi''(x)\,\mathrm{d}x + LV_0\int\limits_{na-\varepsilon}^{na+\varepsilon}\delta(x-na)\,\psi(x)\,\mathrm{d}x = \int\limits_{na-\varepsilon}^{na+\varepsilon}E\psi(x)\,\mathrm{d}x\,.$$

The middle integral can be easily calculated due to the presence of the Dirac distribution. For the others, we take the limit $\varepsilon \to 0$. Thanks to the continuity of ψ, the integral on the rhs yields zero, and the left-hand integral yields the corresponding jump:

$$-\frac{\hbar^2}{2m}\lim_{\varepsilon\to 0}\left[\psi'\right]_{na-\varepsilon}^{na+\varepsilon} + LV_0\,\psi_n(na) = 0\,,$$

from which follows the condition for the jump of the first derivative of ψ at the n^{th} Dirac impulse

$$-\frac{\hbar^2}{2m}\left[\psi_n'(na) - \psi_{n-1}'(na)\right] + LV_0\,\psi_n(na) = 0\,. \tag{2.111}$$

After substituting the solution (2.107), we obtain the second condition:

► $$B_n + A_{n-1}\sin(ka) - B_{n-1}\cos(ka) = \frac{2mLV_0}{k\hbar^2}A_n\,. \tag{2.112}$$

3. Periodicity

The periodicity of the probability density follows from the periodicity of the potential

$$w(x+a) = w(x)\,; \qquad w(x) \equiv \psi^*(x)\,\psi(x)\,. \tag{2.113}$$

It follows that the wave function must satisfy

$$\psi(x+a) = \mathrm{e}^{\mathrm{i}\phi}\,\psi(x)\,, \tag{2.114}$$

where ϕ is some phase shift. For our constants, we then have:

► $$B_n = \mathrm{e}^{\mathrm{i}\phi}\,B_{n-1}\,; \qquad A_n = \mathrm{e}^{\mathrm{i}\phi}\,A_{n-1}\,. \tag{2.115}$$

Now we substitute A_{n-1} and B_{n-1} from equation (2.115) into equations (2.109) and (2.112). This gives us a system of equations for the unknown constants A_n and B_n:

$$\begin{pmatrix} \cos ka - e^{i\phi}, & \sin ka \\ \sin ka - \xi, & -(\cos ka - e^{i\phi}) \end{pmatrix} \begin{pmatrix} A_n \\ B_n \end{pmatrix} = 0 ; \qquad \xi \equiv \frac{2mLV_0}{k\hbar^2} . \tag{2.116}$$

We obtain a non-zero solution only if the determinant of the system is zero, which leads to the condition:

► $$\cos\phi = \cos ka + A\frac{\sin ka}{ka} ; \qquad A \equiv \frac{mLaV_0}{\hbar^2} . \tag{2.117}$$

The rhs of this condition must belong to the interval <–1, 1>, otherwise the angle ϕ on the lhs cannot be determined and no solution exists. The result is energy bands in which the particle can move, and forbidden bands in which no solution exists, i.e., a particle with such energy cannot occur in the periodic potential. Recall that the wave vector **k** in condition (2.117) is related to the energy by the relation (2.107), i.e., $k = 2mE/\hbar^2$.

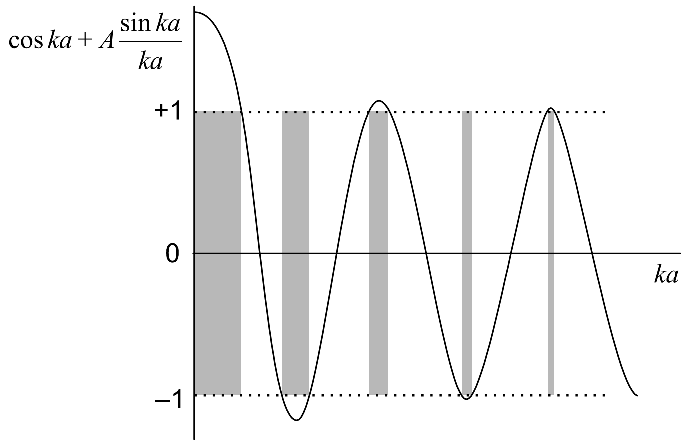


Fig. 2.28: The graph shows the rhs of equation (2.117). Where the curve lies outside the interval <–1,1>, there is no solution, and the forbidden region is marked in gray on the graph.

Forbidden bands are in crystal lattices and semiconductors, and also in the periodic patterns found on butterfly wings, where they produce interesting, almost unnatural colors.

Brillouin zone

Let us now consider the three-dimensional periodicity of a crystal lattice, which repeats itself with every shift by a vector **A**

► $$\mathbf{A} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 , \qquad n_1, n_2, n_3 = 1,2,3\ldots \tag{2.118}$$

where $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are three linearly independent vectors (not lying in the same plane). The volume of the unit cell is given by the volume of a parallelepiped spanned by the basis vectors of the lattice, i.e.

$$V_Z = \left|\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)\right| . \tag{2.119}$$

The potential must satisfy the condition of periodicity

$$V(\mathbf{r}) = V(\mathbf{r}+\mathbf{A}) . \tag{2.120}$$

The lattice periodicity is transferred to the momentum, which changes periodically as particle passes through the lattice. Momentum is associated with the wave vector, i.e.

$$\mathbf{p} = \hbar\mathbf{k} . \tag{2.121}$$

In the k-space, to which we transform via the Fourier transform, periodicity will imply a simple condition on the plane wave

$$e^{i\mathbf{k}\cdot\mathbf{A}} = e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{A}} \text{, tj.} \tag{2.122}$$

$$e^{i\mathbf{G}\cdot\mathbf{A}} = 1 ; \qquad \Rightarrow \qquad \mathbf{G}\cdot\mathbf{A} = 2N\pi ; \qquad N = 1,2,3\ldots \tag{2.123}$$

The periodicity in k-space is therefore **G**; for the energy, e.g., the following holds

► $$E(\mathbf{k}+\mathbf{G}) = E(\mathbf{k}) . \tag{2.124}$$

The vector **G** defines the so-called reciprocal lattice in k-space

► $$\mathbf{G} = m_1\mathbf{g}_1 + m_2\mathbf{g}_2 + m_3\mathbf{g}_3 ; \qquad m_1, m_2, m_3 = 1,2,3\ldots \tag{2.125}$$

The $\mathbf{g}_l$ vectors are linearly independent (do not lie in a plane) and define a reciprocal lattice. Their dimension is m^{-1}. From equation (2.123), the basis reciprocal vectors are

$$\mathbf{g}_1 = 2\pi\frac{\mathbf{a}_2\times\mathbf{a}_3}{V_\mathrm{Z}} ; \qquad \mathbf{g}_2 = 2\pi\frac{\mathbf{a}_3\times\mathbf{a}_1}{V_\mathrm{Z}} ; \qquad \mathbf{g}_2 = 2\pi\frac{\mathbf{a}_1\times\mathbf{a}_2}{V_\mathrm{Z}} . \tag{2.126}$$

A relationship between vectors $\mathbf{a}_k$ of the lattice and $\mathbf{g}_l$ of the reciprocal one is valid

$$\mathbf{a}_k \cdot \mathbf{g}_l = 2\pi\delta_{kl} . \tag{2.127}$$

In a reciprocal lattice, all information is contained within the unit cell, which we call the *Brillouin zone*. By repeating it infinitely, we reconstruct the entire k-space. At the boundary of the Brillouin zones, the quantities may change abruptly. The volume of the Brillouin zone is given by the volume of a parallelepiped spanned by the unit vectors:

$$V_\mathrm{B} = \left|\mathbf{g}_1\cdot(\mathbf{g}_2\times\mathbf{g}_3)\right| . \tag{2.128}$$

The concept of Brillouin zones was developed by the French physicist Léon Brillouin (1889–1969) and has become an integral part of modern solid-state theory. Solving problems in k-space is often simpler and makes the work easier. Here, we have only introduced the so-called first Brillouin zone. By repeating it, we can create additional Brillouin zones. The solution of the Schrödinger equation in the three-dimensional periodic potential of a crystal goes beyond the scope of this introductory textbook on quantum mechanics, and the reader will find it in specialized books, see e.g., [24], [25].

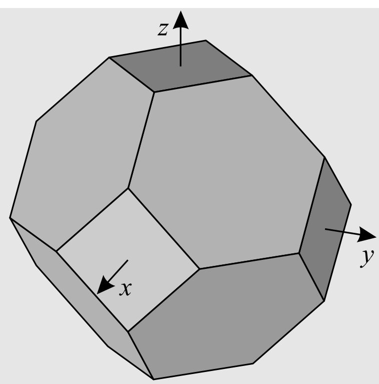


Fig. 2.29: Brillouin zone of a face-centered cubic lattice

2.4.5 Neutron in a Gravity Field

Gravitational interaction is the weakest of the four fundamental forces in nature. We can easily detect it in planets, stars, and galaxies, but for a long time it seemed that any gravitational effects of elementary particles were completely beyond our measuring capabilities. In 2011, a team from the University of Vienna succeeded in measuring the quantum states of a neutron in a gravitational field. This marked the first experimental observation of gravitational effects from an elementary particle.

Consider a particle moving in a homogeneous gravity field. The particle's motion is confined from below by a table. This is a quantum analogy to ping-pong, where the ball can bounce on the table but cannot fall beneath it. The situation is illustrated in Figure 2.30. The classical motion of the particle is confined from below by the table and from above by a maximum height determined by the particle's total energy $E = \tfrac{1}{2}mv^2 + mgy$. From the perspective of quantum theory, this is an infinite triangular well in which the particle has discrete energy states.

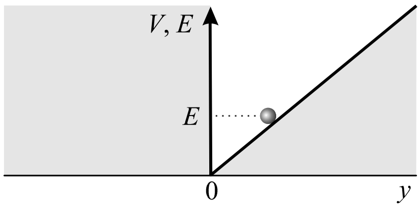


Fig. 2.30: Quantum ping-pong

The potential energy of our problem is given by

► $$V(y) = \begin{cases} \infty\,; & y \le 0\,, \\ mgy\,; & y > 0\,. \end{cases} \tag{2.129}$$

We will solve the Schrödinger equation in the region $y > 0$ and require that $\psi(0) = 0$. The equation itself takes the form

► $$-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2\psi}{\mathrm{d}y^2} + mgy\psi \;=\; E\psi\,. \tag{2.130}$$

The equation can be solved using a procedure similar to that for a harmonic oscillator in. First, we introduce a dimensionless argument for the wave function:

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}\bar{y}^2} - \bar{y}\psi + \lambda\psi = 0\,; \qquad \bar{y} \equiv \frac{y}{y_0}\,; \qquad \lambda \equiv \frac{E}{mg\,y_0}\,; \qquad y_0 \equiv \sqrt[3]{\frac{\hbar^2}{2m^2 g}} \tag{2.131}$$

The eigenvalue λ is a dimensionless energy. We will shift the variable $\bar{y}$ as follows:

$$\xi \equiv \bar{y} - \lambda \tag{2.132}$$

and the resulting equation will be

► $$\psi'' - \xi\psi = 0\,; \qquad {}' \equiv \mathrm{d}/\mathrm{d}\xi\,. \tag{2.133}$$

This is Airi equation, whose solutions are the Airi functions $\mathrm{Ai}(\xi)$ and $\mathrm{Bi}(\xi)$; see, for example, [26], [27]. Both functions can be defined using a series, Bessel functions, or an integral expression:

$$\mathrm{Ai}(\xi) = \frac{1}{\pi}\int_0^\infty \cos\left(t^3/3 + \xi t\right)\mathrm{d}t\,;$$
$$\mathrm{Bi}(\xi) = \frac{1}{\pi}\int_0^\infty \left[\exp\left(-t^3/3 + \xi t\right) + \sin\left(t^3/3 + \xi t\right)\right]\mathrm{d}t\,. \tag{2.134}$$

The function Bi(ξ) diverges for large ξ; therefore, the solution to the problem is

$$\psi(\xi) = \mathrm{Ai}(\xi)\,. \tag{2.135}$$

The boundary condition $\psi(\xi) = 0$ leads to a numerical search for the zeroes of the Airy function and thus to the quantization of the energy, which is a function of the argument of the Airy function, since $\xi = \bar{y} - \lambda(E)$. The values of the first five energy levels, together with the probability density of the particle's occurrence, are shown in the figure:

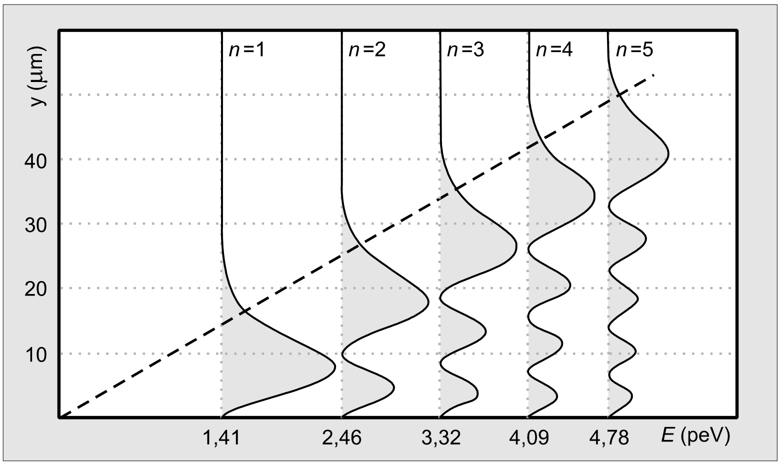


Fig. 2.31: A quantum ball in a gravitational field. The horizontal axis shows the ball's energy, and the permitted energy states in quantum theory are marked. The vertical axis shows the height above the table. The dashed line represents the height that a bouncing ball with a given energy would reach in classical mechanics. The quantum probability of the ball's position is shown in gray (its value increases to the right).

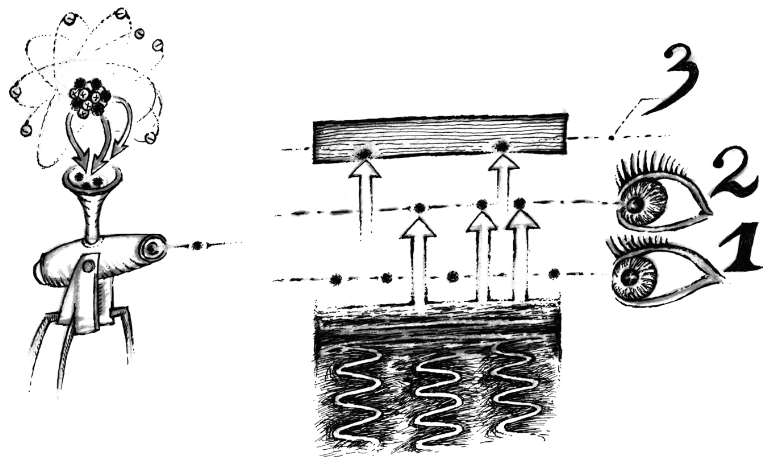
Quantum ping-pong with a neutron

Experiments involving a gravitational field acting on elementary particles were conducted by a group of scientists led by Professor Hartmut Abele of the Vienna University of Technology [28]. The group also included scientists from the Laue-Langevin Institute in Grenoble (ILL), where the experiments were physically carried out. A neutron served as the test particle, as it is minimally affected by ubiquitous electromagnetic forces. Neutrons are very difficult to polarize, so they are not subject to various dipole forces, such as the van der Waals force. The neutron has a sufficient lifetime for gravity experiments; its half-life is over 800 seconds. A source of ultra-cold neutrons with very low energy was used for the experiments. Only with such particles was it possible to measure the quantum states of the neutron in a gravity field. Therefore, a neutron source from the Laue-Langevin Institute in Grenoble was used, which produces neutrons with energies below 300 neV (nanoelectronvolts), corresponding to a temperature below 2 mK (millikelvins) and a velocity below 15 m/s.

If a ball is bouncing on a table, it can reach any height determined by its total energy. A quantum ball in a gravitational field exists only in certain energy states determined by our solution to Schrödinger equation. A quantum ball can rise only to certain heights determined by the possible energy states. The lowest energy state for a bouncing ball is 1.41 peV (picoelectronvolts), the second is 2.46 peV, the third is 3.32 peV, and so on. For a normal ping-pong ball, these states are unmeasurable; for ultracold neutrons, however, it is possible to detect such states. The probability of the ball being at a certain height above the surface is given by the square of the Airy function.

Ultra-cold neutrons were directed between two horizontal plates. The bottom plate served as a surface from which the neutron – which, in the classical case, follows a parabolic path – could be reflected. The upper plate was auxiliary and was designed to absorb neutrons that reached its height. The distance between the plates was approximately 20 to 25 micrometers, so neutrons in the first and second quantum states could pass through the plates without any problems (they did not reach the height of the second plate). Cold neutrons, however, did not have it so easy. The bottom plate vibrated in a controlled manner. It was set into vibration using the piezoelectric effect, and its oscillations were precisely controlled using a laser. When the plate vibrated, it caused a resonant transition of neutrons between the first and third energy states, and most of the neutrons did not pass between the plates, because the third energy state means that the neutron reached the height of the upper plate and was absorbed by it. When the lower plate did not vibrate, most of the neutrons passed between the plates.

For the first time in history, the quantum states of a particle in a gravity field have been measured, and it was possible to alter these states using a vibrating plate. This resonance method could have a profound impact on our understanding of gravitational interactions at small scales, where no measurements have been available until now. We have a good chance of learning how gravity works in the world of elementary particles and whether reality deviates from Newton's and Einstein's ideas or not.



2.5 Spherically Symmetric Potential

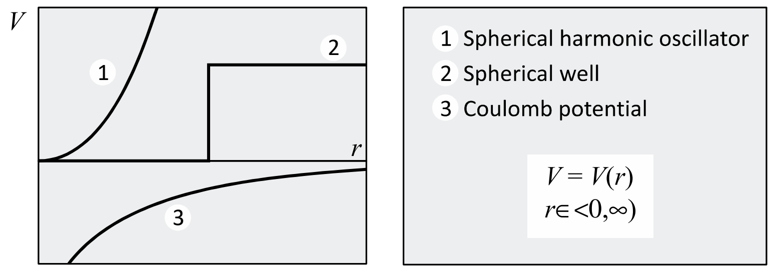


Fig. 2.33: Some spherical potentials

A spherically symmetric (central) potential is one that depends only on the distance from a specific center. The spherical coordinate system is very useful for describing the motion of bodies in such a potential. Among the best-known are the spherical harmonic oscillator, the spherical well, and the Coulomb potential. You can imagine a spherical oscillator as a small body at the origin of the coordinate system, from which springs extend in all directions. Whenever we displace it, a restoring force acts toward the center. The potential energy function is quadratic. The spherical well approximately corresponds to the potential experienced by a neutron trapped in an atomic nucleus. The nuclear forces (derivatives of the potential) at the edge of the well ($r = a$) are significant – in the idealized case (2.136) even infinite – and very weak in other regions. The Coulomb potential applies, for example, in a hydrogen atom, where a lone electron is subject to the action of a single proton in the atomic nucleus. Remember that $r \in <0, \infty)$. The graphs of these well-known potentials are:

$$\begin{aligned} &(1) \qquad V(r) = \frac{1}{2}kr^2, \\ &(2) \qquad V(r) = \begin{cases} 0 & r < a \\ V_0 & r \geq a \end{cases}, \\ &(3) \qquad V(r) = \frac{qQ}{4\pi\varepsilon_0 r} = -\frac{\alpha}{r}. \end{aligned} \tag{2.136}$$

In classical mechanics, a system is described by the Lagrangian, generalized momenta, and energy, as well as the Hamiltonian, in spherical coordinates as follows:

$$L = \frac{1}{2}m\dot{r}^2 + \frac{1}{2}mr^2\sin^2\theta\,\dot{\varphi}^2 + \frac{1}{2}mr^2\dot{\theta}^2 - V(r), \tag{2.137}$$

$$\begin{aligned} p_r &= m\dot{r}, \\ p_\varphi &= mr^2\sin^2\theta\,\dot{\varphi}, \\ p_\theta &= mr^2\dot{\theta}, \end{aligned} \tag{2.138}$$

$$E = \frac{1}{2} m\dot{r}^2 + \frac{1}{2} mr^2 \sin^2\theta\, \dot{\varphi}^2 + \frac{1}{2} mr^2 \dot{\theta}^2 + V(r)\,, \tag{2.139}$$

$$H = \frac{p_r^2}{2m} + \frac{p_\theta^2}{2mr^2} + \frac{p_\varphi^2}{2mr^2 \sin^2\theta} + V(r) = \frac{p_r^2}{2m} + \frac{\mathbf{L}^2}{2mr^2} + V(r)\,. \tag{2.140}$$

We have already shown in classical mechanics that the generalized momenta of angular variables are components of the angular momentum. The rotational terms in the Hamiltonian can be expressed using the angular momentum vector **L** with respect to the z-axis, from which the spherical coordinate system is derived.

Individual components of angular momentum cannot be measured simultaneously and do not commute with one another (2.25). However, we can simultaneously measure the square of the angular momentum (2.26) and any of its components (2.27). For a spherically symmetric problem, we will prefer the third axis and the third component. The z-axis is preferred in a spherical coordinate system; however, it does not matter which of the angular momentum components we choose to include in the set of observables. If an external magnetic field is present in the system, we generally choose the coordinate system such that the third axis points in the direction of this field. If the system is spherically symmetric, then the square of the angular momentum commutes with the Hamiltonian:

$$\left[\hat{\mathbf{L}}^2, \hat{\mathbf{H}}\right] = \left[\hat{\mathbf{L}}^2, \frac{p_r^2}{2m} + \frac{\hat{\mathbf{L}}^2}{2mr^2} + V(r)\right] = \left[\hat{\mathbf{L}}^2, \frac{\hat{\mathbf{L}}^2}{2mr^2}\right] = 0\,.$$

We know, in fact, that generalized coordinates commute only with their generalized momenta. In the commutator between the square of the angular momentum and the Hamiltonian, therefore, the only non-zero terms can be those involving the angular part of the Hamiltonian, but the operator commutes with itself, so the result can only be zero. The situation is similar with the third component of angular momentum:

$$\left[\hat{\mathbf{L}}_3, \hat{\mathbf{H}}\right] = \left[\hat{\mathbf{L}}_3, \frac{p_r^2}{2m} + \frac{\hat{\mathbf{L}}^2}{2mr^2} + V(r)\right] = \left[\hat{\mathbf{L}}_3, \frac{\hat{\mathbf{L}}^2}{2mr^2}\right] = \frac{1}{2mr^2}\left[\hat{\mathbf{L}}_3, \hat{\mathbf{L}}^2\right] = 0\,.$$

We have thus found a set of three independent commuting operators that form the complete set of observables for the non-relativistic spherically symmetric problem (in the relativistic case, spin is added to these variables):

$$[\hat{\mathbf{L}}^2, \hat{\mathbf{L}}_3] = [\hat{\mathbf{L}}^2, \hat{\mathbf{H}}] = [\hat{\mathbf{L}}_3, \hat{\mathbf{H}}] = 0\,. \tag{2.141}$$

For a system of independent, mutually commutative operators, it is possible to find common eigenvectors for all operators. For a spherically symmetric problem, we will therefore solve a system of three equations for the eigenvectors

$$\begin{aligned} \hat{\mathbf{H}}\,|\,v,l,m\,\rangle &= E_v\,|\,v,l,m\,\rangle\,, \\ \hat{\mathbf{L}}^2\,|\,v,l,m\,\rangle &= \lambda_l\,|\,v,l,m\,\rangle\,, \\ \hat{\mathbf{L}}_3\,|\,v,l,m\,\rangle &= \mu_m\,|\,v,l,m\,\rangle\,. \end{aligned} \tag{2.142}$$

The index v numbers the energy states, the index l numbers the states of the square of the angular momentum, and the index m numbers the states of the third angular momentum component. We have denoted the eigenvalues as E, λ, μ. This system must be solved simultaneously. What would happen if, e.g., we solved only the equation for energy? The eigenvalues E would be correct, but for each eigenvalue (each energy value) there would be multiple independent eigenvectors differing from each other by the numbers l and m. We call this type of spectrum a degenerate spectrum. It simply means that there are multiple eigenvectors for a given eigenvalue. We could distinguish them from one another only by using additional operators that commute with the operator whose spectrum we are currently seeking.

In the following two chapters, we will focus on angular momentum, i.e., the second and third equations in (2.142) (2.142). The solution for the angular momentum is the same for all potential energy profiles. In Section 2.5.1, we will find a solution without using a specific representation, and in Section 2.5.2, we will calculate it in the x-representation. We will discuss the first equation in (2.142) in Section 2.5.3. The energy spectrum naturally depends on the potential energy profile and is different, for example, for hydrogen and for a spherical oscillator. Furthermore, the solution for the energy depends on the numbers l and m. This is logical: angular momentum is related to the rotational states of the system, and these contribute to the energy. After all, we see this in the Hamiltonian (2.140), where the rotational part is expressed in terms of $\mathbf{L}^2$.

2.5.1 Angular Momentum

The basic commutation relations for angular momentum are equations (2.25) and (2.27):

$$[\hat{\mathbf{L}}_1, \hat{\mathbf{L}}_2] = \mathrm{i}\hbar \hat{\mathbf{L}}_3,$$
$$[\hat{\mathbf{L}}_2, \hat{\mathbf{L}}_3] = \mathrm{i}\hbar \hat{\mathbf{L}}_1,$$
$$[\hat{\mathbf{L}}_3, \hat{\mathbf{L}}_1] = \mathrm{i}\hbar \hat{\mathbf{L}}_2,$$
$$[\hat{\mathbf{L}}^2, \hat{\mathbf{L}}_3] = 0\,.$$

Let's now introduce the so-called ladder operators

$$\hat{\mathbf{L}}_\pm \equiv \hat{\mathbf{L}}_1 \pm \mathrm{i}\,\hat{\mathbf{L}}_2\,. \tag{2.143}$$

These operators will have a similar meaning to the creation and annihilation operators for the energy of a harmonic oscillator. They will shift us along the angular momentum spectrum. Let's list their important properties (all of which can be easily derived from the definition of shift operators and the commutation relations for angular momentum):

$$\begin{aligned}
&(1)\quad \hat{\mathbf{L}}_1 = \frac{1}{2}\left(\hat{\mathbf{L}}_+ + \hat{\mathbf{L}}_-\right), &&(5)\quad \hat{\mathbf{L}}_-\hat{\mathbf{L}}_+ = \hat{\mathbf{L}}^2 - \hat{\mathbf{L}}_3^2 - \hbar\,\hat{\mathbf{L}}_3\,,\\
&(2)\quad \hat{\mathbf{L}}_2 = \frac{1}{2\mathrm{i}}\left(\hat{\mathbf{L}}_+ - \hat{\mathbf{L}}_-\right), &&(6)\quad \left[\hat{\mathbf{L}}_+, \hat{\mathbf{L}}_-\right] = 2\,\hbar\,\hat{\mathbf{L}}_3\,,\\
&(3)\quad \hat{\mathbf{L}}_\pm^\dagger = \hat{\mathbf{L}}_\mp\,, &&(7)\quad \left[\hat{\mathbf{L}}_3, \hat{\mathbf{L}}_\pm\right] = \pm\,\hbar\,\hat{\mathbf{L}}_\pm\,,\\
&(4)\quad \hat{\mathbf{L}}_+\hat{\mathbf{L}}_- = \hat{\mathbf{L}}^2 - \hat{\mathbf{L}}_3^2 + \hbar\,\hat{\mathbf{L}}_3\,, &&(8)\quad \left[\hat{\mathbf{L}}^2, \hat{\mathbf{L}}_\pm\right] = 0\,.
\end{aligned} \tag{2.144}$$

Given the ladder operators, we can reconstruct the total angular momentum from relations (1), (2), and (6). The problem to solve can be formulated as follows:

$$\hat{\mathbf{L}}^2 \,|\, \lambda,\mu\,\rangle = \lambda |\, \lambda,\mu\,\rangle\,,$$
$$\hat{\mathbf{L}}_3 \,|\, \lambda,\mu\,\rangle = \mu |\, \lambda,\mu\,\rangle\,.$$

First, let's derive three auxiliary lemmas concerning ladder operators:

Lemma 1

Statement: The ladder operators shift the eigenvectors in the third component of angular momentum by Planck's constant:

$$\hat{\mathbf{L}}_\pm \,|\, \lambda,\mu\,\rangle \approx |\, \lambda,\mu \pm \hbar\,\rangle\,.$$

Proof: Let $|\,\psi\,\rangle \equiv \hat{\mathbf{L}}_\pm \,|\, \lambda, \mu\,\rangle$. Apply the operators $\hat{\mathbf{L}}_3$ and $\hat{\mathbf{L}}^2$ to this vector. To express the third component and the square of the angular momentum, we will use a set of basic properties (2.144). From property (7) we determine the third component, and from property (8) the square of the angular momentum:

$$\begin{aligned}\hat{\mathbf{L}}_3 \,|\,\psi\,\rangle &= \hat{\mathbf{L}}_3 \hat{\mathbf{L}}_\pm \,|\, \lambda,\mu\,\rangle = \\ &= (\hat{\mathbf{L}}_\pm \hat{\mathbf{L}}_3 \pm \hbar \hat{\mathbf{L}}_\pm)\,|\, \lambda,\mu\,\rangle = \\ &= (\mu \pm \hbar)\hat{\mathbf{L}}_\pm \,|\, \lambda,\mu\,\rangle = (\mu \pm \hbar)\,|\,\psi\,\rangle\,;\end{aligned}$$

$$\begin{aligned}\hat{\mathbf{L}}^2 \,|\,\psi\,\rangle &= \hat{\mathbf{L}}^2 \hat{\mathbf{L}}_\pm \,|\, \lambda,\mu\,\rangle = \\ &= \hat{\mathbf{L}}_\pm \hat{\mathbf{L}}^2 \,|\, \lambda,\mu\,\rangle = \\ &= \lambda\, \hat{\mathbf{L}}_\pm \,|\, \lambda,\mu\,\rangle = \lambda\,|\,\psi\,\rangle\,.\end{aligned}$$

We see that the ladder operators shift the spectrum of the third-component operator by a constant $\pm\hbar$. In the spectrum of the square of the angular momentum operator, the ladder operators have no effect. Thus, the ladder operators only change the value of the projection of the angular momentum onto the chosen axis. ■

Lemma 2

Statement: For a fixed λ, the spectrum of $\hat{\mathbf{L}}_3$ is bounded, i.e., there exist $\mu_{\min}$ and $\mu_{\max}$.
Proof: In the formulas

$$\hat{\mathbf{L}}^2 \,|\, \lambda, \mu\,\rangle = \lambda \,|\, \lambda, \mu\,\rangle\,;$$
$$\left(\hat{\mathbf{L}}_1^2 + \hat{\mathbf{L}}_2^2\right)|\, \lambda, \mu\,\rangle = \left(\hat{\mathbf{L}}^2 - \hat{\mathbf{L}}_3^2\right)|\, \lambda, \mu\,\rangle = (\lambda - \mu^2)\,|\, \lambda, \mu\,\rangle$$

the operators on the lhs are positive definite. Therefore, $\lambda \geq 0$ and $\lambda - \mu^2 \geq 0$ must hold. It follows that $\mu^2 \leq \lambda \;\wedge\; \lambda \geq 0$, so $\mu \in \langle -\sqrt{\lambda}, +\sqrt{\lambda}\rangle$, and there exist $\mu_{\min}$ a $\mu_{\max}$. ■

Lemma 3

Statement: The spectrum is symmetric about zero, i.e., $\mu_{\min} = -\mu_{\max}$.

Proof: As with the harmonic oscillator, we apply the ladder operator to the first (or last) state. The result of this operation is zero, because there is no subsequent state. We then calculate the square of the norm of this vector and perform some simple adjustments:

$$\hat{\mathbf{L}}_+ \,|\, \lambda,\mu_{\max}\,\rangle = 0\,, \qquad \wedge \qquad \hat{\mathbf{L}}_- \,|\, \lambda,\mu_{\min}\,\rangle = 0\,,$$

$$\langle\, \lambda,\mu_{\max}\,|\,\hat{\mathbf{L}}_- \hat{\mathbf{L}}_+ \,|\, \lambda,\mu_{\max}\,\rangle = 0\,, \qquad \wedge \qquad \langle\, \lambda,\mu_{\min}\,|\,\hat{\mathbf{L}}_+ \hat{\mathbf{L}}_- \,|\, \lambda,\mu_{\min}\,\rangle = 0\,,$$

$$\langle\, \lambda,\mu_{\max}\,|\,\hat{\mathbf{L}}^2 - \hat{\mathbf{L}}_3^2 - \hbar \hat{\mathbf{L}}_3 \,|\, \lambda,\mu_{\max}\,\rangle = 0\,, \quad \wedge \quad \langle\, \lambda,\mu_{\min}\,|\,\hat{\mathbf{L}}^2 - \hat{\mathbf{L}}_3^2 + \hbar \hat{\mathbf{L}}_3 \,|\, \lambda,\mu_{\min}\,\rangle = 0\,,$$

$$(\lambda-\mu_{\max}^2-\mu_{\max}\hbar)\,||\,\lambda,\mu_{\max}\,||^2=0\,,\quad\wedge\quad(\lambda-\mu_{\min}^2+\mu_{\min}\hbar)\,||\,\lambda,\mu_{\min}\,||^2=0\,,$$

$$\lambda=\mu_{\max}^2+\mu_{\max}\hbar\quad\wedge\quad\lambda=\mu_{\min}^2-\mu_{\min}\hbar\,. \tag{2.145}$$

If we denote $\mu_{\max}\equiv a$ and $\mu_{\min}=b$, then the following holds

$$a^2+a\hbar=b^2-b\hbar\qquad\Rightarrow$$

$$b^2-b\hbar-(a^2+a\hbar)=0\qquad\Rightarrow$$

$$b=\frac{1}{2}\left(\hbar\pm\sqrt{\hbar^2+4(a^2+a\hbar)}\right)=\frac{1}{2}\left(\hbar\pm\sqrt{\hbar^2+4a\hbar+4a^2}\right)=$$

$$=\frac{1}{2}\left(\hbar\pm\sqrt{(\hbar+2a)^2}\right)=$$

$$\frac{1}{2}\big(\hbar\pm(\hbar+2a)\big)=\begin{cases}a+\hbar,\\ -a.\end{cases}$$

The first solution contradicts the claim that a is the maximum value; the second solution proves the lemma on the symmetry of the momentum projection spectrum. ■

Spectrum of angular momentum projection

Ladder operators shift the third component of angular momentum by Planck's constant, so the following must hold:

$$\mu=-a,-a+\hbar,-a+2\hbar,\;\ldots,a-\hbar,a\,.$$

Let us introduce the dimensionless parameters $m\equiv\mu/\hbar$ and $l=a/\hbar$. Then

$$m\in\{-l,-l+1,-l+2,\ldots,l-1,l\}\,. \tag{2.146}$$

The number m can therefore take on a total of $2l+1$ different values. The $2l+1$ must be a non-negative integer, and therefore the l itself can only take on half-integer values, i.e.

$$l\in\{0,\frac{1}{2},1,\frac{3}{2},2,\frac{5}{2},\ldots\}\,. \tag{2.147}$$

Eigenvalue $\lambda=\mu_{\max}(\mu_{\max}+\hbar)=l\hbar\,(l\hbar+\hbar)=\hbar^2\,l(l+1)$.

Conclusion

We can summarize the results of the entire derivation as follows:

$$\hat{\mathbf{L}}^2\,|\,l,m\,\rangle=l\,(l+1)\hbar^2\,|\,l,m\,\rangle\,,\qquad l\in\{0,\frac{1}{2},1,\frac{3}{2},2,\frac{5}{2},\ldots\}\,;$$

► $$\hat{\mathbf{L}}_3\,|\,l,m\,\rangle=m\hbar|\,l,m\,\rangle\,,\qquad m\in\{-l,-l+1,\ldots,l-1,l\}\,; \tag{2.148}$$

$$\hat{\mathbf{L}}_\pm\,|\,l,m\,\rangle\approx|\,l,m\pm1\,\rangle\,.$$

Notes on the Solution

The following notes are very important; please read them more carefully than the solution itself!

Note 1: The l quantum number quantizes the magnitude of the angular momentum and is called the *azimuthal* quantum number (the *principal* quantum number quantizes the energy). The number m quantizes the projection of angular momentum onto any axis and is called the *magnetic* quantum number. The name stems from the fact that an electron in an atomic shell has angular momentum proportional to its magnetic moment, so the number m also corresponds to magnetic moment.

Note 2: The possible values for the **L** magnitude and its projection are:

$$|L| = \sqrt{l(l+1)}\,\hbar\,, \qquad l = 0, 1, 2, 3, \ldots; \qquad\qquad (2.149)$$
$$L_3 = m\hbar\,, \qquad m = -l, -l+1, \ldots, l\,.$$

Note 3: The half-integer values derived for the number l are indeed possible. They occur for a spin whose operator has the same commutation structure as the angular momentum. In Schrödinger x-representation (following chapter), we will not obtain these values. The choice of representation here implies the *loss of part of the solution*. The fact that the half-integer values of l are already part of the commutation relations (2.25) was discovered relatively late (in 1968).

Note 4: The true significance of Planck constant follows from the result in (2.149) or (2.148). It represents the *elementary quantum of angular momentum*. When measuring angular momentum, we will always measure the projection of the momentum onto a specific axis, as determined by the measuring device. This projection is a multiple of the reduced Planck constant (the Planck-Dirac constant).

Note 5: We see that states with a specific azimuthal quantum number l are degenerate – there are multiple eigenvectors $|\, l, m \rangle$ that correspond to the same quantum number l. These vectors differ from one another by the quantum number m, and their number is $2l+1$ (degree of degeneracy, which we denote by the symbol #).

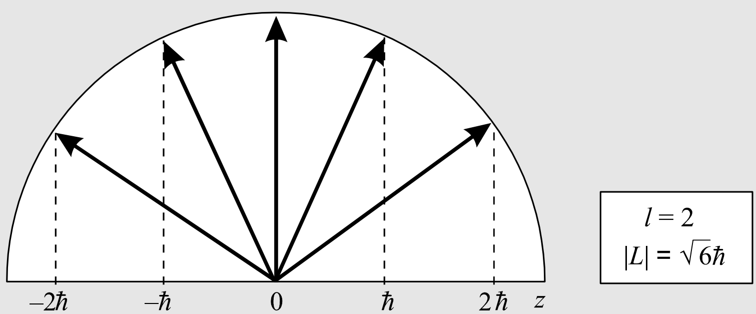


Fig. 2.34: Possible projections of angular momentum for l = 2

Note 6: Historically, the quantum states of the electron's angular momentum in the hydrogen atom's electron shell were denoted by the letters s, p, d, and f, as shown in the table on the following page:

$l=0$	s state	$m=0$	$\#=1$
$l=1$	p state	$m=-1,0,1$	$\#=3$
$l=2$	d state	$m=-2,-1,0,1,2$	$\#=5$
$l=3$	f state	$m=-3,-2,-1,0,1,2,3$	$\#=7$

Note 7: The expression for the square of the angular momentum can also be obtained as the arithmetic mean of possible values. For example, for $l = 2$, the possible values of the L_x, L_y, or L_z are $-2\hbar$, $-\hbar$, 0, $\hbar$, $2\hbar$. The average value of the square is therefore given by the relation (it corresponds to the value from 2.149)

$$\langle L^2 \rangle = \langle L_x^2 \rangle + \langle L_y^2 \rangle + \langle L_z^2 \rangle = 3\langle L_z^2 \rangle =$$

$$= 3\,\frac{4\hbar^2 + \hbar^2 + 0 + \hbar^2 + 4\hbar^2}{5} = 6\hbar^2.$$

Note 8: It is not difficult to count the individual matrix elements

$$\langle l, m' | \hat{\mathbf{L}}_k | l, m \rangle$$

of the angular momentum operator in its eigenrepresentation using ladder operators, similar to the harmonic oscillator, provided we determine the normalization constants. For $l = 0$, m and m' can only be 0, and therefore there is a single element. This matrix acts on scalar quantities; we refer to this as the *scalar representation*. For $l = 1/2$, m and m' can take the values –1/2 and +1/2. These are 2×2 matrices acting on ordered pairs, which we call spinors. This is the so-called *spinor representation*. For $l = 1$, m and m' can take the values –1, 0, and +1. These are 3×3 matrices acting on ordered triples, which we call vectors. This is the so-called *vector representation*. The $\mathbf{L}_3$ matrices are diagonal with eigenvalues on the diagonal.

Spinor representation (l = 1/2)

$$\mathbf{L}_1 = \frac{\hbar}{2}\begin{pmatrix} 0 & +1 \\ +1 & 0 \end{pmatrix}; \quad \mathbf{L}_2 = \frac{\hbar}{2}\begin{pmatrix} 0 & -\mathrm{i} \\ +\mathrm{i} & 0 \end{pmatrix}; \quad \mathbf{L}_3 = \frac{\hbar}{2}\begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}. \qquad (2.150)$$

Vector representation (l = 1)

$$\mathbf{L}_1 = \hbar\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}; \quad \mathbf{L}_2 = \hbar\begin{pmatrix} 0 & -\mathrm{i} & 0 \\ \mathrm{i} & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{pmatrix}; \quad \mathbf{L}_3 = \hbar\begin{pmatrix} +1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \qquad (2.151)$$

Matrices for $l = 1/2$ (without scaling factors) are called Pauli matrices.

Note 9: The well-known statement of Bohr model – that the circumference of an electron's orbit in the atomic shell must be an integer multiple of the wavelength – can be rewritten as follows using equation (2.5):

$$n\lambda = 2\pi r_n \qquad \Rightarrow \qquad n\frac{2\pi\hbar}{m_\mathrm{e} v_n} = 2\pi r_n \qquad \Rightarrow \qquad m_\mathrm{e} v_n r_n = n\hbar$$

and so it is the quantization of the angular momentum projection.

2.5.2 Solution in the *x*-Representation, Spherical Functions

In the *x*-representation, we will solve the spherical potential problem in spherical coordinates (which most closely reflect the symmetry of the potential energy). We need to solve the system of equations (2.142), which will now take the form:

$$\hat{\mathbf{H}}\psi(r,\varphi,\theta) = E_\nu\,\psi(r,\varphi,\theta)\,,$$
$$\hat{\mathbf{L}}^2\psi(r,\varphi,\theta) = \lambda_l\,\psi(r,\varphi,\theta)\,, \qquad (2.152)$$
$$\hat{\mathbf{L}}_3\psi(r,\varphi,\theta) = \mu_m\,\psi(r,\varphi,\theta)\,.$$

►

Let's now rewrite the operators in spherical coordinates. To do this, we first decompose the Laplace operator in spherical coordinates into its radial and angular parts:

$$\nabla^2 = \nabla_r^2 + \frac{1}{r^2}\nabla_{\theta\varphi}^2\,;$$
$$\nabla_r^2 \equiv \frac{1}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial}{\partial r}\,, \qquad (2.153)$$
$$\nabla_{\theta\varphi}^2 \equiv \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\,.$$

We can decompose the kinetic part of the Hamiltonian into radial and angular terms:

$$\hat{\mathbf{H}} = \frac{\hat{\mathbf{p}}_r^2}{2m} + \frac{\hat{\mathbf{L}}^2}{2I} + V(r) = \frac{\hat{\mathbf{p}}_r^2}{2m} + \frac{\hat{\mathbf{L}}^2}{2mr^2} + V(r)\,. \qquad (2.154)$$

In the *x*-representation, the Hamiltonian operator has a simple form

$$\hat{\mathbf{H}} = \frac{-\hbar^2}{2m}\left(\nabla_r^2 + \frac{1}{r^2}\nabla_{\theta\varphi}^2\right) + V(r)\,. \qquad (2.155)$$

From the decomposition of the Laplace operator (2.153), it is clear that its angular part must (up to a constant) correspond to the square of the angular momentum. Comparing the last two relations, we obtain the operator for the square of the angular momentum:

$$\hat{\mathbf{L}}^2 = -\hbar^2\nabla_{\theta\varphi}^2\,. \qquad (2.156)$$

The operator for the third component of angular momentum is a simple generalization of relation (2.47). The operators for the complete set of observables are therefore:

$$\hat{\mathbf{H}} = \frac{-\hbar^2}{2m}\left(\nabla_r^2 + \frac{1}{r^2}\nabla_{\theta\varphi}^2\right) + V(r)\,;$$
$$\hat{\mathbf{L}}^2 = -\hbar^2\nabla_{\theta\varphi}^2\,; \qquad (2.157)$$
$$\hat{\mathbf{L}}_3 = -\mathrm{i}\hbar\frac{\partial}{\partial\varphi}\,.$$

►

In Cartesian coordinates, the Laplace operator decomposes into the sum of second derivatives along each axis; this corresponds to the kinetic energy in various axes. In spherical coordinates, the Laplace operator has radial and angular parts, as well as the kinetic energy. It is precisely the angular component of kinetic energy that is the rotational energy associated with angular momentum; therefore, the square of the angular momentum corresponds to the angular component of the Laplace operator.

The solution $\psi(r, \varphi, \theta)$ we are seeking naturally depends on the quantum numbers v, l, m. We will seek the solution in the separated form

$$\psi(r,\varphi,\theta) = f(r)\, g(\varphi)\, h(\theta)\,. \tag{2.158}$$

First, let's solve the last equation (2.152):

$$-\mathrm{i}\hbar \frac{\partial}{\partial\varphi} f(r) g(\varphi) h(\theta) = \mu_m f(r) g(\varphi) h(\theta) \qquad \Rightarrow$$

$$-\mathrm{i}\hbar \frac{\mathrm{d}g}{\mathrm{d}\varphi} = \mu_m g \qquad \Rightarrow$$

$$g(\varphi) = c \exp\left[\mathrm{i}\frac{\mu_m}{\hbar}\varphi\right].$$

The solution found must be periodic with respect to the angle φ:

$$g(\varphi) = g(\varphi + 2\pi) \qquad \Rightarrow \qquad \mu_m = m\hbar\,; \qquad m = 0,\ \pm 1,\ \pm 2,\ \ldots$$

In the x-representation, we have again derived the quantization of the projection of angular momentum, which can only take on integer multiples of Planck constant. No half-integer solutions can be found in the x-representation. In specific representation, we lose part of the solution. The solution we are seeking now takes the form:

► $$\psi(r, \varphi, \theta) = f(r) \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\mathrm{i}m\varphi}\, h(\theta)\,; \qquad m = 0, \pm 1, \pm 2, \ldots \tag{2.159}$$

We have chosen the constant c such that the solution found is normalized to one. Next, we substitute this solution into the second equation (2.152) using (2.157) and solve it

$$-\hbar^2 \left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\right] \mathrm{e}^{\mathrm{i}m\varphi}\, h(\theta) = \lambda_l\, \mathrm{e}^{\mathrm{i}m\varphi}\, h(\theta) \qquad \Rightarrow$$

$$\frac{1}{\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}h}{\mathrm{d}\theta}\right) + \left(\frac{\lambda_l}{\hbar^2} - \frac{m^2}{\sin^2\theta}\right) h = 0\,.$$

The result is an ordinary differential equation for the function $h(\theta)$, which is solved using standard mathematical methods that go beyond the scope of this textbook. The result is polynomial functions in the variables $\cos\theta$ and $\sin\theta$, which are called the associated Legendre polynomials $P_{lm}(\cos\theta)$ and are defined by the relation

► $$P_{lm}(x) \equiv \frac{(1-x^2)^{m/2}}{2^l\, l!} \frac{\mathrm{d}^{l+m}}{\mathrm{d}x^{l+m}} (x^2 - 1)^l\,; \tag{2.160}$$

$$l = 0, 1, 2, \ldots; \qquad |m| \le l\,; \quad m = 0, \pm 1, \ \ldots$$

When $m = 0$, these functions are called Legendre polynomials. The related eigenvalue is

► $$\lambda_l = l(l+1)\hbar^2 . \tag{2.161}$$

The entire angular part of the solution is called a *spherical harmonic* and is denoted by

► $$Y_{lm}(\varphi,\theta) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\mathrm{i}m\varphi} P_{lm}(\cos\theta) . \tag{2.162}$$

The solution to the remaining two equations of system (2.152) is therefore

$$\psi(r,\varphi,\theta) = f(r)\, Y_{lm}(\varphi,\theta) = \frac{1}{\sqrt{2\pi}} f(r)\, \mathrm{e}^{\mathrm{i}m\varphi} P_{lm}(\cos\theta) ;$$

► $$\lambda_l = l(l+1)\hbar^2 ; \quad l = 0, 1, 2, \ldots \tag{2.163}$$

$$\mu_m = m\hbar ; \quad m = 0, \ \pm 1, \ \ldots; \qquad |m| \le l .$$

The derived quantization of angular momentum is identical to the relations derived by a different method in the previous chapter, except for the absence of half-integer values. For the radial function $f(r)$, the solution can be obtained from the first equation (2.152). This solution depends on the form of the potential energy. For some basic types of the potential energy, the solution will be discussed in the next chapter. Finally, let us give examples of some spherical functions that are excellent bases on the surface of a sphere:

$$\begin{aligned} Y_{00} &= \frac{1}{2\sqrt{\pi}} ; & Y_{1,-1} &= \sqrt{\frac{3}{8\pi}}\, \mathrm{e}^{-\mathrm{i}\varphi} \sin\theta ; \\ Y_{10} &= \sqrt{\frac{3}{4\pi}} \cos\theta ; & Y_{20} &= -\sqrt{\frac{5}{16\pi}}\, (1 - 3\cos^2\theta) ; \\ Y_{11} &= -\sqrt{\frac{3}{8\pi}}\, \mathrm{e}^{\mathrm{i}\varphi} \sin\theta ; & Y_{21} &= -\sqrt{\frac{15}{8\pi}}\, \mathrm{e}^{\mathrm{i}\varphi} \cos\theta \sin\theta . \end{aligned} \tag{2.164}$$

2.5.3 Simple Systems: Oscillator, Hydrogen, Well

Now we will solve the first of the equations (2.152) for the energy, which will give us the energy spectrum and the radial component of the entire solution $\psi(r, \varphi, \theta)$. Both the energy spectrum and the radial component may depend on the quantum numbers l and m from the previous solution and will depend on the form of the potential energy $V(r)$.

In the last equation (2.152), we know the action of the rotational part of the Hamiltonian operator on the wave function. This is given by the action of the square of the angular momentum, according (2.152). We already know the eigenvalue λ_l from equation (2.161). After applying the rotational part, we shorten the angular terms $g(\varphi)$ and $h(\theta)$ on both sides and obtain the equation for the radial part of the solution:

$$\left[-\frac{\hbar^2}{2m} \frac{1}{r^2} \frac{\mathrm{d}}{\mathrm{d}r} r^2 \frac{\mathrm{d}}{\mathrm{d}r} + \frac{\hbar^2\, l(l+1)}{2mr^2} + V(r) \right] f_{\nu l}(r) = E\, f_{\nu l}(r) . \tag{2.165}$$

Note that the equation involves the azimuthal quantum number l; therefore, the energy spectrum depends not only on the radial number ν, which quantizes the energy, but also on the azimuthal quantum number l. Equation (2.165) is solved using standard methods (series expansion, determination of asymptotic behavior, truncation of infinite series). We present the results of calculations for the potential energy of a spherical harmonic oscillator, a spatial well, and the Coulomb potential (2.136).

Harmonic oscillator

The energy spectrum for the potential energy of a harmonic oscillator is given by

► $$V(r)=\frac{1}{2}m\omega^2 r^2 \quad\Rightarrow\quad E_{\nu l}=(2\nu+l+3/2)\,\hbar\omega=(n+3/2)\,\hbar\omega\,. \tag{2.166}$$

The smallest possible energy value (zero-point energy) is $3/2\hbar\omega$. The radial quantum number ν denotes the order of the radial states and, as a rule, also the number of intersections of the radial solution with the x-axis. In most cases, the so-called principal quantum number n is introduced, which actually denotes the energy levels:

$$n\equiv 2\nu+l\ ;\quad n=0,1,2,\dots\quad,\quad l=0,1\ \dots\ n\ . \tag{2.167}$$

The oscillator's spectrum is degenerate (each energy value corresponds to multiple states; each n can be composed of multiple combinations of ν and l). We can easily determine the degree of degeneracy by noting that for each azimuthal quantum number, there are $2l + 1$ values of the magnetic quantum number m:

► $$\#_n=\sum_l 2l+1=\sum_{\nu=0}^{n/2}2(n-2\nu)+1=\sum_{\nu=0}^{n/2}2n-4\nu+1=\frac{(n+1)(n+2)}{2}\,. \tag{2.168}$$

We calculated the series (2.168) as an arithmetic series. Each energy shell n contains $(n + 1)(n + 2)/2$ states.

Coulomb potential

For the Coulomb potential energy, the energy spectrum is given by

► $$\begin{aligned} V(r) &= \frac{qQ}{4\pi\varepsilon_0}\frac{1}{r} = -\frac{\gamma}{r} \quad\Rightarrow \\ E_{\nu l} &= -\frac{\gamma^2 m_\mathrm{e}}{2\hbar^2(\nu+l+1)^2}=-\frac{\gamma^2 m_\mathrm{e}}{2\hbar^2 n^2}\,. \end{aligned} \tag{2.169}$$

We have defined the principal quantum number n for energy levels, using the relation

$$n\equiv\nu+l+1\quad;\qquad n=1,2,\dots\quad,\quad l=0,1\ \dots\ n-1\ . \tag{2.170}$$

The degree of degeneration will be

► $$\#_n=\sum_l 2l+1=\sum_{\nu=0}^{n-1}2(n-\nu-1)+1=\sum_{\nu=0}^{n-1}2n-2\nu-1=n^2\,. \tag{2.171}$$

In the case of a hydrogen atom, each electron can have two additional spin degrees of freedom, $m_\mathrm{s} = \pm 1/2$, and the total number of states in a single energy shell is therefore $2n^2$. These states differ in the values of the quantum numbers l, m, and m_s.

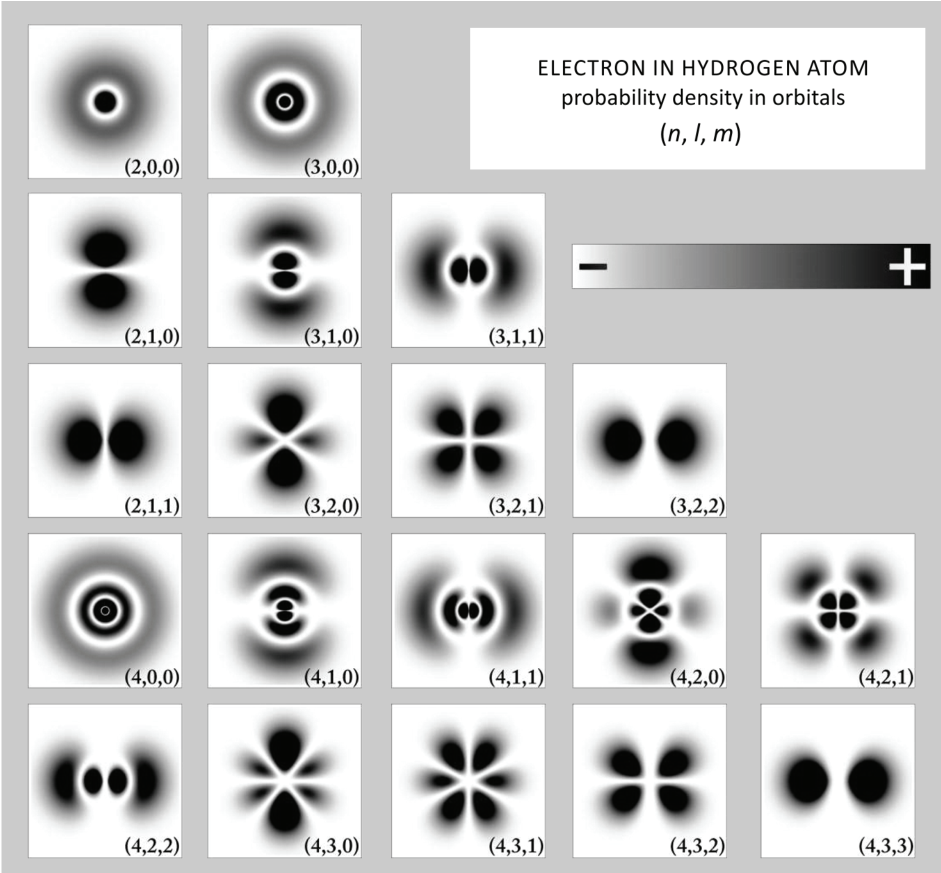


Fig. 2.35: Probability density for hydrogen,
Indiana University Southeast, 2008.

Spherical well

To conclude the chapter on spherical potentials, let us consider a spherical finite quantum well with a potential of the form

$$V(r) = \begin{cases} 0 & r < a\,; \\ V_0 & r \geq a\,. \end{cases} \tag{2.172}$$

Even for such a simple scenario, the problem has no analytical solution. It can only be solved numerically or graphically; see, for example, the publication [16].

2.6 Time Evolution

So far, in quantum theory, we have dealt with stationary states, i.e., states of the system that do not evolve over time. Actual quantum states are linear combinations of stationary states (basis elements), and the coefficients of these combinations change over time. The transition of a state from one time to a later time is performed by the so-called *time-evolution operator*, which we will now examine.

2.6.1 Time-Evolution Operator

The evolution operator maps a known state at time t_0 to the state it evolves into at time t:

$$|\psi(t)\rangle = \hat{\mathbf{U}}(t,t_0)\,|\psi(t_0)\rangle\,. \tag{2.173}$$

An evolutionary operator must meet certain conditions and requirements:

1. Initial condition: The evolution from the initial time to the initial time does not change the state of the system

$$\hat{\mathbf{U}}(t_0,t_0) = \hat{\mathbf{1}}\,. \tag{2.174}$$

2. Semigroup condition: The evolution from time t_1 to t_2 yields the same result whether it is carried out instantaneously or over an intermediate time t. By comparing the two approaches, we obtain the *semigroup condition*:

$$t_1 \to t_2 \qquad \Leftrightarrow \qquad t_1 \to t \to t_2$$

$$|\psi(t_2)\rangle = \hat{\mathbf{U}}(t_2,t_1)\,|\psi(t_1)\rangle \qquad \Leftrightarrow \qquad |\psi(t_2)\rangle = \hat{\mathbf{U}}(t_2,t)\;\hat{\mathbf{U}}(t,t_1)\;\psi(t_1)\rangle$$

$$\hat{\mathbf{U}}(t_2,t_1) = \hat{\mathbf{U}}(t_2,t)\,\hat{\mathbf{U}}(t,t_1)\,. \tag{2.175}$$

3. Unitarity: Changes over time do not alter the normalization of the state:

$$\langle\psi(t_0)|\psi(t_0)\rangle = \langle\psi(t)|\psi(t)\rangle\,,$$

$$\langle\psi_0|\psi_0\rangle = \langle\psi_0|\hat{\mathbf{U}}^{\dagger}\hat{\mathbf{U}}|\psi_0\rangle\,,$$

$$\hat{\mathbf{U}}^{\dagger}\hat{\mathbf{U}} = \hat{\mathbf{1}}\,. \tag{2.176}$$

4. Inversion: The inverse evolution operator has the arguments in reverse order. We derive this from the semigroup condition:

$$\hat{\mathbf{U}}(t,t) = \hat{\mathbf{U}}(t,t_0)\;\hat{\mathbf{U}}(t_0,t) \stackrel{!}{=} \hat{\mathbf{1}} \quad \Rightarrow$$

$$\hat{\mathbf{U}}^{-1}(t_0,t) \;=\; \hat{\mathbf{U}}(t,t_0)\,. \tag{2.177}$$

5. Continuity: The spontaneous evolution of a state (without any measurement) described by an evolution operator must be continuous:

$$\langle\varphi|\,\hat{\mathbf{U}}(t,t_0)\,|\psi(t_0)\rangle \;\text{ je spojité pro }\; \forall\, t_0 \;\text{ a }\; \forall\,|\varphi\rangle \in \mathcal{H}\,.$$

We will now derive the basic equation for the evolution operator. We start from the definition of the mean value of a dynamic variable (see table in Section 2.2.2 on page 124) and take the derivative of this mean value with respect to time:

$$\frac{\mathrm{d}\bar{a}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\langle\psi|\hat{\mathbf{A}}|\psi\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\langle\psi_0|\hat{\mathbf{U}}^{\dagger}\hat{\mathbf{A}}\hat{\mathbf{U}}|\psi_0\rangle =$$

$$= \langle\psi_0|\frac{\mathrm{d}\hat{\mathbf{U}}^{\dagger}}{\mathrm{d}t}\hat{\mathbf{A}}\hat{\mathbf{U}} + \hat{\mathbf{U}}^{\dagger}\hat{\mathbf{A}}\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t}|\psi_0\rangle .$$

Another option is to use the time derivative operator for the dynamic variable:

$$\frac{\mathrm{d}\bar{a}}{\mathrm{d}t} = \langle\psi|\dot{\hat{\mathbf{A}}}|\psi\rangle = \langle\psi_0|\hat{\mathbf{U}}^{\dagger}\dot{\hat{\mathbf{A}}}\hat{\mathbf{U}}|\psi_0\rangle .$$

By comparing the two methods, we obtain the equation

$$\frac{\mathrm{d}\hat{\mathbf{U}}^{\dagger}}{\mathrm{d}t}\hat{\mathbf{A}}\hat{\mathbf{U}} + \hat{\mathbf{U}}^{\dagger}\hat{\mathbf{A}}\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t} = \hat{\mathbf{U}}^{\dagger}\dot{\hat{\mathbf{A}}}\hat{\mathbf{U}} , \tag{2.178}$$

in which we substitute for the time derivative of the dynamic variable operator the time evolution of the dynamic variable written in Poisson brackets, converted to quantum form using the correspondence principle (2.21):

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \{A, H\} \qquad \Rightarrow \qquad \dot{\hat{\mathbf{A}}} = \frac{1}{\mathrm{i}\hbar}\left[\hat{\mathbf{A}}, \hat{\mathbf{H}}\right]. \tag{2.179}$$

This gives us an equation from which we will derive the evolution operator:

$$\frac{\mathrm{d}\hat{\mathbf{U}}^{\dagger}}{\mathrm{d}t}\hat{\mathbf{A}}\hat{\mathbf{U}} + \hat{\mathbf{U}}^{\dagger}\hat{\mathbf{A}}\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t} = \hat{\mathbf{U}}^{\dagger}\frac{1}{\mathrm{i}\hbar}\left[\hat{\mathbf{A}}, \hat{\mathbf{H}}\right]\hat{\mathbf{U}} . \tag{2.180}$$

In all of the following derivations, we use the unitarity $\hat{\mathbf{U}}\hat{\mathbf{U}}^{\dagger} = \hat{\mathbf{U}}^{\dagger}\hat{\mathbf{U}} = \mathbf{1}$. From the last equation, we must eliminate the operator $\hat{\mathbf{U}}^{\dagger}$ and its time derivative, which we obtain by differentiating the definition of unitarity with respect to time and multiplying the result by the operator $\hat{\mathbf{U}}^{\dagger}$ from the right:

$$\hat{\mathbf{U}}^{\dagger}\hat{\mathbf{U}} = \hat{\mathbf{1}} \quad \Rightarrow$$

$$\frac{\mathrm{d}\hat{\mathbf{U}}^{\dagger}}{\mathrm{d}t}\hat{\mathbf{U}} + \hat{\mathbf{U}}^{\dagger}\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t} = 0 \quad \Rightarrow$$

$$\frac{\mathrm{d}\hat{\mathbf{U}}^{\dagger}}{\mathrm{d}t} + \hat{\mathbf{U}}^{\dagger}\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t}\hat{\mathbf{U}}^{\dagger} = 0 \quad \Rightarrow$$

$$\frac{\mathrm{d}\hat{\mathbf{U}}^{\dagger}}{\mathrm{d}t} = -\hat{\mathbf{U}}^{\dagger}\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t}\hat{\mathbf{U}}^{\dagger} .$$

We substitute the result into equation (2.180) and multiply it by $\hat{\mathbf{U}}$ on the left and $\hat{\mathbf{U}}^{\dagger}$ on the right. Through a series of simplifications, we obtain the desired equation for $\hat{\mathbf{U}}$:

$$-\hat{\mathbf{U}}^{\dagger}\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t}\hat{\mathbf{U}}^{\dagger}\hat{\mathbf{A}}\hat{\mathbf{U}} + \hat{\mathbf{U}}^{\dagger}\hat{\mathbf{A}}\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t} = \frac{1}{\mathrm{i}\hbar}\hat{\mathbf{U}}^{\dagger}\left[\hat{\mathbf{A}}, \hat{\mathbf{H}}\right]\hat{\mathbf{U}} \quad \Rightarrow$$

$$-\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t}\hat{\mathbf{U}}^{\dagger}\hat{\mathbf{A}}+\hat{\mathbf{A}}\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t}\hat{\mathbf{U}}^{\dagger}=\frac{1}{\mathrm{i}\hbar}\left[\hat{\mathbf{A}},\hat{\mathbf{H}}\right] \quad \Rightarrow$$

$$\left[\hat{\mathbf{A}},\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t}\hat{\mathbf{U}}^{\dagger}\right]=\left[\hat{\mathbf{A}},\frac{1}{\mathrm{i}\hbar}\hat{\mathbf{H}}\right] \quad \Rightarrow$$

$$\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t}\hat{\mathbf{U}}^{\dagger}=\frac{1}{\mathrm{i}\hbar}\hat{\mathbf{H}} \quad \Rightarrow$$

► $$\mathrm{i}\hbar\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t}=\hat{\mathbf{H}}\hat{\mathbf{U}}. \tag{2.181}$$

The equation just derived is called the *time-evolution equation*. If we apply this operator equation to the initial state $|\,\psi_0\,\rangle$, the evolution operator evolves the state to time t, and the resulting equation for $|\,\psi\,(t)\,\rangle$ is called the *Schrödinger time equation*:

► $$\mathrm{i}\hbar\frac{\mathrm{d}\left|\psi(t)\right\rangle}{\mathrm{d}t}=\hat{\mathbf{H}}\left|\psi(t)\right\rangle. \tag{2.182}$$

So we have two Schrödinger equations. The time-independent Schrödinger equation determines the eigenvalues and eigenstates of the energy operator (the Hamiltonian). From it, we can determine the energy values that can be measured in an experiment.

The time-dependent Schrödinger equation describes the time evolution of any initial state. As we will see, the solution to the time-dependent Schrödinger equation is, in a certain sense, trivial. If the Hamiltonian does not depend explicitly on time and we know the eigenstates and eigenvalues of the energy operator, we can immediately write down the solution to the time-dependent Schrödinger equation.

2.6.2 Time-Dependent Schrödinger Equation

The solution to the time-dependent problem can be found relatively easily if the Hamiltonian is not an explicit function of time, i.e., if it depends only on the generalized coordinate and momentum operators. In such a case, it is advantageous to choose a basis in the Hilbert space of the system as the eigenvectors of the Hamiltonian (2.41):

$$\hat{\mathbf{H}}\,|\,n\,\rangle=E_n\,|\,n\,\rangle \quad ; \quad \langle\,m\,|\,n\,\rangle=\delta_{nm} \quad ; \quad \sum_n|\,n\,\rangle\langle\,n\,|=\hat{\mathbf{1}}\,.$$

We will expand the searched state into these vectors

$$|\psi(t)\,\rangle=\sum_n a_n(t)|\,n\,\rangle\,.$$

We substitute this solution into the time-dependent Schrödinger equation and obtain a linear equation for the time-dependent coefficients $a_n(t)$.

$$\mathrm{i}\hbar\sum_n\frac{\mathrm{d}a_n}{\mathrm{d}t}|\,n\,\rangle=\hat{\mathbf{H}}\sum_n a_n(t)\,|\,n\,\rangle\,;$$

$$\mathrm{i}\hbar\sum_n\frac{\mathrm{d}a_n}{\mathrm{d}t}|\,n\,\rangle=\sum_n a_n(t)E_n\,|\,n\,\rangle\,; \qquad /\ \langle\,k\,|\ \text{ from left}$$

$$\mathrm{i}\hbar\frac{\mathrm{d}a_k}{\mathrm{d}t}=a_k(t)E_k\ ;$$

$$a_k(t)=c_k\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_k(t-t_0)}\ .$$

The solution to the time-dependent problem is therefore:

$$|\psi(t)\rangle=\sum_n c_n\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_n(t-t_0)}\,|\,n\,\rangle\ . \tag{2.183}$$

If we set $t = t_0$, we obtain the expansion of the initial condition:

$$|\psi(t_0)\rangle=\sum_n c_n\,|\,n\,\rangle\ . \tag{2.184}$$

The coefficients c_n are thus the coefficients for expanding the initial condition into a basis of energy eigenvectors. The time evolution differs from the expansion of the initial condition only by an oscillating exponential. Thus, if we know the solution to the time-independent Schrödinger equation, we can immediately write down the solution to the time-dependent Schrödinger equation as well. The time evolution does not play as fundamental a role in quantum theory as it does in Newtonian dynamics; it can always be easily written down if we know the eigenstates and eigenvalues of the Hamiltonian.

A somewhat more elegant solution to the equation for the evolution operator is the operator solution, which can be formally written as

$$\mathrm{i}\,\hbar\,\frac{\mathrm{d}\hat{\mathbf{U}}}{\mathrm{d}t}=\hat{\mathbf{H}}\hat{\mathbf{U}}\quad\Rightarrow$$

$$\hat{\mathbf{U}}(t,t_0)=\mathrm{e}^{\frac{1}{\mathrm{i}\hbar}\hat{\mathbf{H}}(t-t_0)}\ .$$

The evolution operator is a function of the Hamiltonian. If we know the eigenvectors and eigenvalues of the Hamiltonian (the solutions to the time-independent Schrödinger equation), we can express the evolution operator using the spectral expansion theorem:

$$\hat{\mathbf{U}}(t,t_0)=\sum_n \mathrm{e}^{\frac{1}{\mathrm{i}\hbar}E_n(t-t_0)}\,|\,n\,\rangle\langle\,n\,|.$$

Now we apply the evolutionary operator we found to the initial state $|\,\psi_0\,\rangle$:

► $$|\,\psi(t)\,\rangle=\sum_n \mathrm{e}^{\frac{1}{\mathrm{i}\hbar}E_n(t-t_0)}\,|\,n\,\rangle\langle\,n\,|\psi_0\,\rangle\ . \tag{2.185}$$

This immediately gives us a solution to the time-dependent Schrödinger equation:

► $$|\,\psi(t)\,\rangle=\sum_n c_n\,\mathrm{e}^{\frac{1}{\mathrm{i}\hbar}E_n(t-t_0)}\,|\,n\,\rangle\,;\qquad c_n\equiv\langle\,n\,|\psi_0\,\rangle\ . \tag{2.186}$$

Both methods therefore lead to the same solution.

◖ **Example 2.3: Eigenstate time-development**

Suppose the system is in a particular energy state (for example, the third one). Find the probability distribution of the system's state.

Solution: The initial state is equal to

$$\psi_0(x) = c_3\psi_3(x)\,,$$

where c_3 is the normalization coefficient. The time evolution of the state will be

$$\psi(t,x) = c_3\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_3 t}\,\psi_3(x)$$

and the resulting probability density is

$$w(t,x) \equiv \psi^*\psi = \left|c_3\psi_3\right|^2.$$

It is evident that the evolution over time has not affected the probability density; in this sense, therefore, a system in one of its eigenstates does not evolve.

◗

◖ **Example 2.4: Quantum interference**

Find the probability evolution of a system whose initial state is given as a linear combination of two eigenfunctions of the Hamiltonian operator.

Solution: The initial state is a combination of two eigenstates, 1 and 2, of the Hamiltonian operator (these need not, of course, be the first and second energy levels, but can be any two states)

$$\psi_0(x) = c_1\psi_1(x) + c_2\psi_2(x)\,,$$

the evolution over time is

$$\psi(t,x) = c_1\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_1 t}\,\psi_1(x) + c_2\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}E_2 t}\,\psi_2(x)$$

and the resulting probability density is

$$w(t,x) \equiv \psi^*\psi =$$

$$= \left|c_1\psi_1\right|^2 + \left|c_2\psi_2\right|^2 + \left[(c_1\psi_1)(c_2\psi_2)^*\,\mathrm{e}^{\frac{\mathrm{i}}{\hbar}(E_2-E_1)\,t} + (c_1\psi_1)^*(c_2\psi_2)\,\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}(E_2-E_1)\,t}\right].$$

The total probability is the sum of the probabilities that the system is in state 1, in state 2, and the interference term typical of quantum processes. The result can be simply written as follows:

$$w(t,x) = w_1(x) + w_2(x) + A(x)\cos\omega t + B(x)\sin\omega t\,; \qquad \omega \equiv \Delta E/\hbar\,. \tag{2.187}$$

The angular frequency of the time oscillations of the probability corresponds to quantization with an energy equal to the Planck constant, $\Delta E = \hbar\omega$.

◗

2.6.3 Neutrino Oscillation

Neutrinos (electron, muon, and tau) are, in fact, a linear combination of the eigenstates of matter

$$\left|\nu_\alpha\right\rangle = V_{\alpha k}\left|\nu_k\right\rangle, \quad \text{where } \alpha = \mathrm{e}, \mu, \tau; \qquad k = 1, 2, 3.$$

The index α describes the neutrino generations, and the index k describes the mass states. The transformation matrix (known as the *mixing matrix*) is unitary and was first introduced by Ziro Maki, Masami Nakagawa, and Shoichi Sakata in 1962 to explain the neutrino oscillations predicted by Bruno Pontecorvo. To understand the principle of oscillations, let us assume the existence of only two generations of neutrinos and the mixing of their states in the form

$$\left|\nu_\mathrm{e}\right\rangle = +\cos\theta\left|\nu_1\right\rangle + \sin\theta\left|\nu_2\right\rangle,$$

$$\left|\nu_\mu\right\rangle = -\sin\theta\left|\nu_1\right\rangle + \cos\theta\left|\nu_2\right\rangle.$$

We have expressed the unit matrix as a standard rotation matrix using the angle θ. During the flight of neutrinos, the mass states evolve and the mixing ratios changes. However, rather than the evolution over time, we are interested in the evolution of the state along the path of the flying particle. Since the following holds for a plane wavefront

$$\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{x}-\omega t)} = \mathrm{e}^{\frac{\mathrm{i}}{\hbar}(\mathbf{p}\cdot\mathbf{x}-E\,t)},$$

we can express the evolution of the mass states along the neutrino's path as follows:

$$\left|\nu_k(x)\right\rangle = \mathrm{e}^{\frac{\mathrm{i}}{\hbar}p_k x}\left|\nu_k(0)\right\rangle.$$

The state of a neutrino, e.g. electron one, changes during flight according to the formula

$$\left|\nu_\mathrm{e}(x)\right\rangle = +\mathrm{e}^{\frac{\mathrm{i}}{\hbar}p_1 x}\cos\theta\left|\nu_1(0)\right\rangle + \mathrm{e}^{\frac{\mathrm{i}}{\hbar}p_2 x}\sin\theta\left|\nu_2(0)\right\rangle.$$

The probability amplitude that an electron neutrino will appear to an observer as a pure muon neutrino during its flight (given by its initial combination) will be

$$\mathscr{A}_{\nu_\mathrm{e}\to\nu_\mu} = \left\langle\nu_\mu(0)\middle|\nu_\mathrm{e}(x)\right\rangle.$$

Performing the scalar product (the mass states form an orthonormal basis), we have

$$\mathscr{A}_{\nu_\mathrm{e}\to\nu_\mu} = \cos\theta\sin\theta\left[\exp\left(\frac{\mathrm{i}}{\hbar}p_2 x\right) - \exp\left(\frac{\mathrm{i}}{\hbar}p_1 x\right)\right].$$

Neutrinos have low mass and relativistic energy, and therefore can be used expansion

$$p_k = \sqrt{(E/c)^2 - m_k^2 c^2} = \frac{E}{c}\sqrt{1-\left(m_k c^2/E\right)^2} \approx \frac{E}{c} - \frac{m_k^2 c^3}{2E}.$$

We can now easily adjust the probability amplitude

$$\mathscr{A}_{\nu_e \to \nu_\mu} = \cos\theta \sin\theta \, e^{i\frac{Ex}{\hbar c}} \left[\exp\left(-i \frac{m_2^2 c^3}{2\hbar E} x \right) - \exp\left(-i \frac{m_1^2 c^3}{2\hbar E} x \right) \right] =$$

$$= \cos\theta \sin\theta \, e^{i\left(\frac{E}{\hbar c} - \frac{m_1^2 c^3}{2\hbar E}\right)x} \left[\exp\left(-i \frac{\Delta m^2 c^3}{2\hbar E} x \right) - 1 \right] =$$

$$= \cos\theta \sin\theta \, e^{i\left(\frac{E}{\hbar c} - \frac{m_1^2 c^3}{2\hbar E} - \frac{\Delta m^2 c^3}{4\hbar E}\right)x} \left[\exp\left(-i \frac{\Delta m^2 c^3}{4\hbar E} x \right) - \exp\left(+i \frac{\Delta m^2 c^3}{4\hbar E} x \right) \right] =$$

$$= -2i \cos\theta \sin\theta \, e^{i\left(\frac{E}{\hbar c} - \frac{m_1^2 c^3}{2\hbar E} - \frac{\Delta m^2 c^3}{4\hbar E}\right)x} \sin\left(\frac{\Delta m^2 c^3}{4\hbar E} x \right).$$

If the eigenmasses are different (even just one non-zero value is sufficient), *neutrino oscillations* occur. The probability of finding the original electron neutrino as a muon neutrino is a periodic function of the distance from the source

►
$$\mathscr{P}_{\nu_e \to \nu_\mu} = |\mathscr{A}|^2 = \mathscr{A}\mathscr{A}^* = \sin^2 2\theta \sin^2\left(\frac{\Delta m^2 c^3}{4\hbar E} x \right);$$

$$\Delta m^2 \equiv m_2^2 - m_1^2 .$$

Based on various experiments, it is possible to determine the mixing angle and the average distance over which neutrino conversion occurs

►
$$L = \frac{4\pi \hbar E}{\Delta m^2 c^3} .$$

It is clear that only the difference in the squares of the neutrino masses can be determined. Real neutrinos have three generations, and the transformation matrix is 3×3 and contains three angles. The measurements show that the following holds:

$\Delta m_{12}{}^2 = (7{,}41 \pm 0{,}18)\times 10^{-5}\ \text{eV}^2$ JUNO, 2025)

$\Delta m_{23}{}^2 = (2{,}5 \pm 0{,}02)\times 10^{-3}\ \text{eV}^2$ (T2K, NOvA, 2023).

$\theta_{12} \sim 33°$, $\theta_{23} \sim 45°$, $\theta_{13} \sim 8.5°$.

The mixing matrix is thus clearly not similar to a diagonal matrix, as is the case with the corresponding mixing matrix for quarks.

Neutrino oscillations were first observed in 1998 at the Super-Kamiokande detector in Japan. When observing atmospheric neutrinos (which are produced by the interaction of cosmic rays with the upper atmosphere), a different ratio of electron and muon neutrinos was observed from above and below. Neutrinos coming from above did not have enough time to oscillate, whereas neutrinos coming from below passed through the entire Earth and had sufficient time to oscillate. Similar oscillations were also observed at the same time at the Sudbury Neutrino Observatory in the United States.

2.6.4 Double-Slit Experiment, AB Experiment, MZ Interferometer

Let's imagine that a stream of particles strikes two slits. After passing through the slits, a screen records where each particle landed. The result is a classic interference pattern, with the maximum number of hits paradoxically occurring between the two slits. As in the previous chapter, the amplitudes of the probabilities of both possibilities are added together, not the probabilities themselves.

The number of particles does not affect the result: if the flow is very weak and, on average, only one particle appears in the experimental area, we will never be able to determine which slit it passed through. After a sufficiently long time, we obtain a statistical picture of the particles striking the screen, as shown in the figure. We might think, that part of the particle passed through one slit and part through the other, or that it interfered with itself. Such considerations have no real meaning. To evaluate the statistical result of many repeated impacts, the only thing that matters is whether the experimental result agrees with the prediction given by the theory.

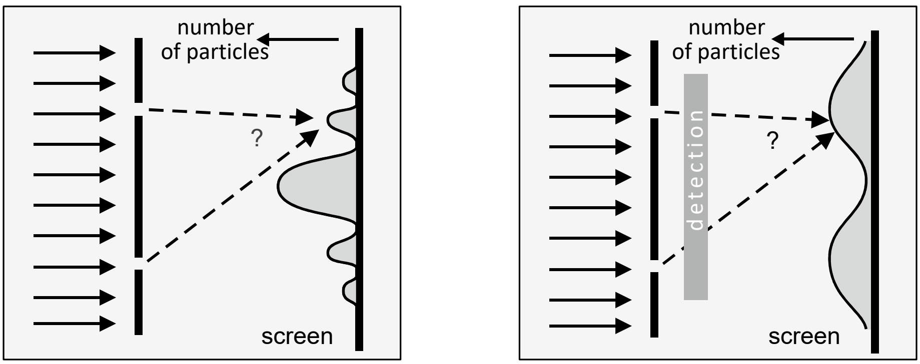


Fig. 2.36: Double-slit experiment

A different picture emerges if we try to determine which path the particle took. If we cover one of the holes, the maximum number of particles will be observed in front of the open hole. We can devise a more sophisticated method, e.g., we can use photons to track the path. If the photon has low energy, it will have too long wavelength to determine the path. However, if the photon has a wavelength short enough for detection, we can actually determine which path the particle took. But a short-wavelength photon has considerable energy and strongly affects the state of the particle and the interference pattern completely disappears. Generally speaking: if we do not attempt detection, the probability amplitudes add up, and the statistics has the character of an interference phenomenon. If we attempt detection, the interference disappears and the probabilities are added together in the classical sense. We don't even have to attempt detection; it is sufficient that it is in principle possible, and quantum interference vanishes. We find this fact difficult to accept. It is a property of the microworld that seems very strange to us. Our experiences from the macroscopic world are based on commutative objects. It is precisely the non-commutativity of phenomena in the microworld that leads to the superposition of probability amplitudes and to the interference phenomenon.

Aharon-Bohm experiment

In classical electrodynamics, an electromagnetic field can be described either in terms of electric intensity and magnetic induction or in terms of the four-potentials (see Section 1.6.3 or [1]). Each of these descriptions has its advantages and disadvantages:

1. Electric and magnetic fields can be measured directly with instruments, whereas potentials cannot. This situation gives the impression that fields are real quantities, while potentials are merely auxiliary mathematical objects.
2. Electromagnetic fields are unique, whereas there are infinitely many potentials for a given problem. This can be used to construct the simplest possible equations for potentials. On the other hand, the ambiguity of potentials again gives the impression that the concept of potentials is a mathematical construct.
3. Maxwell equations in terms of potentials are simpler; they lead to a wave equation with a non-zero right-hand side, for which there are many possible solutions. However, after finding the potentials, we must still determine the fields from the formulas $\mathbf{B} = \text{rot}\,\mathbf{A}$, $\mathbf{E} = -\nabla\phi - \partial\mathbf{A}/\partial t$.
4. Potentials are more suitable when transforming fields into a different coordinate system. The four-potential is transformed by the Lorentz matrix. The transformation of the fields is messy; it's given by the transformation of the field tensor.
5. At first glance, the four-potential of the field (ϕ, $\mathbf{A}$) fits better into the four-dimensional world of relativity than the six values of $\mathbf{E}$ and $\mathbf{B}$. These are actually part of the field tensor $F_{\mu\nu}$, as we saw in Section 1.6.3.

In classical physics, a particle can change its velocity only under the influence of electromagnetic fields. If the fields are zero, no forces act on the particle. In quantum mechanics, the situation is different. We will show that the presence of a non-zero potential changes the phase of the wave function even when the fields themselves are zero (e.g., in the space outside a long coil, the magnetic field is zero and the vector potential is non-zero). The change in the phase of the wave function manifests itself as a change in the interference pattern in the double-slit experiment and is therefore a measurable phenomenon. In this sense, classical Maxwell electrodynamics, supplemented by the Lorentz equation of motion, is an incomplete description of nature, as it does not capture all the measurable processes occurring in nature. Furthermore, field potentials are not merely mathematical constructs but have a real physical impact on the motion.

This fact was first pointed out by the British theorists Werner Ehrenberg and Raymond Siday in 1949, but their work did not gain sufficient traction. A similar phenomenon was predicted again ten years later, in 1959, by the Israeli physicist Yakir Aharonov and the American-British theorist David Bohm. The Aharonov-Bohm effect (AB effect) was experimentally confirmed in 1986 by the Japanese physicist Akira Tonomura.

Change in the impact pattern caused by the presence of a magnetic field

Let us first consider a double-slit experiment with the arrangement shown in Fig. 2.37 on the left. Behind the slits is a narrow band of non-zero magnetic field (of thickness Δl) that is perpendicular to the direction of electron motion. The Lorentz force will be

$$F = evB \tag{2.188}$$

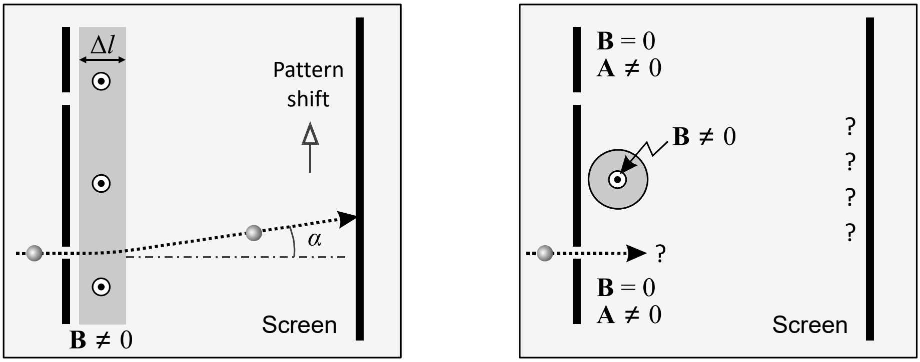
upward. Since the thickness of the non-zero field layer is small, we will assume that the electron beam moves upward under the influence of a constant acceleration $a = F/m$ for a time $\Delta t = \Delta l/v$. In the vertical direction, the electrons will be deflected by a distance

$$\Delta y \cong \frac{1}{2}a(\Delta t)^2 = \frac{1}{2}\frac{evB}{m}\left(\frac{\Delta l}{v}\right)^2 = \frac{eB}{2mv}\Delta l^2. \tag{2.189}$$

The beam deflection angle will be calculated using the standard formula

► $$\operatorname{tg}\alpha \cong \frac{\Delta y}{\Delta l} = \frac{eB\,\Delta l}{2mv}. \tag{2.190}$$

Regardless of the shape of the electron impact pattern on the screen, the magnetic field should cause it to shift upward by an angle α given by equation (2.190).



s

Fig. 2.37: An explanation of the Aharonov–Bohm effect (2.190).

Change in the interference pattern caused by a non-zero potential (AB effect)

Let us now consider a double-slit experiment with the setup shown in Figure 2.37 on the right. Behind the slits is a long solenoid containing a non-zero magnetic field. Outside the solenoid – that is, in the region through which the electrons travel – the magnetic field is zero, and the pattern should not be shifted. Only the vector potential is non-zero here. Let us consider a wave function of the form

$$\psi(t,\mathbf{x}) = A\,\mathrm{e}^{\mathrm{i}\,\varphi(t,\mathbf{x})}. \tag{2.191}$$

The constancy of the amplitude is not relevant to the calculation. We will seek an equation for the phase of the wave function; therefore, we substitute (2.191) into the time-domain Schrödinger equation (2.182) written in the x-representation:

$$i\hbar\frac{\partial\psi}{\partial t} = \left(\frac{-\hbar^2}{2m}\nabla^2 + V\right)\psi. \tag{2.192}$$

The result is the equation

$$-\hbar\frac{\partial\varphi}{\partial t} = \left(-\mathrm{i}\frac{\hbar^2}{2m}\nabla^2\varphi + \frac{\hbar^2}{2m}\left(\frac{\partial\varphi}{\partial\mathbf{x}}\right)^2 + V\right). \tag{2.193}$$

If we isolate only the real part, we get the equation

$$-\frac{\partial\hbar\varphi}{\partial t} = \left(\frac{1}{2m}\left(\frac{\partial\hbar\varphi}{\partial\mathbf{x}}\right)^2 + V\right),$$

which we will transform into the final form

$$\frac{\partial S}{\partial t}+\frac{1}{2m}\left(\frac{\partial S}{\partial \mathbf{x}}\right)^2+V=0\,; \qquad S=\hbar\varphi\,. \tag{2.194}$$

This is the Hamilton-Jacobi equation (1.177), in which S represents the action of the system. The solution to the wave equation is therefore given by the action of the system:

► $$\varphi=\frac{S}{\hbar}\,. \tag{2.195}$$

In the case of the interaction of a charged particle with an electromagnetic field, we know the Lagrangian (1.260)

$$L_{\text{int}} == -Q\phi(t,\mathbf{x})+Q\mathbf{A}(t,\mathbf{x})\cdot\mathbf{v}\,.$$

The phase of the wave function will therefore be

$$\begin{aligned}\varphi&=\frac{S}{\hbar}=\frac{1}{\hbar}\int_{t_0}^{t}L_{\text{int}}\mathrm{d}t=\frac{Q}{\hbar}\int\left(-\phi+\mathbf{A}\cdot\mathbf{v}\right)\mathrm{d}t=\\&=\frac{Q}{\hbar}\int\left(-\phi\mathrm{d}t+\mathbf{A}_k\mathrm{d}x_k\right)=\frac{Q}{\hbar}\int A_\mu\mathrm{d}x^\mu.\end{aligned} \tag{2.196}$$

Whichever formulation we choose, the phase of the wave function is linked to the potentials, which play a primary role here. The entire wave function will be

► $$\psi=A\exp\left[\frac{\mathrm{i}Q}{\hbar}\int A_\mu\mathrm{d}x^\mu\right]. \tag{2.197}$$

In the region through which electrons pass in the Aharonov–Bohm thought experiment, the magnetic field is zero, but the vector potential is nonzero, causing a phase shift in the wave function and thereby a shift in the interference pattern.

The Japanese physicist Akira Tonomura first attempted to measure this phenomenon in 1982 using an electron holographic microscope, which is capable of recording not only the intensity of the electron beam but also the phase of the electrons (which emit a coherent electromagnetic field). The results for the coil used were inconclusive, as the field leaked outside the coil. Therefore, in 1986, Tonomura used a toroidal ferromagnet with a diameter of 6 μm as the field source. The surface was coated with superconducting niobium, which perfectly shielded the magnetic field. The temperature was maintained at 5 K. The shift of the interference fringes was measured between the electron beam passing through the interior of the toroid and the electron beam passing outside the magnet. In these regions, the magnetic field is zero, but the vector potential differs. The fringes were shifted by an amount predicted by the Aharonov–Bohm effect [29].

Note: Feynman imagined that between the start and end of a particle's motion there are an infinite number of plates and slits. Every trajectory is possible and yields a probability amplitude $\psi = A\exp(\mathrm{i}S/\hbar)$. Summing all the probability amplitudes led Feynman to introduce the so-called *path integral*. Quantum interference cancels out most of the amplitudes. Only those with similar phases are amplified, i.e., those for which the action S does not change significantly (near the minimum or maximum). Therefore, the highest probability will be associated with trajectories with extremal action, which is precisely *Hamilton principle of least action*!

Mach-Zehnder interferometer

A similar variant of the double-slit experiment is the Mach–Zehnder interferometer. It is named after the Austrian physicist Ludwig Mach (1868–1951), who was the son of the famous Ernst Mach, and the Swiss physicist Ludwig Zehnder (1858–1949), considered the inventor of the interferometer. Instead of electrons, light is used here, which again has the option of traveling to the detector via two paths. This time, however, there are not two slits, but two interference arms formed by two fully reflective and two partially reflective mirrors. We will assume that the phase shifts by 90° with each reflection.

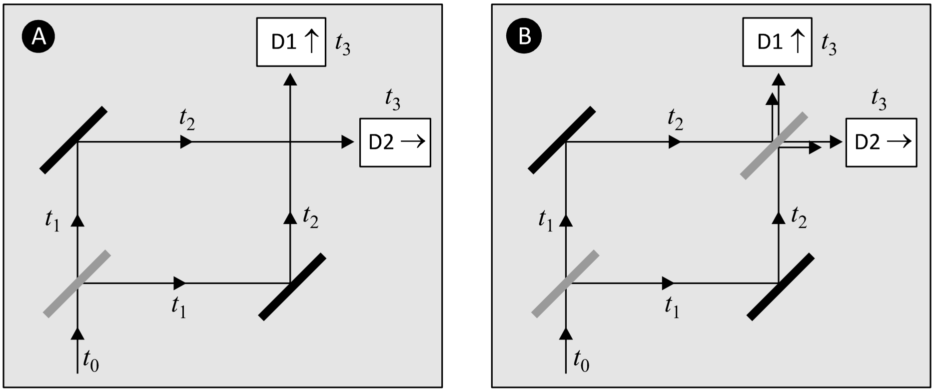


Fig. 2.38: Mach-Zehnder interferometer

Let us first consider the situation shown on the left. Light from the source splits into two beams (↑ and →) at the half-silvered mirror; which no longer interact with each other and enter detectors D1 (vertical-beam detector) and D2 (horizontal-beam detector):

$$\left|\psi(t_0)\right\rangle = \left|\uparrow\right\rangle;$$
$$\left|\psi(t_1)\right\rangle = \frac{1}{\sqrt{2}}\left|\uparrow\right\rangle + \frac{\mathrm{i}}{\sqrt{2}}\left|\rightarrow\right\rangle; \tag{2.198}$$
$$\left|\psi(t_2)\right\rangle = \left|\psi(t_3)\right\rangle = \frac{\mathrm{i}}{\sqrt{2}}\left|\rightarrow\right\rangle + \frac{\mathrm{i\,i}}{\sqrt{2}}\left|\uparrow\right\rangle = \frac{\mathrm{i}}{\sqrt{2}}\left|\rightarrow\right\rangle - \frac{1}{\sqrt{2}}\left|\uparrow\right\rangle.$$

At time t_1, the beam (or a single photon) is in a superposition of two states (vertical and horizontal paths). The superposition coefficients ensure normalization (the sum of all probabilities equals 1). The probabilities are given by the squares of the coefficients for the individual states of the superposition, i.e., 0.5 for both beams. At times t_2 and t_3, the situation is the same – each beam has undergone a single reflection, which is represented by a phase shift of 90° (multiplication by i). Detectors D1 and D2 have the same probability of detecting photons; they will register the same light intensity. In the situation on the right (B), a single half-silvered mirror is added, on which each of the states at time t_2 splits into another superposition:

$$\left|\psi(t_3)\right\rangle = \frac{\mathrm{i}}{\sqrt{2}}\left[\frac{1}{\sqrt{2}}\left|\rightarrow\right\rangle + \frac{\mathrm{i}}{\sqrt{2}}\left|\uparrow\right\rangle\right] - \frac{1}{\sqrt{2}}\left[\frac{1}{\sqrt{2}}\left|\uparrow\right\rangle + \frac{\mathrm{i}}{\sqrt{2}}\left|\rightarrow\right\rangle\right] = -\left|\uparrow\right\rangle. \tag{2.199}$$

We see that destructive interference occurs for the horizontal beam at detector D2 (it detects no signal, as the beams arrive out of phase) and constructive interference at D1 (it detects all photons in phase, i.e., 100 % of the light's intensity). This is another example of the non-local behavior of quantum theory caused by superposition of states.

If we wanted to describe the result in the classical sense, we can consider only the phase shifts of the upper and lower arms of the interferometer. The beam from the upper arm reaches detector D1 after two reflections, i.e., with a phase shift of 180°. The beam from the lower arm also arrives here after two reflections, again with a phase shift of 180°. It is clear that both beams arrive at D1 in phase and will interfere constructively.

In detector D2 the situation is different. The beam from the upper branch has undergone a single reflection, so its phase shift is only 90°. The beam from the lower branch undergoes three reflections and has a phase shift of 270°. Both beams are thus in antiphase and cancel each other out through interference. However, this classical interpretation does not answer the question of where the energy of the two waves, which arrive at the detector in antiphase, has gone. The answer lies in the nonlocality of quantum theory: at time t_3, the photon is in a superposition of multiple states and was detected by detector D1. In D2, the probability of detecting this photon is zero. We will discuss further details of interference in the Mach-Zehnder interferometer in Sec. 2.9.1 and 2.9.2

2.6.5 Ehrenfest Theorems, Virial Theorem

In this chapter, we will discuss three fundamental theorems concerning time evolution.

Ehrenfest first theorem

The first theorem concerns the time evolution of the coordinate operator. For simplicity, we derive it in the one-dimensional case, starting from the principle of correspondence and the time evolution (1.40):

$$\begin{aligned}\frac{\mathrm{d}\hat{\mathbf{X}}}{\mathrm{d}t} &= \frac{1}{\mathrm{i}\hbar}\left[\hat{\mathbf{X}},\hat{\mathbf{H}}\right] = \frac{1}{\mathrm{i}\hbar}\left[\hat{\mathbf{X}},\frac{\hat{\mathbf{P}}^2}{2m}+V(\hat{\mathbf{X}})\right] = \\ &= \frac{1}{2m\mathrm{i}\hbar}\left[\hat{\mathbf{X}},\hat{\mathbf{P}}^2\right] + \frac{1}{\mathrm{i}\hbar}\left[\hat{\mathbf{X}},V(\hat{\mathbf{X}})\right] = \\ &= \frac{1}{2m\mathrm{i}\hbar}\left(\hat{\mathbf{P}}\left[\hat{\mathbf{X}},\hat{\mathbf{P}}\right]+\left[\hat{\mathbf{X}},\hat{\mathbf{P}}\right]\hat{\mathbf{P}}\right) = \\ &= \frac{1}{2m\mathrm{i}\hbar}\left(\mathrm{i}\hbar\hat{\mathbf{P}}+\mathrm{i}\hbar\hat{\mathbf{P}}\right) = \frac{\hat{\mathbf{P}}}{m}.\end{aligned}$$

Ehrenfest first theorem is analogous to the definition of momentum in classical case:

► $$\frac{\mathrm{d}\hat{\mathbf{X}}}{\mathrm{d}t} = \frac{\hat{\mathbf{P}}}{m}. \tag{2.200}$$

Ehrenfest second theorem

Ehrenfest second theorem concerns the time evolution of the momentum operator. We will proceed in a similar manner to the previous case:

$$\frac{\mathrm{d}\hat{\mathbf{P}}}{\mathrm{d}t} = \frac{1}{\mathrm{i}\hbar}\left[\hat{\mathbf{P}},\hat{\mathbf{H}}\right] = \frac{1}{\mathrm{i}\hbar}\left[\hat{\mathbf{P}},\frac{\hat{\mathbf{P}}^2}{2m}+V(\hat{\mathbf{X}})\right] =$$

$$= \frac{1}{2m\mathrm{i}\hbar}\left[\hat{\mathbf{P}}, \hat{\mathbf{P}}^2\right] + \frac{1}{\mathrm{i}\hbar}\left[\hat{\mathbf{P}},\ V(\hat{\mathbf{X}})\right] =$$

$$= \frac{1}{\mathrm{i}\hbar}\left[\hat{\mathbf{P}},\ V(\hat{\mathbf{X}})\right].$$

We determine the value of the last commutator as follows: First, we find the commutator of the momentum operator with an arbitrary power of the position operator (by induction), and then we apply the result term by term to the potential operator expanded into a power series:

$$\left[\hat{\mathbf{P}}, \hat{\mathbf{X}}\right] = -\left[\hat{\mathbf{X}}, \hat{\mathbf{P}}\right] = -\mathrm{i}\hbar\,\hat{\mathbf{1}},$$

$$\left[\hat{\mathbf{P}}, \hat{\mathbf{X}}^2\right] = \hat{\mathbf{X}}\left[\hat{\mathbf{P}}, \hat{\mathbf{X}}\right] + \left[\hat{\mathbf{P}}, \hat{\mathbf{X}}\right]\hat{\mathbf{X}} = -\mathrm{i}\hbar 2\hat{\mathbf{X}}$$

$$\vdots$$

$$\left[\hat{\mathbf{P}}, \hat{\mathbf{X}}^n\right] = -\mathrm{i}\hbar n\hat{\mathbf{X}}^{n-1},$$

$$\left[\hat{\mathbf{P}}, \hat{\mathbf{X}}^{n+1}\right] = \hat{\mathbf{X}}\left[\hat{\mathbf{P}}, \hat{\mathbf{X}}^n\right] + \left[\hat{\mathbf{P}}, \hat{\mathbf{X}}\right]\hat{\mathbf{X}}^n = -\mathrm{i}\hbar(n+1)\hat{\mathbf{X}}^n,$$

$$\left[\hat{\mathbf{P}}, V(\hat{\mathbf{X}})\right] = -\mathrm{i}\hbar\frac{\partial V}{\partial\hat{\mathbf{X}}}.$$

The fundamental assumption is that the potential energy can be expanded into a Taylor series. Substituting the calculated commutator yields Ehrenfest second theorem:

► $$\frac{\mathrm{d}\hat{\mathbf{P}}}{\mathrm{d}t} = -\frac{\partial V}{\partial\hat{\mathbf{X}}}, \tag{2.201}$$

which is essentially a quantum analogy of Newton equations of motion (the negative gradient of the potential energy is the acting force).

Virial theorem

The Virial theorem is useful not only in quantum theory but also in statistical physics. It determines the average value of the kinetic energy contained in a system from the form of its potential energy. Let us first determine the matrix elements of the commutator of the dynamical variable A with the Hamiltonian operator in the energy representation:

$$\langle n|[\hat{\mathbf{A}},\hat{\mathbf{H}}]|m\rangle =$$

$$= \langle n|\hat{\mathbf{A}}\hat{\mathbf{H}} - \hat{\mathbf{H}}\hat{\mathbf{A}}|m\rangle =$$

$$= (E_m - E_n)\langle n|\hat{\mathbf{A}}|m\rangle =$$

$$= (E_m - E_n)A_{nm}.$$

For $n = m$ we have

$$\langle n|[\hat{\mathbf{A}},\hat{\mathbf{H}}]|n\rangle = 0.$$

Let's choose the product of the coordinate and momentum as the operator for A:

$$\langle n|[\hat{\mathbf{X}}\hat{\mathbf{P}},\hat{\mathbf{H}}]|n\rangle = 0\,,$$

$$\langle n|[\hat{\mathbf{X}},\hat{\mathbf{H}}]\hat{\mathbf{P}}|n\rangle + \langle n|\hat{\mathbf{X}}[\hat{\mathbf{P}},\hat{\mathbf{H}}]|n\rangle = 0\,,$$

$$\langle n|\frac{\mathrm{d}\hat{\mathbf{X}}}{\mathrm{d}t}\,\hat{\mathbf{P}}|n\rangle + \langle n|\hat{\mathbf{X}}\frac{\mathrm{d}\hat{\mathbf{P}}}{\mathrm{d}t}|n\rangle = 0\,.$$

From Ehrenfest theorems, we obtain the time evolution of position and momentum:

$$\langle n|\frac{\hat{\mathbf{P}}^2}{2m}|n\rangle = \langle n|\frac{1}{2}\,\hat{\mathbf{X}}\,\frac{\partial V}{\partial \hat{\mathbf{X}}}|n\rangle\,.$$

In 3D, the result is the sum of the contributions along each axis. On the lhs is the mean value of the system's kinetic energy, and on the rhs is the so-called virial operator:

► $$\langle n|\hat{\mathbf{T}}|n\rangle = \langle n|\hat{\mathscr{V}}|n\rangle\,; \qquad \hat{\mathscr{V}} \equiv \frac{1}{2}\,\hat{\mathbf{X}}_k\,\frac{\partial V}{\partial \hat{\mathbf{X}}_k}\,. \tag{2.202}$$

For a one-dimensional harmonic oscillator, the virial operator has a straightforward interpretation – it is directly equal to the potential energy:

$$V(\hat{\mathbf{X}}) = \frac{1}{2}k\,\hat{\mathbf{X}}^2 \qquad \Rightarrow \qquad \hat{\mathscr{V}} = \frac{1}{2}\,\hat{\mathbf{X}}\,\frac{\partial V}{\partial \hat{\mathbf{X}}} = \frac{1}{2}\,k\,\hat{\mathbf{X}}^2\,.$$

The average values of kinetic and potential energy are therefore equal in every state.

Note: In 1933, Fritz Zwicky pointed out that the motion of galaxies in the Coma Cluster is greater than would be predicted by the virial theorem for gravitation. The solution lies in the existence of additional invisible (dark) matter within this cluster. In 1968, Vera Rubin discovered a similar problem regarding the orbital velocities of stars in the outer regions of spiral galaxies. The solution lies in the existence of a dark matter halo surrounding the galaxy. The virial theorem can therefore be useful even for macroscopic non-quantum systems. Luminous (detected) matter in galaxies accounts for only about 5 %. In 2000, using the Hubble Space Telescope, it was shown that up to 50 percent of the Galaxy's mass may be concentrated in very old and faint white dwarfs that had not been observable until then. They likely belonged to the first generation of stars some 12 billion years ago and fill the entire halo of the Galaxy. The situation is likely similar in other galaxies. White dwarfs are far from solving the dark matter problem. Most likely, it involves an unknown form of non-baryonic matter, which is being sought in many underground laboratories around the world. One example is the Italian laboratory beneath Mount Gran Sasso and its DAMA/Libra experiment; another is the CoGeNT experiment in the Soudan Mine in the US. Measurements from the Planck probe indicate that 27 % of the universe consists of dark matter (out of the total amount of matter and energy). This data is from 2018.

2.7 Relativistic Quantum Theory, Spin

2.7.1 Spatial rotation and the Lorentz transformation

The Lorentz transformation describes the transition between two inertial coordinate systems that move uniformly in a straight line relative to each other. It forms the basis of Albert Einstein special theory of relativity; readers can find the details of its derivation in [1] or in the excellent textbook [21]. The Lorentz transformation is a unitary transformation and belongs to the group of rotations. Therefore, we will first familiarize ourselves with ordinary rotation in three-dimensional space.

Spatial rotation

If we rotate the coordinate system around the z-axis by an angle φ, the transformation can be simply written as (see Section 3.2.3 Rotation in a Plane)

$$\begin{aligned} \tilde{t} &= t\,, \\ \tilde{x} &= x\cos\varphi + y\sin\varphi\,, \\ \tilde{y} &= -x\sin\varphi + y\cos\varphi\,, \\ \tilde{z} &= z\,. \end{aligned} \tag{2.203}$$

The time coordinate is at the zero position; time remains unchanged during spatial rotation. We will describe the transformation using the rotation matrix $\mathbf{R}_z$. Rotations around the other coordinate axes are given by cycling through $x \to y \to z \to x$:

►
$$\mathbf{R}_x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\varphi & \sin\varphi \\ 0 & 0 & -\sin\varphi & \cos\varphi \end{pmatrix}, \quad \mathbf{R}_y = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\varphi & 0 & -\sin\varphi \\ 0 & 0 & 1 & 0 \\ 0 & \sin\varphi & 0 & \cos\varphi \end{pmatrix},$$
$$\mathbf{R}_z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi & 0 \\ 0 & -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.204}$$

Rotation is a unitary transformation. Recall that unitary operators are invariant under scalar multiplication, so the following holds

► $$\mathbf{U}^\dagger\mathbf{U} = \mathbf{1} \;\Rightarrow\; \det\mathbf{U}^\dagger \det\mathbf{U} = 1 \;\Rightarrow\; \left(\det\mathbf{U}\right)^* \left(\det\mathbf{U}\right) \;\Rightarrow\; |\det\mathbf{U}|^2 = 1\,. \tag{2.205}$$

The determinant of all real unitary transformations is either +1 (*rotation*) or –1 (*reflection*). It is easy to verify that the determinant of all three rotation matrices is equal to +1. Rotational symmetry is associated with the conservation of angular momentum. This quantity is defined by the given symmetry (see Noether theorem, Section 1.2.1).

Lorentz transformation

A transformation related to rotations is the Lorentz transformation, which describes the transition between two inertial coordinate systems moving uniformly relative to each other; let us assume along the x-axis (for a derivation, see, for example, [1], [21], [7])):

$$\tilde{t} = \frac{t - vx/c^2}{\sqrt{1-v^2/c^2}} ,$$

► $$\tilde{x} = \frac{x - vt}{\sqrt{1-v^2/c^2}} , \tag{2.206}$$

$$\tilde{y} = y ,$$

$$\tilde{z} = z .$$

This transformation can be written much more elegantly in matrix form. If we introduce relativistic variables $x_0 \equiv ct$, $x_1 \equiv x$, $x_2 \equiv y$, $x_3 \equiv z$ and relativistic coefficients

$$\beta \equiv \frac{v}{c}; \qquad \gamma \equiv \frac{1}{\sqrt{1-\beta^2}} , \tag{2.207}$$

the matrices of the Lorentz transformation (the matrices along the other axes are obtained by cyclic permutation) will have the form

► $$\Lambda_x = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \Lambda_y = \begin{pmatrix} \gamma & 0 & -\gamma\beta & 0 \\ 0 & 1 & 0 & 0 \\ -\gamma\beta & 0 & \gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$
$$\Lambda_z = \begin{pmatrix} \gamma & 0 & 0 & -\gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma\beta & 0 & 0 & \gamma \end{pmatrix}. \tag{2.208}$$

The determinant of the transformation matrices is equal to

$$\det \Lambda = \gamma^2 - \gamma^2\beta^2 = \gamma^2(1-\beta^2) = 1 , \tag{2.209}$$

so these are again rotations, this time in a plane defined by the time axis and one spatial axis. The nature of the rotations becomes clearer if we use a substitution

$$\gamma = \operatorname{ch} u , \qquad \gamma\beta = \operatorname{sh} u , \tag{2.210}$$

where the quantity u is called the *rapidity* and is defined by the relation

► $$u \equiv \operatorname{ath} \frac{v}{c} . \tag{2.211}$$

Using rapidity, the Lorentz transformation takes on an even clearer form

$$\Lambda_x = \begin{pmatrix} \operatorname{ch} u & -\operatorname{sh} u & 0 & 0 \\ -\operatorname{sh} u & \operatorname{ch} u & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Lambda_y = \begin{pmatrix} \operatorname{ch} u & 0 & -\operatorname{sh} u & 0 \\ 0 & 1 & 0 & 0 \\ -\operatorname{sh} u & 0 & \operatorname{ch} u & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

►

$$\Lambda_z = \begin{pmatrix} \operatorname{ch} u & 0 & 0 & -\operatorname{sh} u \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\operatorname{sh} u & 0 & 0 & \operatorname{ch} u \end{pmatrix}.$$

Note that the unit determinant of all these three matrices is now immediately apparent ($\operatorname{ch}^2 u - \operatorname{sh}^2 u = 1$). Lorentz symmetry (the experiment yields the same result in two inertial frames moving uniformly in a straight line relative to each other) is associated with the existence of a new conserved quantity called *spin*.

2.7.2 Spin

The Lorentz transformation plays a role similar to that of spatial rotation. It is also a rotation, but in a plane defined by a time coordinate and one spatial coordinate, by an imaginary angle called the rapidity. Rotational symmetry corresponds to the symmetry of a system with respect to rotation; Lorentz symmetry corresponds to the same behavior of the system in different inertial coordinate systems moving uniformly relative to one another. Both symmetries are associated with corresponding conservation laws:

Rotational symmetry	→	Angular momentum **L**
Lorentz symmetry	→	Spin **S**

Spin has properties very similar to those of angular momentum, but it is very difficult to visualize. It is quite imprecise, yet still illustrative, to imagine a particle orbiting around a center while simultaneously rotating around its own axis. In this classic model, angular momentum corresponds to orbital rotation, and spin corresponds to self-rotation.

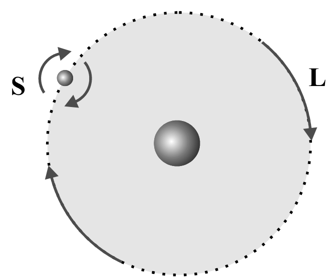


Fig. 2.39: The inaccurate idea of spin as rotation around an axis and angular momentum as circulation around a center

Real particles neither orbit around a center nor rotate around their own axis. Their overall rotational state is determined by two quantities – angular momentum (orbital momentum) and spin (intrinsic momentum). Both quantities can combine; in this case, we speak of spin-orbit interaction, also known as *LS* coupling. The spin operator has the same commutation relations as angular momentum (2.25), (2.27).

$$\blacktriangleright \quad \begin{aligned} &[\hat{\mathbf{S}}_1, \hat{\mathbf{S}}_2] = \mathrm{i}\hbar\hat{\mathbf{S}}_3 \quad + \quad \text{cyclic permutations,} \\ &[\hat{\mathbf{S}}^2, \hat{\mathbf{S}}_3] = 0\,. \end{aligned} \tag{2.212}$$

For spin, we introduce two quantum numbers (just as for angular momentum): 1) the spin number, or spin, which determines the magnitude, and 2) the magnetic spin number m_S, which determines the projection of the spin onto the third axis. For spin, we can derive the relations (2.149) using ladder operators, just as we did for angular momentum.

$$\blacktriangleright \quad \begin{aligned} &|S| = \sqrt{s(s+1)}\,\hbar\,, \qquad && s = 0, 1/2, 1, 3/2, \ldots; \\ &S_3 = m_s\hbar\,, && m_s = -s, -s+1, \ldots, s\,. \end{aligned} \tag{2.213}$$

This time, however, we also consider the half-integer values that we derived earlier for the commutation relations (2.212) and (2.25), respectively. The spin value *s* is an invariant property of elementary particles, just like the electric charge *Q* or the rest mass m_0.

Spin of some particles	
Leptons (electron, tauon, muon, neutrinos)	1/2
Quarks (d, u, s, c, b, t)	1/2
Scalar mesons (pions π, kaons K)	0
Vector mesons (rhoons ρ, kaons K)	1
Hadrons (neutron, proton, Λ hyperon)	1/2
Hadrons (Δ, Ω)	3/2
Field bosons (γ, $W^{\pm}$, Z^0, gluons)	1
Gravitons	2

The presence of spin increases the degree of degeneracy of energy levels. For example, an electron in an atomic shell, whose energy state is determined by the principal quantum number, no longer has a degeneracy of n^2, but $2n^2$. This is because the electron has a spin of 1/2, and its states are determined by the four numbers *n*, *l*, *m*, and m_S. The spin projection m_S can take on two values, ±1/2, and the number of states doubles.

Particles with non-zero spin exhibit a magnetic moment even without having orbital angular momentum. The magnetic properties of particles therefore need not be related solely to rotational motion of the particles, but also to their “intrinsic momentum” – spin. In the presence of a non-homogeneous magnetic field, particles respond to this field. States that originally corresponded to a single energy level split into multiplets of closely spaced energy sublevels. The degree of degeneracy decreases; states with different *m* and m_S values have different energies. We refer to this as the lifting of degeneracy in the presence of a magnetic field.

Spin was first observed in the Stern–Gerlach experiment in 1922. Silver atoms evaporated from a furnace were collimated into a beam passing through a non-uniform magnetic field. A force $\mathbf{F} = -\mu\nabla B$ acts on magnetic moments in the inhomogeneous field (see, e.g. [2]). The magnetic moment of each state is different, and therefore the resulting force and energy of a given state are also different. If spin did not exist, the $l = 0$ state would not split at all ($m = 0$), the $l = 1$ state would split into three different sub-states ($m = 0, \pm 1$), and one or three silver spots would form on the screen. However, two silver spots were observed on the screen, which indicates an electron with an orbital state $l = 0$ and a spin state $s = 1/2$ (the magnetic properties are determined by *two* projections $m_S = \pm 1/2$). An even number of projections implies a half-integer solution to the commutation relations (2.212) and (2.25), respectively. The hypothesis regarding the existence of an intrinsic electron spin, which has properties similar to those of orbital angular momentum, was proposed by Ralph Kronig, George Uhlenbeck, and Samuel Goudsmit in 1925, even before the theoretical explanation of spin. Spin, as a consequence of Lorentz symmetry, was theoretically explained by Paul Dirac in 1927.

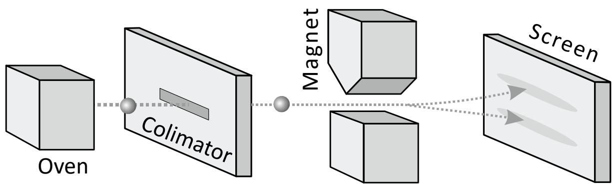


Fig. 2.40: Schematic diagram of the Stern-Gerlach experiment

Isospin

In the mid-20th century, a large number of elementary particles were known. Their research revealed that some behave almost identically and form a kind of related group, or multiplet. Examples include the doublet of the neutron and proton or the quadruplet of Δ particles. Heisenberg had already noticed the similar properties of the neutron and proton under strong interactions as early as 1932, and it occurred to him that it might be possible to understand these particles as two different quantum states of a single particle, the nucleon. The term *isospin* was proposed by Eugene Wigner in 1937; it was an abbreviation of the words *isotopic spin*. The concept worked similarly to spin. The neutron and proton are two particles, so their isospin $I = ½$, and the particles differ in the third projection of isospin I_3, which can take the values +½ (proton) or −½ (neutron).

The introduction of isospin foreshadowed the discovery of the internal structure not only of the neutron and proton, but also of particles in other multiplets. It turned out that they are composed of “d” and “u” quarks, which behave very similarly under the strong interaction. This is why multiplets arise, whose particles have similar internal compositions. Just as a particle with spin s has $2s+1$ possible projections, a multiplet with isospin I has a total of $2I+1$ members; that is, a multiplet with N members has isospin

$$I = \frac{N-1}{2}. \tag{2.214}$$

The third component is then determined by the internal structure of the particle, specifically by the number of "d" and "u" quarks. Each "u" quark contributes a value of +½ to the isospin projection, and each "d" quark contributes a value of −½; the antiquarks, of course, contribute the opposite values:

► $$I_3 = \tfrac{1}{2}N_\mathrm{u} - \tfrac{1}{2}N_\mathrm{d} - \tfrac{1}{2}N_{\bar{\mathrm{u}}} + \tfrac{1}{2}N_{\bar{\mathrm{d}}} \tag{2.215}$$

The so-called *Gell-Mann–Nishijima formula* is very useful; it can be easily derived from the last equation if we know the charges of the individual quarks:

► $$I_3 = Q - \tfrac{1}{2} Y\,, \tag{2.216}$$

where Y is the so-called hypercharge (the average electric charge of the multiplet). This is best illustrated in the following figure.

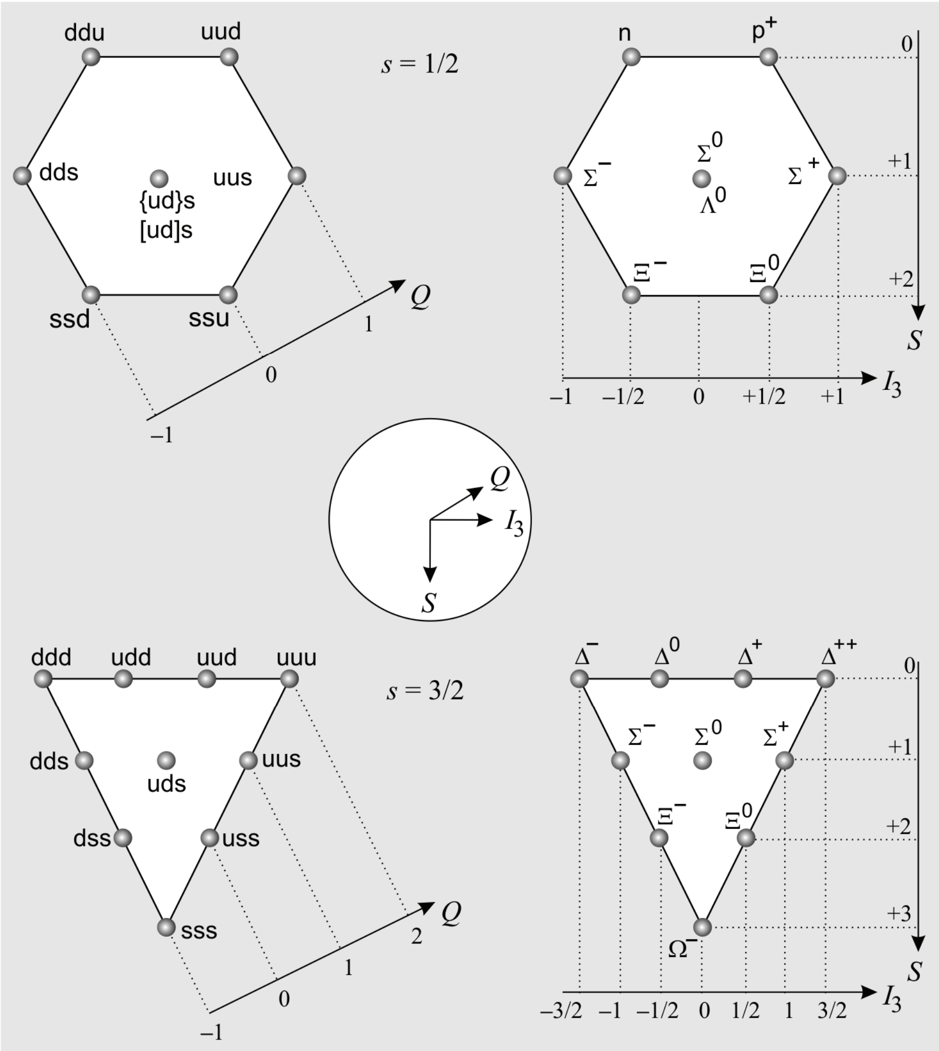


Fig. 2.41: An illustration of the formation of multiplets in baryons (particles composed of three quarks). In the lower part, all three quarks have the same spin orientation, so the resulting spin of the particles is 3/2. In the upper part, the baryons have a resulting spin of 1/2 (one quark has a spin orientation opposite). On the left is the quark structure of the particles; on the right are the names of the individual particles. The electric charge increases diagonally upward to the right; the number of strange quarks (strangeness S) increases downward; and the isospin projection in the multiplet increases to the right. All particles of a multiplet are at the same height. The diagrams show from top to bottom: the nucleon doublet (neutron and proton), the Σ (sigma) triplet with spin 1/2, the Ξ (xi) doublet with spin 1/2, the Δ (delta) quadruplet, the Σ triplet with spin 3/2, the Ξ doublet with spin 3/2, and the singlet particle Ω (omega).

2.7.3 Klein-Gordon Equation

Schrödinger equation is not relativistic and therefore cannot correctly describe spin. In deriving it, we used the non-relativistic form of the Hamiltonian. The result was Schrödinger time equation (2.182) , which takes the following form in x-representation

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t}=\Big(-\frac{\hbar^2}{2m}\nabla^2+V\Big)\psi=0\,.$$

The equation contains first-order time derivatives and second-order spatial derivatives; time and space are not treated equally, so the equation is clearly not relativistic. A relativistic formulation can be constructed in both second (Klein-Gordon equation) and first (Dirac equation) derivatives. In this chapter, we will focus on constructing the correct equation in second derivatives.

Klein-Gordon equation

The equation for the wave function of a particle in second derivatives was derived by Oskar Klein and Walter Gordon. We are looking for a linear equation that, in the limit of small velocities, reduces to the Schrödinger equation. From the principle of superposition, we know that every solution can be expressed as a superposition of plane waves

$$\psi_k(x)=a(k)\,\mathrm{e}^{\mathrm{i}[k\cdot x]}=a(k)\,\mathrm{e}^{\mathrm{i}[k^\alpha x_\alpha]}=a(\omega,\mathbf{k})\,\mathrm{e}^{\mathrm{i}[\mathbf{k}\cdot\mathbf{x}-\omega t]}\,. \tag{2.217}$$

3D vectors are shown in bold. The components of the wave vector k^α must necessarily be dependent, since even the partial waves (2.217) must satisfy the desired equation. Such a dependence is called the dispersion relation, and we can write it in implicit form

$$\phi(\omega,\mathbf{k})=0\,. \tag{2.218}$$

In some cases, an explicit relationship $\omega=\omega(\mathbf{k})$ can be found. The general wave function will be a superposition

$$\psi(x)=\int a(k)\,\mathrm{e}^{\mathrm{i}[k\cdot x]}\,\delta(\phi)\,\mathrm{d}^4k=\int a(\omega,\mathbf{k})\,\mathrm{e}^{\mathrm{i}[\mathbf{k}\cdot\mathbf{x}-\omega(\mathbf{k})t]}\,\mathrm{d}^3\mathbf{k}\,. \tag{2.219}$$

The Dirac distribution ensures that the dispersion relation (2.218) is automatically satisfied. Partial (plane) waves can be easily derived:

$$\partial^\alpha\psi_k(x)=\mathrm{i}\,k^\alpha\psi_k(x) \tag{2.220}$$

and the partial derivatives correspond to algebraic expressions

$$\partial^\alpha \quad\leftrightarrow\quad \mathrm{i}\,k^\alpha \tag{2.221}$$

Using the duality (2.4) we have

$$\hbar\partial^\alpha \quad\leftrightarrow\quad \mathrm{i}\,p^\alpha\,. \tag{2.222}$$

The most natural transition from a commutative to a non-commutative description is therefore the introduction of operators on $\mathcal{L}^2$ via the following definition

$$\begin{aligned}\hat{p}^\alpha&\equiv-\mathrm{i}\,\hbar\partial^\alpha\,;\\ \hat{x}^\alpha&\equiv x^\alpha\,.\end{aligned}\,. \tag{2.223}$$

It is easy to see that the operators defined in this way satisfy commutation relations that are consistent with the correspondence principle

$$\begin{aligned}\left[\hat{p}^{\alpha},\hat{p}^{\alpha}\right]&=\left[\hat{x}^{\alpha},\hat{x}^{\alpha}\right]=0\,,\\ \left[\hat{x}^{\alpha},\hat{p}^{\beta}\right]&=\mathrm{i}\hbar\,g^{\alpha\beta}\,.\end{aligned}\tag{2.224}$$

In (3+1)-dimensional formalism, the first of the relations (2.223)can be written as

$$\begin{aligned}\hat{E}&\equiv+\mathrm{i}\hbar\partial/\partial t\,,\\ \hat{\mathbf{p}}&\equiv-\mathrm{i}\hbar\partial/\partial\mathbf{x}\,.\end{aligned}\tag{2.225}$$

The different sign in the time variable is related to the relativistic transformation properties of four-vectors. We have already used the second relation in the x-representation of the momentum operator; see (2.48). The magnitude of the four-momentum will be

$$p^{\alpha}\equiv\begin{pmatrix}E/c\\ \mathbf{p}\end{pmatrix}\qquad\Rightarrow\qquad p_{\alpha}p^{\alpha}=-\frac{E^2}{c^2}+\mathbf{p}^2.\tag{2.226}$$

This value must be the same in all coordinate systems, and we can determine it in the particle's rest frame, where $E=m_0c^2$, $\mathbf{p}=0$:

$$p_{\alpha}p^{\alpha}=-m_0^2c^2\,.\tag{2.227}$$

In the (3+1) formalism, this refers to the well-known Pythagorean theorem for energy

$$E^2=\mathbf{p}^2c^2+m_0^2c^4.$$

This relation is the correct relativistic relation for the energy of a free particle, and therefore the derivation of the relativistic version of the Schrödinger equation must be based on it. Let us therefore rewrite (2.227) in operator form:

$$\left(\hat{p}_{\alpha}\hat{p}^{\alpha}+m_0^2c^2\right)\psi=0\,;\qquad \hat{p}^{\alpha}\equiv-\mathrm{i}\hbar\partial^{\alpha}\,.\tag{2.228}$$

Equation (2.228) is the Klein–Gordon equation for a free particle. Substituting for the operators yields another commonly used form of the Klein–Gordon equation

► $$\left(\Box-\kappa^2\right)\psi=0\,;\qquad \kappa\equiv\frac{m_0c}{\hbar}\,.\tag{2.229}$$

The Klein–Gordon equation is the relativistic analogue of the Schrödinger equation for a free particle. At low velocities, it asymptotically reduces to the non-relativistic Schrödinger equation. It is a linear equation, and every “reasonable” solution can be expressed using the Fourier transform as a superposition of plane waves. In the natural unit system (c = 1, $\hbar$ = 1), the constant κ is equal to the rest mass of the particle.

Non-relativistic limit

We can write Klein–Gordon equation (2.228) in operator form as

$$\hat{E}^2=\hat{\mathbf{p}}^2c^2+m_0^2c^4\hat{\mathbf{1}}\,.\tag{2.230}$$

We take the square root of both sides. We interpret the square root as a function of the operator in the sense of (3.263) or (3.296):

$$\hat{E} = \pm\sqrt{\hat{\mathbf{p}}^2c^2 + m_0^2c^4\hat{\mathbf{1}}} = m_0c^2\sqrt{\hat{\mathbf{1}} + \frac{\hat{\mathbf{p}}^2}{m_0^2c^2}} \approx$$

$$\approx m_0c^2\left(\hat{\mathbf{1}} + \frac{\hat{\mathbf{p}}^2}{2m_0^2c^2} + \cdots\right) \qquad \Rightarrow$$

$$\hat{E} \approx m_0c^2\hat{\mathbf{1}} + \frac{\hat{\mathbf{p}}^2}{2m_0}.$$

We have omitted the negative sign in front of the square root for now, as it is not physically meaningful, and we will address it in the chapter on the Dirac equation. The first term can be interpreted as constant/zero potential energy (shifting by a constant does not change the potential energy), and the second is the standard kinetic energy of the particle. Substituting for the operators from (2.225), we obtain the time-dependent Schrödinger equation (2.182) with zero, or constant, potential energy. For small velocities (momenta), the Klein–Gordon equation reduces to the Schrödinger equation.

Probabilistic interpretation

The density ρ and the probability flux $\mathbf{j}$ of particle occurrence should satisfy the continuity equation (the law of conservation of particle occurrence probability) in the form

$$\partial_\alpha j^\alpha = 0\,; \qquad j^\alpha \equiv \begin{pmatrix} \rho c \\ \mathbf{j} \end{pmatrix}. \tag{2.231}$$

Let us show that such a conservation law is contained in the Klein–Gordon equation. Let us find the combination ψ*(2.229)–ψ(2.229)*:

$$\psi^*\left(\Box - \kappa^2\right)\psi \;-\; \psi\left(\Box - \kappa^2\right)\psi^* = 0 \quad \Rightarrow$$

$$\psi^*\Box\psi \;-\; \psi\Box\psi^* = 0 \quad \Rightarrow$$

$$\psi^*\partial_\alpha\partial^\alpha\psi - \psi\partial_\alpha\partial^\alpha\psi^* = 0\,.$$

Now, in both expressions, we will use the identity $f\partial_\alpha g = \partial_\alpha(fg) - (\partial_\alpha f)g$:

$$\partial_\alpha\left(\psi^*\partial^\alpha\psi\right) - \left(\partial_\alpha\psi^*\right)\left(\partial^\alpha\psi\right) - \partial_\alpha\left(\psi\partial^\alpha\psi^*\right) + \left(\partial_\alpha\psi\right)\left(\partial^\alpha\psi^*\right) = 0\,.$$

If we increase the first index and decrease the second in the last term of the expression, it cancels out with the second term, and we get:

$$\partial_\alpha\left(\psi^*\partial^\alpha\psi\right) - \partial_\alpha\left(\psi\partial^\alpha\psi^*\right) = 0 \quad \Rightarrow$$

► $$\partial_\alpha j^\alpha = 0\,; \qquad j^\alpha \equiv \psi^*\partial^\alpha\psi - \psi\partial^\alpha\psi^*\,. \tag{2.232}$$

The four-vector j^α represents the unnormalized probability of a particle's occurrence. Unfortunately, the probability density j^0 (in SI units, j^0/c) is not positive definite, and the Klein–Gordon equation allows for negative probability densities. We will address this issue (which leads to the existence of antiparticles) in the chapter on the Dirac equation.

Dispersion relation

Substituting the plane wave (2.217) into the Klein–Gordon equation yields the dispersion relation

$$\omega^2 = c^2k^2 + c^2\kappa^2 \qquad \Rightarrow \qquad \omega = \pm\sqrt{c^2k^2 + c^2\kappa^2}\ . \tag{2.233}$$

We will again consider the negative solution (with negative energy $\hbar\omega$) to be non-physical. We will determine the phase and group velocities using the standard procedure:

$$\begin{aligned} v_\mathrm{f} &= \frac{\omega}{k} = c\sqrt{1+\frac{\kappa^2}{k^2}} = c\sqrt{1+\frac{\kappa^2\lambda^2}{4\pi^2}}\ , \\ v_\mathrm{g} &= \frac{\partial\omega}{\partial k} = \frac{c}{\sqrt{1+\dfrac{\kappa^2}{k^2}}} = \frac{c}{\sqrt{1+\dfrac{\kappa^2\lambda^2}{4\pi^2}}}\ . \end{aligned}$$

At first glance, it is clear that the group velocity is always subluminal. From Hamilton equations of mechanics

$$v_\mathrm{g} = \frac{\partial\omega}{\partial k} = \frac{\partial\hbar\omega}{\partial\hbar k} = \frac{\partial H}{\partial p} = \dot{x}$$

it follows that the group velocity of a wave packet is analogous to the mechanical velocity of a moving particle. In contrast, the phase velocity is always superluminal and has no significance for the transmission of information. There is a simple relationship between the two velocities: $v_\mathrm{f}v_\mathrm{g} = c^2$. Both velocities depend on the wavelength of the partial wave, i.e., dispersion occurs.

The Klein–Gordon equation for a charged particle in an electromagnetic field

In the presence of electromagnetic field, the canonical (generalized) momentum $\mathbf{p} - Q\mathbf{A}$ appeared in the Hamiltonian (1.50). The same must hold for the Klein–Gordon equation (2.228), which takes the form for a charged particle in an electromagnetic field:

$$\left[\left(\hat{p}_\alpha - QA_\alpha\right)\left(\hat{p}^\alpha - QA^\alpha\right) + m_0^2c_0^2\right]\psi = 0\,; \qquad \hat{p}^\alpha \equiv -\mathrm{i}\hbar\partial^\alpha\ . \tag{2.234}$$

After substituting for the momentum operator and multiplying all terms, we have

$$-\hbar^2\Box\psi + Q^2A_\alpha A^\alpha\psi + m_0^2c^2\psi + \mathrm{i}\hbar Q\partial_\alpha A^\alpha\psi + 2\mathrm{i}\hbar QA^\alpha\partial_\alpha\psi = 0\,. \tag{2.235}$$

If we use the calibration condition (1.256), i.e., if we set $\partial_\alpha A^\alpha = 0$, we obtain the resulting equation

$$► \quad \begin{aligned} &\left[\Box - \kappa^2 - \frac{Q^2}{\hbar^2}A_\alpha A^\alpha - 2\mathrm{i}\frac{Q}{\hbar}A^\alpha\partial_\alpha\right]\psi = 0\,; \\ &\kappa \equiv \frac{m_0c}{\hbar} \end{aligned} \tag{2.236}$$

for describing a charged particle in an electromagnetic field.

Hydrogen atom

Let us now outline how to find the spectrum of the hydrogen atom using the Klein–Gordon equation. An electron with charge $Q = -e$ is in the field of the nucleus, which can be expressed by the following relations (for hydrogen a number of electrons $Z = 1$)

$$A^0 = \frac{\phi}{c} = +\frac{Ze}{4\pi\varepsilon_0 rc}\,, \qquad \mathbf{A} = 0\,. \tag{2.237}$$

The Klein–Gordon equation takes the form

$$\left[\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \kappa^2 + \left(\frac{e\phi}{\hbar c}\right)^2 + 2\mathrm{i}\frac{e\phi}{\hbar c}\frac{\partial}{c\partial t}\right]\psi(t,r,\theta,\varphi) = 0\,.$$

We decompose the Laplace operator into radial and angular parts, just as in the non-relativistic case (2.153). We will seek a stationary solution; i.e., we will assume the time component of the wave function to be $\exp(-\mathrm{i}\omega t) = \exp(-\mathrm{i}Et/\hbar)$, and we will write the spatial component as the product of the radial and angular parts as earlier, see (2.162):

$$\left[\nabla_r^2 - \frac{\hat{\mathbf{L}}^2}{\hbar^2 r^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \kappa^2 + \frac{Z^2\alpha^2}{r^2} + 2\mathrm{i}\frac{Z\alpha}{r}\frac{\partial}{c\partial t}\right]\mathrm{e}^{-\frac{\mathrm{i}}{\hbar}Et}R(r)Y_{lm}(\theta,\varphi) = 0\,,$$

where we have denoted

$$\alpha \equiv \frac{e^2}{4\pi\varepsilon_0\hbar c} \tag{2.238}$$

the so-called fine-structure constant. After taking the time derivatives, we obtain

$$\left[\nabla_r^2 - \frac{\hat{\mathbf{L}}^2}{\hbar^2 r^2} + \frac{E^2 - m_0^2c^4}{\hbar^2c^2} + \left(\frac{Z\alpha}{r}\right)^2 + 2\frac{Z\alpha}{r}\frac{E}{\hbar c}\right]R(r)Y_{lm}(\theta,\varphi) = 0\,.$$

Now we apply the operator $\mathbf{L}^2$ to the angular part of the wave function according to equation (2.148) and express the radial part of the Laplace operator

$$\left[\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}r^2\frac{\mathrm{d}}{\mathrm{d}r} - \frac{l(l+1) - Z^2\alpha^2}{r^2} + 2\frac{Z\alpha}{r}\frac{E}{\hbar c} + \frac{E^2 - m_0^2c^4}{\hbar^2c^2}\right]R(r) = 0\,.$$

This is an ordinary differential equation that is solved using standard methods (asymptotic behavior, series expansion, truncation). The results are the so-called Laguerre polynomials and the energy spectrum

► $$E_{nl} = m_0c^2 - m_0c^2\left[\frac{Z^2\alpha^2}{2n^2} + \frac{Z^4\alpha^4}{(2l+1)n^4}\left(n - \frac{3}{8}(2l+1)\right) + \mathcal{O}\left((Z\alpha)^6\right)\right]. \tag{2.239}$$

The principal quantum number is defined in the same way as in the non-relativistic case; the second term in square brackets represents the first relativistic correction and, at the same time, the removal of spectral line degeneracy.

Problems

The Klein–Gordon equation has three fundamental problems:

1. The second time derivatives require initial conditions not only for the wave function (which represents the state of the system) but also for the first time derivative of the wave function, which is difficult to interpret physically.
2. The probability density is not positive definite.
3. The Klein-Gordon equation also yields negative energy states.

2.7.4 Dirac Equation

The correct relativistic quantum equation for a charged particle, containing only first derivatives, was derived by Paul Adrien Maurice Dirac (1902–1984) in 1928. It turned out to be a much more suitable equation for the electron than the Klein–Gordon equation. Since the equation involves only first derivatives, it suffices to specify the initial value of the wave function, and the need to specify the first derivative of the wave function is automatically eliminated. In the Dirac equation, the probability density is positive definite, thus eliminating the second fundamental problem of the Klein–Gordon equation. However, the problem of negative energy states persists, and Dirac interpreted these states as belonging to the antiparticle of the electron – the positron. The positron was not discovered until 1932 by Carl Anderson.

Dirac equation

Let us find an equation that has the same form as the Schrödinger equation, but in which the Hamiltonian operator is a linear function of spatial derivatives:

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t}=\hat{H}\psi\,;$$
$$\hat{H}\equiv a^1\partial_1+a^2\partial_2+a^3\partial_3+b\,. \tag{2.240}$$

For dimensional reasons, instead of the coefficients a^k and b, we will seek the dimensionless coefficients α^k and β, which are dimensionless:

$$\hat{H}\equiv-\mathrm{i}\,\hbar c\left(\alpha^1\partial_1+\alpha^2\partial_2+\alpha^3\partial_3\right)+\beta m_0c^2\,. \tag{2.241}$$

There are two basic conditions for the coefficients:

1. The square of the Hamiltonian must yield the rhs of equation (2.230), i.e.,

$$\hat{H}^2=\hat{\mathbf{p}}^2c^2+m_0^2c^4\,\hat{\mathbf{1}}\,. \tag{2.242}$$

   Thus, every solution to the Dirac equation will be a solution to the Klein–Gordon equation (but not vice versa; the second derivatives add some solutions).
2. The new equation must be relativistically covariant (i.e., its form must not change after applying a Lorentz transformation to the coordinates and fields).

We will soon see that no scalar coefficients satisfy these conditions, and that the sought-after quantities α^k and β must be matrices. Let us start with condition (2.242), into which we substitute the Hamiltonian (2.241) and the momentum operator from (2.225):

$$\hat{H}^2 = \hat{\mathbf{p}}^2 c^2 + m_0^2 c^4 \hat{\mathbf{1}}\,;$$

$$\left(-\mathrm{i}\hbar c\alpha^k \partial_k + \beta m_0 c^2\right)\left(-\mathrm{i}\hbar c\alpha^l \partial_l + \beta m_0 c^2\right) = \hat{\mathbf{p}}^2 c^2 + m_0^2 c^4 \hat{\mathbf{1}}\,;$$

$$-\hbar^2 c^2 \alpha^k \alpha^l \partial_k \partial_l - \mathrm{i}\hbar c m_0 c^2 \left(\alpha^k \beta + \beta \alpha^k\right)\partial_k + \beta^2 m_0^2 c^4 = -\hbar^2 c^2 \nabla^2 + m_0^2 c^4 \hat{\mathbf{1}}$$

By comparing the terms on the left and right sides, we obtain the following coefficients:

$$\alpha^k \alpha^l \partial_k \partial_l = \nabla^2\,,$$
$$\alpha^k \beta + \beta \alpha^k = 0\,, \tag{2.243}$$
$$\beta^2 = \hat{\mathbf{1}}\,.$$

We can easily rewrite the first expression as

$$\frac{1}{2}\left(\alpha^k \alpha^l + \alpha^l \alpha^k\right)\partial_k \partial_l = \nabla^2 \quad \Rightarrow$$
$$\alpha^k \alpha^l + \alpha^l \alpha^k = \begin{cases} 0 & \text{pro } k \neq l\,, \\ 2 & \text{pro } k = l\,. \end{cases} \tag{2.244}$$

No real or complex numbers satisfy the conditions (2.243) or (2.244). We will therefore look for a system of four matrices, whose interesting properties we will first list clearly and then prove

1. The matrices α^k and β are anticommutative (any two of them):

$$\left\{\alpha^k, \alpha^l\right\} = \left\{\alpha^k, \beta\right\} = 0\,; \quad k \neq l\,. \tag{2.245}$$

2. The squares of the matrices α^k and β yield the identity matrix:

$$\left(\alpha^1\right)^2 = \left(\alpha^2\right)^2 = \left(\alpha^3\right)^2 = \left(\beta\right)^2 = \hat{\mathbf{1}}\,. \tag{2.246}$$

3. The matrices α^k and β are Hermitian:

$$\left(\alpha^k\right)^\dagger = \alpha^k\,, \qquad \beta^\dagger = \beta\,. \tag{2.247}$$

4. The eigenvalues of the matrices α^k and β can only take the values +1 and −1.

5. The trace of the matrices α^k and β is zero.

6. The matrices α^k and β are independent.

Let's now prove each of these statements

Re 1. The anticommutation relations for α^k and β follow immediately from relations (2.243) and (2.244). Note that the anticommutator is defined as $\{A, B\} \equiv AB + BA$.

Re 2. Once again, this statement follows directly from relations (2.243) and (2.244). It is clear that the squares of all Dirac matrices yield the identity matrix.

Re 3. The Hermitian nature of the matrices α^k and β follows from the requirement that the energy operator be Hermitian. Dirac matrices are therefore Hermitian.

Re 4. It follows from condition (2.246) that the eigenvalues of the matrices α^k and β lie on the unit circle in the complex plane, i.e., $|\lambda| = 1$. However, Hermitian matrices have real eigenvalues, so only the values $\lambda = \pm 1$ are possible.

Re 5. The trace of a matrix is defined as the sum of its diagonal elements

$$\mathrm{Tr}(\mathbf{A}) = A^k{}_k \,. \tag{2.248}$$

The trace of a matrix does not change under matrix cyclic permutation:

$$\mathrm{Tr}(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_N) = \mathrm{Tr}(\mathbf{A}_2\cdots\mathbf{A}_N\mathbf{A}_1)\,, \tag{2.249}$$

That is, we can move the first matrix to the last position in the product (or the last one to the first). Now we can easily prove that the trace of the matrices in question is zero:

$$\mathrm{Tr}\left(\alpha^k\right) = \mathrm{Tr}\left(\beta^2\alpha^k\right) = \mathrm{Tr}\left(\beta\beta\alpha^k\right) =$$

$$\mathrm{Tr}\left(\beta\alpha^k\beta\right) = -\mathrm{Tr}\left(\beta\beta\alpha^k\right) = -\mathrm{Tr}\left(\alpha^k\right).$$

First, we added β^2, which is, however, the identity matrix. Then we moved one β matrix to the end using a cyclic permutation and returned it to its original position using the anticommutativity of α^k and β. If we read the beginning and the end, we have

$$\mathrm{Tr}\left(\alpha^k\right) = -\mathrm{Tr}\left(\alpha^k\right) \quad \Rightarrow \quad 2\,\mathrm{Tr}\left(\alpha^k\right) = 0 \quad \Rightarrow$$

$$\mathrm{Tr}\left(\alpha^k\right) = 0\,.$$

We can proceed in a similar procedure for the matrix β:

$$\mathrm{Tr}\left(\beta\right) = \mathrm{Tr}\left((\alpha^k)^2\beta\right) = \mathrm{Tr}\left(\alpha^k\beta\alpha^k\right) = -\mathrm{Tr}\left(\alpha^k\alpha^k\beta\right) = -\mathrm{Tr}\left(\beta\right) \quad \Rightarrow$$

$$\mathrm{Tr}\left(\beta\right) = 0\,.$$

Ad 6. Let us assume that the matrices are dependent; that is, for example, the matrix β can be expressed as a linear combination of the others:

$$\beta = \sum c_k\alpha^k \,.$$

Let's multiply the relation from the left by the matrix β:

$$\mathbf{1} = \sum c_k\beta\alpha^k \quad \Rightarrow$$

$$\mathrm{Tr}(\mathbf{1}) = \sum c_k\,\mathrm{Tr}(\beta\alpha^k) = \frac{1}{2}\sum c_k\,\mathrm{Tr}(\beta\alpha^k + \alpha^k\beta) = \frac{1}{2}\sum c_k\,\mathrm{Tr}(0) = 0\,.$$

This is a contradiction, since the trace of the left unit matrix is nonzero. Therefore, the matrices must be independent. This proves all the statements (1 through 6).

■

The trace is an invariant – it remains the same in all bases. If we choose the eigenvectors of a Hermitian matrix as a basis, the matrix will be diagonal, and its eigenvalues will lie on the diagonal. The trace is therefore the sum of its eigenvalues. In our case, the eigenvalues are +1 or −1, the trace of the matrix is zero, and therefore the matrices must have an even dimension (so that the sum of +1 and −1 can equal zero).

Solution for *N* = 2

In the chapter on angular momentum, we derived spinor representation (2.150). Spin matrices without the associated coefficients are called Pauli matrices:

$$\sigma^1 = \begin{pmatrix} 0 & +1 \\ +1 & 0 \end{pmatrix}; \quad \sigma^2 = \begin{pmatrix} 0 & -\mathrm{i} \\ +\mathrm{i} & 0 \end{pmatrix}; \quad \sigma^3 = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.250}$$

Pauli matrices have the properties we are looking for. They are Hermitian, anticommute with each other, their squares are identity matrices, their eigenvalues are ±1 and the sum of the diagonal elements is zero. But there are only three of them. We are looking for a system of four independent anticommuting matrices. However, such a system does not exist in two dimensions. Identity matrix is independent, but commutes with them, not anticommutes. Furthermore, the sum of its diagonal entries is not zero.

Solution for *N* = 4

In four dimensions, there are a total of 16 independent matrices, and it is indeed possible to select 4 anticommuting matrices with the desired properties from among them. This is the smallest number of dimensions in which Dirac problem can be solved. There are several ways to select the desired set of anticommuting matrices. Dirac assembled them block by block from Pauli matrices and found a solution

$$\beta = \sigma^3 \otimes \mathbf{1} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix}; \quad \alpha^k = \sigma^1 \otimes \sigma^k = \begin{pmatrix} \mathbf{0} & \sigma^k \\ \sigma^k & \mathbf{0} \end{pmatrix}. \tag{2.251}$$

Each element of the matrix represents a 2×2 block. The resulting Dirac matrices are:

$$\begin{aligned} \beta &= \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}; \quad \alpha^1 = \begin{pmatrix} 0 & 0 & 0 & +1 \\ 0 & 0 & +1 & 0 \\ 0 & +1 & 0 & 0 \\ +1 & 0 & 0 & 0 \end{pmatrix}; \\ \alpha^2 &= \begin{pmatrix} 0 & 0 & 0 & -\mathrm{i} \\ 0 & 0 & +\mathrm{i} & 0 \\ 0 & -\mathrm{i} & 0 & 0 \\ +\mathrm{i} & 0 & 0 & 0 \end{pmatrix}; \quad \alpha^3 = \begin{pmatrix} 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & -1 \\ +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}. \end{aligned} \tag{2.252}$$

All matrices are Hermitian, have eigenvalues of ±1, the sum of the elements on the diagonal is 0, their squares yield the identity matrix, and each matrix anticommutes with every other matrix. The Dirac equation for a free particle now takes a simple form:

$$\mathrm{i}\hbar \frac{\partial \psi}{\partial t} = \left(-\mathrm{i}\hbar c \alpha^k \partial_k + m_0 c^2 \beta\right)\psi; \quad \psi = \begin{pmatrix} \psi_1(t,\mathbf{x}) \\ \psi_2(t,\mathbf{x}) \\ \psi_3(t,\mathbf{x}) \\ \psi_4(t,\mathbf{x}) \end{pmatrix}. \tag{2.253}$$

The coefficients of the equation are 4×4 matrices; therefore, the wave function consists of a set of four functions (not a four-vector!). A different choice of four Dirac matrices would lead to the same physical solution.

Velocity operator, negative energies

Let us define the particle velocity operator as the operator of the time evolution of position according to the correspondence principle (2.179):

$$\hat{\upsilon}^k = \frac{\mathrm{d}x^k}{\mathrm{d}t} = \frac{1}{\mathrm{i}\hbar}\left[x^k, \hat{H}\right] = \frac{1}{\mathrm{i}\hbar}\left[x^k, -\mathrm{i}\hbar c\alpha^l\partial_l + m_0c^2\beta\right] =$$

$$= \frac{1}{\mathrm{i}\hbar}\left[x^k, -\mathrm{i}\hbar c\alpha^l\partial_l\right] = \frac{c}{\mathrm{i}\hbar}\alpha^l\left[x^k, -\mathrm{i}\hbar\partial_l\right] = \frac{c}{\mathrm{i}\hbar}\alpha^l\left[x^k, p_l\right] =$$

$$= \frac{c}{\mathrm{i}\hbar}\alpha^l\,\mathrm{i}\hbar\delta^k{}_l = c\alpha^k\ .$$

The matrices α^k thus represent (up to a constant c) the velocity operator:

$$\hat{\upsilon}^k = c\alpha^k\ . \tag{2.254}$$

Formally, all three relations can be written together

$$\hat{\vec{\mathbf{v}}} = c\hat{\vec{\boldsymbol{\alpha}}}\ . \tag{2.255}$$

Using the velocity and momentum operators, the Dirac equation (2.253) takes on a simple form:

$$\begin{aligned} \mathrm{i}\hbar\frac{\partial\psi}{\partial t} &= \left(\hat{\vec{\mathbf{v}}}\cdot\hat{\vec{\mathbf{p}}} + m_0c^2\beta\right)\psi\ ; \\ \hat{\vec{\mathbf{v}}} \equiv c\hat{\vec{\boldsymbol{\alpha}}}\ , &\qquad \hat{\vec{\mathbf{p}}} \equiv -\mathrm{i}\hbar\vec{\nabla}\ . \end{aligned} \tag{2.256}$$

Let us now solve Dirac equation for a particle at rest, i.e., with a zero velocity operator

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\begin{pmatrix}\psi_1\\ \psi_2\\ \psi_3\\ \psi_4\end{pmatrix} = m_0c^2\begin{pmatrix}+\psi_1\\ +\psi_2\\ -\psi_3\\ -\psi_4\end{pmatrix}.$$

The solution is:

$$\psi_1(t,\mathbf{x}) = A_1(\mathbf{x})\exp\left[-\mathrm{i}\frac{m_0c^2}{\hbar}t\right],$$

$$\psi_2(t,\mathbf{x}) = A_2(\mathbf{x})\exp\left[-\mathrm{i}\frac{m_0c^2}{\hbar}t\right],$$

$$\psi_3(t,\mathbf{x}) = A_3(\mathbf{x})\exp\left[+\mathrm{i}\frac{m_0c^2}{\hbar}t\right],$$

$$\psi_4(t,\mathbf{x}) = A_4(\mathbf{x})\exp\left[+\mathrm{i}\frac{m_0c^2}{\hbar}t\right].$$

Usually the time component of a plane wave is $\exp[-\mathrm{i}\omega t] = \exp[-\mathrm{i}(E/\hbar)t]$. It is clear that the second two solutions correspond to negative energy $E = -\,m_0c^2$. Thus, Dirac equation did not solve the problem of negative energy states.

It turned out that the Dirac equation describes the behavior of particles with spin 1/2 (such as the electron). The quadruple ψ is called a *bispinor*. It has special transformation properties. The top two components describe the states of a particle with spin projections +1/2 and −1/2 and have positive energy. The lower two components have negative energy, and Dirac interpreted them as states of *antiparticles* with spin projections of +1/2 and −1/2. In a certain sense, the Dirac equation is the "square root" of the Klein–Gordon equation based on the relation $E^2 = p^2c^2 + m_0{}^2c^4$. Therefore, states with negative energy are not surprising. What was elegant, however, was Dirac explanation: All negative-energy states are occupied (Dirac "sea of electrons" with negative energy, see Section 2.7.5) and behave like a "hole," which Dirac interpreted as a positive-energy antiparticle. Dirac performed this analysis in 1928 and theoretically predicted the existence of the positron even before its experimental discovery in 1932 (Carl Anderson).

Probabilistic interpretation

When deriving the continuity equation for probability, we will proceed in the same way as for the Klein-Gordon equation, except that instead of complex conjugation, we will use Hermitian conjugation of the individual matrices and of the basis bispinor that forms the wave function. The Hermitian-conjugate bispinor has the form

$$\psi^\dagger = \begin{pmatrix} \psi_1^* & \psi_2^* & \psi_3^* & \psi_4^* \end{pmatrix}. \tag{2.257}$$

Let's now find a combination $\psi^\dagger$(2.253) – ψ (2.253)†:

$$\psi^\dagger\,\mathrm{i}\hbar\frac{\partial\psi}{\partial t} + \mathrm{i}\hbar\frac{\partial\psi^\dagger}{\partial t}\psi = \psi^\dagger\left(-\mathrm{i}\hbar c\alpha^k\partial_k + m_0c^2\beta\right)\psi - \left(\mathrm{i}\hbar c\left(\alpha^k\partial_k\psi\right)^\dagger + m_0c^2\left(\beta\psi\right)^\dagger\right)\psi,$$

$$\psi^\dagger\,\mathrm{i}\hbar\frac{\partial\psi}{\partial t} + \mathrm{i}\hbar\frac{\partial\psi^\dagger}{\partial t}\psi = -\mathrm{i}\hbar c\psi^\dagger\alpha^k\partial_k\psi + m_0c^2\psi^\dagger\beta\psi - \mathrm{i}\hbar c\left(\partial_k\psi^\dagger\right)\alpha^k\psi - m_0c^2\psi^\dagger\beta\psi\,,$$

$$\psi^\dagger\frac{\partial\psi}{\partial t} + \frac{\partial\psi^\dagger}{\partial t}\psi = -c\psi^\dagger\alpha^k\partial_k\psi - c\left(\partial_k\psi^\dagger\right)\alpha^k\psi\,,$$

$$\frac{\partial}{\partial t}\left(\psi^\dagger\psi\right) = -\partial_k\left(\psi^\dagger c\alpha^k\psi\right),$$

$$\frac{\partial}{\partial ct}\left(c\psi^\dagger\psi\right) + \partial_k\left(\psi^\dagger c\alpha^k\psi\right) = 0\,.$$

We have thus obtained the continuity equation in the form

$$\blacktriangleright \qquad \begin{gathered} \partial_\mu j^\mu = 0\,; \\ j^0 = c\psi^\dagger\psi\,, \qquad \vec{\mathbf{j}} = \psi^\dagger\,\hat{\vec{\mathbf{v}}}\,\psi\,, \qquad \hat{\vec{\mathbf{v}}} = c\hat{\vec{\boldsymbol{\alpha}}}\,. \end{gathered} \tag{2.258}$$

The probability density is given by the equation

$$\rho_\mathrm{P} = j^0/c = \psi^\dagger\psi = \psi_1^*\psi_1 + \psi_2^*\psi_2 + \psi_3^*\psi_3 + \psi_4^*\psi_4 \geq 0 \tag{2.259}$$

and is therefore positive definite. The probability flux is a generalization of the classical flux (density×velocity), where velocity is replaced by its operator. However, the resulting probability flux is an ordinary vector, since each of its components is the product of a row, square, and column matrix, i.e., it yields an ordinary number.

Dirac equation for a charged particle in an electromagnetic field

We generalize from a free particle to a particle in a field in the same way as in the Klein–Gordon equation, i.e., we replace

$$\hat{p}^{\alpha} \to \hat{p}^{\alpha} - QA^{\alpha} . \tag{2.260}$$

In the (3+1) symbolism, we have

$$\begin{aligned} \mathrm{i}\hbar\frac{\partial}{\partial ct} \quad &\to \quad \mathrm{i}\hbar\frac{\partial}{\partial ct} - Q\frac{\phi}{c} ; \\ -\mathrm{i}\hbar\frac{\partial}{\partial \mathbf{x}} \quad &\to \quad -\mathrm{i}\hbar\frac{\partial}{\partial \mathbf{x}} - Q\mathbf{A} . \end{aligned} \tag{2.261}$$

The Dirac equation (2.256) now takes the form

$$\mathrm{i}\hbar\frac{\partial\psi}{\partial t} - Q\phi\psi = \left[\hat{\mathbf{v}}\cdot\left(\hat{\mathbf{p}} - Q\mathbf{A}\right) + m_0 c^2\beta\right]\psi \text{ , or}$$

► $$\mathrm{i}\hbar\frac{\partial\psi}{\partial t} = \left[\hat{\mathbf{v}}\cdot\hat{\mathbf{p}} + m_0 c^2\beta\right]\psi + \left[Q\phi - Q\mathbf{A}\cdot\hat{\mathbf{v}}\right]\psi . \tag{2.262}$$

Compared to the free particle, an interaction Hamiltonian has been added on the right-hand side according to the relation

$$\begin{gathered} \mathrm{i}\hbar\frac{\partial\psi}{\partial t} = \hat{H}_0\psi + \hat{H}_{\mathrm{I}}\psi ; \\ \hat{H}_0 \equiv \hat{\mathbf{v}}\cdot\hat{\mathbf{p}} + m_0 c^2\beta , \qquad \hat{H}_{\mathrm{I}} \equiv Q\phi - Q\mathbf{A}\cdot\hat{\mathbf{v}} . \end{gathered} \tag{2.263}$$

In this case, too, it is simply a direct generalization of the interaction term familiar from the Lagrangian in classical mechanics.

Covariant form of the Dirac equation

Let's multiply Dirac equation (2.253) from the left by the matrix β and then move all terms to the left side of the equation:

$$\mathrm{i}\hbar\beta\frac{\partial\psi}{\partial t} = \left(-\mathrm{i}\hbar c\beta\alpha^k\partial_k + m_0 c^2\right)\psi \quad \Rightarrow$$

$$\left(\mathrm{i}\hbar\beta\frac{\partial\psi}{\partial ct} + \mathrm{i}\hbar\beta\alpha^k\partial_k - m_0 c\right)\psi = 0 .$$

This gives us the most well-known form of the Dirac equation

► $$\begin{gathered} \left(\mathrm{i}\hbar\gamma^{\mu}\partial_{\mu} - m_0 c\right)\psi = 0 ; \\ \gamma^0 \equiv \beta , \\ \gamma^k \equiv \beta\alpha^k , \end{gathered} \tag{2.264}$$

in which the coefficients of the derivatives form the so-called gamma matrix

$$\gamma^0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}; \quad \gamma^1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix};$$

► (2.265)

$$\gamma^2 = \begin{pmatrix} 0 & 0 & 0 & -\mathrm{i} \\ 0 & 0 & \mathrm{i} & 0 \\ 0 & \mathrm{i} & 0 & 0 \\ -\mathrm{i} & 0 & 0 & 0 \end{pmatrix}; \quad \gamma^3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix};$$

There are simple relationships between the original matrices and the gamma matrices:

$$\begin{aligned} \gamma^0 &= \beta , \\ \gamma^k &= \beta\alpha^k , \\ \alpha^k &= \beta\gamma^k . \end{aligned} \tag{2.266}$$

The first two define Dirac matrices; the third relation follows from the second after multiplying by the matrix β from the left. For the spatial part of both sets of matrices, a simple rule applies: multiplying by the matrix β from the left yields the corresponding matrix from the second set. The gamma matrices are again anti-commutative, but they are no longer Hermitian, and the squares of the spatial matrices do not yield the identity matrix, but rather the negative identity matrix:

$$\left(\gamma^1\right)^2 = \gamma^1\gamma^1 = \beta\alpha^1\beta\alpha^1 = -\alpha^1\beta\beta\alpha^1 = -\alpha^1\alpha^1 = -\mathbf{1} .$$

We can proceed in the same way with the other matrices; therefore, the following holds

$$\left(\gamma^k\right)^2 = -\mathbf{1} ; \qquad k = 1,2,3 . \tag{2.267}$$

Let's now introduce two useful and frequently used operations. The first of these is the so-called Dirac product:

► $$\bar{\mathbf{A}} \equiv \mathbf{A}^\dagger \gamma^0 . \tag{2.268}$$

This is a Hermitian conjugate multiplied by the matrix γ^0 from the right. The second operation is the *Feynman slash*:

► $$\not{K} \equiv \gamma^\alpha K_\alpha . \tag{2.269}$$

Using these operations, we can elegantly express all components of the probability four-flux (2.258)

$$\begin{aligned} j^0 &= c\psi^\dagger\psi = c\psi^\dagger\gamma^0\gamma^0\psi = c\bar{\psi}\gamma^0\psi , \\ j^k &= \psi^\dagger c\alpha^k\psi = \psi^\dagger c\gamma^0\gamma^k\psi = c\bar{\psi}\gamma^k\psi . \end{aligned}$$

So, we can consistently write

► $$\partial_\mu j^\mu = 0 ; \qquad j^\mu \equiv c\bar{\psi}\gamma^\mu\psi . \tag{2.270}$$

The Dirac equation (2.264) can be rewritten in a "concise" form

$$\left(\gamma^{\mu}\hat{p}_{\mu}+m_0c\right)\psi=0\,,\quad \text{or} \tag{2.271}$$

$$\left(\hat{\not p}+m_0c\right)\psi=0\,. \tag{2.272}$$

For a charged particle in an electromagnetic field, the Dirac equation will now take a very simple form

► $$\left(\hat{\not p}-Q\not A+m_0c\right)\psi=0\,. \tag{2.273}$$

Spin became an integral part of the relativistic equations of quantum theory. The Klein–Gordon equation ultimately proved to be the appropriate equation for scalar particles (with spin 0), while the Dirac equation applies to particles with spin ½ (electrons, neutrinos, quarks). Modern quantum electrodynamics is based precisely on this equation.

◖ **Example 2.5:**

$$\left\{\gamma^{\alpha},\gamma^{\beta}\right\}=\begin{cases}0\,; & \alpha\neq\beta\\ +2; & \alpha=\beta=0\\ -2; & \alpha=\beta=1,2,3\end{cases}\quad\Rightarrow\quad\left\{\gamma^{\alpha},\gamma^{\beta}\right\}=-2g^{\alpha\beta}. \tag{2.274}$$

The square of the matrices α and β is equal to the identity matrix. This is not the case for the matrices γ: $(\gamma^0)^2=1$, but $(\gamma^k)^2=-1$ for $k=1,2,3$. This is natural; in the Minkowski metric, the spatial component always behaves differently from the temporal component. In equation (2.274), the result on the right-hand side is always multiplied by the identity matrix, but it is not customary to write this out.

◗

◖ **Example 2.6:**

$$\not K\not K\equiv\gamma^{\alpha}K_{\alpha}\gamma^{\beta}K_{\beta}=\gamma^{\alpha}\gamma^{\beta}K_{\alpha}K_{\beta}=$$

$$=\frac{1}{2}\left\{\gamma^{\alpha},\gamma^{\beta}\right\}K_{\alpha}K_{\beta}=-g^{\alpha\beta}K_{\alpha}K_{\beta}=K_0^2-\mathbf{K}^2.$$

◗

◖ **Example 2.7:**

$$\not\partial\not\partial\equiv\gamma^{\alpha}\partial_{\alpha}\gamma^{\beta}\partial_{\beta}=\gamma^{\alpha}\gamma^{\beta}\partial_{\alpha}\partial_{\beta}=$$

$$=\frac{1}{2}\left\{\gamma^{\alpha},\gamma^{\beta}\right\}\partial_{\alpha}\partial_{\beta}=-g^{\alpha\beta}\partial_{\alpha}\partial_{\beta}=-\Box.$$

◗

◖ **Example 2.8:**

$$\not p\equiv\gamma^{\alpha}p_{\alpha}=\gamma^0p_0+\gamma^kp_k=\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\frac{E}{c}+\begin{pmatrix}\mathbf{0} & \sigma^k\\ -\sigma^k & \mathbf{0}\end{pmatrix}p_k=\begin{pmatrix}(E/c)\mathbf{1} & p_k\sigma^k\\ -p_k\sigma^k & -(E/c)\mathbf{1}\end{pmatrix}.$$

◗

◖ **Example 2.9:**

Let's prove that the following relation holds (which is useful when calculating the conjugate of the matrices $\gamma^{\mu\dagger}$)

$$\gamma^{\mu\dagger} = \gamma^0 \gamma^\mu \gamma^0 . \tag{2.275}$$

We multiply the left and right sides of the equation (2.275) by γ^0 and prove that $\gamma^0\gamma^{\mu\dagger}\gamma^0 = \gamma^\mu$ holds: We expand the conjugate matrix $\gamma^{\mu\dagger}$ using the set (2.266) of Hermitian matrices α and β:

$$\gamma^0\gamma^{0\dagger}\gamma^0 = \gamma^0\beta\gamma^0 = \gamma^0\gamma^0\gamma^0 = \gamma^0 ;$$

$$\gamma^0\gamma^{k\dagger}\gamma^0 = \gamma^0\left(\gamma^0\alpha^k\right)^\dagger\gamma^0 = \gamma^0\alpha^k\gamma^0\gamma^0 = \gamma^0\alpha^k = \gamma^k .$$ ◗

Matrix C

There are a number of other interesting matrices that can be derived from the basic set of γ^μ matrices. Let us first introduce the **C** matrix, which will be useful for charge conjugation (the transition from particles to antiparticles) and also for calculating the transposed γ^μ matrices $\gamma^{\mu\mathrm{T}}$:

$$\mathbf{C} \equiv \mathrm{i}\,\gamma^2\gamma^0 = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & +1 & 0 \\ 0 & -1 & 0 & 0 \\ +1 & 0 & 0 & 0 \end{pmatrix} . \tag{2.276}$$

The matrix has elements only on the secondary diagonal, alternating between +1 and –1. The trace of the matrix is zero. It is clear that **C** has an interesting property:

$$\mathbf{C}^\mathrm{T} = \mathbf{C}^\dagger = \mathbf{C}^{-1} = -\mathbf{C} . \tag{2.277}$$

If we want to find the inverse, transpose, or Hermitian conjugate of a matrix, we simply need to change the sign of the matrix (by swapping the positions of +1 and −1 on the secondary diagonal). If we need to find the transpose $\gamma^{\mu\mathrm{T}}$, we can use the matrix **C**:

$$\gamma^{\mu\,\mathrm{T}} = \mathbf{C}\gamma^\mu\mathbf{C} ; \qquad \text{resp.} \qquad \gamma^{\mu\,\mathrm{T}} = -\mathbf{C}^{-1}\gamma^\mu\mathbf{C} . \tag{2.278}$$

The claim can be proven simply by transforming the matrices γ^μ into a set of matrices α and β that remain unchanged after transposition.

Matrix γ^5

Another important matrix used to describe left-right symmetry is the matrix

$$\gamma^5 \equiv \mathrm{i}\,\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{pmatrix} . \tag{2.279}$$

This matrix is Hermitian, its square is equal to the identity matrix, it is linearly independent of the other γ matrices, and it anticommutes with them:

$$\gamma^{5\dagger} = \gamma^5 ,$$
$$\left(\gamma^5\right)^2 = \mathbf{1} , \tag{2.280}$$
$$\left\{\gamma^5, \gamma^\mu\right\} = 0 ; \quad \mu = 0,1,2,3 .$$

Matrix *Σ* and basis *Γ^k*

From the γ^μ matrices, we can construct Σ matrices, which are useful for defining a basis on the matrix space and for investigating the transformation properties of antisymmetric tensors. We define these matrices as commutators

$$\Sigma^{\mu\nu} \equiv \frac{\mathrm{i}}{2}\left[\gamma^\mu, \gamma^\nu\right]. \tag{2.281}$$

Obviously, the following applies

$$\Sigma^{\alpha\beta} = \begin{cases} 0 ; & \alpha = \beta , \\ -\Sigma^{\beta\alpha} ; & \alpha \neq \beta . \end{cases} \tag{2.282}$$

There are a total of six independent matrices $\Sigma^{\alpha\beta}$, e.g., Σ^{01}, Σ^{02}, Σ^{03}, Σ^{12}, Σ^{13}, and Σ^{23}. The remaining elements are either zero or can be derived from antisymmetry. An interesting basis for the space of 4×4 matrices consists of the following 16 matrices:

$$\Gamma^k = \left\{\mathbf{1}, \gamma^5, \gamma^\mu, \gamma^5\gamma^\mu, \Sigma^{\mu\nu}\right\}. \tag{2.283}$$

It is easy to show that the matrices Γ^k are independent. Since there are 16 of them, they form a basis for the space of 4×4 matrices. Their squares are ±1; by multiplying them by ±i, one could ensure that the square is always equal to the identity matrix, but this is unnecessary. The trace of all, with the exception of the identity matrix Γ^1, is zero. The product of any two distinct matrices Γ^k is, up to a sign, equal to some other matrix Γ:

$$\begin{aligned} &1) \quad \Gamma^k \text{ are linearly independent,} \\ &2) \quad \left(\Gamma^k\right)^2 = \pm 1, \\ &3) \quad \mathrm{Tr}\,(\Gamma^k) = \begin{cases} 4 & k = 1; \\ 0 & k \neq 1, \end{cases} \\ &4) \quad \forall k,l;\ k \neq l\ \exists m \neq k,l : \qquad \Gamma^k \Gamma^l = \pm \Gamma^m . \end{aligned} \tag{2.284}$$

It is clear from the continuity equation (2.270) that the quantity $\bar{\psi}\gamma^\mu\psi$ is a four-vector. Similarly, other quantities that transform characteristically can be constructed using the basis vectors Γ^k:

$$\begin{array}{ll} \bar{\psi}\psi & \text{scalar} \\ \bar{\psi}\gamma^5\psi & \text{pseudoscalar} \\ \bar{\psi}\gamma^\mu\psi & \text{vector} \\ \bar{\psi}\gamma^5\gamma^\mu\psi & \text{pseudovector} \\ \bar{\psi}\Sigma^{\mu\nu}\psi & \text{antisymmetric tensor} \end{array} \tag{2.285}$$

2.7.5 Positron, Charge Symmetry

Let us describe the reasoning that led Paul Dirac to predict the existence of the positron, and then the mathematical transformation (charge conjugation) that converts Dirac equation for the electron into the equation for the positron. Dirac predicted the existence of the positron in 1928; it was discovered by Carl Anderson in cosmic rays in 1932.

Dirac sea

As we have seen, the Dirac equation yields negative energy states. However, negative energies do not occur in nature, so Dirac reasoned that these states are all occupied by electrons and none of them are free, which is why we do not observe them. According to this concept, the vacuum consists of a sea of electrons in negative energy states, known as the Dirac sea.

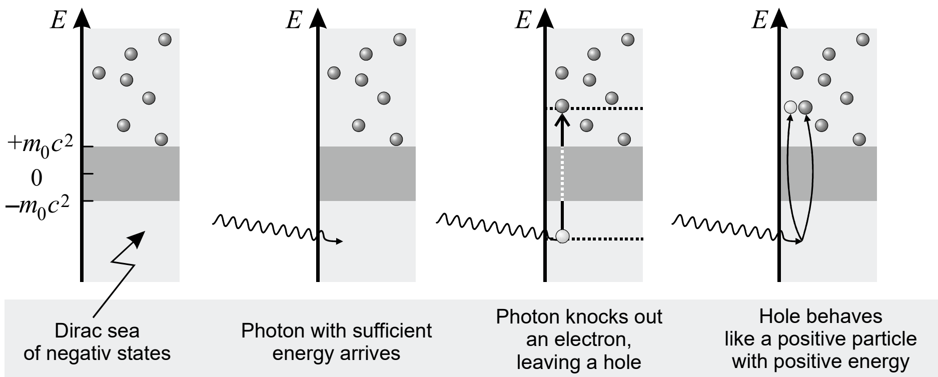


Fig. 2.42: Dirac sea

Let's imagine that a photon with energy greater than twice the rest energy of an electron flies into this sea.

$$E_\gamma > 2m_0c^2.$$

It can then knock out an electron from the Dirac sea and transfer it to an energy state with positive energy. A hole remains in the negative Dirac sea – an empty energy state that appears to the surrounding environment as a positively charged region with positive energy (mass). In 1928, Dirac interpreted this hole as a positively charged particle that otherwise has the same properties as an electron, and he called it a *positron*. Externally, it therefore appears as though the original photon has decayed into an electron-positron pair. In 1929, Dirac extended this concept to all particles and introduced the term "*anti-particle*" – an object that has opposite values for all quantum charges compared to the original particle. This was at a time when the neutron had not yet been discovered.

The motion of a free electron should be influenced by the Dirac sea. The electron interacts with nearby electrons in negative states, pushing them apart, and from a distance it appears as if it has a smaller charge than it actually does. From a greater distance, therefore, we do not see the electron's true charge, but rather the charge shielded by the Dirac sea. The closer we get to the moving electron, the more we perceive its true, bare charge.

Charge conjugation

The Dirac equation for an electron in an external field takes the form (2.273)

$$\left(\hat{p\!\!/} - Q\,A\!\!\!/ + m_0 c\right)\psi = 0 \quad \Rightarrow$$

$$\left(-i\hbar\gamma^\mu \partial_\mu - Q\,\gamma^\mu A_\mu + m_0 c\right)\psi = 0 \quad \Rightarrow$$

$$\left(i\hbar\gamma^\mu \partial_\mu + Q\,\gamma^\mu A_\mu - m_0 c\right)\psi = 0 \quad \Rightarrow$$

$$\left(i\hbar\gamma^\mu \partial_\mu - e\,\gamma^\mu A_\mu - m_0 c\right)\psi = 0\,. \tag{2.286}$$

The equation for the positron should be

$$\left(i\hbar\gamma^\mu \partial_\mu + e\,\gamma^\mu A_\mu - m_0 c\right)\psi_C = 0\,, \tag{2.287}$$

where ψ_C is the wave function of the positron solution. We first perform a Hermitian conjugation on the Dirac equation and then transpose it. After these two operations, the equation for the electron transforms into the equation for the positron. The operations of Hermitian conjugation and transposition satisfy property (3.267), see Section 3.4.2:

$$(\mathbf{AB})^\dagger = \mathbf{B}^\dagger \mathbf{A}^\dagger\,; \qquad (\mathbf{AB})^\mathrm{T} = \mathbf{B}^\mathrm{T} \mathbf{A}^\mathrm{T}. \tag{2.288}$$

Let’s perform Hermitian conjugate of Dirac equation for the electron (2.286):

$$\psi^\dagger \left(i\hbar\gamma^\mu \underset{\leftarrow}{\partial_\mu} - e\,\gamma^\mu A_\mu - m_0 c \right)^\dagger = 0\,.$$

The derivative of the wavefunction acts to the left (this is indicated by the arrow below). Let’s now apply the Hermitian conjugate to the individual terms in parentheses

$$\psi^\dagger \left(-\mathrm{i}\hbar \underset{\leftarrow}{\partial_\mu}\, \gamma^{\mu\dagger} - e\,A_\mu \gamma^{\mu\dagger} - m_0 c \right) = 0$$

and we express the conjugate Dirac gamma matrices using the relation (2.275)

$$\psi^\dagger \left(-\mathrm{i}\hbar \underset{\leftarrow}{\partial_\mu}\, \gamma^0 \gamma^\mu \gamma^0 - e\,A_\mu \gamma^0 \gamma^\mu \gamma^0 - m_0 c \right) = 0\,.$$

We multiply the equation from the right by a matrix γ^0:

$$\psi^\dagger \gamma^0 \left(-\mathrm{i}\hbar \underset{\leftarrow}{\partial_\mu}\, \gamma^\mu - e\,A_\mu \gamma^\mu - m_0 c \right) = 0 \quad \Rightarrow$$

$$\overline{\psi} \left(-\mathrm{i}\hbar \underset{\leftarrow}{\partial_\mu}\, \gamma^\mu - e\,A_\mu \gamma^\mu - m_0 c \right) = 0\,.$$

After the Hermitian conjugation, the sign of the first term changes; the wave function is on the left and takes the form of a Dirac-conjugated bispinor. Now we will return the wave function to the right using the transposition operation

$$\left(-\mathrm{i}\hbar\partial_\mu\gamma^\mu - eA_\mu\gamma^\mu - m_0c\right)^{\mathrm{T}}(\bar{\psi})^{\mathrm{T}} = 0\,.$$

We will transpose all the terms in parentheses

$$\left(-\mathrm{i}\hbar\gamma^{\mu\mathrm{T}}\partial_\mu - e\gamma^{\mu\mathrm{T}}A_\mu - m_0c\right)(\bar{\psi})^{\mathrm{T}} = 0$$

and rewrite the transposed gamma matrix using the formula (2.278)

$$\left(-\mathrm{i}\hbar\mathbf{C}\gamma^\mu\mathbf{C}\partial_\mu - e\mathbf{C}\gamma^\mu\mathbf{C}A_\mu - m_0c\mathbf{1}\right)(\bar{\psi})^{\mathrm{T}} = 0\,.$$

We multiply the entire equation by the matrix $\mathbf{C}^{-1}$ from the left

$$\left(-\mathrm{i}\hbar\gamma^\mu\mathbf{C}\partial_\mu - e\gamma^\mu\mathbf{C}A_\mu - m_0c\mathbf{C}^{-1}\right)(\bar{\psi})^{\mathrm{T}} = 0$$

and we use equation (2.277) for the inverse matrix $\mathbf{C}^{-1} = -\mathbf{C}$:

$$\left(-\mathrm{i}\hbar\gamma^\mu\partial_\mu - e\gamma^\mu A_\mu + m_0c\right)\mathbf{C}(\bar{\psi})^{\mathrm{T}} = 0\,.$$

Transposition therefore changed the sign of the last term of the equation, that is

$$\left(\mathrm{i}\hbar\gamma^\mu\partial_\mu + e\gamma^\mu A_\mu - m_0c\right)\psi_{\mathrm{C}} = 0\,; \qquad \psi_{\mathrm{C}} \equiv \mathbf{C}(\bar{\psi})^{\mathrm{T}}\,. \tag{2.289}$$

We obtained the desired equation for the positron, which has the opposite charge and the same mass. If we know the solution ψ for the electron in a given situation, the corresponding solution for the positron will be the wave function $\psi_{\mathrm{C}} = \mathbf{C}(\bar{\psi})^{\mathrm{T}}$.

The solution for the positron is therefore not a new solution; it is contained in the solution for the electron. If we perform the charge conjugation, or C transformation

$$A_\mu \to -A_\mu\,; \qquad \psi \to \mathbf{C}(\bar{\psi})^{\mathrm{T}}\,, \tag{2.290}$$

The Dirac equation does not change its form (it is covariant). In the original Dirac equation, the two solutions with positive energy correspond to the electron with spin projections of ±½ (which is why the solution is double), and the solution with negative energy corresponds to the positron with spin projections of ±½. The two pairs combined into a set of four wave functions are called *bispinors*, see p. 210. After performing the transformation (2.290), the positron solutions have positive energy and the electron solutions have negative energy, so we can view the situation in reverse and interpret the electron as a hole in a sea of positrons occupying the negative energy states.

2.7.6 Electron and Its Field, U(1) Symmetry

Charged particles in nature generate electromagnetic fields described by Maxwell equations (or quantum electrodynamics) and move within these fields in accordance with the Lorentz equation of motion (or the Dirac equation). In this chapter, we will first focus on the complete Lagrangian description of the field-electron system and then turn our attention to U(1) symmetry, from which the necessity of the existence of a field in the vicinity of the electron directly follows.

Lagrange formulation

The total density of the Lagrangian function for the interaction between a charged particle and an electromagnetic field is given by

$$\mathscr{L} = \mathscr{L}_{\text{Field}} + \mathscr{L}_{\text{Int}} + \mathscr{L}_{\text{Part}} . \tag{2.291}$$

We are familiar with the field term from theoretical mechanics; see the equation (1.261)

$$\mathscr{L}_{\text{Field}} = -\frac{1}{4\mu_0} F_{\mu\nu} F^{\mu\nu} ; \qquad F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu . \tag{2.292}$$

We also encountered the interaction term. If particles are described by the four-momentum j^μ and fields by the four-potential A^μ, the simplest scalar is $j^\mu A_\mu$:

$$\mathscr{L}_{\text{Int}} = j^\mu A_\mu . \tag{2.293}$$

In classical physics, the particle charge flux is given by the expression $j^\mu = (\rho_Q c, \mathbf{j})$; in quantum theory, the particle flux must follow the probability of their occurrence and must therefore be proportional to the probability flux (2.270). In the case of charge, the proportionality coefficient is the charge itself:

$$j^\mu = Q\, \bar{\psi} \gamma^\mu \psi . \tag{2.294}$$

The interaction term then takes the form:

$$\mathscr{L}_{\text{Int}} = j^\mu A_\mu = Q\, \bar{\psi} \gamma^\mu \psi A_\mu . \tag{2.295}$$

All that remains is to find the density of the Lagrangian for electrons, from which Dirac equation follows. A simple scalar expression derived directly from Dirac equation is

$$\mathscr{L}_{\text{Part}} = \bar{\psi} \left(\mathrm{i}\hbar \gamma^\mu \partial_\mu - m_0 c \right) \psi . \tag{2.296}$$

If we interpret the fields $\bar{\psi}, \psi$ as independent, the density of the Lagrangian function is

$$\mathscr{L}_{\text{Part}} = \mathscr{L}_{\text{Part}}(\psi, \partial_\alpha \psi, \bar{\psi})$$

and the corresponding Lagrange equations give

$$\partial_\alpha \frac{\partial \mathscr{L}_{\text{Part}}}{\partial(\partial_\alpha \bar{\psi})} - \frac{\partial \mathscr{L}_{\text{Part}}}{\partial \bar{\psi}} = 0 \qquad \Rightarrow \qquad \left(\mathrm{i}\hbar \gamma^\mu \underset{\rightarrow}{\partial_\mu} - m_0 c \right) \psi = 0 ;$$

$$\partial_\alpha \frac{\partial \mathscr{L}_{\text{Part}}}{\partial(\partial_\alpha \psi)} - \frac{\partial \mathscr{L}_{\text{Part}}}{\partial \psi} = 0 \qquad \Rightarrow \qquad \bar{\psi} \left(\mathrm{i}\hbar \gamma^\mu \underset{\leftarrow}{\partial_\mu} + m_0 c \right) = 0 .$$

The first equation is Dirac equation for a charged particle; the second equation is, for now, merely an auxiliary equation for the Dirac-conjugated field. Note that the mass term has changed sign. In Feynman diagrams, this corresponds to the inflow (or outflow) of mass into (from) a given vertex. In the derivatives, the direction of their action is indicated in both cases. We can now write down the complete density of the Lagrangian for the particle and the electromagnetic field:

$$\mathscr{L} = -\frac{1}{4\mu_0} F_{\mu\nu} F^{\mu\nu} \; + \; j^{\mu} A_{\mu} \; + \; \bar{\psi}\left(\mathrm{i}\hbar\gamma^{\mu}\partial_{\mu} - m_0 c\right)\psi \, ;$$

► $$F^{\mu\nu} \equiv \partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu} \, ; \tag{2.297}$$

$$j^{\mu} \equiv Q\,\bar{\psi}\gamma^{\mu}\psi \, .$$

Note 1: The first term corresponds to the free field, the second to the interaction between the field and a particle, and the third to a free particle.

Note 2: The Lagrangian is a function of the fields $A_{\mu}, \psi, \bar{\psi}$ and their derivatives. The Lagrangian equations for the fields A_{μ} yield Maxwell equations, while the Lagrangian equations for the fields $\bar{\psi}$ yield Dirac equation.

Note 3: If we retain only the first term, we obtain Maxwell equations in a vacuum from the density of the Lagrangian. If we retain only the last term, we obtain Dirac free-particle equation. The first and second terms yield Maxwell equations with source terms (the field interacts with particles), while the second and third terms yield Dirac equation for a particle in the presence of an electromagnetic field (the particle interacts with the field).

Note 4: The interaction term (second) and the particle term (third) can be combined into a form that yields the Dirac equation with an electromagnetic field:

$$\mathscr{L}_{\mathrm{Dir}} = \bar{\psi}\left(i\hbar\gamma^{\mu}\partial_{\mu} + Q\,\gamma^{\mu} A_{\mu} - m_0 c\right)\psi \tag{2.298}$$

U(1) symmetry

The density of the Lagrangian function of the Dirac equation and the four-flux

► $$\mathscr{L}_{\mathrm{Dir}} = \bar{\psi}\left(i\hbar\gamma^{\mu}\partial_{\mu} + Q\,\gamma^{\mu} A_{\mu} - m_0 c\right)\psi \, ;$$

$$j^{\mu} = Q\bar{\psi}\gamma^{\mu}\psi$$

will not change during the transformation

$$\begin{aligned} \psi \quad &\rightarrow \quad \psi' = \psi\,\mathrm{e}^{\mathrm{i}\alpha} \, , \\ \bar{\psi} \quad &\rightarrow \quad \bar{\psi}' = \bar{\psi}\,\mathrm{e}^{-\mathrm{i}\alpha} \, , \\ A^{\mu} \quad &\rightarrow \quad A'^{\mu} = A^{\mu} \, . \end{aligned} \tag{2.299}$$

This transformation does not alter the field tensor $F^{\mu\nu}$, and thus does not alter the density of the Lagrangian $\mathscr{L}_{\mathrm{Field}}$. The entire theory represented by the Lagrangian (2.297) is covariant with respect to the transformation (2.299), which consists of rotating the wave function at every point in spacetime by the same angle α. It is a unitary operation (it does not change the scalar product) with a single parameter (the angle α). This transformation is denoted U(1). It represents the so-called *internal symmetry* of the theoretical description of the field-particle interaction. A consequence of this symmetry is the existence of electric charge, which is conserved. U(1) symmetry in other theories (Lagrangians) leads to the existence of other quantum charges, which are also conserved.

$U(1)_{loc}$ symmetry

Let us now examine how the Lagrangian of a free particle would change if we allowed the rotation angle α of the wave function to be different at every point in spacetime. Imagine an infinite spatial lattice with identical balls at each point. U(1) symmetry in our model corresponds to rotating all the balls around their centers simultaneously by the same angle. After this transformation, the lattice will look the same as before. Now let's imagine that we rotate different balls by different angles at different times. The result? The spatial lattice will still look the same as it did before the rotations began. And it is precisely this kind of symmetry that we call $U(1)_{loc}$ symmetry:

$$
\blacktriangleright \qquad \begin{aligned}
\psi &\;\rightarrow\; \psi'(x) = \psi(x)\,\mathrm{e}^{\mathrm{i}\alpha(x)}\,, \\
\bar{\psi} &\;\rightarrow\; \bar{\psi}'(x) = \bar{\psi}(x)\,\mathrm{e}^{-\mathrm{i}\alpha(x)}\,.
\end{aligned}
$$

The variable x represents the entire event $x = (t, \mathbf{x})$, i.e., a four-vector. This transformation again does not change the four-flux (2.294). But how does the Lagrangian of a free particle change? Let us substitute the quantities with prime into the Lagrangian (2.296) and take the derivatives of the wave and exponential functions:

$$
\begin{aligned}
\mathscr{L}'_{\text{Part}} &= \bar{\psi}'\left(i\hbar\gamma^\mu\partial_\mu - m_0 c\right)\psi' = \bar{\psi}\,\mathrm{e}^{-\mathrm{i}\alpha(x)}\left(i\hbar\gamma^\mu\partial_\mu - m_0 c\right)\psi\,\mathrm{e}^{\mathrm{i}\alpha(x)} = \\
&= \bar{\psi}\left(i\hbar\gamma^\mu\partial_\mu - \hbar\gamma^\mu\alpha_{,\mu} - m_0 c\right)\psi \neq \mathscr{L}_{\text{part}}\,.
\end{aligned}
$$

The density of the Lagrangian has changed its form after the transformation. A term $\hbar\gamma^\mu\alpha_{,\mu}$ has been added, with derivatives of the rotation angle, which mirrors the field term in the density of the charged particle's Lagrangian in the presence of a field (2.298). The result is very interesting. If we insist that the density of the Lagrangian for a free particle satisfy $U(1)_{loc}$ symmetry, we must add an electromagnetic field to the theory. The requirement that the Dirac equation be covariant with respect to $U(1)_{loc}$ symmetry leads to the requirement for the existence of an electromagnetic field in the vicinity of the particle! The electromagnetic field itself changes under the transformation in such a way as to compensate for the newly arising term $\hbar\gamma^\mu\alpha_{,\mu}$. Let us therefore consider the transformation

$$
\begin{aligned}
\psi &\;\rightarrow\; \psi' = \psi(x)\,\mathrm{e}^{\mathrm{i}\alpha(x)}\,, \\
\bar{\psi} &\;\rightarrow\; \bar{\psi}' = \bar{\psi}(x)\,\mathrm{e}^{-\mathrm{i}\alpha(x)}\,, \\
A^\mu &\;\rightarrow\; A'^\mu = A^\mu(x) + \delta A^\mu(x)
\end{aligned}
$$

and once again we will calculate the primes Lagrangian, this time with the electromagnetic field:

$$
\begin{aligned}
\mathscr{L}'_{\text{Part}} &= \bar{\psi}'\left(i\hbar\gamma^\mu\partial_\mu + Q\gamma^\mu A'_\mu - m_0 c\right)\psi' = \\
&= \bar{\psi}\,\mathrm{e}^{-\mathrm{i}\alpha(x)}\left(i\hbar\gamma^\mu\partial_\mu + Q\gamma^\mu(A_\mu + \delta A_\mu) - m_0 c\right)\psi\,\mathrm{e}^{\mathrm{i}\alpha(x)} = \\
&= \bar{\psi}\left(i\hbar\gamma^\mu\partial_\mu + Q\gamma^\mu A_\mu + \gamma^\mu(-\hbar\alpha_{,\mu} + Q\,\delta A_\mu) - m_0 c\right)\psi\,.
\end{aligned}
$$

The density of the Lagrangian function remains unchanged if the inner round bracket is zero, i.e.,

$$\delta A_\mu = \frac{\hbar}{Q}\alpha_{,\mu}\,.$$

This gives us the formula for the correct transformation of the electromagnetic field. The total $U(1)_{loc}$ transformation will therefore take the form:

$$\begin{aligned} \psi &\to \psi' = \psi(x)\,e^{i\alpha(x)}\,, \\ \bar{\psi} &\to \bar{\psi}' = \bar{\psi}(x)\,e^{-i\alpha(x)}\,, \\ A^\mu &\to A'^\mu = A^\mu(x) + \frac{\hbar}{Q}\partial^\mu\alpha\,. \end{aligned} \tag{2.300}$$

As we have shown, the sum $\mathscr{L}_{\text{int}} + \mathscr{L}_{\text{Dir}}$ remains unchanged under this transformation. We can easily verify that the U(1)loc transformation does not affect the electromagnetic field tensor and, consequently, the field part of the Lagrangian, $\mathscr{L}_{\text{Field}}$. The transformation also does not affect the probability current (2.270). The entire theory is thus covariant with respect to the $U(1)_{loc}$ transformation. The quantum theory of the electromagnetic field is therefore usually referred to as a $U(1)_{loc}$ theory. The $U(1)_{loc}$ symmetry ensures the coupling between the charged particle (electron) and the electromagnetic field.

If we apply the Klein-Gordon or Dirac equation to a system of particles, we find that the statistical behavior of multiple particles differs for each equation. The Klein-Gordon equation is suitable for particles with spin 0 that do not satisfy the Pauli exclusion principle. The Dirac equation, on the other hand, is suitable for particles with spin ½ that do satisfy the Pauli exclusion principle. We will discuss the behavior of a system of identical particles in quantum theory in the following chapter.

2.8 System of Identical Particles

We refer particles with the same parameters (mass, charge, spin, etc.) as identical. In theoretical mechanics, the trajectory of these particles is determined by Hamilton equations, and if we know the initial positions and velocities of the particles, we can precisely predict their future positions and determine which particle is which.

In quantum theory, we can only predict the probability of a particle's occurrence at a certain location and time. This probability reaches a maximum at the point of the classical trajectory, generally decreases exponentially with distance from it, and, although very small quite far from the classical trajectory, is not zero. If we have two identical particles, we can never be certain which particle is which. The probability of one particle being found at the location of the other is nonzero. We say that identical particles are indistinguishable in quantum theory. The Hamiltonian operator does not change when two identical particles are interchanged:

$$\hat{\mathbf{H}}_{12} = \hat{\mathbf{H}}_{21}. \tag{2.301}$$

2.8.1 Exchange Operator

For simplicity, we will consider only two particles, for which we observe the dynamic variable A (preferably the entire set of observables). We will denote the state in which the first particle has the value a_1 and the second particle has the value a_2 as

$$|\psi\rangle = |a_1, a_2\rangle.$$

Let's denote reverse situation, where the first particle has value a_2 and the second a_1, as

$$|\varphi\rangle = |a_2, a_1\rangle.$$

Because identical particles are indistinguishable in quantum mechanics, the two states must be dependent (they actually represent one and the same quantum state), therefore

$$|a_2, a_1\rangle = \beta\,|a_1, a_2\rangle. \tag{2.302}$$

Let us now define the particle exchange operator using the relation

► $$\hat{\mathbf{P}}_{12}\,|a_2, a_1\rangle \equiv |a_1, a_2\rangle \tag{2.303}$$

and examine its properties:

$$\begin{aligned} &(1)\quad \hat{\mathbf{P}}^2 = \hat{\mathbf{1}}, \\ &(2)\quad \lambda_{1,2} = \pm 1, \\ &(3)\quad \left[\hat{\mathbf{P}}, \hat{\mathbf{H}}\right] = 0. \end{aligned} \tag{2.304}$$

Proof (1): A double exchange of particles results in the original configuration.

Proof (2): The eigenvectors are the vectors $|\psi\rangle$ and $|\varphi\rangle$ defined above:

$$\hat{\mathbf{P}}_{12}\,|a_1, a_2\rangle \equiv |a_2, a_1\rangle = \beta\,|a_1, a_2\rangle. \tag{2.305}$$

The number β is the eigenvalue of the permutation operator. Let us now perform a double permutation, first using the relation (2.304) and then according to (2.305):

$$P^2 \,|\, a_1\,,\, a_2 > = \begin{cases} |\, a_1\,,\, a_2 > \\ \beta^2 \,|\, a_1\,,\, a_2 > \end{cases} \quad \Rightarrow \quad \beta^2 = 1 \quad \Rightarrow \quad \beta = \pm 1\,.$$

The eigenvalues of the exchange operator are evident from the first equation (2.304). It is a unitary and Hermitian operator. The eigenvalues must lie on the unit circle in the complex plane and, at the same time, be real. The only such values are ±1.

Proof (3): In this proof, we will use the time-dependent Schrödinger equation (2.182):

$$\begin{aligned} &\hat{\mathsf{H}}_{12}\,\hat{\mathsf{P}}_{12}\,|\,a_1,a_2\,\rangle = \hat{\mathsf{H}}_{12}\,|\,a_2,a_1\,\rangle = \hat{\mathsf{H}}_{21}\,|\,a_2,a_1\,\rangle = \\ &= \mathrm{i}\hbar\frac{\mathrm{d}\,|\,a_2,a_1\,\rangle}{\mathrm{d}t} = \mathrm{i}\hbar\,\hat{\mathsf{P}}_{12}\,\frac{\mathrm{d}\,|\,a_1,a_2\,\rangle}{\mathrm{d}t} = \\ &= \hat{\mathsf{P}}_{12}\,\mathrm{i}\hbar\,\frac{\mathrm{d}\,|\,a_1,a_2\,\rangle}{\mathrm{d}t} = \hat{\mathsf{P}}_{12}\,\hat{\mathsf{H}}_{12}\,|\,a_1,a_2\,\rangle\,. \end{aligned}$$

2.8.2 Bosons and Fermions, Pauli Exclusion Principle

It is clear from the previous analysis that

$$|\,a_2,a_1\,\rangle = \hat{\mathsf{P}}\,|\,a_1,a_2\,\rangle = \pm\,|\,a_1,a_2\,\rangle\,. \tag{2.306}$$

The wave function of two particles can be either symmetric or antisymmetric. There is nothing in between. With respect to the exchange of arguments, particles can be of only two types: those with symmetric wave functions (*bosons*) or those with antisymmetric wave functions (*fermions*). This property cannot be altered even by time evolution, because the particle exchange operator according to the third relation (2.304) commutes with the Hamiltonian operator, and its time evolution is therefore zero. If a particle emerges as a fermion or a boson, it remains so until its annihilation.

Bosons

Bosons have a symmetric wave function

► $$|\,a_2,a_1\,\rangle = |\,a_1,a_2\,\rangle\,. \tag{2.307}$$

If both states are identical, i.e., $a_1 = a_2 = a$, we obtain the relation $|a, a\,\rangle = |a, a\,\rangle$, which is always satisfied; therefore, multiple bosons can exist in the same quantum state. At low temperatures, bosons even tend to cumulate in the lowest possible energy state and form what is known as a boson condensate. This is particularly well-known in superfluidity and superconductivity. The statistics governing a system of bosons is called Bose–Einstein statistics, and we discuss it in the follow-up textbook [1]. Further developments in quantum mechanics have shown that bosons are always particles with integer spin (0, 1, 2,…) and that creation operators satisfying simple commutation relations can be introduced for these particles (see the following chapter). The most typical representatives of this family are scalar ($s = 0$) and vector ($s = 1$) mesons, as well as all intermediate particles (the photon, W^+, Z^0, and gluons with spin 1).

Fermions

Fermions have an antisymmetric wave function

$$| a_2, a_1 \rangle = -| a_1, a_2 \rangle \,. \tag{2.308}$$

If both states are identical, i.e., $a_1 = a_2 = a$, we obtain the relation $|a, a\rangle = -|a, a\rangle$, which is never satisfied; therefore, no more than one fermion can exist in the same quantum state. This fact is known as *Pauli exclusion principle*. At low temperatures, fermions occupy individual energy levels sequentially; for example, in the atomic shell, there can be only as many electrons on each level as there are quantum states represented by that level (this is determined by the degree of degeneracy). Thus, in the atomic shell, there cannot be two electrons with the same quantum numbers n, l, m, m_s. The statistics governing a system of fermions is called Fermi-Dirac statistics, and we will discuss it in the next textbook [1]. Fermions are always particles with half-integer spin (1/2, 3/2,…) and for these particles, creation operators satisfying simple anticommutation relations can be introduced (see the following chapter). The most typical representatives of this family of particles are leptons (electron, muon, tauon, and neutrinos with spin 1/2), quarks (d, u, s, c, b, t with spin 1/2), and particles composed of three quarks, or baryons (neutron, proton, Λ hyperons with spin 1/2, and, for example, Δ baryons with spin 3/2).

	BOSONS	FERMIONS
Spin	Integer	Half-integer
Wave function	Symmetric	Antisymmetric
Statistics	Bose-Einstein	Fermi-Dirac
Pauli principle	Do not satisfy	Satisfy
Creation operators	Satisfy commutation relations	Satisfy anticommutation relations

2.8.3 Second Quantization

Let's imagine we have N identical particles occupying the states of a certain dynamic variable. N_1 particles are in the first state (value a_1), N_2 particles are in the second state (value a_2), and so on. We call the numbers N_k the occupation numbers of state k. The sum of all occupation numbers is equal to the number of particles:

$$\sum_k N_k = N \,. \tag{2.309}$$

For bosons, $N_k = 0, 1, 2, 3,\ldots$ For fermions, the situation is simpler. There can be at most one fermion in a given state, i.e., $N_k = 0, 1$. We denote the corresponding state of a system of N identical particles with given occupation numbers as

$$|\psi\rangle = | N_1, N_2, \ldots, N_k, \ldots \rangle \,. \tag{2.310}$$

We call this notation the *occupation number representation*, and we refer to the corresponding states as *Fock states*. The situation differs for bosons and fermions.

Bosons

Just as with the harmonic oscillator, let us introduce *creation and annihilation operators* into the state k via the defining relations (we will keep the normalization constants the same as for the harmonic oscillator):

$$\begin{aligned} \hat{a}_k^\dagger \,|\, N_1, N_2, \ldots, N_k, \ldots \rangle &\equiv \sqrt{N_k + 1}\,|\, N_1, N_2, \ldots, N_k + 1, \ldots \rangle\,, \\ \hat{a}_k \,|\, N_1, N_2, \ldots, N_k, \ldots \rangle &\equiv \sqrt{N_k}\,|\, N_1, N_2, \ldots, N_k - 1, \ldots \rangle\,. \end{aligned} \tag{2.311}$$

Directly from these defining relations (by acting on the state vector (2.310), we can easily compute the commutation relations for the creation and annihilation operators:

$$\begin{aligned} [\hat{a}_k\,, \hat{a}_l] &= 0\,, \\ [\hat{a}_k^\dagger\,, \hat{a}_l^\dagger] &= 0\,, \\ [\hat{a}_k\,, \hat{a}_l^\dagger] &= \delta_{kl}\,. \end{aligned} \tag{2.312}$$

Let's introduce another operator

$$\hat{N}_k \equiv \hat{a}_k^\dagger\, \hat{a}_k\,. \tag{2.313}$$

The operator is called (by analogy with the harmonic oscillator) the *particle number operator*, because acting on the state vector yields the number of particles in state k:

$$\hat{a}_k^\dagger\, \hat{a}_k \,|\, N_1, N_2, \ldots, N_k, \ldots \rangle = \sqrt{N_k}\;\; \hat{a}_k^\dagger \,|\, N_1, N_2, \ldots, N_k - 1, \ldots \rangle =$$

$$\sqrt{N_k}\,\sqrt{N_k}\,|\, N_1, N_2, \ldots, N_k\;, \ldots \rangle = N_k \,|\, N_1, N_2, \ldots, N_k\;, \ldots \rangle\,.$$

The total particle number operator is then

$$\hat{N} \equiv \sum_k \hat{a}_k^\dagger\, \hat{a}_k\,. \tag{2.314}$$

If the entire set of observables is continuous, we can repeat the entire procedure for continuous variables. For example, in the x-representation, we can introduce

$\hat{\psi}^\dagger(x)$	Creation operator into position x,
$\hat{\psi}(x)$	Annihilation operator from position x.

The commutation relations will be similar, except that the Dirac delta function appears on the right-hand side instead of the Kronecker delta:

$$\begin{aligned} [\hat{\psi}(x), \hat{\psi}(y)] &= 0\,, \\ [\hat{\psi}^\dagger(x), \hat{\psi}^\dagger(y)] &= 0\,, \\ [\hat{\psi}(x), \hat{\psi}^\dagger(y)] &= \delta(x - y)\,. \end{aligned} \tag{2.315}$$

The particle number density operator is introduced by the relation

$$\hat{\mathcal{N}}_k \equiv \hat{\psi}^\dagger(x)\,\hat{\psi}(x)\,. \tag{2.316}$$

The operator for the number of particles found in the interval <a, b> is

$$\hat{N}(a,b) \equiv \int_a^b \hat{\psi}^\dagger(x)\hat{\psi}(x)\,\mathrm{d}x \tag{2.317}$$

and the operator for the total number of particles is

$$\hat{N} \equiv \int_{-\infty}^{+\infty} \hat{\psi}^\dagger(x)\hat{\psi}(x)\,\mathrm{d}x\,. \tag{2.318}$$

A similar procedure would be followed in three dimensions. The entire transition from the physics of a single particle to the physics of many identical particles can be formally carried out by replacing the wave function with creation and annihilation operators and replacing the probability density with the particle number density operator:

$$\begin{aligned} \psi(x) \quad &\to \quad \hat{\psi}(x)\,; \\ w(x) \equiv \psi^*(x)\,\psi(x) \quad &\to \quad \hat{\mathcal{N}}(x) \equiv \hat{\psi}^\dagger(x)\,\hat{\psi}(x)\,. \end{aligned} \tag{2.319}$$

This process is called *second quantization*; the wave functions describing the system become operators, and quantum theory turns into *quantum field theory*, in which the quantities describing classical continuous fields are replaced by operators. The second line of the assignment (2.319) has another important implication: for a system of identical particles, we express the probability of an event using the particle number density operator, as is the case with real systems (such as a beam of identical particles). For a single particle, we can speak of the probability density of its occurrence $\psi^*(x)\psi(x)$. The total probability is equal to one, as corresponds to the normalization of the state vector.

Fermions

For fermions, the second quantization proceeds similarly. Once again, we introduce creation and annihilation operators $\hat{b}_k^\dagger$, $\hat{b}_l$ for the k and l states. Due to the antisymmetry of the wave functions, these operators must satisfy the anticommutation relations:

$$|\,k,l\,\rangle = -|\,l,k\,\rangle \quad \Rightarrow$$

$$|\,k,l\,\rangle + |\,l,k\,\rangle = 0 \quad \Rightarrow$$

$$\hat{b}_k^\dagger \hat{b}_l^\dagger + \hat{b}_l^\dagger \hat{b}_k^\dagger = 0\,.$$

We denote anticommutators with curly braces, and relation (2.312), which holds for bosons, takes the following form for fermions:

► $$\begin{aligned} \{\hat{b}_k\,,\hat{b}_l\} &= 0\,, \\ \{\hat{b}_k^\dagger\,,\hat{b}_l^\dagger\} &= 0\,, \\ \{\hat{b}_k\,,\hat{b}_l^\dagger\} &= \delta_{kl}\,. \end{aligned} \tag{2.320}$$

The definitions of continuous operators and the particle number density operator remain the same. For fermions, the commutation relations are replaced everywhere by anticommutation relations. In many situations, the behavior of fermions and bosons differs only in sign (wave function symmetry; commutation and anticommutation relations; Bose-Einstein and Fermi-Dirac statistics).

2.8.4 Example of Second Quantization for Klein-Gordon field

Let us consider the simplest form of the real Klein-Gordon field for a free particle with the Lagrangian

$$\mathscr{L} = \frac{1}{2}(\partial_\alpha \varphi)\left(\partial^\alpha \varphi\right) + \frac{1}{2}\kappa^2 \varphi^2 \tag{2.321}$$

and the field equation

$$\left(\Box - \kappa^2\right)\varphi = 0\,. \tag{2.322}$$

If we transfer to a system of identical particles, the field becomes an operator

$$\varphi \quad \to \quad \hat{\varphi} \tag{2.323}$$

with the following properties

$$\begin{gathered} \left[\hat{\varphi}(x),\ \hat{\varphi}(y)\right] = \left[\hat{\varphi}^\dagger(x), \hat{\varphi}^\dagger(y)\right] = 0\,; \\ \left[\hat{\varphi}(x), \hat{\varphi}^\dagger(y)\right] = \delta(x-y)\,. \end{gathered} \tag{2.324}$$

The variables x and y represent the entire event (time and space). Let us now expand the field operator into plane waves (we will denote the positive-frequency and negative-frequency parts separately):

$$\hat{\varphi}(x) = \int C(\mathbf{k})\left[\hat{a}^\dagger(\mathbf{k})\,\mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{x}-\omega t)} + \hat{a}(\mathbf{k})\,\mathrm{e}^{-\mathrm{i}(\mathbf{k}\cdot\mathbf{x}-\omega t)}\right]\mathrm{d}^3\mathbf{k}\,. \tag{2.325}$$

In the given relation, the time component of the four-vector k^μ is linked to the spatial component via the dispersion relation $\omega = \omega(\mathbf{k})$, so that the integration actually takes place over all four components. The constant $C(\mathbf{k})$ is a normalization constant that ensures that the expansion coefficients (operators $\hat{a}$, $\hat{a}^\dagger$) satisfy the relations for creation and annihilation operators. If we substitute the expansion of the field operator (2.325) into the commutation relations (2.324), we immediately obtain (the correct choice of C ensures a coefficient for the delta function in the second relation to be 1)

$$\begin{gathered} \left[\hat{a}(\mathbf{k}),\ \hat{a}(\mathbf{k}')\right] = \left[\hat{a}^\dagger(\mathbf{k}),\ \hat{a}^\dagger(\mathbf{k}')\right] = 0\,; \\ \left[\hat{a}(\mathbf{k}), \hat{a}^\dagger(\mathbf{k}')\right] = \delta(\mathbf{k}-\mathbf{k}')\,. \end{gathered} \tag{2.326}$$

If we substitute the expansion of the field operator (2.325) into the definitions of the Hamiltonian (1.230) and the momentum (1.229), we obtain, after some elementary manipulations,

$$\begin{gathered} \hat{H} = \frac{1}{2}\int E(\mathbf{k})\left(\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a}\right)\mathrm{d}^3\mathbf{k}\,; \\ \hat{\mathbf{P}} = \frac{1}{2}\int \mathbf{p}(\mathbf{k})\left(\hat{a}\hat{a}^\dagger + \hat{a}^\dagger\hat{a}\right)\mathrm{d}^3\mathbf{k}\,, \end{gathered} \tag{2.327}$$

where we denoted

$$\mathbf{p}(\mathbf{k}) = \hbar\mathbf{k} ;$$
$$E(k) = \hbar\omega(k) = \hbar\sqrt{c^2k^2 + c^2\kappa^2} = \sqrt{p^2c^2 + m_0^2c^4} . \tag{2.328}$$

The relationship $\omega(k)$ is given by the dispersion relation (2.233). Using the commutation relations (2.326), we obtain after substitution the following relations

$$\begin{aligned} \left[\hat{H}, \hat{a}(\mathbf{k})\right] &= -E(\mathbf{k})\hat{a}(\mathbf{k}) ; \\ \left[\hat{H}, \hat{a}^\dagger(\mathbf{k})\right] &= +E(\mathbf{k})\hat{a}^\dagger(\mathbf{k}) ; \\ \left[\hat{\mathbf{P}}, \hat{a}(\mathbf{k})\right] &= -\mathbf{p}(\mathbf{k})\hat{a}(\mathbf{k}) ; \\ \left[\hat{\mathbf{P}}, \hat{a}^\dagger(\mathbf{k})\right] &= +\mathbf{p}(\mathbf{k})\hat{a}^\dagger(\mathbf{k}) . \end{aligned} \tag{2.329}$$

From these relations, the meaning of the operators $\hat{a}$ and $\hat{a}^\dagger$ is clear: the field φ is quantized, and the operator $\hat{a}^\dagger$ creates a field quantum with energy $E(k)$ and momentum $\mathbf{p}(\mathbf{k})$, while the operator $\hat{a}$ annihilates the same quantum. We can interpret this quantum as a zero-spin boson (the field has a single wave function corresponding to a single spin projection). The Klein-Gordon equation acquires a clear interpretation after second quantization. It describes a field that can be understood as a system of excitations – bosons with zero spin.

Complex field

If the Klein–Gordon field were complex, i.e.,

$$\hat{\varphi} = \hat{\varphi}_1 + \mathrm{i}\,\hat{\varphi}_2 ; \qquad \hat{\varphi}^\dagger = \hat{\varphi}_1 - \mathrm{i}\,\hat{\varphi}_2 ,$$

it can be shown that the excitations of such a field correspond to two types of scalar bosons (with spin 0) that are each other's antiparticles.

Normal ordering of operators

There is a problem with the relations for energy and momentum (2.327). If we were to calculate the average values of energy and momentum in the vacuum state, we would obtain infinite values. This is due to the first term, in which the creation operator is on the right-hand side and assigns a non-zero value to the vacuum state:

$$\langle 0|\hat{H}|0\rangle = \frac{1}{2}\int E(\mathbf{k})\left\{\langle 0|\hat{a}\hat{a}^\dagger|0\rangle + \langle 0|\hat{a}^\dagger\hat{a}|0\rangle\right\}\mathrm{d}^3\mathbf{k} =$$
$$= \frac{1}{2}\int E(\mathbf{k})\langle 0|\hat{a}\hat{a}^\dagger|0\rangle\mathrm{d}^3\mathbf{k} \sim \int E(\mathbf{k})\langle 0|0\rangle\mathrm{d}^3\mathbf{k} \sim \int E(\mathbf{k})\,\mathrm{d}^3\mathbf{k} \;\rightarrow\; \infty .$$

This problem stems from the principle of correspondence, which does not address the correct order of the operators in the product. If we have two dynamic variables A and B, we can assign two possible orders of operators to the product AB in quantum theory:

$$AB \quad \rightarrow \quad \begin{cases} \hat{A}\hat{B} , \\ \hat{B}\hat{A} . \end{cases} \tag{2.330}$$

However, only one order will correspond to what happens in nature. This order is not determined by the principle of correspondence; rather, we must choose it so that the results are consistent with our observations. The correct order of operators is called the *normal arrangement*, and we denote it with a colon, i.e.,

► $$:\hat{A}\hat{B}:$$

This notation means that the order of the operators between the colons is not uniquely determined, and we must choose it based on experimental results, for example:

1. We express the operators using the appropriate creation and annihilation operators (boson operators commute, while fermion operators anticommute).
2. In products, we will shift the annihilation operators to the right according to the following rules:
   - If there are two boson operators side by side, we move the annihilation operator to the right;
   - If there is one boson operator and one fermion operator next to each other, we shift the annihilation operator to the right;
   - If there are two fermion operators next to each other, we move the annihilation operator to the right and *change the sign* of that term.

◖ **Example 2.10:**

Find the correct ordering of operators in the Hamiltonian of the Klein-Gordon field:

$$\hat{H} = :\frac{1}{2}\int E(\mathbf{k})\left(\hat{a}\hat{a}^{\dagger}+\hat{a}^{\dagger}\hat{a}\right)\mathrm{d}^3\mathbf{k}: = \frac{1}{2}\int E(\mathbf{k}):\left(\hat{a}\hat{a}^{\dagger}+\hat{a}^{\dagger}\hat{a}\right):\mathrm{d}^3\mathbf{k} =$$
$$= \frac{1}{2}\int E(\mathbf{k})\left(\hat{a}^{\dagger}\hat{a}+\hat{a}^{\dagger}\hat{a}\right)\mathrm{d}^3\mathbf{k} = \int E(\mathbf{k})\,\hat{a}^{\dagger}\hat{a}\,\mathrm{d}^3\mathbf{k} =$$
$$= \int E(\mathbf{k})\,\hat{\mathcal{N}}(\mathbf{k})\,\mathrm{d}^3\mathbf{k}.$$

The average value of the Hamiltonian in the vacuum state no longer diverges; moreover, the structure of the Hamiltonian operator is quite clear: $\hat{\mathcal{N}}(\mathbf{k})$ is the particle number density operator with wave vector **k**. We will similarly modify equation (2.327) for momentum. The correct relations are therefore:

$$\hat{H} = \int E(\mathbf{k})\,\hat{a}^{\dagger}\hat{a}\,\mathrm{d}^3\mathbf{k}\,; \qquad \hat{\mathbf{P}} = \int \mathbf{p}(\mathbf{k})\,\hat{a}^{\dagger}\hat{a}\,\mathrm{d}^3\mathbf{k}\,. \tag{2.331}$$ ◗

◖ **Example 2.11:**

Determine the correct order of the operators (a are boson operators, b are fermion one)

$$\hat{A} = :\frac{1}{2}\left(\hat{a}\hat{a}^{\dagger}+\hat{a}^{\dagger}\hat{a}-\hat{b}\hat{b}^{\dagger}+\hat{b}^{\dagger}\hat{b}+\hat{a}\hat{b}^{\dagger}-\hat{b}^{\dagger}\hat{a}\right):$$

By applying the above rules, we can easily obtain the result

$$\hat{A} = \frac{1}{2}\left(\hat{a}^{\dagger}\hat{a}+\hat{a}^{\dagger}\hat{a}+\hat{b}^{\dagger}\hat{b}+\hat{b}^{\dagger}\hat{b}+\hat{b}^{\dagger}\hat{a}-\hat{b}^{\dagger}\hat{a}\right) = \hat{a}^{\dagger}\hat{a}+\hat{b}^{\dagger}\hat{b} = \hat{N}_a+\hat{N}_b\,.$$ ◗

2.9 Measurement and Hidden Parameters

Quantum theory brings with it a number of unusual phenomena that seem incomprehensible at first glance. People intuitively reject what they cannot imagine, which is why quantum theory has many opponents who take issue primarily with the probabilistic nature of measurement results, the superposition of states, and the nonlocality of quantum theory. Even at the dawn of quantum mechanics, a debate arose between two groups. The leading voice of the first group was Niels Bohr, who held the view that quantum theory is what it is. We must accept that not all mathematical objects used to describe reality are conceivable and that the measurement process includes a measuring instrument that selects a single state from the superposition of states. The wave function, which was a superposition of multiple states before the measurement, describes a single state after the measurement. We call this change the *collapse of the wave function*. Subsequent measurements will always yield the same value (provided we have not completely destroyed the object through measurement). The collapse of the wave function is a nonlinear phenomenon, and this nonlinearity is somehow related to the measurement. The principle of superposition of states leads to the non-locality; e.g., in the double-slit experiment, an electron passes through both the first and second slits simultaneously, and in a Mach–Zehnder interferometer, a photon is present in both arms. It is not a localized particle; however, at the moment of measurement, it is an indivisible object, and the measuring instrument always records it at only a single location.

The second group was represented by Albert Einstein, who could not accept the aforementioned properties of quantum theory and assumed that quantum theory was incomplete. The randomness of the measurement result could be related to the fact that the system has some additional, so-called *hidden parameters*, the ignorance of which leads to the seemingly random result of the measurement. The existence of hidden parameters could also explain the non-local behavior of quantum theory.

In the end, the outcome between the two approaches was decided by experiments conducted in the 1980s, which unequivocally disproved the hidden-parameter theory and demonstrated that we must accept quantum theory along with its strange features.

2.9.1 Measurement and Decoherence

In Section 2.6.4, which discusses the double-slit experiment, we mentioned that measurements taken to determine which slit a particle passed through always disrupt the interference pattern. This is a fundamental property of quantum theory; it is not even necessary to actually perform the measurement – the mere possibility of such a measurement is sufficient for the interference pattern to disappear. Let's demonstrate this property using a Mach–Zehnder interferometer. It is clear that if we place a detector in one of the arms, the constructive and destructive interference responsible for all photons striking only detector D1 will vanish.

Let's imagine that one of the arms does not contain the entire detection apparatus, but only a microscopic object M (such as an atom), which interacts with the photon without destroying it, allowing the photon to continue on its path. Such non-destructive measurements can indeed be performed. Serge Haroche of the French ENS (*École Normale Supérieure*) received the 2012 Nobel Prize in Physics for this work [36]. From a human perspective, this is not a true measurement, since as observers we never learn the result of the photon's interaction with object M.

When a photon interacts with object M, their states become entangled. The two subsystems (the photon and object M) become a single system whose Hilbert space is the direct product $\mathcal{H} = \mathcal{H}_\gamma \otimes \mathcal{H}_M$ of the Hilbert spaces of the two subsystems (this is analogous to the decomposition of a two-dimensional space of real pairs into two Cartesian axes). In Dirac notation, we will write the state of both objects as

$$|\psi\rangle = |\gamma\rangle|M\rangle, \tag{2.332}$$

where the first part describes a photon (γ) and the second a microscopic object M. We refer to this as *entanglement*.

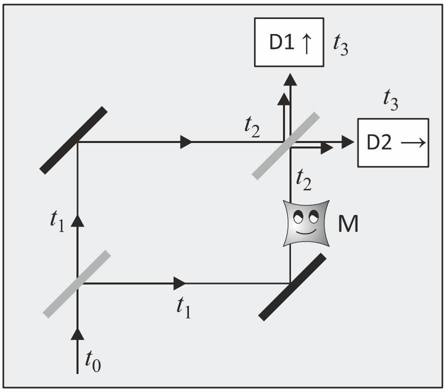


Fig. 2.43: A Mach–Zehnder interferometer with a quantum object M in one of the arms. The basic version of the experiment is shown in Figure 2.38 on page 190.

Let us assume that, as long as object M does not interact with a photon, it is described by the state $|M_0\rangle = |0\rangle$, and after the interaction, its state changes to $|M_1\rangle = |1\rangle$. We will now proceed in the same way as when describing the states of a photon at individual times in Section 2.6.4 on page 190:

$$\begin{aligned}
|\psi(t_0)\rangle &= |\uparrow\rangle|0\rangle;\\
|\psi(t_1)\rangle &= \frac{1}{\sqrt{2}}|\uparrow\rangle|0\rangle + \frac{i}{\sqrt{2}}|\rightarrow\rangle|0\rangle;\\
|\psi(t_2)\rangle &= \frac{i}{\sqrt{2}}|\rightarrow\rangle|0\rangle - \frac{1}{\sqrt{2}}|\uparrow\rangle|1\rangle;\\
|\psi(t_3)\rangle &= \frac{i}{\sqrt{2}}\left[\frac{1}{\sqrt{2}}|\rightarrow\rangle + \frac{i}{\sqrt{2}}|\uparrow\rangle\right]|0\rangle - \frac{1}{\sqrt{2}}\left[\frac{1}{\sqrt{2}}|\uparrow\rangle + \frac{i}{\sqrt{2}}|\rightarrow\rangle\right]|1\rangle.
\end{aligned} \tag{2.333}$$

The resulting state will therefore be

$$|\psi(t_3)\rangle = \frac{i}{2}|\rightarrow\rangle|0\rangle - \frac{i}{2}|\rightarrow\rangle|1\rangle - \frac{1}{2}|\uparrow\rangle|0\rangle - \frac{1}{2}|\uparrow\rangle|1\rangle. \tag{2.334}$$

The coefficients for a photon moving horizontally $|\rightarrow\rangle$ do not cancel each other out, and photons will strike both detectors. This results is known as *decoherence*, i.e., the disappearance of the interference phenomenon due to the presence of object M in the interferometer. Any possibility of interaction between the object and its surroundings leads to the entanglement of states with the surroundings and to quantum decoherence.

2.9.2 Hidden Parameters

Some physicists criticized the non-locality of quantum theory and the randomness of its results. It seemed that these problems could be solved by introducing so-called hidden parameters, whose values we do not know, and which is why we obtain seemingly random results during measurements. Let's illustrate this construct using the Mach–Zehnder interferometer. Suppose that the photons traveling through both arms have some additional unknown property described by parameter p that determines their behavior when interacting with the mirrors. Let's simplify the situation as follows::

1. The parameter p can take only two values: 0 and 1;
2. When interacting with a normal mirror, the value of p does not change;
3. When interacting with a half-transparent mirror, the value of p changes;
4. Photon with $p = 0$ passes through the mirror; a photon with $p = 1$ is reflected.

The parameter p, defined in this way, easily explains all the behavior of photons in a Mach–Zehnder interferometer, as shown in the following figure. On the left is the setup with two half-silvered mirrors (interferometric configuration); on the right, the second half-silvered mirror is missing. Two photons enter the interferometer, one with a hidden parameter $p = 0$ and the other with a hidden parameter $p = 1$. In the interferometric setup, both photons end up in detector D1; in the setup with a single half-transparent mirror, each photon arrives at a different detector. If a stream of photons with a completely random hidden parameter value strikes the apparatus, we obtain results consistent with the quantum description, without resorting to the superposition of states or the non-local behavior of the photon in the interferometer. This design looks promising, but as we will show in Section 2.9.4, the concept of hidden parameters contradicts other experiments. The theory of hidden parameters was definitively refuted in the 1980s.

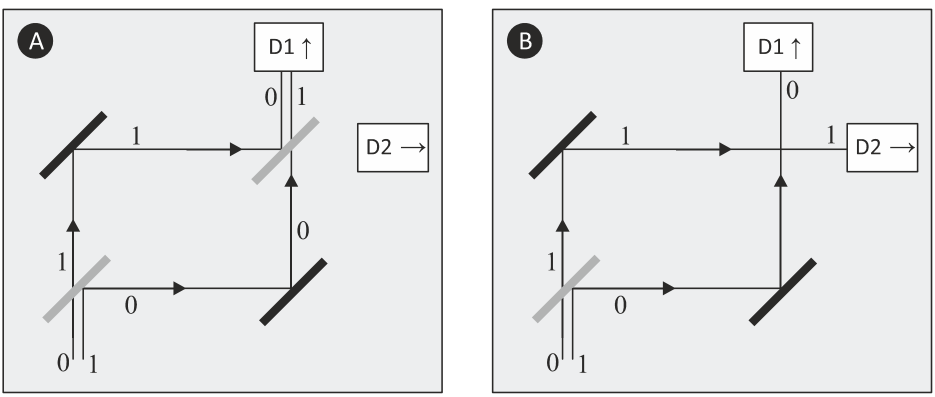


Fig. 2.44: Mach–Zehnder interferometer with hidden photon parameters

2.9.3 EPR Paradox

The rejection of the quantum theory by some physicists led, in 1935, to the formulation of a thought experiment [32] intended to demonstrate the incompleteness of quantum theory and to show that quantum theory is internally inconsistent and will have to be replaced by a better theory. Albert Einstein, the Russian-American physicist Boris Podolsky, and the American-Israeli physicist Nathan Rosen were at the origin of this thought experiment. It is referred to as the EPR paradox, named after their initials.

Today, the formulation by David Bohm from 1951, is most commonly used. Let us imagine a particle with total angular momentum 0 that decays into particles A and B with spin ½, flying apart from each other. The orbital angular momentum of both particles is zero (they are moving apart), and therefore the law of conservation of total angular momentum leads to the condition that if we measure the projection of the spin of one of the particles onto any axis as ½, the other particle must have a projection onto the same axis of −½, and vice versa. Both particles are described by an entangled state

$$|\psi\rangle = \alpha\,|+½\rangle_A\,|-½\rangle_B + \beta\,|-½\rangle_A\,|+½\rangle_B\,, \tag{2.335}$$

which takes both possibilities into account: either particle A has spin ½ and particle B has spin –½, or the opposite is true.

An apparent paradox arises because by measuring the spin projection of one particle, we immediately learn the spin projection of the other particle, no matter how far away it is. It seems as though information travels instantaneously, which contradicts the principle of causality. Again, the much-discussed non-local behavior of particles is to blame. When a measurement is made on one particle, the wave function collapses throughout space, and this manifests itself in the subsequent measurement on the second particle. Opponents of quantum theory claim that both particles possess some hidden parameter that determined in advance how the spin projection measurement would turn out in the future. Proponents of the Copenhagen interpretation, on the other hand, argue that a measurement on either particle can yield any result, and only at the moment of measurement does one of the two possible outcomes occur. In the process, the wave function collapses throughout the entire space, and the subsequent measurement on the remaining particle yields a complementary result. Even in this case, there is no violation of causality; performing a measurement on one particle does not transfer any mass or energy to the other particle, and both observers must in any case use subluminal communication when checking their results, which ensures the causality of both measurements. It will be helpful to reformulate the problem using the photon polarizations.

The polarization of a photon is the plane of oscillation of the electric field, which can generally rotate or remain fixed (plane polarization). Photons have two independent, mutually perpendicular plane polarizations. The actual state of a photon is linear combination of both polarization states. The polarization of a photon can be measured, for example, using a prism made of Icelandic limestone, which is birefringent, and therefore splits a light beam into ordinary and extraordinary rays. Photons traveling in the directions of the ordinary and extraordinary rays have mutually perpendicular polarization. Icelandic limestone can function as a detection device. Let us choose the axes (basis) such that the photon moves in the direction of the z-axis; if it continues along the path of the ordinary beam, it has polarization in the direction of the x-axis; if it follows the path of the extraordinary beam, it has polarization in the direction of the y-axis.

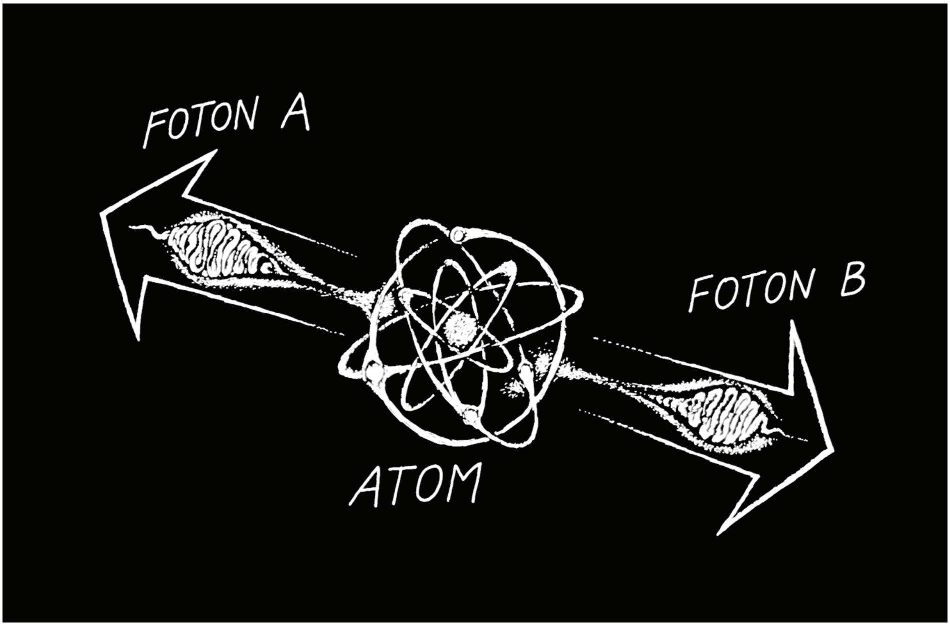


Fig. 2.45: EPR paradox

Let us consider a simplified experiment: suppose that an atom has zero total angular momentum and, upon excitation, emits two photons whose polarizations are correlated; we will even assume that the photons have opposite polarizations. If we measure "horizontal" polarization on one photon in some basis (along the x-axis), the polarization of the second photon will be "vertical" (along the y-axis), and vice versa. It is clear that this is the same formulation as before, but the spin projection is replaced by the photon's polarization. Let us assume that the entangled state of both photons has the form

► $$|\psi\rangle = \frac{1}{\sqrt{2}}|x\rangle_A |y\rangle_B - \frac{1}{\sqrt{2}}|y\rangle_A |x\rangle_B . \qquad (2.336)$$

The first term indicates that, in our reference frame, photon A has polarization $|x\rangle$ and photon B has polarization $|y\rangle$, while the second term describes the opposite situation. Therefore, the measurement can yield only two possible outcomes:

1. A has polarization $|x\rangle$, B has polarization $|y\rangle$;
2. A has polarization $|y\rangle$, B has polarization $|x\rangle$.

The normalization coefficients for the superposition of states are chosen so that both possibilities have the same probability (which is the square of the coefficient, i.e., 1/2) and the sum of both probabilities equals 1. The minus sign in the second term is not essential for our considerations; it merely reflects the fact that if we substituted the same polarization into both arguments, we would obtain a zero wave function, i.e., such a result is not possible. The result is not predetermined; it is entirely random. However, as soon as we perform a measurement on one photon, the state collapses into one of the two possibilities, and the measurement on the second photon yields only a complementary result, regardless of where this photon is physically located. This is a consequence of the entanglement of the two states. Note that an entangled state cannot be expressed as the product of one term with the parameters of photon A and the second one with the parameters of photon B. This is characteristic of entangled states.

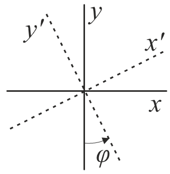


Fig. 2.46: Measuring the polarization of a photon in a rotated basis

What would the result of our measurement be if we chose a different reference frame (rotated by an angle φ), i.e., if the measuring prism made of Icelandic limestone was oriented differently? Let us assume a standard rotation transformation, i.e.,

$$\begin{aligned} |x\rangle &= \cos\varphi\,|x'\rangle - \sin\varphi\,|y'\rangle\,, \\ |y\rangle &= \sin\varphi\,|x'\rangle + \cos\varphi\,|y'\rangle\,, \end{aligned} \tag{2.337}$$

and substitute it into the entangled state (2.336):

$$\begin{aligned} |\psi\rangle = &\frac{1}{\sqrt{2}}\left(\cos\varphi\,|x'\rangle_{\mathrm{A}} - \sin\varphi\,|y'\rangle_{\mathrm{A}}\right)\left(\sin\varphi\,|x'\rangle_{\mathrm{B}} + \cos\varphi\,|y'\rangle_{\mathrm{B}}\right) - \\ &-\frac{1}{\sqrt{2}}\left(\sin\varphi\,|x'\rangle_{\mathrm{A}} + \cos\varphi\,|y'\rangle_{\mathrm{A}}\right)\left(\cos\varphi\,|x'\rangle_{\mathrm{B}} - \sin\varphi\,|y'\rangle_{\mathrm{B}}\right). \end{aligned}$$

After multiplying all the terms, we get

$$|\psi\rangle = \frac{1}{\sqrt{2}}\,|x'\rangle_{\mathrm{A}}\,|y'\rangle_{\mathrm{B}} - \frac{1}{\sqrt{2}}\,|y'\rangle_{\mathrm{A}}\,|x'\rangle_{\mathrm{B}}\,. \tag{2.338}$$

The result is extremely interesting. In the dashed basis, the form of the state vector has not changed at all; expressions (2.336) and (2.338), which describe the state in both bases, are identical. It therefore does not matter how we rotate the measuring prism; the polarization measurement result will, for any rotation of the prism, have a fifty-percent probability that we will measure photon A with "horizontal" polarization and photon B with "vertical" polarization, and a fifty-percent probability of the opposite result.

2.9.4 Bell Inequalities

In 1964, the Irish physicist *John Stewart Bell* showed that the statistical properties of photon polarization measurements would differ under a hidden-parameter theory compared to the standard quantum interpretation. Let us introduce the quantity P, which will have a value of +1 if "horizontal" polarization (in the x direction) was measured in a given basis, and –1 if "vertical" polarization (in the y direction) was measured:

$$P = \begin{cases} +1: & \text{Horizontal polarization;} \\ -1: & \text{Vertical polarization.} \end{cases} \tag{2.339}$$

Suppose that the measurements on photons A and B are performed using prisms oriented at different angles. Their orientation relative to the laboratory frame is given by the angles α and β. If photon A is polarized in the direction α, the result of the polarization measurement will be $P_A = +1$; if it is polarized perpendicular, $P_A = -1$. Similarly, for photon B, the measurement will yield either $P_B = +1$ (polarization is in the direction of angle β) or $P_B = -1$ (perpendicular direction). Let's complicate the situation by assuming that we have two measuring prisms available for each of the photons. For photon A, the prisms will be oriented at angles α, α'; for photon B, at angles β, β'. We will perform the measurement with both prisms available – first with one, then with the other. We will repeat the polarization measurement of both photons with both prisms many times (i.e., four measurements). Let us construct for each measurement a quantity

$$\Gamma \equiv P_A(\alpha)P_B(\beta) + P_A(\alpha)P_B(\beta') + P_A(\alpha')P_B(\beta) - P_A(\alpha')P_B(\beta') . \tag{2.340}$$

From a mathematical point of view, the quantity Γ has one very interesting property: it is a function of four variables $P_A(\alpha)$, $P_A(\alpha')$, $P_B(\beta)$, $P_B(\beta')$, which, in principle, can only take the values +1 or −1. However, the function Γ is constructed such that, when any combination of inputs (even non-physical ones) is substituted, the result is always +2 or −2; there is no other possibility. Try it out.

We will keep records of the measurements and calculate the average values of the measured quantities. If the hidden-parameter theory holds, the measurement result is predetermined by the value of the hidden parameter, and there is no element of randomness whatsoever. The illusory randomness is dictated by our ignorance of the value of the hidden parameter. In many repeated measurements, we will obtain for Γ a sequence composed of the values +2 and −2, e.g., +2, +2, +2, −2, −2, +2… We will calculate the arithmetic mean, which will necessarily lead to the inequality

$$-2 \le \langle \Gamma \rangle \le 2 , \tag{2.341}$$

which, using the definition of Γ, we rewrite as

► $$-2 \le \langle P_A P_B \rangle + \langle P_A P'_B \rangle + \langle P'_A P_B \rangle - \langle P'_A P'_B \rangle \le 2 . \tag{2.342}$$

This is Bell's inequality, which the mean values of repeated measurements must satisfy if there are hidden parameters and the measurement results are predetermined.

However, quantum theory yields a different result when calculating the average value. Let us assume that in an entangled state (2.336)

$$|\psi\rangle = \frac{1}{\sqrt{2}}|x\rangle_A |y\rangle_B - \frac{1}{\sqrt{2}}|y\rangle_A |x\rangle_B$$

we perform a measurement on photon A using a prism oriented at an angle α, and on photon B using a prism oriented at an angle β. We therefore substitute the corresponding transformations into the state:

$$\begin{aligned} |x\rangle_A &= \cos\alpha\,|x'\rangle_A - \sin\alpha\,|y'\rangle_A , \\ |y\rangle_A &= \sin\alpha\,|x'\rangle_A + \cos\alpha\,|y'\rangle_A , \\ |x\rangle_B &= \cos\beta\,|x'\rangle_B - \sin\beta\,|y'\rangle_B , \\ |y\rangle_B &= \sin\beta\,|x'\rangle_B + \cos\beta\,|y'\rangle_B , \end{aligned} \tag{2.343}$$

and obtain

$$|\psi\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} -\sin(\alpha-\beta)\,|x'\rangle_\mathrm{A}\,|x'\rangle_\mathrm{B} + \cos(\alpha-\beta)\,|x'\rangle_\mathrm{A}\,|y'\rangle_\mathrm{B} - \\ -\cos(\alpha-\beta)\,|y'\rangle_\mathrm{A}\,|x'\rangle_\mathrm{B} - \sin(\alpha-\beta\,|y'\rangle_\mathrm{A}\,|y'\rangle_\mathrm{B} \end{bmatrix}. \tag{2.344}$$

The probabilities of the individual outcomes are given by squares of coefficients, i.e.,

Photon polarization P_A	Photon polarization P_B	Probability w_AB
+	+	½ $\sin^2(\alpha-\beta)$
+	−	½ $\cos^2(\alpha-\beta)$
−	+	½ $\cos^2(\alpha-\beta)$
−	−	½ $\sin^2(\alpha-\beta)$

We have used the symbol "+" for "horizontal" polarization and the symbol "–" for "vertical" one. You may be surprised that the probability of finding both photons with the same polarization is nonzero. However, this is due to the fact that the measuring prisms are not oriented identically, and these are projections onto different bases. If you substitute $\alpha=\beta$, the probabilities of both photons having the same polarization will be zero. Using the table we will calculate the weighted average value of $\langle P_\mathrm{A}P_\mathrm{B}\rangle$:

$$\langle P_\mathrm{A}P_\mathrm{B}\rangle = \sum P_\mathrm{A}P_\mathrm{B}w_\mathrm{AB} = w_{++} - w_{+-} - w_{-+} + w_{--} \qquad \Rightarrow$$

$$\langle P_\mathrm{A}P_\mathrm{B}\rangle = \sin^2(\alpha-\beta) - \cos^2(\alpha-\beta) \qquad \Rightarrow$$

$$\langle P_\mathrm{A}P_\mathrm{B}\rangle = -\cos(2\alpha-2\beta)\,. \tag{2.345}$$

In quantum computation, therefore, the following will hold for the average value of Γ:

$$\langle \Gamma\rangle = \langle P_\mathrm{A}P_\mathrm{B}\rangle + \langle P_\mathrm{A}P'_\mathrm{B}\rangle + \langle P'_\mathrm{A}P_\mathrm{B}\rangle - \langle P'_\mathrm{A}P'_\mathrm{B}\rangle \qquad \Rightarrow$$

► $$\langle \Gamma\rangle = -\cos(2\alpha-2\beta) - \cos(2\alpha-2\beta') - \cos(2\alpha'-2\beta) + \cos(2\alpha'-2\beta')\,. \tag{2.346}$$

For various cases, this result clearly violates Bell's inequalities; for example, let $\alpha = 0°$, $\alpha' = 45°$, $\beta = 112{,}5°$, $\beta' = 67{,}5°$. In this case, $\langle\Gamma\rangle = 2\sqrt{2}$, which is in clear contradiction with Bell's inequalities. By testing Bell's inequalities, it is possible to rule out the existence of hidden parameters in quantum theory.

* * *

The first experiments that violated Bell's inequalities were conducted as early as 1972, but were not accepted. Convincing proof of the invalidity of Bell's inequalities was finally provided by Alain Aspect's group in Orsay, France, in experiments conducted between 1976 and 1983. In these experiments, they excited calcium atoms using laser pulses. The excited electron returned to its original energy level via an intermediate state; during the first transition, it emitted a photon with a wavelength of 551.3 nm, and during the second, a photon with a wavelength of 422.7 nm. Both the excited and ground states had zero total angular momentum, while the intermediate state had non-zero total angular momentum, which led to a certain correlation between the polarizations of the two emitted photons. The situation was not as simple as in our illustrative example, but the principle of the experiments was the same. It turned out that Bell's inequalities do not hold, and the random results of the experiments are not a consequence of hidden parameters, but a fundamental property of nature itself.

2.9.5 What's Next?

This concludes our overview of quantum theory. We have presented a selection of fundamental principles and phenomena upon which the thoughtful reader can build further. Quantum theory is one of the most successful theories ever developed by humankind. The Standard Model of elementary particles and forces has become an accurate quantum description of processes governed by electromagnetic, strong, and weak interactions. Quantum electrodynamics sparked the electronic revolution, the results of which we encounter at every turn. In recent decades, quantum theory has increasingly influenced our understanding of information storage and processing. Quantum encryption has become a commercial reality, quantum teleportation is in the stage of successful experiments, and quantum computers are a clearly defined goal awaiting sufficient technological infrastructure. The exploration of the universe cannot do without quantum theory either. Matter in white dwarfs and neutron stars is governed by quantum laws, and in the early moments of the Big Bang, under extreme densities and temperatures, the quantum properties influenced the universe's evolution.

The fundamentals of quantum theory and some model calculations should hopefully enable the reader to pursue further study of specialized fields that lie beyond the scope of this textbook. We will encounter quantum processes influencing the behavior of large numbers of particles again in statistical physics in the follow-up textbook [1].

Fig. 2.47. The founders of quantum theory at the Solvay Conference (1927). First row: I. Langmuir, M. Planck, M. Curie, H. Lorentz, A. Einstein, P. Lanagevin, C. Guye, C. Wilson, O. W. Richardson. Second row: P. Debye, M. Knudsen, W. Bragg, H. Kramers, P. Dirac, A. Compton, L. Broglie, M. Born, N. Bohr. Třetí řada: A. Piccard, E. Henriot, P. Ehrenfest, E. Herzen, T. Donder, E. Schrödinger, E. Verschaffelt, W. Pauli, W. Heisenberg, R. H. Fowler, L. Brillouin.

3. Mathematics for Physics

3.1 Einstein Summation Convention

3.1.1 Introduction of the Summation Convention

If two identical indices appear in an expression, we automatically sum over them. We will denote summation indices with lowercase letters ($i, j, k, ...$):

$$a_1 b_1 + a_2 b_2 + \cdots + a_N b_N = \sum_{k=1}^{N} a_k b_k = a_k b_k \, . \tag{3.1}$$

The labeling of the summation index doesn't matter; we can change it as we like:

$$a_k b_k = a_l b_l = a_n b_n = a_1 b_1 + a_2 b_2 + \cdots + a_N b_N \, . \tag{3.2}$$

In the following examples, take a close look at how the summation convention is used. At the same time, you'll review some basic mathematical concepts. We purposely use different notations for summation indices in the various examples; this is the only way to get used to this useful notation.

3.1.2 Simple Examples

Scalar product of two vectors

$$\begin{aligned} &\mathbf{a} = (a_1, a_2, \ldots, a_N) \, ; \quad \mathbf{b} = (b_1, b_2, \ldots, b_N) \, ; \\ &\mathbf{a} \cdot \mathbf{b} \equiv a_1 b_1 + a_2 b_2 + \cdots + a_N b_N = a_j b_j \, . \end{aligned} \tag{3.3}$$

Divergence

$$\begin{aligned} &\mathbf{T} \equiv (T_1 \, , T_2 \, , T_3) \, ; \\ &\operatorname{div} \mathbf{T} = \frac{\partial T_1}{\partial x_1} + \frac{\partial T_2}{\partial x_2} + \frac{\partial T_3}{\partial x_3} = \frac{\partial T_i}{\partial x_i} \, . \end{aligned} \tag{3.4}$$

Matrix multiplication

$$\begin{aligned} &\mathbf{A} = \{ a_{ij} \} \, ; \qquad \mathbf{B} = \{ b_{ij} \} ; \\ &\{ \mathbf{A} \cdot \mathbf{B} \}_{ij} = \sum_{k=1}^{N} a_{ik} b_{kj} = a_{ik} b_{kj} \, . \end{aligned} \tag{3.5}$$

Free indices are indices that appear on both sides of an equality (here, i and j). The sum does not pass over a free index. A dummy (bound) index is a pair of identical indices in a single mathematical term over which the sum passes (here, k).

Small finite increment of a single-variable function

Let's consider a single-variable real function $f(q)$ that assigns the value f to the value q:

► $$f(q): \quad q \to f \; ; \qquad \text{thereafter } \Delta f \cong \frac{\mathrm{d}f}{\mathrm{d}q} \Delta q \, . \tag{3.6}$$

The validity of this approximation is precisely defined by Lagrange theorem of increments. Let's illustrate its application using a sphere of radius r, whose volume is

$$V(r) = \frac{4}{3} \pi r^3 \, . \tag{3.7}$$

We change the radius by Δr. For small Δr, volume changes approximately by the value

$$\Delta V \cong \frac{\mathrm{d}V}{\mathrm{d}r} \Delta r = 4\pi r^2 \Delta r \, . \tag{3.8}$$

The meaning is clear: $4\pi r^2$ is the surface area of a sphere with radius r, and Δr is the thickness of this surface. The product represents small change in the sphere's volume.

Small finite increment of a multivariable function

Let f be a function of several real variables ($q_1, q_2, \ldots q_N$), which assigns the value f to the values of $\mathbf{q}$:

$$f(q_1, \ldots, q_N): \qquad q_1, \ldots, q_N \to f \, . \tag{3.9}$$

Lagrange theorem regarding the increment of this function can be simply written as (without higher-order terms)

► $$\Delta f \cong \frac{\partial f}{\partial q_1} \Delta q_1 + \frac{\partial f}{\partial q_2} \Delta q_2 + \cdots + \frac{\partial f}{\partial q_N} \Delta q_N = \frac{\partial f}{\partial q_k} \Delta q_k \tag{3.10}$$

In the final part of the formula, we used Einstein summation convention. Let's see how this theorem applies to the change in the volume of a cylinder. The volume of a cylinder is a function of two variables (the radius of the base and the height):

$$V(r,h) = \pi r^2 h \, . \tag{3.11}$$

We can easily calculate the increment

$$\Delta V \cong \frac{\partial f}{\partial q_k} \Delta q_k = \frac{\partial V}{\partial r} \Delta r + \frac{\partial V}{\partial h} \Delta h = 2\pi r h \Delta r + \pi r^2 \Delta h \, . \tag{3.12}$$

The first contribution comes from a change in the base radius, and the second from a change in the cylinder height.

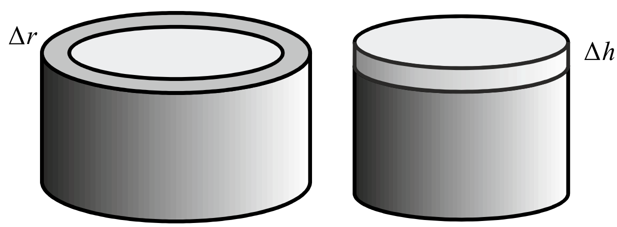


Fig. 3.1: The increment theorem applied to a cylinder

Infinitesimal (infinitely small) increment of a multivariable function

If we introduce infinitesimal changes instead of small increments, we obtain the so-called first differential of the function

► $$\mathrm{d}f = \frac{\partial f}{\partial q_1}\cdot \mathrm{d}q_1 + \cdots + \frac{\partial f}{\partial q_N}\cdot \mathrm{d}q_N = \frac{\partial f}{\partial q_j}\cdot \mathrm{d}q_j \,. \qquad (3.13)$$

> **Note:** These relationships can be formulated more precisely using Lagrange increment theorem and the first differential theorem. For our purposes, however, it suffices to remember that Lagrange theorem deals with a finite increment and is an approximate relationship, whereas the first differential deals with an infinitely small increment and is an exact relationship.

Derivative of a composite function:

If the internal variable q_i depends on time, then the total time derivative is given by:

► $$f = f(q_1, q_2, \cdots, q_N)\,;$$
$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial q_1}\cdot\frac{\mathrm{d}q_1}{\mathrm{d}t} + \cdots + \frac{\partial f}{\partial q_N}\cdot\frac{\mathrm{d}q_N}{\mathrm{d}t} = \frac{\partial f}{\partial q_k}\cdot\frac{\mathrm{d}q_k}{\mathrm{d}t} = \frac{\partial f}{\partial q_k}\cdot\dot{q}_k \,. \qquad (3.14)$$

We will illustrate both the first differential and the derivative of a composite function using the transformation between polar and Cartesian coordinates:

$$x(t) = r(t)\cos\varphi(t)\,,$$
$$y(t) = r(t)\sin\varphi(t)\,; \qquad (3.15)$$

$$\mathrm{d}x = \frac{\partial x}{\partial r}\mathrm{d}r + \frac{\partial x}{\partial \varphi}\mathrm{d}\varphi = \cos\varphi\,\mathrm{d}r - r\sin\varphi\;\;\mathrm{d}\varphi\,,$$
$$\mathrm{d}y = \frac{\partial y}{\partial r}\mathrm{d}r + \frac{\partial y}{\partial \varphi}\mathrm{d}\varphi = \sin\varphi\,\mathrm{d}r + r\;\cos\varphi\,\mathrm{d}\varphi\,; \qquad (3.16)$$

$$\dot{x} = \dot{r}\cos\varphi - r\dot{\varphi}\sin\varphi,$$
$$\dot{y} = \dot{r}\sin\varphi + r\dot{\varphi}\cos\varphi\,. \qquad (3.17)$$

On the symbolism of Cartesian coordinates

Let's now look at the various ways of writing the same expression. It's not important to dwell on which notation to choose, but rather to understand what the notations mean. For example, in printed publications, a vector is denoted by bold type, but on a blackboard, where this isn't possible, an arrow is placed above the symbol:

$$\mathbf{x} \equiv \vec{x}\,; \qquad \mathbf{a}\cdot\mathbf{b} \equiv \vec{a}\cdot\vec{b}\,. \qquad (3.18)$$

For $f = f(\mathbf{x}, \mathbf{v})$, we can express the gradients (spatial and velocity) in many ways:

$$\nabla f \equiv \nabla_{\mathbf{x}} f \equiv \frac{\partial f}{\partial \vec{x}} \equiv \frac{\partial f}{\partial \mathbf{x}} \equiv \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right);$$
$$\nabla_{\mathbf{v}} f \equiv \frac{\partial f}{\partial \vec{v}} \equiv \frac{\partial f}{\partial \mathbf{v}} \equiv \left(\frac{\partial f}{\partial v_x}, \frac{\partial f}{\partial v_y}, \frac{\partial f}{\partial v_z}\right). \tag{3.19}$$

A spatial gradient is often written using only its components:

$$\frac{\partial f}{\partial x_k} \equiv \partial_k f \equiv f_{,k} \; . \tag{3.20}$$

The square of a vector's magnitude, such as velocity, can also be written in many ways:

$$v^2 = \mathbf{v}^2 = \mathbf{v} \cdot \mathbf{v} = v_j v_j = v_1^2 + v_2^2 + v_3^2 \; . \tag{3.21}$$

Let's now find the rate of change of the expression $f(\mathbf{v}) = \mathbf{v}^2$:

$$\frac{\partial f}{\partial v_i} = \frac{\partial}{\partial v_i} v_j v_j = \delta_{ji} v_j + v_j \delta_{ji} = v_i + v_i = 2 v_i \; . \tag{3.22}$$

Faster symbolic notation is (as if we were differentiating with respect to the symbol **v**):

$$\frac{\partial f}{\partial \mathbf{v}} = \frac{\partial \mathbf{v}^2}{\partial \mathbf{v}} = 2\mathbf{v} \; . \tag{3.23}$$

Note 1: The gradient operator points in the direction of the greatest increase in the function and is perpendicular to the isosurfaces (surfaces where the function takes a constant value). This is evident from the decomposition.

$$f(\mathbf{x}) = \text{const} \quad \Rightarrow \quad \frac{\partial f}{\partial x_k} \mathrm{d}x_k = 0 \quad \Rightarrow \quad \nabla f \cdot \mathrm{d}\mathbf{x} = 0 \; . \tag{3.24}$$

The vector $\mathrm{d}\mathbf{x}$ is in the isosurface, and ∇f is perpendicular to it (zero dot product).

Note 2: The gradient operator need not apply only to spatial variables; as seen in the previous examples, it can also apply to the velocity field or other variables. We will explore the detailed application of this operator and the history of its introduction in mathematics and physics in Chapter 3.5 *From Gradient to Helicity*.

Note 3: The scalar product of the nabla operator with a vector field is called the divergence; div $\mathbf{K} \equiv \nabla \cdot \mathbf{K} = \partial K_k / \partial x_k$. This is a test to determine whether the field has a source at a given point. If div $\mathbf{K} > 0$, the field is emerging at that point; if div $\mathbf{K} < 0$, the field is vanishing; and if div $\mathbf{K} = 0$, the field is simply passing.

Note 4: The vector product of the nabla operator with a vector field is called the rotation; rot $\mathbf{K} \equiv \nabla \times \mathbf{K}$. This is a test to determine whether there is a vortex at a given point. It must be a vector test – that is, three tests related to viewing the vortex from the direction of the coordinate axes. If the only nonzero component of rot **K** is nonzero, there is a vortex center and its axis points along the vector rot **K**.

Note 5: The scalar product of a vector with its rotation is used to detect helices; readers can find details in the chapter 3.5 *From Gradient to Helicity*.

3.1.3 Line Element

We define a line element as the square of the infinitesimal distance between two points. In the Cartesian coordinate system, the Pythagorean theorem holds, and it does not matter whether the distance is finite or infinitesimal; in both cases, the exact relationship holds:

$$\Delta l^2 = \Delta x^2 + \Delta y^2 \ ; \qquad \mathrm{d}l^2 = \mathrm{d}x^2 + \mathrm{d}y^2 \ . \tag{3.25}$$

In the polar coordinate system, the situation is different. For finite increments, only an approximate relationship holds, since we have replaced one side of a right-angled triangle with an arc (see Fig. 3.2):

$$\Delta l_{\mathrm{pol}}^2 \doteq \Delta r^2 + r^2 \Delta\varphi^2 \ ; \qquad \mathrm{d}l_{\mathrm{pol}}^2 = \mathrm{d}r^2 + r^2 \mathrm{d}\varphi^2 \ ; \tag{3.26}$$

For infinitesimally small distances, the approximate equalities once again become exact.

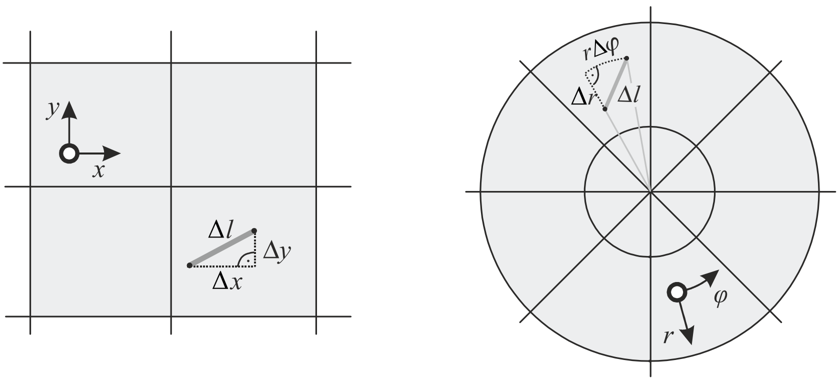


Fig. 3.2: Line element in Cartesian and polar coordinate systems

In orthogonal coordinate systems (where the coordinate axes are mutually perpendicular), a line element can generally be expressed by the equation

► $$\mathrm{d}l^2 = g_{11}\mathrm{d}q_1^2 + g_{22}\mathrm{d}q_2^2 + g_{33}\mathrm{d}q_3^2 \ , \tag{3.27}$$

In non-orthogonal systems, it generally holds that the line element is a quadratic function of the increments:

► $$\mathrm{d}l^2 = g_{ij}\mathrm{d}q_i\mathrm{d}q_j \ . \tag{3.28}$$

Note that the summation convention applies. The g_{ij} coefficients are called the *metric tensor*. They can be determined either geometrically (see the Figure above) or from the differentials of the transformation relations for the coordinates. For polar coordinates, these relations are (3.16). We proceed analogously for other coordinate systems:

Polar coordinates:

$$\begin{aligned} x &= r\cos\varphi \\ y &= r\sin\varphi \end{aligned} \quad ; \qquad \mathrm{d}l^2 = \mathrm{d}r^2 + r^2\mathrm{d}\varphi^2 \,. \tag{3.29}$$

Spherical coordinates:

$$\begin{aligned} x &= r\cos\varphi\sin\theta \\ y &= r\sin\varphi\sin\theta \\ z &= r\cos\theta \end{aligned} \quad ; \qquad \mathrm{d}l^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\varphi^2 \,. \tag{3.30}$$

Cylindrical coordinates:

$$\begin{aligned} x &= r\cos\varphi \\ y &= r\sin\varphi \\ z &= z \end{aligned} \quad ; \qquad \mathrm{d}l^2 = \mathrm{d}r^2 + r^2\mathrm{d}\varphi^2 + \mathrm{d}z^2 \,. \tag{3.31}$$

We can then easily determine the kinetic energy of the system in generalized coordinates using a line element from the following equation:

$$T(\mathbf{q},\dot{\mathbf{q}}) = \frac{1}{2}\,m\frac{\mathrm{d}l^2}{\mathrm{d}t^2} = \frac{1}{2}m\,g_{ij}\,\frac{\mathrm{d}q_i\,\mathrm{d}q_j}{d\,t^2} = \frac{1}{2}\,m\,g_{ij}\,\dot{q}_i\,\dot{q}_j \,. \tag{3.32}$$

Specifically for the previous coordinates, the following applies:

$$\begin{aligned}
&\text{Cartesian} && T(\dot{x},\dot{y},\dot{z}) = \frac{1}{2}m\,(\dot{x}^2+\dot{y}^2+\dot{z}^2) \\
&\text{Polar} && T(r,\dot{r},\dot{\varphi}) = \frac{1}{2}m\,(\dot{r}^2+r^2\dot{\varphi}^2) \\
&\text{Spherical} && T(r,\theta,\dot{r},\dot{\varphi},\dot{\theta}) = \frac{1}{2}m\,(\dot{r}^2+r^2\dot{\theta}^2+r^2\sin^2\theta\,\dot{\varphi}^2) \\
&\text{Cylindrical} && T(r,\dot{r},\dot{\varphi},\dot{z}) = \frac{1}{2}m\,(\dot{r}^2+r^2\dot{\varphi}^2+\dot{z}^2)
\end{aligned} \tag{3.33}$$

►

In each coordinate system, kinetic energy breaks down into the sum of terms corresponding to the individual degrees of freedom. For example, in polar coordinates, kinetic energy consists of a radial component T_r and a rotational component T_φ.

Note: The magnitude of kinetic energy cannot depend on the choice of coordinate system; kinetic energy is a scalar function of generalized coordinates. Another scalar function is, for example, potential energy.

3.2 Complex Numbers and Functions

Complex numbers were first used by Italian mathematicians in the 17th century to solve algebraic equations. The golden age of complex numbers was the 18th century, when they became an integral part of mathematical and physical procedures. The French mathematician Abraham de Moivre (1667–1754), the Swiss mathematician Johann Bernoulli (1667–1748), and his student Leonhard Euler (1707–1803), who introduced the well-known symbol “i” for √(−1) and began interpreting complex numbers as points in the plane, and, of course, the German mathematician Carl Friedrich Gauss (1777–1855), who perfected this concept. The Irish mathematician William Rowan Hamilton (1805–1865) contributed most significantly to the generalization of complex numbers to quaternions (using four axes).

3.2.1 Complex Number Representation

Algebraic, rectangular, and polar form

We most often think of complex numbers as an extension of the real axis by multiples of the imaginary unit i, whose fundamental property is

$$\mathrm{i}^2 = -1 \,, \tag{3.34}$$

The algebraic form of a complex number is therefore

$$f = x + \mathrm{i}\, y \,. \tag{3.35}$$

Multiplying two complex numbers in algebraic form results in a somewhat complicated expression

$$f_1 f_2 = (x_1 + \mathrm{i}\, y_1)(x_2 + \mathrm{i}\, y_2) = (x_1 x_2 - y_1 y_2) + \mathrm{i}\,(x_1 y_2 + y_1 x_2) \,. \tag{3.36}$$

A complex number represents a pair of real numbers that can be interpreted as points in the plane. We can always construct a complex number from these points, or conversely, separate the two coordinates from a complex number. We can therefore write

$$f = x + \mathrm{i}\, y = (x, y) \,. \tag{3.37}$$

If we use polar coordinates, we can represent a complex number using a different pair of values – the distance from the origin A (amplitude) and the azimuth φ (phase):

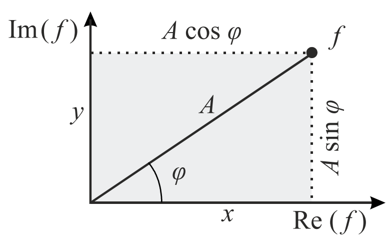


Fig. 3.3: Rectangular and polar representations of a complex number

So at this point, we have three representations of a complex number: algebraic, rectangular, and polar:

$$f = x + \mathrm{i}\, y = (x, y) = [A, \varphi]\,. \tag{3.38}$$

In any case, a complex number is always a pair of real numbers. If we know the polar coordinates (magnitude and phase), we can easily determine the Cartesian coordinates (real and imaginary parts):

► $$\begin{aligned} x &= A\cos\varphi, \\ y &= A\sin\varphi. \end{aligned} \tag{3.39}$$

The inverse transformation can be derived from the Pythagorean theorem and the definition of the tangent of an angle:

► $$\begin{aligned} A &= \sqrt{x^2 + y^2}\,, \\ \varphi &= \operatorname{atg}(y/x). \end{aligned} \tag{3.40}$$

Complex conjugation

A reflection is a transformation in which a complex number is symmetrically reflected across the real (horizontal) axis:

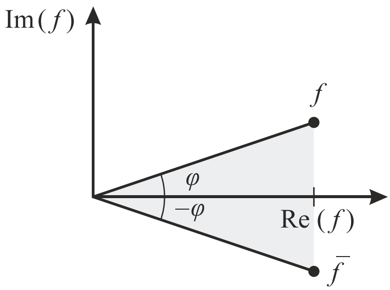


Fig. 3.4: Reflection, or complex conjugate

In this transformation, the imaginary part of the number Im f changes to −Im f. We call the result of this reflection a *complex conjugate*, denoted by a bar or an asterisk:

► $$\bar{f} = f^{*} = x - \mathrm{i}\, y\,. \tag{3.41}$$

Complex conjugation is a useful transformation; we can use it to easily find the amplitude of a complex number, since the following holds:

► $$\bar{f}\, f = (x - \mathrm{i}\, y)(x + \mathrm{i}\, y) = x^2 + y^2 = A^2\,. \tag{3.42}$$

Using conjugation, we can also easily express the real and imaginary parts:

► $$\operatorname{Re} f = \frac{1}{2}\left(f + \bar{f}\right), \tag{3.43}$$

► $$\operatorname{Im} f = \frac{1}{2\mathrm{i}}\left(f - \bar{f}\right), \tag{3.44}$$

3.2.2 Exponential Form

A complex number can be expressed using the exponential function. Let's first define this function as an infinite series. Let's find a function whose derivative is equal to the function itself:

$$F'(x) = F(x) \tag{3.45}$$

Then the second, third, and any higher derivative will also be equal to the original function. In short, this function will be invariant under differentiation. Let us seek such a special function as an infinite series

$$F(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + c_5 x^5 + \cdots . \tag{3.46}$$

We will take the derivative term by term:

$$F'(x) = c_1 + 2c_2 x + 3c_3 x^2 + 4c_4 x^3 + 5c_5 x^4 + \cdots . \tag{3.47}$$

If the last two functions are to be equal (a function is equal to its first derivative), the following must hold:

$$c_1 = c_0 , \quad 2c_2 = c_1 , \quad 3c_3 = c_2 , \quad 4c_4 = c_3 , \quad 5c_5 = c_4 , \quad \cdots \tag{3.48}$$

If we choose a constant c_0, we can calculate all the expansion coefficients. Choosing $c_0 = 0$ will result in a zero function; any nonzero number will generate the function we are looking for. The value of c_0 is irrelevant and will merely serve as a scaling factor for this function. Therefore, we choose $c_0 = 1$:

$$c_0 = 1 , \quad c_1 = 1 , \quad c_2 = \frac{1}{2} , \quad c_3 = \frac{1}{3 \cdot 2} , \quad c_4 = \frac{1}{4 \cdot 3 \cdot 2} , \quad c_5 = \frac{1}{5 \cdot 4 \cdot 3 \cdot 2} \cdots \tag{3.49}$$

It's fairly easy to come up with a general formula:

$$c_n = \frac{1}{n!} . \tag{3.50}$$

The function found is called the exponential function, and its expansion is given by

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots . \tag{3.51}$$

For the zero argument, the function we found takes the value 1. For the argument 1, the result is what is known as Euler's number:

$$\exp(0) = 1 , \tag{3.52}$$

$$\exp(1) = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \cdots = 2{,}71828183\ldots = \mathrm{e} . \tag{3.53}$$

Alternative method for finding the exponential

We can also find a function "immune" to differentiation from a differential equation

$$\frac{\mathrm{d}F}{\mathrm{d}x} = F \quad \Rightarrow \quad \frac{\mathrm{d}F}{F} = \mathrm{d}x \quad \Rightarrow \quad \ln F = x + C \quad \Rightarrow \quad F(x) = K\,\mathrm{e}^x . \tag{3.54}$$

Since $F(0) = 1$, K must be 1. We can therefore see that the series we have found is an exponential function with a base equal to Euler's number:

► $$\exp(x) = \mathrm{e}^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots. \quad (3.55)$$

In mathematics, it is very common for functions to be defined using infinite series, and their values are usually calculated from such series (for example, on your calculator). If we select only the even powers from the expansion of the exponential, we get the hyperbolic cosine (remember that it has a value of 1 at zero and is turned upward, similar to a parabola). If we select only the even powers and alternate their signs, we get the ordinary cosine. The alternating signs will cause the polynomials forming the series to alternate between turning downward and upward, resulting in a periodic function. If we select the odd powers, the function is called the hyperbolic sine, and if we select the odd powers and alternate their signs, we obtain the standard sine:

► $$\begin{aligned} \exp x &\equiv 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots, \\ \mathrm{ch}\, x &\equiv 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \frac{x^8}{8!} + \cdots, \\ \cos x &\equiv 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} \pm \cdots, \\ \mathrm{sh}\, x &\equiv x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \frac{x^9}{9!} \pm \cdots, \\ \sin x &\equiv x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} - \frac{x^9}{9!} \pm \cdots \end{aligned} \quad (3.56)$$

The coincidence with their classical definition from trigonometry can be proven using Taylor series expansions. There are a number of interesting relationships among the functions in the table; one of the best known is Euler's formula. Let's try to find the exponential function with a purely imaginary argument (using its series):

$$\begin{aligned} \exp(\mathrm{i}\varphi) &= 1 + (\mathrm{i}\varphi) + \frac{(\mathrm{i}\varphi)^2}{2!} + \frac{(\mathrm{i}\varphi)^3}{3!} + \frac{(\mathrm{i}\varphi)^4}{4!} + \frac{(\mathrm{i}\varphi)^5}{5!} + \cdots = \\ &= 1 + \mathrm{i}x - \frac{\varphi^2}{2!} - \mathrm{i}\frac{\varphi^3}{3!} + \frac{\varphi^4}{4!} + \mathrm{i}\frac{\varphi^5}{5!} + \cdots = \\ &= \left(1 - \frac{\varphi^2}{2!} + \frac{\varphi^4}{4!} \pm \cdots\right) + \mathrm{i}\left(x - \frac{\varphi^3}{3!} + \frac{\varphi^5}{5!} \mp \cdots\right). \end{aligned}$$

The first parentheses contain the cosine term, and the second the sine term. Therefore:

► $$\mathrm{e}^{\mathrm{i}\varphi} = \cos\varphi + \mathrm{i}\sin\varphi. \quad (3.57)$$

Euler formula is extremely useful when expressing complex numbers:

$$f = x + \mathrm{i}y = A\cos\varphi + \mathrm{i}\,A\sin\varphi = A\left(\cos\varphi + \mathrm{i}\sin\varphi\right) = A\,\mathrm{e}^{\mathrm{i}\varphi}. \tag{3.58}$$

This is known as the exponential representation of a complex number. We can now write a complex number in four different ways:

► $$f = x + \mathrm{i}\,y = (x, y) = [A, \varphi] = A\,\mathrm{e}^{\mathrm{i}\varphi}. \tag{3.59}$$

In exponential form, we can easily interpret the multiplication of two complex numbers:

$$f\,g = A\,\mathrm{e}^{\mathrm{i}\varphi}\,B\,\mathrm{e}^{\mathrm{i}\psi} = AB\,\mathrm{e}^{\mathrm{i}(\varphi+\psi)}. \tag{3.60}$$

The product of two complex numbers has an amplitude equal to the product of their amplitudes and a phase equal to the sum of the phases of the original numbers:

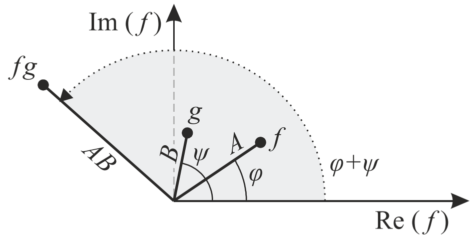


Fig. 3.5: Graphical representation of the multiplication of two complex numbers

A very special meaning has a number in the form

$$g = \mathrm{e}^{\mathrm{i}\alpha}. \tag{3.61}$$

The magnitude of this number is equal to one, meaning the number lies on the unit circle in the complex plane. This is the name given to the plane in which we plot the real and imaginary parts of a complex number on the Cartesian coordinate system. If we multiply any other complex number by this number (the so-called complex unit), the result is its rotation by an angle α in the positive mathematical direction (counterclockwise):

$$f\,\mathrm{e}^{\mathrm{i}\alpha} = A\,\mathrm{e}^{\mathrm{i}\varphi}\,\mathrm{e}^{\mathrm{i}\alpha} = A\,\mathrm{e}^{\mathrm{i}(\varphi+\alpha)}. \tag{3.62}$$

In exponential form, complex numbers can be easily multiplied and rotated around the origin of the coordinate system. It is also possible to derive various relationships very elegantly; as an example, let's consider one of the sum formulas:

$$\cos(\alpha+\beta) = \mathrm{Re}\left(\mathrm{e}^{\mathrm{i}(\alpha+\beta)}\right) = \mathrm{Re}\left(\mathrm{e}^{\mathrm{i}\alpha}\,\mathrm{e}^{\mathrm{i}\beta}\right) =$$

$$= \mathrm{Re}\left[\left(\cos\alpha + \mathrm{i}\sin\alpha\right)\left(\cos\beta + \mathrm{i}\sin\beta\right)\right] =$$

$$= \cos\alpha\cos\beta - \sin\alpha\sin\beta$$

We can derive other relationships in a similar manner; it suffices to apply Euler's relation (3.57). A summary of useful expressions can be found in the following box:

$$\blacktriangleright \quad \begin{gathered} e^{ix} = \cos x + i\sin x\,, \\ \cos(x+y) = \cos x\cos y - \sin x\sin y\,, \\ \sin(x+y) = \sin x\cos y + \sin y\cos x\,, \\ \operatorname{ch} x = \frac{e^{x}+e^{-x}}{2}\,, \qquad \operatorname{sh} x = \frac{e^{x}-e^{-x}}{2}\,, \\ \cos x = \frac{e^{ix}+e^{-ix}}{2}\,, \quad \sin x = \frac{e^{ix}-e^{-ix}}{2i}\,, \\ \cos(-x) = \cos x\,, \qquad \sin(-x) = -\sin x\,. \end{gathered} \tag{3.63}$$

3.2.3 Rotation in a Plane

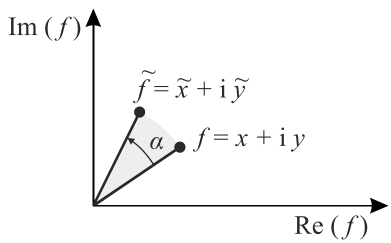


Fig. 3.6: Rotation of a point in the complex plane

We can express the rotation of a point in a plane very simply using complex numbers

$$\tilde{f} = \tilde{x} + i\,\tilde{y} = f\,e^{i\alpha} = (x+i\,y)(\cos\alpha + i\sin\alpha) = (x\cos\alpha - y\sin\alpha) + i(x\sin\alpha + y\cos\alpha)\,.$$

By comparing the start and end of the formula, it is clear that the coordinates of the rotated point are (separating the real and imaginary parts)

$$\begin{aligned} \tilde{x} &= x\cos\alpha - y\sin\alpha, \\ \tilde{y} &= x\sin\alpha + y\cos\alpha. \end{aligned} \tag{3.64}$$

If it is not the point but the coordinate system that rotates, it is sufficient to change the angle α to $-\alpha$. The resulting relationship will be in matrix form

$$\blacktriangleright \quad \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{3.65}$$

This is a typical rotation matrix. Note that its determinant is equal to one – this is characteristic for rotation transformations. Now let's write this transformation for a very small angle (preferably infinitesimal). From the expansion of the trigonometric functions, it suffices to keep only the first nonzero term:

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = \begin{pmatrix} 1 & \alpha \\ -\alpha & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{3.66}$$

We can now decompose this matrix into an identity matrix and a remainder:

$$\mathbb{R}_{\text{inf}} = \begin{pmatrix} 1 & \alpha \\ -\alpha & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \alpha \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{3.67}$$

The first matrix is the identity matrix and does nothing to the coordinates; the second matrix swaps the coordinates (x, y) and reverses the sign of the second coordinate. It is common practice to factor out the imaginary unit from the second matrix:

$$\mathbb{R}_{\text{inf}} = \begin{pmatrix} 1 & \alpha \\ -\alpha & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \mathrm{i}\alpha \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}. \tag{3.68}$$

There are two reasons for this adjustment: 1) the coefficients of the matrices correspond to the first two terms of the exponential expansion $\exp[\mathrm{i}\alpha] = 1 + \mathrm{i}\alpha$; 2) the new matrix has real eigenvalues and orthogonal eigenvectors. We call it the *generator of rotation.*

► $$\mathbb{M} = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}. \tag{3.69}$$

The following holds for a rotation transformation in the plane:

► $$\mathbb{R}_{\text{inf}} = \mathbb{I} + \mathrm{i}\alpha\mathbb{M}\,. \tag{3.70}$$

Let's now construct a final rotation by combining many small rotations. We'll divide the total angle into n equal small angles ($n \to \infty$)

$$\alpha = n\frac{\alpha}{n} \tag{3.71}$$

and repeat n times the rotation transformation (3.70) by an angle of α/n:

$$\mathbb{R} = \lim_{n\to\infty} \mathbb{R}_{\text{inf}}^{n} = \lim_{n\to\infty} \left(\mathbb{I} + \mathrm{i}\frac{\alpha}{n}\mathbb{M} \right)^{n}. \tag{3.72}$$

Since $(1+x/n)^n \to \mathrm{e}^x$, it follows that

$$\mathbb{R} = \mathrm{e}^{\mathrm{i}\alpha\mathbb{M}}\,. \tag{3.73}$$

This is an elegant way to express a rotation by a finite angle α using a rotation generator. The exponential function of the matrix is defined by the expansion (3.55), i.e.

► $$\mathbb{R} = \mathrm{e}^{\mathrm{i}\alpha\mathbb{M}} = \mathbb{I} + \mathrm{i}\alpha\mathbb{M} + \frac{(\mathrm{i}\alpha\mathbb{M})^2}{2!} + \frac{(\mathrm{i}\alpha\mathbb{M})^3}{3!} + \cdots. \tag{3.74}$$

Expressing rotations in terms of generators is particularly useful when we need to combine multiple rotations around different axes and at different angles. A similar approach is also useful for the Lorentz transformation (see [1]). In fact, the general Lorentz transformation can be written using the generators of this transformation.

◖ Example 3.1: The motion of a charged particle in a magnetic field

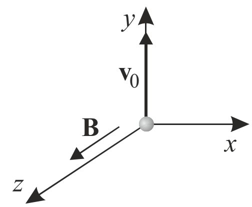


Fig. 3.7: Choice of coordinate system

In this example, we will demonstrate the use of complex numbers in solving differential equations that describe the motion of a charged particle in a homogeneous magnetic field. Suppose we have chosen a coordinate system such that the field points along the *z*-axis, and we have released the charged particle with a velocity v_0 across the magnetic field lines (if it did not have a nonzero velocity, the magnetic field would not act on it). The magnetic field, initial conditions, and Lorentz equation of motion will be:

$$\mathbf{B} = (0, 0, B) , \tag{3.75}$$

$$\mathbf{r}(0) = (0, 0, 0) , \tag{3.76}$$

$$\mathbf{v}(0) = (0, v_0, 0) , \tag{3.77}$$

$$m\ddot{\mathbf{r}} = Q\dot{\mathbf{r}} \times \mathbf{B} . \tag{3.78}$$

We will break down the equation of motion into its components. The motion will take place in the (x, y) plane, so we will limit ourselves to just these two components:

$$\ddot{x} = \frac{QB}{m}\dot{y} , \tag{3.79}$$

$$\ddot{y} = -\frac{QB}{m}\dot{x} . \tag{3.80}$$

These are two ordinary second-order differential equations in which the variables $x(t)$ and $y(t)$ are combined. Both equations can be solved in various ways. Probably the quickest way is Landau's method, which uses complex numbers: we multiply the second equation by the imaginary unit and add it to the first. We denote QB/m by ω_c. We will see that this is the particle's orbital frequency, called the cyclotron frequency.

$$\ddot{x} + \mathrm{i}\,\ddot{y} = -\mathrm{i}\,\omega_c(\dot{x} + \mathrm{i}\,\dot{y}) \tag{3.81}$$

By combining the equations, we have not lost any information. We can separate the real and imaginary parts at any time and return to the original equations. Now we simply need to introduce the complex variable $f = x + \mathrm{i}y$ and solve the simple equation

$$\ddot{f} = -\mathrm{i}\,\omega_c\,\dot{f} \tag{3.82}$$

in the complex numbers. After the first integration, we have:

$$\dot{f} + \mathrm{i}\,\omega_\mathrm{c}\, f = C_1 \,. \tag{3.83}$$

Now let's find the homogeneous and particular solutions:

$$f(t) = C_2\, \mathrm{e}^{-\mathrm{i}\,\omega_\mathrm{c}\, t} - \frac{\mathrm{i}}{\omega_\mathrm{c}} C_1 \,. \tag{3.84}$$

We determine the integration constants from the initial conditions, i.e., from

$$\begin{aligned} f(0) &= x(0) + \mathrm{i}\, y(0) = 0 \,, \\ \dot{f}(0) &= \dot{x}(0) + \mathrm{i}\, \dot{y}(0) = \mathrm{i}\, v_0 \,. \end{aligned} \tag{3.85}$$

By simply substituting into the solution (3.84), we see that:

$$C_1 = \mathrm{i}\, v_0 \,, \qquad C_2 = -v_0/\omega_c \,. \tag{3.86}$$

The overall solution is therefore given by

$$f(t) \equiv x(t) + \mathrm{i} y(t) = -\frac{v_0}{\omega_\mathrm{c}} \mathrm{e}^{-\mathrm{i}\,\omega_\mathrm{c}\, t} + \frac{v_0}{\omega_\mathrm{c}} . \tag{3.87}$$

We used complex numbers to solve this problem. Now it's time to separate the real and imaginary parts to obtain the coordinates of the moving charged particle:

$$\begin{aligned} x(t) &= R_\mathrm{L} - R_\mathrm{L} \cos \omega_\mathrm{c} t \,, \\ y(t) &= R_\mathrm{L} \sin \omega_\mathrm{c}\, t \,, \end{aligned} \tag{3.88}$$

where we marked

$$R_\mathrm{L} \equiv \frac{m v_0}{QB} ; \qquad \omega_\mathrm{c} \equiv \frac{QB}{m} \tag{3.89}$$

the Larmor radius R_L and the cyclotron frequency ω_c. Now we eliminate time (we keep the sine and cosine terms on the rhs, square both equations, and add them together):

$$\left(x - R_\mathrm{L}\right)^2 + y^2 = R_\mathrm{L}^2 \,. \tag{3.90}$$

We see that the motion follows a circle with radius R_L and center $S = [\, R_\mathrm{L}, 0\,]$. The position of the center depends on the sign of the particle's charge, as does the cyclotron frequency of the orbit.

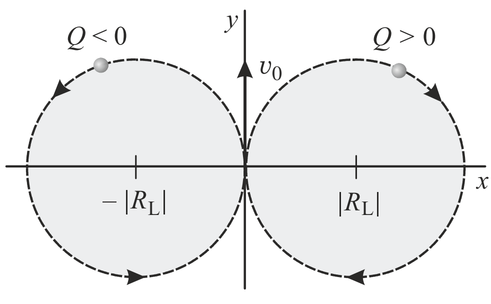


Fig. 3.8: The motion of a charged particle in a magnetic field

3.2.4 Quaternions

Everything we know about complex numbers can be summarized in a simple definition table, which lists all possible combinations of the real part (represented by the unit 1) and the imaginary part (represented by the imaginary unit i):

► $$\begin{array}{c|cc} & 1 & \mathrm{i} \\ \hline 1 & 1 & \mathrm{i} \\ \mathrm{i} & \mathrm{i} & -1 \end{array} \tag{3.91}$$

Complex numbers represent an ordered pair of real numbers that can be identified with the Cartesian coordinates of a point in the plane. By the mid-19th century, it was clear that some extension of complex numbers to four axes would be useful, since most phenomena in nature are described by four-tuples (time and space, energy and momentum, etc.). The concept of such an extension gradually took shape. It would be a "world" with one real axis (corresponding to time) and three imaginary axes (corresponding to space). Such a "complex" number (*quaternion*) would have the form

$$f = s + \mathrm{i}\,x + \mathrm{j}\,y + \mathrm{k}\,z\,, \tag{3.92}$$

where s is the real (or scalar) part, (x, y, z) is the imaginary (or vector) part, and i, j, k are the three imaginary units that satisfy the relation

$$\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1. \tag{3.93}$$

Just as with complex numbers, it should be possible to construct a quaternion from a quadruple of real numbers at any time, and conversely, to separate all four parts from a quaternion at any time and construct an ordered quadruple, or a four-vector, i.e., to switch between the two representations

► $$f = \begin{pmatrix} s \\ x \\ y \\ z \end{pmatrix} = s + \mathrm{i}\,x + \mathrm{j}\,y + \mathrm{k}\,z\,. \tag{3.94}$$

It was necessary to properly define multiplication between individual imaginary units. The Irish mathematician Willard Rowan Hamilton – whom readers will know from Hamilton's principle, Hamilton's equations, and the Hamilton operator – devoted himself to this task. Hamilton is considered the father of quaternions, but other mathematicians also contributed to this theory, such as the German mathematician Adolf Hurwitz, who proved that a meaningful extension of complex numbers can only be achieved in four (*quaternions*) and eight (*octonions*) dimensions, as well as the British mathematicians Arthur and Augustus de Morgan, the Graves brothers, and others. Hamilton introduced non-commutative multiplication of imaginary units, which mirrors the vector products of unit vectors along the imaginary axes:

$$\begin{aligned} \mathrm{i}\,\mathrm{j} &= \mathrm{k}, & \mathrm{j}\,\mathrm{i} &= -\mathrm{k}\,, \\ \mathrm{j}\,\mathrm{k} &= \mathrm{i}, & \mathrm{k}\,\mathrm{j} &= -\mathrm{i}\,, \\ \mathrm{k}\,\mathrm{i} &= \mathrm{j}, & \mathrm{i}\,\mathrm{k} &= -\mathrm{j}\,. \end{aligned} \tag{3.95}$$

These relationships can be derived from one another through cyclic permutation:

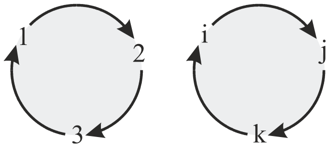


Fig. 3.9: Rotating a point in the complex plane

The table of basic operations between real and imaginary units is therefore

$$\begin{array}{c|cccc} & 1 & \mathrm{i} & \mathrm{j} & \mathrm{k} \\ \hline 1 & 1 & \mathrm{i} & \mathrm{j} & \mathrm{k} \\ \mathrm{i} & \mathrm{i} & -1 & \mathrm{k} & -\mathrm{j} \\ \mathrm{j} & \mathrm{j} & -\mathrm{k} & -1 & \mathrm{i} \\ \mathrm{k} & \mathrm{k} & \mathrm{j} & -\mathrm{i} & -1 \end{array} \tag{3.96}$$

►

It is said that Hamilton came up with these operations on October 16, 1843, while hurrying to a meeting of the Royal Society and crossing Brougham Bridge in Dublin. Apparently, he used a pocket knife to carve the equation $\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = \mathrm{ijk} = -1$ into the bridge. Today, there is a commemorative plaque on the bridge bearing this equation. As we can see, acts of vandalism can be committed not only by nameless hooligans, but also by venerable mathematicians with noble motives. If we multiply two quaternions according to these rules

$$f = \begin{pmatrix} \phi \\ A_x \\ A_y \\ A_z \end{pmatrix} = \begin{pmatrix} \phi \\ \mathbf{A} \end{pmatrix}; \qquad g = \begin{pmatrix} \psi \\ B_x \\ B_y \\ B_z \end{pmatrix} = \begin{pmatrix} \psi \\ \mathbf{B} \end{pmatrix}, \tag{3.97}$$

we get

$$\begin{aligned} fg &= (\phi + \mathrm{i}\,A_x + \mathrm{j}\,A_y + \mathrm{k}\,A_z)(\psi + \mathrm{i}\,B_x + \mathrm{j}\,B_y + \mathrm{k}\,B_z) = \cdots \\ &= \left(\phi\psi - A_x B_x - A_y B_y - A_z B_z\right) + \\ &+ \mathrm{i}\left(\phi B_x + \psi A_x + A_y B_z - A_z B_y\right) + \\ &+ \mathrm{j}\left(\phi B_y + \psi A_y + A_z B_x - A_x B_z\right) + \\ &+ \mathrm{k}\left(\phi B_z + \psi A_z + A_x B_y - A_y B_x\right). \end{aligned} \tag{3.98}$$

We can write the result in a more compact form

$$f\,g = \begin{pmatrix} \phi\psi - \mathbf{A}\cdot\mathbf{B} \\ \phi\mathbf{B} + \psi\mathbf{A} + \mathbf{A}\times\mathbf{B} \end{pmatrix} \tag{3.99}$$

For purely real quaternions (lacking an imaginary, i.e., vector, part), the following holds

$$f\,g = \phi\psi \tag{3.100}$$

and the multiplication of quaternions reduces to the multiplication of two real numbers. For purely imaginary quaternions (those with a zero scalar component), we then have

$$fg = \begin{pmatrix} -\mathbf{A}\cdot\mathbf{B} \\ \mathbf{A}\times\mathbf{B} \end{pmatrix}. \qquad (3.101)$$

Thus, the scalar part contains the negative scalar product, and the vector part contains the vector product. Imagine that the vector **A** is the gradient operator, i.e., $\mathbf{A} = \boldsymbol{\nabla}$; then, the scalar part of the relation (3.101) naturally contains the divergence, and the vector part contains the rotation of the field **B**. The properties of quaternions correspond well with scalar and vector products in the description of the electromagnetic field; therefore, James Clerk Maxwell used them in 1873 in the final formulation of his equations. He understood vectors as purely imaginary quaternions (with only a vector part) and scalars as purely real quaternions (with only a scalar part). For example, the vector **A** and the gradient can be understood as purely imaginary quaternions

$$\mathcal{A} = \mathrm{i}A_x + \mathrm{j}A_y + \mathrm{k}A_z = \begin{pmatrix} 0 \\ \mathbf{A} \end{pmatrix}; \qquad \nabla = \mathrm{i}\partial_x + \mathrm{j}\partial_y + \mathrm{k}\partial_z = \begin{pmatrix} 0 \\ \partial/\partial\mathbf{x} \end{pmatrix} \qquad (3.102)$$

Just as we define real and imaginary parts for complex numbers, we can define scalar and vector parts for quaternions:

► $$f = \begin{pmatrix} \phi \\ \mathbf{A} \end{pmatrix}; \qquad \mathrm{S}.\, f = \phi; \qquad \mathrm{V}.\, f = \mathbf{A}. \qquad (3.103)$$

We can now express the divergence and curl of the field as the corresponding components of the product of two quaternions:

$$\mathrm{div}\,\mathbf{A} = -\mathrm{S}.\,\nabla\mathcal{A}; \qquad \mathrm{rot}\,\mathbf{A} = \mathrm{V}.\,\nabla\mathcal{A} \qquad (3.104)$$

The table below shows, on the left, the historical quaternion notation of some of Maxwell's equations from his 1873 work *A Treatise on Electricity and Magnetism* (Volume II, page 257 and following), and on the right, the current notation:

$\mathcal{B} = \mathrm{V}.\,\nabla\mathcal{A}$	$\mathbf{B} = \mathrm{rot}\,\mathbf{A}$
$\mathcal{E} = -\dot{\mathcal{A}} - \nabla\psi$	$\mathbf{E} = -\partial\mathbf{A}/\partial t - \nabla\phi$
$e = \mathcal{S}.\,\nabla\mathcal{D}$	$\mathrm{div}\,\mathbf{D} = \rho_Q$
$0 = \mathcal{S}.\,\nabla\mathcal{B}$	$\mathrm{div}\,\mathbf{B} = 0$
$\mathcal{F} = \mathcal{V}.\,\mathcal{GB} - \dot{\mathcal{A}} - \nabla\psi$	$\mathbf{F} = Q\,\mathbf{v}\times\mathbf{B} + Q\,\mathbf{E}$
$\mathcal{C} = \mathcal{K} + \dot{\mathcal{D}}$	$\mathbf{j}_{\mathrm{tot}} = \mathbf{j} + \partial\mathbf{D}/\partial t$
$4\pi\mathcal{C} = \mathrm{V}.\,\nabla\mathcal{H}$	$\mathrm{rot}\,\mathbf{H} = \mathbf{j} + \partial\mathbf{D}/\partial t$

3.2.5 Holomorphic Functions

In the following text, we will introduce basic concepts and procedures from complex analysis. Let us assume that the function $f(z)$ is a function of a complex variable, i.e.

$$f(z): \mathbb{C} \to \mathbb{C} \,. \tag{3.105}$$

Let us say that a complex function $f(x, y) = u(x, y) + \mathrm{i}v(x, y)$ is *holomorphic* on an open set if it has a derivative at every point of the set. The open set is essential in the definition because there must exist a neighborhood of every point of the set in which the derivative can be defined. The existence of a complex derivative is a very strong requirement, and if the function is holomorphic, it has interesting properties:

CR conditions. The so-called Cauchy-Riemann (CR) conditions apply:

► $$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}\,; \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \,. \tag{3.106}$$

The validity of the Cauchy-Riemann conditions is evident from the fact that the derivative must yield the same result whether we approach a given point along the real or imaginary axis; that is, the equality $\partial f/\partial x = \partial f/\partial \mathrm{i}y$ must hold. By separating the real and imaginary parts, we obtain the Cauchy-Riemann conditions.

Harmonicity. It is immediately clear from CR conditions that both the real and imaginary parts of a holomorphic function are harmonic, i.e., they satisfy Laplace equation:

► $$\nabla^2 u = 0\,; \qquad \nabla^2 v = 0 \,. \tag{3.107}$$

The situation can also be reversed. If we take some harmonic function as the real part of a complex function, we can use the Cauchy-Riemann conditions to determine its imaginary part; thus, every harmonic function corresponds to a complex function.

Taylor expansion. A Taylor expansion can always be found for a holomorphic function at every point in a set. Later in this chapter, we will show how to elegantly find its coefficients; see the formula (3.146).

Line integral. Let γ be a closed simple curve (exactly one loop). If f is holomorphic on the curve and inside the curve, then *Cauchy fundamental theorem* holds:

► $$\oint_\gamma f(z)\,\mathrm{d}z = 0 \,. \tag{3.108}$$

Since the result of the integration remains unchanged under any continuous deformation of the curve (we will not prove this claim here), we will demonstrate the validity of Cauchy theorem for a circle in the complex plane parameterized by the equation

$$z = z_0 + R\,\mathrm{e}^{\mathrm{i}\varphi}\,; \qquad \varphi \in \langle 0, 2\pi) \,. \tag{3.109}$$

For the calculation, we will use the Taylor series expansion of the function:

$$\oint\limits_{z=z_0+R\,\mathrm{e}^{\mathrm{i}\varphi}} f(z)\,\mathrm{d}z = \oint\limits_{z=z_0+R\,\mathrm{e}^{\mathrm{i}\varphi}} \sum_{k=0}^{\infty} c_k (z-z_0)^k \,\mathrm{d}z = \sum_{k=0}^{\infty} \int_0^{2\pi} c_k R^k \,\mathrm{e}^{\mathrm{i}k\varphi}\, R\,\mathrm{i}\,\mathrm{e}^{\mathrm{i}\varphi}\,\mathrm{d}\varphi =$$

$$\sum_{k=0}^{\infty} \mathrm{i}\, c_k R^{k+1} \int_0^{2\pi} \mathrm{e}^{\mathrm{i}(k+1)\varphi}\, \mathrm{d}\varphi = 0\,.$$

The result is zero, since for any k, this is a periodic function whose integral is zero (the areas above and below the axis are equal).

3.2.6 Laurent Expansion and the Residue Theorem

Laurent expansion

Laurent expansion of a complex function $f(z)$ in the neighborhood of z_0 is called a series

► $$f(z) = \sum_{k=-\infty}^{+\infty} c_k (z - z_0)^k \quad . \tag{3.110}$$

The French mathematician Pierre Alphonse Laurent studied this series. The sum of the negative terms of the series ($k < 0$) is called the *principal part* of the Laurent series, while the sum of the non-negative terms ($k \geq 0$) is called the *regular part*.

If a function in the complex plane has poles (isolated points at which the function diverges, but in whose annular neighborhood the function is holomorphic; see below), it is always possible to find some annuli $\mathcal{K}$ centered at a given point z_0 in which the function is holomorphic. For these annuli, it is possible to uniquely determine the coefficients c_k of the series such that the Laurent series is convergent on these annuli. The series will have different coefficients for different annuli. In the following figure, the poles are at the points z_1, z_2, and z_3, and there are four subcircles in which coefficients c_k can be found such that the Laurent series converges to the original function:

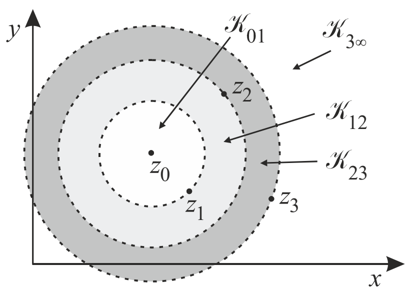


Fig. 3.10: Convergence in annuli

Residue theorem

Let us find the line integral of a complex function along a simple, positively oriented closed curve γ (the region is looped around exactly once). The function must be holomorphic at every point on the curve, but there may be poles in the region bounded by the curve, and therefore the line integral will not be zero, since the assumption of Cauchy's fundamental theorem regarding the holomorphy throughout the entire region does not hold. Let us first consider a simple situation with a single pole at z_0, around which there exists a ring-shaped neighborhood (the point z_0 does not belong to it) on which f is holomorphic. Let us find the integral along the circle around the point z_0:

$$\oint_{z=z_0+R\,\mathrm{e}^{\mathrm{i}\varphi}} f(z)\,\mathrm{d}z = \oint_{z=z_0+R\,\mathrm{e}^{\mathrm{i}\varphi}} \sum_k c_k (z-z_0)^k \,\mathrm{d}z = \sum_k \int_0^{2\pi} c_k R^k \,\mathrm{e}^{\mathrm{i}k\varphi}\, R\,\mathrm{i}\,\mathrm{e}^{\mathrm{i}\varphi}\,\mathrm{d}\varphi =$$

$$= \sum_k \mathrm{i}\,c_k R^{k+1} \int_0^{2\pi} \mathrm{e}^{\mathrm{i}(k+1)\varphi}\,\mathrm{d}\varphi = \sum_k \mathrm{i}\,c_k R^{k+1} 2\pi\delta_{k,-1} = 2\pi\,\mathrm{i}\,c_{-1}\,.$$

It is clear that the only non-zero term is the one with $k = -1$. We therefore call the coefficient c_{-1} the *residue* of the function f at the point z_0, denoted by Res(f; z_0). We can proceed similarly for a general curve; the result is the *residue theorem*

► $$\oint_\gamma f(z)\,\mathrm{d}z = 2\pi\,\mathrm{i} \sum_{z_k \in\, \mathrm{Int}\,\gamma} \mathrm{Res}\left(f, z_k\right)\,. \tag{3.111}$$

The integral of a simple, positively oriented closed curve is equal to 2π times the sum of all residues of the function lying inside the curve. This theorem allows efficient calculation of curve integrals in the complex plane, as well as integrals along the real axis, which we regard as a part of a curve in the complex plane. If the curve were oriented in the opposite direction (clockwise), a minus sign would appear on the right-hand side.

Pole

We say that the function $f(z)$ has a pole at the point z_0 if it satisfies

1. $\lim\limits_{z\to z_0} f(z) = \infty$,
2. Function is holomorphic in the ring neighborhood of z_0.

Order of a pole

We say that the pole z_0 of the function $f(z)$ has multiplicity (order) k if the coefficients of the Laurent series expansion in the ring neighborhood of the point z_0 satisfy

1. $c_{-k} \neq 0$,
2. $c_l = 0\,; \quad l < -k$.

The simple formulas given below can be used to calculate the residues for poles of low order.

Residue at the pole of first order

The residue at the pole of the first order can be determined from the following relation (which follows immediately from Laurent's expansion)

$$\mathrm{Res}\left(f, z_0\right) = \lim_{z\to z_0} \left[(z-z_0)f(z)\right]\,. \tag{3.112}$$

It is clear from the relation that for a holomorphic function g, the following holds

► $$\mathrm{Res}\left(\frac{g(z)}{(z-z_0)}, z_0\right) = g(z_0)\,. \tag{3.113}$$

An example of this is

$$\operatorname{Res}\left(\frac{\sin(z)}{(z+\mathrm{i})(z-\mathrm{i})},\mathrm{i}\right)=\lim_{z\to\mathrm{i}}\left[(z-\mathrm{i})\frac{\sin(z)}{(z+\mathrm{i})(z-\mathrm{i})}\right]=$$

$$=\lim_{z\to\mathrm{i}}\left[\frac{\sin(z)}{(z+\mathrm{i})}\right]=\frac{\sin(\mathrm{i})}{2\mathrm{i}}.$$

Residue at the pole of the k^th order

$$\operatorname{Res}\left(f,z_0\right)=\frac{1}{(k-1)!}\lim_{z\to z_0}\left[(z-z_0)^k f(z)\right]^{(k-1)} . \tag{3.114}$$

Residue in infinity

For a holomorphic function, we can define a Laurent series in a ring neighborhood of infinity

$$f(z)=\sum_{-\infty}^{+\infty}\frac{b_k}{z^k} . \tag{3.115}$$

We then define the residue at infinity by the relation

$$\operatorname{Res}(f,\infty)=-b_1 . \tag{3.116}$$

The sign is defined as negative so that, for a function that is holomorphic except at a finite number of poles z_k, the sum of all residues is zero:

$$\operatorname{Res}(f,\infty)\ +\ \sum_{z_k}\operatorname{Rez}(f,z_k)=0\ . \tag{3.117}$$

This relationship makes it possible to calculate the residue at infinity without using the defining equation (3.114).

3.2.7 Examples of Integral Calculations

◖ **Example 3.2**

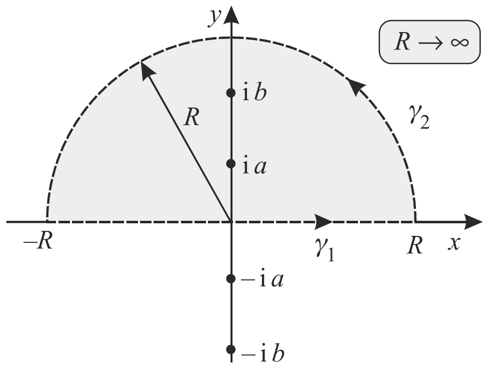
Calculate the integral

$$I=\int_{-\infty}^{+\infty}\frac{x^2}{\left(x^2+a^2\right)\left(x^2+b^2\right)}\,\mathrm{d}x . \tag{3.118}$$

Instead of the real integral (3.118), we will evaluate the integral of a complex function (replacing x with z)

$$f(z)=\frac{z^2}{\left(z^2+a^2\right)\left(z^2+b^2\right)}=\frac{z^2}{(z+\mathrm{i}a)(z-\mathrm{i}a)(z+\mathrm{i}b)(z-\mathrm{i}b)} , \tag{3.119}$$

that has four poles, none of which lie on the real axis. We will trace the integral along the curve shown in the following figure; the poles are marked with black circles:



To perform the calculation, we will use the residue theorem; there are only two poles inside the dashed contour:

$$\oint_\gamma f(z)\,\mathrm{d}z = \int_{\gamma_1} f(z)\,\mathrm{d}z + \int_{\gamma_2} f(z)\,\mathrm{d}z = 2\pi i\left[\mathrm{Res}\left(f(z),\mathrm{i}a\right)+\mathrm{Res}\left(f(z),\mathrm{i}b\right)\right]. \tag{3.120}$$

As $R \to \infty$, the integral over the curve γ_1 converges to the integral I we are seeking, while the integral over the curve γ_2 approaches zero as $R \to \infty$, since the integrand tends to zero at infinity on the curve γ_2 (in any direction). Therefore, we have

$$I = 2\pi i\left[\mathrm{Res}\left(f(z),\mathrm{i}a\right)+\mathrm{Res}\left(f(z),\mathrm{i}b\right)\right] = \frac{\pi}{a+b}. \tag{3.121}$$

◗

◖ Example 3.3

Calculate the integral

$$I = \int_{-\infty}^{+\infty} \frac{\cos mx}{1+x^2}\,\mathrm{d}x\,; \qquad m>0. \tag{3.122}$$

The integral certainly exists, since the numerator is a bounded function and the denominator behaves like $1/x^2$ at $\pm\infty$. We will determine the integral as the real part of the integral in which we replace the cosine function with an oscillating exponential

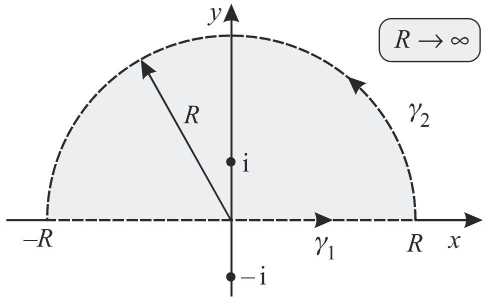
$$I = \mathrm{Re}\int_{-\infty}^{+\infty} \frac{\mathrm{e}^{\mathrm{i}mx}}{1+x^2}\,\mathrm{d}x\,; \qquad m>0. \tag{3.123}$$

The next step will be similar to the previous example. We will calculate the integral of a complex function

$$f(z) = \frac{\mathrm{e}^{\mathrm{i}mz}}{1+z^2} = \frac{\mathrm{e}^{\mathrm{i}mz}}{(z+\mathrm{i})(z-\mathrm{i})}, \tag{3.124}$$

which has poles at the points +i and −i. We choose the black dashed curve in the following figure as the integrating curve. From the residue theorem, we then have

$$\int_\gamma \frac{\mathrm{e}^{\mathrm{i}mz}}{1+z^2}\,\mathrm{d}z = \int_{\gamma_1} \frac{\mathrm{e}^{\mathrm{i}mz}}{1+z^2}\,\mathrm{d}z + \int_{\gamma_2} \frac{\mathrm{e}^{\mathrm{i}mz}}{1+z^2}\,\mathrm{d}z = 2\pi i\,\mathrm{Res}\left(f(z),\mathrm{i}\right). \tag{3.125}$$



As R approaches infinity, the first integral will be integrated over the entire real axis; in the second integral, the integrand will converge to zero for $m > 0$. If $m < 0$, we would have to integrate over the lower half-plane. The result is therefore

$$\int_{-\infty}^{+\infty} \frac{e^{imx}}{1+x^2}\, dx = 2\pi i \operatorname{Res}\big(f(z), i\big) = 2\pi i \frac{e^{-m}}{2i} = \pi e^{-m} . \quad (3.126)$$

After separating the real and imaginary parts, we get the result

$$\int_{-\infty}^{+\infty} \frac{\cos mx}{1+x^2}\, dx = \pi e^{-m} ; \quad (3.127)$$

$$\int_{-\infty}^{+\infty} \frac{\sin mx}{1+x^2}\, dx = 0 . \quad (3.128)$$

◗

◖ Example 3.4

Calculate the integral

$$\int_{-\infty}^{+\infty} \frac{1}{x^3}\, dx. \quad (3.129)$$

Calculating this integral will cause problems near zero, where the integrand diverges on both sides (see figure), and integration cannot be performed at such a point. Nevertheless, the integration can be done in a certain sense. We divide the integration into two parts, in which we omit the “problematic” point and approach it from the left and right as the limit. We call this procedure integration in the sense of the *principal Cauchy value* and denote it

$$V.P. \int_{-\infty}^{+\infty} \frac{1}{x^3}\, dx \equiv \lim_{\varepsilon \to 0} \left[\int_{-\infty}^{-\varepsilon} \frac{1}{x^3}\, dx + \int_{\varepsilon}^{+\infty} \frac{1}{x^3}\, dx \right]. \quad (3.130)$$

The notation V. P. stands for *Value Principal*. In our case, the first integral is negative and the second is positive, and they cancel each other, so the result is zero. In other cases involving asymmetric functions, the result may be nonzero. The result is therefore

$$V.P. \int_{-\infty}^{+\infty} \frac{1}{x^3}\, dx = 0 . \quad (3.131)$$

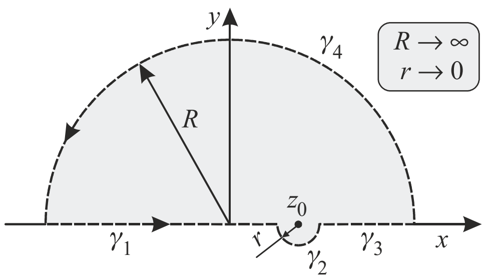
◗

◖ **Example 3.5**

In physics, it is quite common to calculate integrals with a simple pole on the real axis

$$\int_{-\infty}^{+\infty} \frac{g(x)}{x-x_0}\,\mathrm{d}x\,. \tag{3.132}$$

Various problems on resonance lead to such an integral. If the function $g(z)$ is holomorphic in the complex plane and approaches zero at both the real (∞) and imaginary ($\mathrm{i}\infty$) infinities fast enough for the integral to converge, we use the residue theorem for the dashed curve in the figure to compute it:



We rewrite integral (3.132) in terms of a complex variable and perform integration along the curves in the mathematically positive direction, i.e., counterclockwise:

$$\int_{\gamma_1} \frac{g(z)}{z-z_0}\,\mathrm{d}z + \int_{\gamma_2} \frac{g(z)}{z-z_0}\,\mathrm{d}z + \int_{\gamma_3} \frac{g(z)}{z-z_0}\,\mathrm{d}z + \int_{\gamma_4} \frac{g(z)}{z-z_0}\,\mathrm{d}z = 2\pi \mathrm{i} g(x_0)\,.$$

In the integrals, we take both limits as $R \to \infty$ and $r \to 0$. Given the assumptions, the last integral will tend to zero as $R \to \infty$. The individual integrals will successively yield

$$\lim_{\substack{R\to\infty \\ r\to 0}} \left(\int_{-R}^{x_0-r} \frac{g(x)}{x-x_0}\,\mathrm{d}x + \int_{\pi}^{2\pi} \frac{g(z_0 + r\,\mathrm{e}^{\mathrm{i}\varphi})}{r\,\mathrm{e}^{\mathrm{i}\varphi}}\,\mathrm{i}r\,\mathrm{e}^{\mathrm{i}\varphi}\,\mathrm{d}\varphi + \int_{x_0+r}^{R} \frac{g(x)}{x-x_0}\,\mathrm{d}x \right) + 0 = 2\pi\mathrm{i}g(x_0)$$

$$\lim_{\substack{R\to\infty \\ r\to 0}} \left(\int_{-R}^{x_0-r} \frac{g(x)}{x-x_0}\,\mathrm{d}x + \int_{x_0+r}^{R} \frac{g(x)}{x-x_0}\,\mathrm{d}x \right) + \lim_{r\to 0} \int_{\pi}^{2\pi} \frac{g(z_0 + r\,\mathrm{e}^{\mathrm{i}\varphi})}{r\,\mathrm{e}^{\mathrm{i}\varphi}}\,\mathrm{i}r\,\mathrm{e}^{\mathrm{i}\varphi}\,\mathrm{d}\varphi = 2\pi\mathrm{i}g(x_0)$$

$$\mathrm{V.P.}\int_{-\infty}^{+\infty} \frac{g(x)}{x-x_0}\,\mathrm{d}x + \mathrm{i}\int_{\pi}^{2\pi} g(x_0)\,\mathrm{d}\varphi = 2\pi\mathrm{i}g(x_0)\,.$$

► $$\mathrm{V.P.}\int_{-\infty}^{+\infty} \frac{g(x)}{x-x_0}\,\mathrm{d}x = \pi\mathrm{i}g(x_0)\,. \tag{3.133}$$

If the integral does not converge at the imaginary infinity, we will use other integration paths, such as rectangles or other suitable shapes. ◗

◖ **Example 3.6**

Find the integrals

$$\int_{-\infty}^{+\infty} \frac{\sin x}{x}\,\mathrm{d}x. \tag{3.134}$$

$$\int_{-\infty}^{+\infty} \frac{\cos x}{x}\,\mathrm{d}x. \tag{3.135}$$

The first of these integrals will be key to the theory of distributions, and we will encounter it again later. The integrand behaves "normally" at the beginning, since

$$\lim_{x\to 0} \frac{\sin x}{x} = 1. \tag{3.136}$$

Unfortunately, the same cannot be said for the second integrand; the following holds

$$\lim_{x\to 0^{\pm}} \frac{\cos x}{x} = \pm\infty \tag{3.137}$$

and the integral can only be evaluated using Cauchy's principal value. We will evaluate both integrals at once using exponential notation, i.e., we will determine the integral

$$I = \int_{-\infty}^{+\infty} \frac{\mathrm{e}^{\mathrm{i}x}}{x}\,\mathrm{d}x \tag{3.138}$$

and then separate the real and imaginary parts. We will follow the general procedure from the previous example. We will replace the integrand with a complex function

$$f(z) = \frac{\mathrm{e}^{\mathrm{i}z}}{z}, \tag{3.139}$$

which has a single pole on the real axis at $z = 0$. In the upper half-plane, the integrand converges to zero at infinity, so we can directly apply equation (3.133) from the previous example:

$$\text{V.P.}\int_{-\infty}^{+\infty} \frac{\mathrm{e}^{\mathrm{i}x}}{x}\,\mathrm{d}x = \pi\mathrm{i}\,\mathrm{e}^{0} = \mathrm{i}\pi. \tag{3.140}$$

If we now separate the real and imaginary parts, we have

$$\text{V.P.}\int_{-\infty}^{+\infty} \frac{\cos x}{x}\,\mathrm{d}x = 0. \tag{3.141}$$

► $$\int_{-\infty}^{+\infty} \frac{\sin x}{x}\,\mathrm{d}x = \pi. \tag{3.142}$$

In the second integral, we omitted the symbol V.P., since the integrand is finite at $x = 0$ and the principal value coincides with the standard meaning of integration. We will need this derived relation later. ◗

3.2.8 Cauchy Integral Formula and the Holographic Principle

The values of a holomorphic function inside any closed simple curve can be calculated from the values on that curve (the function must be holomorphic throughout the domain) using *Cauchy's integral formula*:

$$f(z_0) = \frac{1}{2\pi \mathrm{i}} \oint_\gamma \frac{f(z)}{z - z_0} \, \mathrm{d}z \, . \tag{3.143}$$

The integrand itself is not holomorphic at the point $z = z_0$. We can immediately obtain a proof of this statement from the residue theorem. Let us now find the result of integration along a circle: Without loss of generality, we replace the curve by a circle centered at the point z_0 (a continuous deformation of the curve in a holomorphic region does not change the value of the line integral; therefore, we must not deform the curve through the central point z_0, where the integrand is not holomorphic).

The center of the circle is at the point where we evaluate the function; the radius of the circle does not matter (circles with different radii can be continuously deformed into one another). We expand a holomorphic function into a Laurent series around the point z_0, which, due to holomorphy, will have only non-negative terms (i.e., a Taylor series):

$$f(z) = \sum_{k=0}^{\infty} c_k \left(z - z_0\right)^k \, . \tag{3.144}$$

Let us now integrate any term from the series on the right-hand side of equation (3.143):

$$\frac{1}{2\pi \mathrm{i}} \oint_\gamma \frac{c_k \left(z - z_0\right)^k}{z - z_0} \mathrm{d}z = \frac{1}{2\pi \mathrm{i}} \oint_\gamma c_k \left(z - z_0\right)^{k-1} \mathrm{d}z =$$

$$= \frac{1}{2\pi \mathrm{i}} \int_0^{2\pi} c_k R^{k-1} \, \mathrm{e}^{\mathrm{i}(k-1)\varphi} \, \mathrm{i} R \mathrm{e}^{\mathrm{i}\varphi} \, \mathrm{d}\varphi = \tag{3.145}$$

$$= \frac{\mathrm{i} c_k R^k}{2\pi \mathrm{i}} \int_0^{2\pi} \mathrm{e}^{\mathrm{i}k\varphi} \, \mathrm{d}\varphi = \frac{\mathrm{i} c_k R^k}{2\pi \mathrm{i}} 2\pi \delta_{k0} = c_0 \delta_{k0} = f(z_0) \delta_{k0} \, .$$

The only non-zero term therefore has a zero expansion term, which is directly equal to the value we are seeking. Equation (3.144) can easily be generalized to the coefficients of the Laurent series. For $z_0 \neq \infty$, the coefficients of the series can be determined from the equation

► $$c_k = \frac{1}{2\pi \mathrm{i}} \oint_\gamma \frac{f(z)}{(z - z_0)^{k+1}} \mathrm{d}z \, ; \qquad \gamma \subset \mathscr{K} \, . \tag{3.146}$$

The curve γ is a closed simple curve and lies entirely within the given annulus. It could be, for example, a circle with center at z_0 and an appropriate radius. The proof of relation (3.146) is entirely analogous to the proof of (3.145), i.e., we simply substitute the parameterization of the circle (3.109) and the desired expansion of the function. Only one term will be nonzero, and that term will give the coefficient c_k. For the regular part of the series, the coefficients ck become the usual coefficients of the Taylor series

► $$c_k = \frac{f^{(k)}(z_0)}{k!}; \qquad k = 1, 2, \ldots \tag{3.147}$$

There are a number of tricks for expanding a function into a Laurent series that do not require calculating the coefficients using the formula (3.146).

* * *

A very interesting property of holomorphic functions is relation (3.143), which allows us to compute the values of a holomorphic function based on its behavior at the boundary of a set. We encounter a similar principle in physics, where it is known as the holographic principle. Essentially, it states that an N-dimensional reality can be encoded into information on an $N-1$-dimensional set. So far, we encounter the holographic principle in three areas of physics. The first is holography, i.e., the creation of a 3D image (hologram) on a 2D medium. Using interference, a hologram records not only information about the intensity of light reflected from an object – as in a conventional photograph – but also its phase, which is encoded as a pattern created by the interference of the reflected beam with a reference beam. A three-dimensional "photograph" of the original object can then be reconstructed from the hologram using laser beams. Holography was discovered in 1947 by the English physicist of Hungarian origin, Denis Gabor (1900–1979).

The second area is the thermodynamics of black holes. The work of Israeli-American physicist Jacob Bekenstein (1947–2015) and British theorist Stephen Hawking (*1942) suggests that information about a black hole's interior is encoded on its event horizon. The surface of a black hole can be assigned not only a temperature but also entropy, which is normally a quantity tied to volume.

Thirdly, the Dutch theorist Erik Verlinde (*1962) applied the holographic principle in his alternative theory of gravity, in which he does not regard gravity as a separate interaction, but as a macroscopic result of the quantum behavior of the microworld (a so-called entropic force belonging to the same category as diffusion or elasticity). Verlinde's theory employs an unspecified projection surface on which all information about the objects enclosed within the surface is encoded. Such a surface could be, for example, the worldline of a string corresponding to some elementary particle. From the above, it is evident that the holographic principle – that is, the claim that information about N dimensions can be encoded on an $N-1$-dimensional set – likely has a deeper meaning in nature, the significance of which we do not yet fully understand.

3.3 Vectors and Tensors

3.3.1 Linear Vector Space

We can imagine a vector as an arrow (distinguishing between the start and end points). Arrows shifted parallel to each other are considered to be the same vector. If we shift the vector to the origin, we can "subtract" the so-called coordinates of the vector from the position of the endpoint. These coordinates always depend on the choice of coordinate system. Therefore, it is not enough to specify a vector as an ordered triple; we must supplement its definition with transformation rules, i.e., how the vector coordinates change when moving to a different coordinate system:

$$\mathbf{f} = (f_1, f_2, f_3)\,, \tag{3.148}$$

$$\tilde{f}_k = \sum_{l=1}^{3} A_{kl} f_l\,, \tag{3.149}$$

where **A** is the transformation matrix from the untilded to the tilded system. Structured objects are usually denoted in bold type. Sometimes it is convenient to write the components of a vector in a row (to save space), and other times as columns. If we transform the components like in (3.149), the column notation is necessary:

$$\begin{pmatrix} \tilde{f}_1 \\ \tilde{f}_2 \\ \tilde{f}_3 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \cdot \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}. \tag{3.150}$$

We can perform two operations on vectors: *scaling* by a real number and *addition*:

► $$\alpha\mathbf{f} \equiv (\alpha f_1, \alpha f_2, \alpha f_3)\,, \tag{3.151}$$

► $$\mathbf{f} + \mathbf{g} \equiv (f_1 + g_1, f_2 + g_2, f_3 + g_3)\,. \tag{3.152}$$

We perform both scaling and addition on all components. The result of scaling is a vector in the same ($\alpha > 0$) or opposite ($\alpha < 0$) direction, which is α times longer. The result of the addition is a vector that arises as the diagonal of a parallelogram stretched along both vectors (in physics well-known parallelogram of forces).

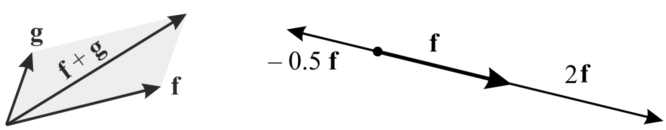


Fig. 3.14: Vector addition and scalar multiplication

These two operations satisfy rules that may apply not only to our example "arrows" but also to other objects. We will therefore abstract the concept of a vector and focus primarily on the properties of the scalar multiplication and addition operations. Let us therefore introduce a linear vector space.

Let's say that $\mathcal{V}$ is a linear vector space over the set of real (complex) numbers, provided that operations are defined for the elements of this space

Labeling	From → to	Name	Notation
+	$\mathcal{V}\times\mathcal{V}\to\mathcal{V}$	Vector addition	$\mathbf{h}=\mathbf{f}+\mathbf{g}$
·	$\mathcal{R}(\mathcal{C})\times\mathcal{V}\to\mathcal{V}$	Vector scalar multiplication	$\mathbf{g}=\alpha\cdot\mathbf{f}$

These operations have the following properties:

$$
\begin{array}{lll}
1) & \mathbf{f}+\mathbf{g}=\mathbf{g}+\mathbf{f}\,, & \mathbf{f}+(\mathbf{g}+\mathbf{h})=(\mathbf{f}+\mathbf{g})+\mathbf{h}\,,\\
2) & \alpha(\mathbf{f}+\mathbf{g})=\alpha\mathbf{f}+\alpha\mathbf{g}\,, & (\alpha+\beta)\mathbf{f}=\alpha\mathbf{f}+\beta\mathbf{f}\,,\\
3) & \alpha(\beta\mathbf{f})=(\alpha\beta)\mathbf{f}\,, & 1\mathbf{f}=\mathbf{f}\,,\\
4) & \mathbf{f}+\mathbf{g}=\mathbf{f}+\mathbf{h} \quad\Rightarrow\quad \mathbf{g}=\mathbf{h}\,. &
\end{array}
\tag{3.153}
$$

The first property states that commutativity and associativity hold; the second defines linearity. At first glance, it is clear that physical vectors (such as force) possess these properties – we can add and subtract them, and we can visualize them as arrows. On the other hand, completely different objects (matrices, solutions to equations) can also have similar properties if we appropriately define these operations for them. What matters are the properties of the objects (3.153), not the objects themselves.

3.3.2 Scalar Product

Vector norm

For now, let's focus on ordered pairs or triples of numbers that represent the coordinates of a vector. Later, we will extend our discussion to more general objects. We can easily determine the magnitude (norm) of a vector using the Pythagorean theorem. As shown in the figure, the following holds for the two-dimensional and three-dimensional cases:

$$
\begin{aligned}
\|\mathbf{f}\|_{2D} &= |\mathbf{f}| = \sqrt{f_x^2+f_y^2} = \sqrt{x^2+y^2}\,,\\
\|\mathbf{f}\|_{3D} &= |\mathbf{f}| = \sqrt{l_0^2+y^2} = \sqrt{x^2+z^2+y^2}\,.
\end{aligned}
\tag{3.154}
$$

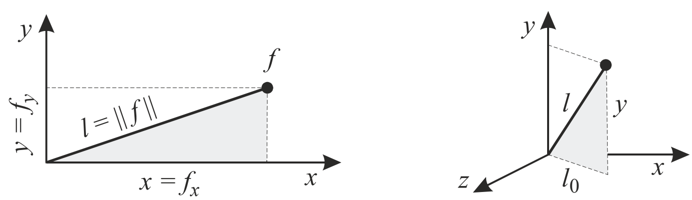


Fig. 3.15: The magnitude of a vector for ordered pairs and triples

It is clear that for higher dimensions, the Euclidean norm can be defined by the relation

$$
\|\mathbf{f}\| = \sqrt{x_1^2+x_2^2+\cdots+x_N^2}\,. \tag{3.155}
$$

A norm is denoted by a single vertical bar (as an absolute value) or a double vertical bar. The concept of a norm can also be extended to the so-called *p-q* norm, defined by

$$\|\mathbf{f}\| = \left(|x_1|^{p_1} + |x_2|^{p_2} + \cdots + |x_N|^{p_N}\right)^{1/q}. \qquad (3.156)$$

For $p_k = q = 2$, this norm is the Euclidean distance. Of course, it would be necessary to examine in detail whether a norm defined in this way satisfies our expectations regarding vector magnitude, such as the triangle inequality, etc. The reader can find such an analysis in mathematics textbooks. Let us show what the unit circle looks like (the set of points whose distance from the origin is equal to one) in various *p-q* norms:

$$|x|^{p_1} + |y|^{p_2} = 1 \qquad (3.157)$$

The following figure shows these "circles" for various values of the exponents.

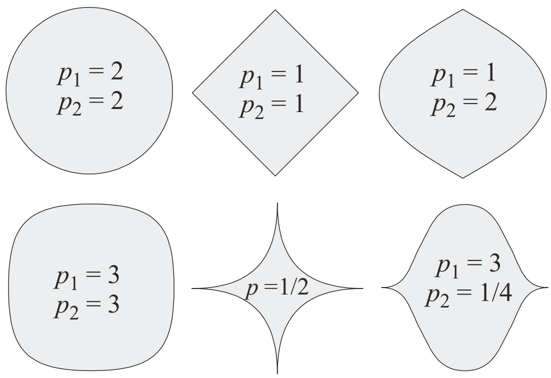


Fig. 3.16: Unit circles in various norms

Scalar product

With vectors, we can perform not only scaling and vector addition, but also other operations. If we have two vectors **f** and **g**, we can keep their indices different and thus create a two-dimensional array of numbers that satisfies certain transformation rules; we call this a second-order tensor

► $$T_{kl} \equiv f_k g_l\,; \qquad \mathbf{T} \equiv \mathbf{f} \otimes \mathbf{g}\,. \qquad (3.158)$$

We call this operation the *tensor produc*t. It is important in describing electromagnetic fields, gravity, inertia, permittivity, etc. The first notation is component-based, while the second is symbolic. The tensor product is denoted by a cross inside a circle.

◖ **Example 3.7:**

$$\mathbf{f} = (1,2,3)\,; \qquad \mathbf{g} = (-7,-1,2)\,;$$

$$\mathbf{T} = \begin{pmatrix} f_1g_1 & f_1g_2 & f_1g_3 \\ f_2g_1 & f_2g_2 & f_2g_3 \\ f_3g_1 & f_3g_2 & f_3g_3 \end{pmatrix} = \begin{pmatrix} -7 & -1 & 2 \\ -14 & -2 & 4 \\ -21 & -3 & 6 \end{pmatrix}.$$

◗

The next option is to add the two indices together; this operation is the *scalar product*:

► $$\mathbf{f}\cdot\mathbf{g} \equiv (\mathbf{f}\,|\,\mathbf{g}) \equiv \langle \mathbf{f}|\mathbf{g}\rangle = \sum_{k=1}^{3} f_k g_k = f_1 g_1 + f_2 g_2 + f_3 g_3 \,. \qquad (3.159)$$

We denote the scalar product either with a centered dot, round brackets, or angle brackets. The sum is taken over both indices, and the result is a single real number that can be positive, zero, or negative. For common transformations, such as rotation of the coordinate system or inversion, the scalar product does not depend on the choice of coordinate system. We call such numbers scalars.

◖ **Example 3.8**

$$\mathbf{f} = (1,2,3)\,; \qquad \mathbf{g} = (-7,-1,2)\,;$$

$$\mathbf{f}\cdot\mathbf{g} = f_1 g_1 + f_2 g_2 + f_3 g_3 = -7-2+6 = -3\,.$$

◗

Let's consider any two vectors **f** and **g** and choose a coordinate system as simply as possible. We'll let the x-axis point in the direction of the first vector, the y-axis lie in the plane of both vectors, and the z-axis be perpendicular to them (the choice of coordinate system is always up to us):

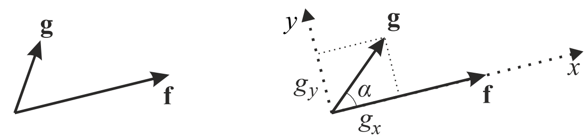


Fig. 3.17: Choice of coordinate system when interpreting the scalar product

If we denote the magnitudes of the vectors f and g and the angle between them as α, their coordinates will be

$$\begin{aligned}\mathbf{f} &= (f_1, f_2, f_3) = (f, 0, 0)\,;\\ \mathbf{g} &= (g_1, g_2, g_3) = (g\cos\alpha,\, g\sin\alpha,\, 0)\,.\end{aligned} \qquad (3.160)$$

We can now easily find the scalar product of the two vectors using the definition:

$$\mathbf{f}\cdot\mathbf{g} \equiv f_1 g_1 + f_2 g_2 + f_3 g_3 = f\,g\cos\alpha = \|\,\mathbf{f}\,\|\;\|\,\mathbf{g}\,\|\cos\alpha\,. \qquad (3.161)$$

The scalar product of two vectors is therefore equal to the product of their magnitudes and the cosine of the angle between them. This is very important. Let's reiterate that this result will always be the same, regardless of the choice of coordinate system. So what is the scalar product good for? We can use it to determine the angle between two vectors:

► $$\cos\alpha = \frac{\mathbf{f}\cdot\mathbf{g}}{\|\,\mathbf{f}\,\|\;\|\,\mathbf{g}\,\|}\,. \qquad (3.162)$$

It is also easy to determine the magnitude of a vector. It suffices to set both vectors in equation (3.161) equal, yielding $\mathbf{f}\cdot\mathbf{f} = \|\,\mathbf{f}\,\|^2$. From this, we can determine the magnitude (which is given by the square root of the scalar product of the vector with itself).

► $$\|\,\mathbf{f}\,\| = \sqrt{\mathbf{f}\cdot\mathbf{f}}\,. \qquad (3.163)$$

The scalar product also allows us to calculate the mechanical work done along a given path γ when a force **F** acts on an object:

► $$\Delta A = \int_\gamma \mathbf{F}\cdot \mathrm{d}\mathbf{s}\,. \tag{3.164}$$

◖ **Example 3.9**

Problem: Find magnitudes and the angle between vectors **f** = (1, 3, 0) and **g** = (2, 2, 0).

Solution: First, we find the magnitudes of both vectors:

$$\|\mathbf{f}\| = \sqrt{\mathbf{f}\cdot\mathbf{f}} = \sqrt{f_x^2 + f_y^2 + f_z^2} = \sqrt{10}\,;$$
$$\|\mathbf{g}\| = \sqrt{\mathbf{g}\cdot\mathbf{g}} = \sqrt{g_x^2 + g_y^2 + g_z^2} = \sqrt{8}\,.$$

Now we can easily find the angle between the two vectors:

$$\cos\alpha = \frac{\mathbf{f}\cdot\mathbf{g}}{\|\mathbf{f}\|\,\|\mathbf{g}\|} = \frac{f_x g_x + f_y g_y + f_z g_z}{\sqrt{10}\sqrt{8}} = \frac{8}{\sqrt{80}} \doteq 0{,}89\,.$$

The corresponding angle is approximately 27°. Draw both vectors and verify the calculation visually. ◗

Schwarz's lemma holds for the scalar product; it is useful in various estimates and follows immediately from relation (3.161):

► $$|\,\mathbf{f}\cdot\mathbf{g}\,| \le \|\mathbf{f}\|\,\|\mathbf{g}\|\,. \tag{3.165}$$

- Scalar product is given by the equation $\mathbf{f}\cdot\mathbf{g} \equiv f_1 g_1 + f_2 g_2 + f_3 g_3$.
- Scalar product can also be written as $\mathbf{f}\cdot\mathbf{g} = \|\mathbf{f}\|\,\|\mathbf{g}\|\cos\alpha$.
- Schwarz lemma holds for the scalar product $|\,\mathbf{f}\cdot\mathbf{g}\,| \le \|\mathbf{f}\|\,\|\mathbf{g}\|$.
- The scalar product does not depend on the choice of coordinate system.
- Scalar product allows us to determine the angle between vectors $\cos\alpha = \mathbf{f}\cdot\mathbf{g}/(\|\mathbf{f}\|\,\|\mathbf{g}\|)$.
- Magnitude of a vector is always equal to $\|\mathbf{f}\| = \sqrt{\mathbf{f}\cdot\mathbf{f}}$.

* * *

Using relations (3.153), we have extended the concept of a vector to objects more general than ordered triples and introduced the concept of a linear vector space. Now we will extend the concept of the scalar product to different linear vector spaces.

$\mathcal{R}^N$ space of real *N*-tuples

The transition from ordered triples to a larger number of dimensions is straightforward. All derived properties remain intact, whether it be the magnitude of a vector – defined as the square root of the scalar product of an element with itself – the definition of the angle between two elements, Schwarz's lemma, or other relationships. In the scalar product, the summation will extend from three to N:

$$\mathbf{f} = (f_1, \dots, f_N) \quad , \qquad \mathbf{g} = (g_1, \dots, g_N) \quad ; \qquad f_l\,, g_l \in \mathcal{R}, \tag{3.166}$$

► $$\mathbf{f} \cdot \mathbf{g} \;\equiv\; f_1 g_1 + \dots + f_N g_N \;=\; \sum_{k=1}^{N} f_k g_k \,. \tag{3.167}$$

C^N space of complex N-tuples

Let us now assume that the components of the vector are complex numbers. This assumption poses a problem when defining the magnitude of the vector, since the definition of the scalar product (3.167) does not guarantee that the magnitude of the vector will be a non-negative real number. However, from the properties of complex numbers, we know that the following holds for the magnitude of a complex number:

$$\| f \| = \sqrt{\bar{f} f} = \sqrt{(x - \mathrm{i}y)(x + \mathrm{i}y)} = \sqrt{x^2 + y^2} \,. \tag{3.168}$$

For a complex N-tuple, it would therefore be natural for the following to hold

$$\| \mathbf{f} \| = \sqrt{\bar{f}_1 f_1 + \dots + \bar{f}_N f_N} = \sqrt{| f_1 |^2 + \dots + | f_N |^2} \,. \tag{3.169}$$

We must adapt the definition of the scalar product accordingly and treat one of the arguments of the scalar product as complex conjugate (by convention, the left one):

$$\mathbf{f} = (f_1, \dots, f_N) \quad , \qquad \mathbf{g} = (g_1, \dots, g_N) \quad ; \qquad f_l\,, g_l \in C, \tag{3.170}$$

► $$\mathbf{f} \cdot \mathbf{g} \equiv \bar{f}_1 g_1 + \dots + \bar{f}_N g_N = \sum_{k=1}^{N} \bar{f}_k g_k . \tag{3.171}$$

The scalar product defined in this way is generally a complex number. However, the scalar product of an element with itself is always a non-negative real number, so the definition of the norm as the square root can be preserved

► $$\| \mathbf{f} \| \equiv \sqrt{\mathbf{f} \cdot \mathbf{f}} = \sqrt{\bar{f}_1 f_1 + \dots + \bar{f}_N f_N} = \sqrt{| f_1 |^2 + \dots + | f_N |^2} \tag{3.172}$$

For real N-tuples, our new definition coincides with the original relation (3.167), so this is a direct extension of that definition. Schwarz lemma holds again, and in the same way we can introduce a definition of the angle between two elements (in this case, complex numbers), and so on. Everything we know from the original "arrows" remains valid.

◖ **Example 3.10**

Find the magnitude and the dot product of the vectors **f** = (i, 3) and **g** = (1, 2i).

$$\begin{aligned} \| \mathbf{f} \| &\equiv \sqrt{\mathbf{f} \cdot \mathbf{f}} = \sqrt{\bar{f}_1 f_1 + \bar{f}_2 f_2} = \sqrt{-\mathrm{i}\,\mathrm{i} + 3^2} = \sqrt{10} \,; \\ \| \mathbf{g} \| &\equiv \sqrt{\mathbf{g} \cdot \mathbf{g}} = \sqrt{\bar{g}_1 g_1 + \bar{g}_2 g_2} = \sqrt{1^2 + (-2\mathrm{i})(2\mathrm{i})} = \sqrt{5} \,; \\ \mathbf{f} \cdot \mathbf{g} &\equiv \bar{f}_1 g_1 + \bar{f}_2 g_2 = (\mathrm{i})(1) + (3)(2\mathrm{i}) = 7\,\mathrm{i} \,. \end{aligned} \tag{3.173}$$

◗

ℓ^2 Space of sequences

Another generalization will be the transition from complex N-tuples to complex sequences with an infinite number of elements. We denote such spaces by ℓ^2 ("el two"). We will extend the scalar product in a straightforward manner:

$$\mathbf{f} = \{f_1, \dots, f_n, \dots\} = \{f_l\}_{l=1}^{\infty}, |\mathbf{g}> = \{g_l\}_{l=1}^{\infty} ; f_l, g_l \in C, \tag{3.174}$$

► $$\mathbf{f} \cdot \mathbf{g} \equiv f_1^* g_1 + \cdots + f_n^* g_n + \cdots = \sum_{k=1}^{\infty} f_k^* g_k = f_k^* g_k . \tag{3.175}$$

The scalar product defined in this way makes sense only for convergent sequences. We can include in the space ℓ^2 only those elements **f** for which $|| \mathbf{f} || < \infty$, i.e., we require

$$\sum_{k=1}^{\infty} \bar{f}_k f_k < \infty \qquad \text{for} \quad \forall \quad \mathbf{f} \in l^2 . \tag{3.176}$$

Then the following applies

$$|\mathbf{f} \cdot \mathbf{g}| = |\sum_{k=1}^{\infty} \bar{f}_k g_k| \le ||\mathbf{f}|| \cdot ||\mathbf{g}|| < \infty \qquad \text{for} \quad \forall \quad |\mathbf{f}>, |\mathbf{g}> \in l^2 , \tag{3.177}$$

since Schwarz lemma also holds for infinite sequences. Thus, the space ℓ^2 contains only sequences that are summable to a square, hence the symbol "2" in the name of the space. For example, the sequence **f** = {1, 2, 3, 4...} does not belong to the space ℓ^2, as

$$\|\mathbf{f}\| \equiv \sqrt{\mathbf{f} \cdot \mathbf{f}} = \sqrt{1 + 2^2 + 3^2 \cdots} = \infty .$$

On the other hand, the sequence **f** = {1, 1/2, 1/4, 1/8...} belongs to the space ℓ^2, because

$$\|\mathbf{f}\| \equiv \sqrt{\mathbf{f} \cdot \mathbf{f}} = \sqrt{1^2 + (1/2)^2 + (1/4)^2 \cdots} < \infty .$$

$\mathcal{L}^2$ Space of complex functions of a real variable

In a further generalization of the space ℓ^2, we can assume that the index k is continuous. Instead of k, we will write x: f_x. However, it is nothing other than a complex function of a real variable, which is customarily written in the form $f(x)$, i.e.

$$\mathbf{f} \equiv f_x \equiv f(x), \qquad \mathbf{g} \equiv g_x \equiv g(x), \quad ; \qquad x \in \mathcal{R}, \quad f, g \in C, \tag{3.178}$$

► $$\mathbf{f} \cdot \mathbf{g} \equiv \int_{-\infty}^{+\infty} f^*(x) \, g(x) \, \mathrm{d}x. \tag{3.179}$$

We denote this space by $\mathcal{L}^2$. By analogy with $\mathcal{L}^2$, the space should include only elements with a norm satisfying $|| \mathbf{f} || < \infty$, i.e.,

$$\int_{-\infty}^{+\infty} \bar{f}(x) \, f(x) \, \mathrm{d}x = \int_{-\infty}^{+\infty} | f(x) |^2 \, \mathrm{d}x < \infty \qquad \text{pro} \quad \forall \quad f(x) \in \mathcal{L}^2. \tag{3.180}$$

According to Schwarz lemma, which also holds for integrals,

$$|\mathbf{f}\cdot\mathbf{g}| \;=\leq\; \|\mathbf{f}\|\cdot\|\mathbf{g}\| \;<\; \infty \qquad \text{pro} \quad \forall \quad \mathbf{f},\mathbf{g}\in \mathcal{L}^2 \tag{3.181}$$

and the scalar product is well-defined. $\mathcal{L}^2$ is sometimes called the *space of square-integrable functions*. It can also be defined for a domain other than $(-\infty, \infty)$; in that case, we write $\mathcal{L}^2(\mathcal{M})$, where M is the definition domain of the functions $f(x) \in \mathcal{L}^2(\mathcal{M})$. We can now generalize definition of spaces with a scalar product (so-called unitary spaces).

◖ **Example 3.11**

Problem: In space $\mathcal{L}^2(0,1)$ of functions integrable with the square on the interval <0, 1> find the magnitudes of functions $f(x) = 1$, $g(x) = x^2$ and an angle between them.

Solution: The procedure is very straightforward:

$$\|f\|^2 = \mathbf{f}\cdot\mathbf{f} = (f\,|\,f) = \int_0^1 1\;\mathrm{d}x = 1\,; \qquad \Rightarrow \qquad \|f\| = 1\,;$$

$$\|g\|^2 = \mathbf{g}\cdot\mathbf{g} = (g\,|\,g) = \int_0^1 x^2 x^2\;\mathrm{d}x = \frac{1}{5}\,; \qquad \Rightarrow \qquad \|g\| = 1/\sqrt{5}\,;$$

$$\mathbf{f}\cdot\mathbf{g} = (f\,|\,g) = \int_0^1 \overline{f}\,g\;\mathrm{d}x = \int_0^1 1\cdot x^2\;\mathrm{d}x = 1/3\,;$$

$$\cos\alpha = \frac{(f\,|\,g)}{\|f\|\|g\|} = \frac{1/3}{1\cdot 1/\sqrt{5}} = \sqrt{5}/3 \qquad \Rightarrow \qquad \alpha \approx 42^\circ\,.$$ ◗

◖ **Example 3.12**

Problem: Find the scalar product of the functions $f(x) = \sin x$, $g(x) = \sin 2x$ in $\mathcal{L}^2(0,2\pi)$.

Solution: The procedure is again straightforward

$$(f\,|\,g) = \int_0^{2\pi} \overline{f}\,g\;\mathrm{d}x = \int_0^{2\pi} \sin x \sin 2x\;\mathrm{d}x = \int_0^{2\pi} \sin x\; 2\sin x\cos x\;\mathrm{d}x =$$

$$= 2\int_0^{2\pi} \sin^2 x\cos x\;\mathrm{d}x = \left|\begin{matrix}\text{substitution}\\ \sin x \equiv \xi\end{matrix}\right| = \left[2\frac{\xi^3}{3}\right]_{\dots}^{\dots} = \left[2\frac{\sin^3(x)}{3}\right]_0^{2\pi} = 0\,.$$

This result means that the functions f and g are orthogonal to each other. In fact, the functions {1, sin x, cos x, sin 2x, cos 2x…} are all orthogonal to each other and form an orthogonal basis in the space $\mathcal{L}^2(0,2\pi)$. Try to prove for yourself that $\|\sin x\| = \sqrt{\pi}$. ◗

Unitary space (space with a scalar product)

We call a linear vector space $\mathcal{V}$ a unitary space $\mathcal{U}$ if it has the following operations

Labeling	From → to	Name	Notation
+	$\mathcal{V}\times\mathcal{V}\rightarrow\mathcal{V}$	Vector addition	$\mathbf{h} = \mathbf{f} + \mathbf{g}$
·	$\mathcal{R}(\mathcal{C})\times\mathcal{V}\rightarrow\mathcal{V}$	Vector scalar multiplication	$\mathbf{g} = \alpha\cdot\mathbf{f}$
(|)	$\mathcal{V}\times\mathcal{V}\rightarrow\mathcal{R}(\mathcal{C})$	Scalar product	$\alpha = (\mathbf{f}\,\|\,\mathbf{g})$

The operations satisfy the properties of the scalar product derived in the preceding text:

$$\begin{aligned}
&1)\quad (\mathbf{f}\,|\,\mathbf{g}+\mathbf{h}) = (\mathbf{f}\,|\,\mathbf{g})+\mathbf{f}\,|\,\mathbf{h}),\\
&2)\quad (\mathbf{f}\,|\,\alpha\,\mathbf{g}) = \alpha\,(\mathbf{f}\,|\,\mathbf{g}),\\
&3)\quad (\mathbf{g}\,|\,\mathbf{f}) = \overline{(\mathbf{f}\,|\,\mathbf{g})} \qquad \Rightarrow (\alpha\,\mathbf{f}\,|\,\mathbf{g}) = \bar{\alpha}\,(\mathbf{f}\,|\,\mathbf{g}),\\
&4)\quad (\mathbf{f}\,|\,\mathbf{f}) \geq 0 \quad ; \quad (\mathbf{f}\,|\,\mathbf{f}) = 0 \quad \Leftrightarrow \quad \mathbf{f}=0.
\end{aligned} \tag{3.182}$$

The first two conditions imply linearity in the right-hand argument. The third operation implies antilinearity in the left-hand argument (additivity plus the extraction of a complex-conjugated constant). If a scalar product space is complete – i.e., if every convergent sequence converges to an element of the space – we call it a *Hilbert space*.

Convolution

Let's now try to rewrite the linear matrix transformation

$$g_k = \sum_l A_{kl} f_l \tag{3.183}$$

into the function space. We'll replace the discrete indices k and l with continuous indices x and y, and change the summation to an integration. The result will be:

$$g(x) = \int A(x,y) f(y)\,\mathrm{d}y\,. \tag{3.184}$$

This operation is called *convolution*, and the function of two variables $A(x, y)$ is known as the convolution *kernel*. Most integral transforms are convolutions. Depending on the choice of the function A, we can have the Fourier transform, the Laplace transform, the Abell transform, and many others. Convolution is nothing more than matrix multiplication with continuous indices, which is why we symbolically write it as

$$g = A * f\,. \tag{3.185}$$

Convolution maps a function f to a new function g using "rule" A. If the function A has only one variable, we automatically assume integration in the context of convolution

$$g(x) = \int A(x-y) f(y)\,\mathrm{d}y\,. \tag{3.186}$$

Using convolution with the so-called Green's function, it is easy to find solutions to certain partial differential equations. We will return to this topic in Section 3.8.4.

◖ **Example 3.13**

Problem: Find the convolution $A(x, y) = (x-y)^2$ with the function $f(x) = x$ in $\mathcal{L}^2\,(0,1)$

Solution:

$$\begin{aligned}
g(x) &= A * f = \int_0^1 A(x,y) f(y)\,\mathrm{d}y = \\
&= \int_0^1 (x-y)^2\, y\,\mathrm{d}y = \int_0^1 (x^2 y - 2xy^2 + y^3)\,\mathrm{d}y = \\
&= \left[x^2 \frac{y^2}{2} - 2x\frac{y^3}{3} + \frac{y^4}{4} \right]_0^1 = \frac{1}{4} - \frac{2}{3}x + \frac{1}{2}x^2.
\end{aligned}$$

◗

3.3.3 Vector Product

We already know how to stretch and add elements of a linear vector space and how to take their scalar product. In physics, however, another operation is also important: the vector product, which assigns an ordered triple to two vectors, whose properties resemble those of vectors in some respects. We will proceed similarly to the scalar product. We will introduce this operation, show where it can be useful, then examine its properties, and finally generalize the operation to any objects satisfying these properties.

Symmetric and antisymmetric matrices

In matrix theory, symmetric and antisymmetric matrices are very important. Symmetric matrices satisfy the relation

► $$S_{kl} = S_{lk}, \tag{3.187}$$

i.e., swapping the indices has no effect on the matrix. The elements resulting from flipping the matrix across the diagonal are the same both below and above the diagonal. An example is the matrix

$$\mathbb{S} = \begin{pmatrix} 1 & 2 & 4\mathrm{i} \\ 2 & \mathrm{i} & -1 \\ 4\mathrm{i} & -1 & 5 \end{pmatrix}. \tag{3.188}$$

For symmetric matrices, it is enough to enter the elements on the diagonal and above it (marked as solid circles). The remaining elements can be easily calculated:

$$(\bullet)\,; \quad \begin{pmatrix} \bullet & \bullet \\ \circ & \bullet \end{pmatrix}; \quad \begin{pmatrix} \bullet & \bullet & \bullet \\ \circ & \bullet & \bullet \\ \circ & \circ & \bullet \end{pmatrix}; \quad \begin{pmatrix} \bullet & \bullet & \bullet & \bullet \\ \circ & \bullet & \bullet & \bullet \\ \circ & \circ & \bullet & \bullet \\ \circ & \circ & \circ & \bullet \end{pmatrix}. \tag{3.189}$$

In one dimension, symmetric matrices have a single independent element; in two dimensions, three elements; in three dimensions, six elements; and in four dimensions, ten. In general relativity, the gravitational field is described by a 4×4 symmetric matrix (the so-called metric coefficients, which describe the curvature of spacetime); therefore, general relativity is based on ten partial differential equations for these coefficients. The second group consists of antisymmetric matrices whose elements satisfy the relation

► $$A_{kl} = -A_{lk}. \tag{3.190}$$

An antisymmetric matrix has zeros on the diagonal; for example, for the element A22, the following holds when the indices are swapped

$$A_{22} = -A_{22} \quad \Rightarrow \quad 2A_{22} = 0 \quad \Rightarrow \quad A_{22} = 0. \tag{3.191}$$

An example of an antisymmetric matrix is

$$\mathbb{A} = \begin{pmatrix} 0 & \mathrm{i} & 5 \\ -\mathrm{i} & 0 & -1 \\ -5 & +1 & 0 \end{pmatrix}. \tag{3.192}$$

For an antisymmetric matrix, it is sufficient to specify only the elements above the diagonal, and the entire matrix is determined:

$$(0)\,; \qquad \begin{pmatrix} 0 & \bullet \\ \circ & 0 \end{pmatrix}; \qquad \begin{pmatrix} 0 & \bullet & \bullet \\ \circ & 0 & \bullet \\ \circ & \circ & 0 \end{pmatrix}; \qquad \begin{pmatrix} 0 & \bullet & \bullet & \bullet \\ \circ & 0 & \bullet & \bullet \\ \circ & \circ & 0 & \bullet \\ \circ & \circ & \circ & 0 \end{pmatrix}. \tag{3.193}$$

In one dimension, an antisymmetric matrix has no independent elements; in two dimensions, it has a single independent element; in three dimensions, it has three; and in four dimensions, there are six independent elements. The electromagnetic field in spacetime is described by the electromagnetic field tensor (1.246), a 4×4 antisymmetric matrix; therefore, to describe the electromagnetic field, we need six quantities – three components of the electric field and three components of the magnetic field.

If we sum the products of all corresponding elements of an antisymmetric and a symmetric matrix (of the same form), the result is always zero:

► $$A_{kl}S_{kl} = 0. \tag{3.194}$$

We used the summation convention, i.e., the sum is taken over both pairs of indices k and l. If we expand all the terms, there will always be one positive and one negative term, and they will cancel each other. For example, the sum will include terms $A_{25}S_{25}$ and $A_{52}S_{52}$, which has the opposite sign due to the antisymmetry of A. A general proof can be obtained by renaming the indices, using symmetry, and renaming them again:

$$A_{kl}S_{kl} = A_{op}S_{op} = -A_{po}S_{po} = -A_{kl}S_{kl}. \tag{3.195}$$

If we read the beginning and the end, we see that

$$A_{kl}S_{kl} = -A_{kl}S_{kl} \qquad \Rightarrow \qquad 2A_{kl}S_{kl} = 0 \qquad \Rightarrow \qquad A_{kl}S_{kl} = 0. \tag{3.196}$$

If we have a general matrix that exhibits neither symmetry nor antisymmetry, we can always decompose it into a symmetric and an antisymmetric part as follows:

► $$M_{kl} = \frac{1}{2}\left(M_{kl} + M_{lk}\right) + \frac{1}{2}\left(M_{kl} - M_{lk}\right). \tag{3.197}$$

The first part is clearly a symmetric matrix, and the second part is an antisymmetric matrix. Every matrix can therefore be decomposed into a symmetric and an antisymmetric part. Let us now assume that the matrix is the tensor product of two vectors:

$$M_{kl} = f_k g_l\,. \tag{3.198}$$

The decomposition into symmetric and antisymmetric parts will now be

$$M_{kl} = \frac{1}{2}\left(f_k g_l + f_l g_k\right) + \frac{1}{2}\left(f_k g_l - f_l g_k\right). \tag{3.199}$$

Antisymmetric expressions in parentheses, such as $f_2 g_3 - f_3 g_2$, are in fact components of a vector product. Among the most useful symmetric matrices is the Kronecker delta, and among the antisymmetric matrices is the Levi-Civita symbol:

► $$\delta_{kl} = \begin{cases} 0\,; & k \neq l\,, \\ 1\,; & k = l\,. \end{cases} \tag{3.200}$$

$$\varepsilon_{klm} = -\varepsilon_{lkm}\,,\quad \varepsilon_{klm} = -\varepsilon_{mlk}\,,\quad \varepsilon_{klm} = -\varepsilon_{kml}\,.\qquad \varepsilon_{123} = 1\,; \tag{3.201}$$

The Kronecker delta consists of the elements of the identity matrix. It has ones on the diagonal and zeros off the diagonal. The Levi-Civita symbol looks more complicated, but it is not. It is a totally antisymmetric tensor with respect to all pairs of indices. Thus, swapping any two indices results in a change of sign. Such a symbol has a single independent element, and all others can be calculated. The independent element is chosen to be $\varepsilon_{123} = 1$. Any elements of the Levi-Civita symbol with two identical indices are zero (this is due to antisymmetry), for example

$$\varepsilon_{112} = \varepsilon_{322} = \varepsilon_{222} = \varepsilon_{313} = \cdots = 0\,. \tag{3.202}$$

Non-zero elements have all three indices distinct. They can be derived from ε_{123}, e.g.:

$$\begin{aligned} \varepsilon_{123} &= 1\,, \\ \varepsilon_{213} &= -\varepsilon_{123} = -1\,, \\ \varepsilon_{312} &= -\varepsilon_{132} = +\varepsilon_{123} = +1 \\ &\text{atd.} \end{aligned} \tag{3.203}$$

The values of the Levi-Civita symbol are therefore only 0, +1, or −1.

Vector product

We denote the vector (cross) product of two vectors **f** and **g** as **h** = **f**×**g**, or **h** = [**f**, **g**], or **h** = [**f** | **g**]. Its components are defined by the following relations:

$$\begin{aligned} h_1 &= f_2 g_3 - f_3 g_2\,, \\ h_2 &= f_3 g_1 - f_1 g_3\,, \\ h_3 &= f_1 g_2 - f_2 g_1\,. \end{aligned} \tag{3.204}$$

At first glance, this definition might seem a bit scary, but it’s actually simple. You just need to remember the relationship for the first component. After the 1 on the left, the indices 2 and 3 follow on the right (and vice versa for minus). If you remember this relationship, you’ve got it. Everything else follows from a cyclic permutation: after one comes two, after two comes three, and after three comes one again. Or, after index x comes y, then z, and after that x:

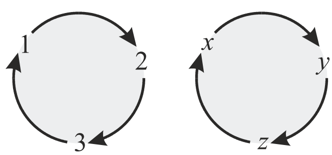


Fig. 3.18: Cyclic permutation

The result is a triple of numbers that has properties similar to those of vectors. We sometimes refer to this structure as a *pseudovector*.

◖ **Example 3.14**

Problem: Find the cross product of the vectors **f** = (1, 2, 3) and **g** = (4, 5, 6).

Solution: Let's start directly from the definition:

$$\mathbf{f}\times\mathbf{g} \equiv (f_2 g_3 - f_3 g_2,\ f_3 g_1 - f_1 g_3,\ f_1 g_2 - f_2 g_1) =$$
$$= (2\cdot 6 - 3\cdot 5, 3\cdot 4 - 1\cdot 6, 1\cdot 5 - 2\cdot 4) = (-3, 6, -3).$$

◗

Calculating the vector product using the determinant

The vector product can be calculated using the expansion of the following determinant with respect to the first row:

$$\mathbf{f}\times\mathbf{g} = \det\begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ f_1 & f_2 & f_3 \\ g_1 & g_2 & g_3 \end{pmatrix} =$$
$$= \mathbf{e}_1(f_2 g_3 - f_3 g_2) + \mathbf{e}_2(\cdots) + \mathbf{e}_3(\cdots) = \tag{3.205}$$
$$= (f_2 g_3 - f_3 g_2,\ f_3 g_1 - f_1 g_3,\ f_1 g_2 - f_2 g_1).$$

Definition of the vector product using the Levi-Civita symbol

In theoretical physics, the definition of the vector product using the Levi–Civita symbol is commonly used (various vector identities can be easily derived from this definition):

► $$h_k = \varepsilon_{klm} f_l g_m \tag{3.206}$$

The first index (k) is free and appears on both sides of the equality. The indices l and m are summation indices. Let's calculate the first term, for example:

$$h_1 = \varepsilon_{1lm} f_l g_m .$$

Given the properties of the Levi-Civita symbol, the only non-zero terms in the sum will be those in which no indices are repeated, that is

$$h_1 = \varepsilon_{123} f_2 g_3 + \varepsilon_{132} f_3 g_2 = f_2 g_3 - f_3 g_2 .$$

We would calculate the other components in the same way.

Transformational properties of the vector product

It is clear from definition (3.204) that the result of the vector product is not transformed as vectors, but as the products of the components of two vectors. The individual parts of the vector product **h** = **f**×**g** are the elements of the matrix

$$h_{kl} \equiv f_k g_l - f_l g_k = \begin{pmatrix} 0 & f_1 g_2 - f_2 g_1 & f_1 g_3 - f_3 g_1 \\ f_2 g_1 - f_1 g_2 & 0 & f_2 g_3 - f_3 g_2 \\ f_3 g_1 - f_1 g_3 & f_3 g_2 - f_2 g_3 & 0 \end{pmatrix}, \tag{3.207}$$

$$h_{kl} = \begin{pmatrix} 0 & +h_3 & -h_2 \\ -h_3 & 0 & +h_1 \\ +h_2 & -h_1 & 0 \end{pmatrix} . \tag{3.208}$$

The vector product will not transform as vectors do, but its components will transform as the products of vector components, i.e., as a second-order tensor:

$$\tilde{f}_k = A_{kl} f_l \, ; \qquad \tilde{g}_k = A_{kl} f_l \, ; \qquad \tilde{h}_{kl} = A_{ko} A_{lp} f_o g_p \, . \tag{3.209}$$

The vector product is therefore a second-order antisymmetric tensor (3.207). Its matrix has three independent components that can be arranged as an ordered triple, which we call a *pseudovector*. We can easily see that it is not a vector by considering a spatial inversion of the coordinates (the new system will have axes $-x$, $-y$, $-z$). Under this transformation, **f** becomes **−f** and **g** becomes **−g**, but the result of the product $\mathbf{h} = \mathbf{f}\times\mathbf{g}$ remains unchanged; thus, it has different transformation properties than vectors.

Meaning of the vector product

Let us now find the value of the vector product. To do this, we will use the same coordinate system as we did for the scalar product, i.e., the vectors will have components (3.160). By definition, the vector product is then

$$\mathbf{f}\times\mathbf{g} = (0, 0, f\,g\sin\alpha) \tag{3.210}$$

The vector product has a component only along the z-axis, meaning it points perpendicular to both original vectors. Its magnitude is equal to $fg \sin \alpha$, i.e., the area of the parallelogram “stretched” across both vectors.

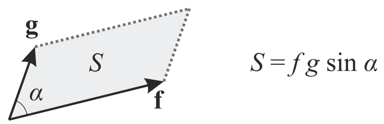


Fig. 3.19: The vector product and the area of a parallelogram

Corkscrew rule: *If we place the tip of a wine corkscrew at the intersection of the vectors and turn it from the first vector to the second, the corkscrew will move in the direction of the vector product. Using this rule, we can easily determine which of the two possible perpendicular directions is the correct one.*

And what is the vector product good for? We can easily find the perpendicular to two vectors. The vector product is also useful for calculating the area of a parallelogram. In physics, we use the vector product to describe the angular momentum of a body, the torque of a force, or the motion of an electrically charged particle in a magnetic field.

◖ **Example 3.15**

Problem: Find the area of the parallelogram spanned on $\mathbf{f} = (1, 2, 3)$ and $\mathbf{g} = (4, 5, 6)$.

Solution: From the previous example, we know that the vector product of these vectors is $(-3, 6, -3)$. The area (magnitude of this vector) is $(9+36+9)^{1/2}$, i.e., ~ 7.35. ◗

- Vector product is defined by $\mathbf{f} \times \mathbf{g} \equiv (f_2\, g_3 - f_3\, g_2,\; f_3\, g_1 - f_1\, g_3, f_1\, g_2 - f_2\, g_1)$.
- Vector product can also be written as $(\mathbf{f}\times\mathbf{g})_k = \varepsilon_{klm}\, f_l\, g_m$.
- Vector product can be calculated using determinants.
- Vector product is perpendicular to both vectors **f**, **g**.
- Vector product magnitude is the area of a parallelogram formed by vectors **f**, **g**.
- Vector product can be used to easily construct perpendicular vectors.
- Vector product is an antisymmetric tensor of the second-order.
- Vector product result is called a pseudovector.

3.3.4 Vector Identities

When describing physical laws, various vector identities are often used. We can easily derive them if we know that a simple relationship holds between the components of the Levi-Civita tensor and the Kronecker delta:

► $$\varepsilon_{klm}\varepsilon_{kop} = \delta_{lo}\delta_{mp} - \delta_{lp}\delta_{mo}\,. \tag{3.211}$$

This key relationship can most easily be derived by direct verification (on both sides, the sums consist of zeros, ones, and negative ones). However, there are also more sophisticated derivations, for example via symmetries or from the properties of orthogonal transformations. If the reader is interested, he can read the derivation of this identity based on orthonormal transformations. Let us consider an orthonormal transformation (the rows of the matrix consist of unit vectors that are mutually orthogonal)

$$\tilde{f}_k = a_{kl}\, f_l\,. \tag{3.212}$$

We can now write the Levi-Civita symbol using the coefficients of this transformation

$$\varepsilon_{klm} = \det\begin{pmatrix} a_{1k} & a_{1l} & a_{1m} \\ a_{2k} & a_{2l} & a_{2m} \\ a_{3k} & a_{3l} & a_{3m} \end{pmatrix}. \tag{3.213}$$

Verify that the Levi-Civita tensor written in this form possesses all of its properties. If any two indices are the same, the determinant has two identical columns and is zero. If we swap two columns, the sign of the determinant changes (this corresponds to swapping two indices), i.e., the Levi-Civita tensor is totally antisymmetric. Now we will apply the following in turn: 1) the determinant of a transposed matrix remains unchanged; 2) the product of two determinants is equal to the determinant of the product of the matrices; 3) the transformation is orthogonal:

$$\varepsilon_{klm}\varepsilon_{opq} = \det\begin{pmatrix} a_{1k} & a_{1l} & a_{1m} \\ a_{2k} & a_{2l} & a_{2m} \\ a_{3k} & a_{3l} & a_{3m} \end{pmatrix} \det\begin{pmatrix} a_{1o} & a_{1p} & a_{1q} \\ a_{2o} & a_{2p} & a_{2q} \\ a_{3o} & a_{3p} & a_{3q} \end{pmatrix} =$$

$$= \det\begin{pmatrix} a_{1k} & a_{2k} & a_{3k} \\ a_{1l} & a_{2l} & a_{3l} \\ a_{1m} & a_{2m} & a_{3m} \end{pmatrix} \det\begin{pmatrix} a_{1o} & a_{1p} & a_{1q} \\ a_{2o} & a_{2p} & a_{2q} \\ a_{3o} & a_{3p} & a_{3q} \end{pmatrix} =$$

$$= \det \begin{pmatrix} \delta_{ko} & \delta_{kp} & \delta_{kq} \\ \delta_{lo} & \delta_{lp} & \delta_{lq} \\ \delta_{mo} & \delta_{mp} & \delta_{mq} \end{pmatrix}.$$

Now the derivation is straightforward; we expand the determinant, for example, with respect to the first row, and in the result we equate the first two indices:

$$\varepsilon_{klm}\varepsilon_{opq} = \delta_{ko}\left(\delta_{lp}\delta_{mq} - \delta_{lq}\delta_{mp}\right) - \delta_{kp}\left(\delta_{lo}\delta_{mq} - \delta_{lq}\delta_{mo}\right) + \delta_{kq}\left(\delta_{lo}\delta_{mp} - \delta_{lp}\delta_{mo}\right) \Rightarrow$$

$$\begin{aligned} \varepsilon_{klm}\varepsilon_{kpq} &= \delta_{kk}\left(\delta_{lp}\delta_{mq} - \delta_{lq}\delta_{mp}\right) - \delta_{kp}\left(\delta_{lk}\delta_{mq} - \delta_{lq}\delta_{mk}\right) + \delta_{kq}\left(\delta_{lk}\delta_{mp} - \delta_{lp}\delta_{mk}\right) = \\ &= 3\delta_{lp}\delta_{mq} - 3\delta_{lq}\delta_{mp} - \delta_{kp}\delta_{lk}\delta_{mq} + \delta_{kp}\delta_{lq}\delta_{mk} + \delta_{kq}\delta_{lk}\delta_{mp} - \delta_{kq}\delta_{lp}\delta_{mk} = \\ &= 3\delta_{lp}\delta_{mq} - 3\delta_{lq}\delta_{mp} - \delta_{lp}\delta_{mq} + \delta_{mp}\delta_{lq} + \delta_{ql}\delta_{mp} - \delta_{mq}\delta_{lp} = \\ &= \delta_{lp}\delta_{mq} - \delta_{lq}\delta_{mp} . \end{aligned}$$

This completes the derivation of the desired relationship.

Examples of vector identities

We will now derive some important vector identities as simple examples, and summarize them in a clear table at the end (we will denote 3D vectors here by **A**, **B**, **C**, etc.)

◖ **Example 3.16:** div grad f

$$\operatorname{div}\operatorname{grad} f = \nabla\cdot(\nabla f) = \partial_k(\partial_k)f = \frac{\partial^2 f}{\partial x_k\,\partial x_k} = \Delta f .$$

The calculation resulted in the sum of second derivatives, which is the Laplace operator. ◗

◖ **Example 3.17:** rot grad f

$$\left[\operatorname{rot}\operatorname{grad} f\right]_k = \left[\nabla\times(\nabla f)\right]_k = \varepsilon_{klm}\partial_l\left(\partial_m f\right) = \varepsilon_{klm}\partial^2{}_{lm} f = 0.$$

The next-to-last expression is obtained by double contraction of the antisymmetric and symmetric matrices with respect to the indices l and m. According to equation (3.194), the result must be zero. ◗

◖ **Example 3.18: A×(B×C)**

$$\left[\mathbf{A}\times(\mathbf{B}\times\mathbf{C})\right]_k = \varepsilon_{klm}A_l(\mathbf{B}\times\mathbf{C})_m = \varepsilon_{klm}A_l\varepsilon_{mop}B_oC_p = \varepsilon_{mkl}\varepsilon_{mop}A_lB_oC_p .$$

In the last expression, we moved the index m in the first Levi–Civita tensor forward twice (changing the sign each time) so that we could apply Theorem (3.211):

$$\left[\mathbf{A}\times(\mathbf{B}\times\mathbf{C})\right]_k = \left(\delta_{ko}\delta_{lp} - \delta_{kp}\delta_{lo}\right)A_lB_oC_p = A_lB_kC_l - A_lB_lC_k = B_k(\mathbf{A}\cdot\mathbf{C}) - C_k(\mathbf{A}\cdot\mathbf{B}) .$$

When applying the Kronecker symbol, we must take care to ensure that the index k, which appears on the lhs of the equality (the so-called free index), always remains in the expressions. In vector terms, we can now write the well-known “bac cab” rule:

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B}) .$$ ◗

◖ **Example 3.19:** div (**A**×**B**)

$$\begin{aligned}\operatorname{div}(\mathbf{A}\times\mathbf{B}) &= \nabla\cdot(\mathbf{A}\times\mathbf{B}) = \partial_k(\mathbf{A}\times\mathbf{B})_k = \\ &= \partial_k\varepsilon_{klm}(A_l B_m) = \varepsilon_{klm}(\partial_k A_l)B_m + \varepsilon_{klm}A_l(\partial_k B_m) = \\ &= B_m\varepsilon_{mkl}(\partial_k A_l) - A_l\varepsilon_{lkm}(\partial_k B_m) = \\ &= B_m(\operatorname{rot}\mathbf{A})_m - A_l(\operatorname{rot}\mathbf{B})_l = \mathbf{B}\cdot\operatorname{rot}\mathbf{A} - \mathbf{A}\cdot\operatorname{rot}\mathbf{B}.\end{aligned}$$

In the expression, we differentiated the product and then rearranged the index m so that we could write the scalar product of a vector with a rotation. The derived expression is useful, e.g., in deriving the law of conservation of energy from Maxwell equations.

◗

◖ **Example 3.20:** rot rot **A**

$$\begin{aligned}[\operatorname{rot}\operatorname{rot}\mathbf{A}]_k &= [\nabla\times(\nabla\times\mathbf{A})]_k = \varepsilon_{klm}\partial_l(\nabla\times\mathbf{A})_m = \\ &= \varepsilon_{klm}\partial_l\varepsilon_{mop}\partial_o A_p = \varepsilon_{mkl}\varepsilon_{mop}\partial_l\partial_o A_p = \\ &= (\delta_{ko}\delta_{lp} - \delta_{kp}\delta_{lo})\partial_l\partial_o A_p = \partial_l\partial_k A_l - \partial_l\partial_l A_k = \partial_k\partial_l A_l - \partial_l\partial_l A_k = \\ &= \partial_k \operatorname{div}\mathbf{A} - \Delta A_k \qquad \Rightarrow \\ &\operatorname{rot}\operatorname{rot}\mathbf{A} = \operatorname{grad}\operatorname{div}\mathbf{A} - \Delta\mathbf{A}.\end{aligned}$$

◗

Similarly, we can derive other useful identities; the derivations are as alike as two peas in a pod. In this epsilon-delta gymnastics, we must follow certain basic rules, and the adjustments then follow automatically: we must never rename a free index on one side of the equality; conversely, we can rename bound (summation) indices however we like. The Kronecker symbol means that the indices it contains are equal (otherwise it is zero). Let’s list the derived identities in a simple table:

► $$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B}), \tag{3.214}$$

► $$(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D}) = (\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D}) - (\mathbf{A}\cdot\mathbf{D})(\mathbf{B}\cdot\mathbf{C}), \tag{3.215}$$

► $$\operatorname{div}\operatorname{rot}\mathbf{A} = 0, \tag{3.216}$$

► $$\operatorname{rot}\operatorname{grad} f = 0, \tag{3.217}$$

► $$\operatorname{rot}\operatorname{rot}\mathbf{A} = \operatorname{grad}\operatorname{div}\mathbf{A} - \Delta\mathbf{A}, \tag{3.218}$$

► $$\operatorname{rot}(\mathbf{A}\times\mathbf{B}) = (\mathbf{B}\cdot\nabla)\mathbf{A} - (\mathbf{A}\cdot\nabla)\mathbf{B} + \mathbf{A}\cdot\operatorname{div}\mathbf{B} - \mathbf{B}\cdot\operatorname{div}\mathbf{A}, \tag{3.219}$$

► $$\operatorname{div}(\mathbf{A}\times\mathbf{B}) = \mathbf{B}\cdot\operatorname{rot}\mathbf{A} - \mathbf{A}\cdot\operatorname{rot}\mathbf{B}, \tag{3.220}$$

► $$\operatorname{div}\operatorname{grad} f = \Delta f, \tag{3.221}$$

► $$\mathbf{A}\times\operatorname{rot}\mathbf{A} = \nabla\left(A^2/2\right) - (\mathbf{A}\cdot\nabla)\mathbf{A}, \tag{3.222}$$

► $$\partial r/\partial x_k = x_k/r, \tag{3.223}$$

► $$\varepsilon_{klm}\varepsilon_{kop} = \delta_{lo}\delta_{mp} - \delta_{lp}\delta_{mo}. \tag{3.224}$$

3.3.5 Lie Algebra

A linear vector space equipped with the operations of vector addition and scalar multiplication is called a Lie algebra if one additional operation [,] is defined on it:

Labeling	From → to	Name	Notation
+	$\mathcal{V}\times\mathcal{V}\to\mathcal{V}$	Vector addition	$\mathbf{h}=\mathbf{f}+\mathbf{g}$
·	$\mathcal{V}\times\mathcal{R}(\mathcal{C})\to\mathcal{V}$	Vector scalar multiplication	$\mathbf{g}=\alpha\cdot\mathbf{f}$
[,]	$\mathcal{V}\times\mathcal{V}\to\mathcal{V}$	Lie operation	$\mathbf{h}=[\mathbf{f},\mathbf{g}]$

The properties of the Lie operation are:

1) $[\mathbf{f},\mathbf{g}]=-[\mathbf{g},\mathbf{f}]$ antisymmetry (3.225)

2) $[\mathbf{f}+\mathbf{g},\mathbf{h}]=[\mathbf{f},\mathbf{h}]+[\mathbf{g},\mathbf{h}]$ linearity (3.226)

3) $[\alpha\mathbf{f},\mathbf{g}]=\alpha[\mathbf{f},\mathbf{g}]$ linearity (3.227)

4) $\big[\mathbf{f},[\mathbf{g},\mathbf{h}]\big]+\big[\mathbf{g},[\mathbf{h},\mathbf{f}]\big]+\big[\mathbf{h},[\mathbf{f},\mathbf{g}]\big]=0$ Bianci identity (3.228)

This is another mapping in which we assign a vector to a pair of vectors. In the case of the vector product, however, this "vector" has somewhat different transformation properties than ordinary vectors. Linearity in the first argument and antisymmetry immediately imply linearity in the second argument. The last relation arises from a cyclic permutation of the first term. Three typical examples of Lie algebras are:

◖ **Example 3.21: Vector product of ordered triples**

$$\mathbf{f}\equiv(f_1,f_2,f_3)\,,\quad \mathbf{g}=(g_1,g_2,g_3)\,;\qquad f_i,\,g_i\in C(R)\qquad \alpha\in C(R)$$

$$\begin{aligned} + &: \quad \mathbf{f}+\mathbf{g}\equiv(f_1+g_1,f_2+g_2,f_3+g_3)\,,\\ \cdot &: \quad \alpha\cdot\mathbf{f}\equiv(\alpha f_1,\alpha f_2,\alpha f_3)\,,\\ [\,,\,] &: \quad [\mathbf{f}\,,\mathbf{g}]\equiv\mathbf{f}\times\mathbf{g}\,. \end{aligned}$$

The vector product serves as a Lie algebra on ordered triples. Verify that the vector product satisfies all the properties of a Lie algebra (3.225) through (3.228). Here, we will simply show that the last relation holds:

$$\big[\mathbf{f},[\mathbf{g},\mathbf{h}]\big]+\big[\mathbf{g},[\mathbf{h},\mathbf{f}]\big]+\big[\mathbf{h},[\mathbf{f},\mathbf{g}]\big]=$$

$$=\mathbf{f}\times(\mathbf{g}\times\mathbf{h})+\mathbf{g}\times(\mathbf{h}\times\mathbf{f})+\mathbf{h}\times(\mathbf{f}\times\mathbf{g})=$$

$$=\mathbf{g}(\mathbf{f}\cdot\mathbf{h})-\mathbf{h}(\mathbf{f}\cdot\mathbf{g})+\mathbf{h}(\mathbf{g}\cdot\mathbf{f})-\mathbf{f}(\mathbf{g}\cdot\mathbf{h})+\mathbf{f}(\mathbf{h}\cdot\mathbf{g})-\mathbf{g}(\mathbf{h}\cdot\mathbf{f})=0\,.$$

This example shows that a Lie algebra is, in a certain sense, a generalization of the vector product. Other structures also possess properties similar to the vector product ones. ◗

◖ **Example 3.22: Square matrix commutator**

For certainty, we will consider 2×2 matrices and define the addition, scalar multiplication, and Lie operation according to the following rules

$$+ \quad : \quad \mathbb{A}+\mathbb{B} \equiv \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{pmatrix} ,$$

$$\cdot \quad : \quad \alpha \cdot \mathbb{A} \equiv \begin{pmatrix} \alpha a_{11} & \alpha a_{12} \\ \alpha a_{21} & \alpha a_{22} \end{pmatrix} ,$$

$$[\,,] \quad : \quad [\mathbb{A}, \mathbb{B}] \equiv \mathbb{A}\mathbb{B} - \mathbb{B}\mathbb{A} .$$

A Lie algebra is defined using matrix multiplication as the so-called *commutator*. If $\mathbb{A}\mathbb{B} = \mathbb{B}\mathbb{A}$, the matrices commute and the commutator $[\mathbb{A}, \mathbb{B}]$ is equal to zero. Verify that the commutator satisfies all the properties of a Lie algebra (3.225) through (3.228). Once again, we will show that the last of these properties holds:

$$[\mathbb{A},[\mathbb{B},\mathbb{C}]]+[\mathbb{B},[\mathbb{C},\mathbb{A}]]+[\mathbb{C},[\mathbb{A},\mathbb{B}]] =$$
$$= \mathbb{A}(\mathbb{B}\mathbb{C}-\mathbb{C}\mathbb{B})-(\mathbb{B}\mathbb{C}-\mathbb{C}\mathbb{B})\mathbb{A}+\cdots = \mathbb{A}\mathbb{B}\mathbb{C}-\mathbb{A}\mathbb{C}\mathbb{B}-\mathbb{B}\mathbb{C}\mathbb{A}\pm\cdots = 0 .$$

◗

◖ **Example 3.23: Poisson brackets**

The following operations can be defined for real functions of real variables $f(q,p)$

$$+ \quad : \quad f+g \equiv f(q,p)+g(q,p) \quad ,$$

$$\cdot \quad : \quad \alpha \cdot f \equiv \alpha f(q,p) \quad ,$$

$$[\,,] \quad : \quad [f, g] \equiv \frac{\partial f}{\partial q}\frac{\partial g}{\partial p} - \frac{\partial f}{\partial p}\frac{\partial g}{\partial q} .$$

Once again, we can easily show that Poisson brackets satisfy the properties of a Lie algebra (3.225) through (3.228).

* * *

Once again, we have found that what matters is not the choice of objects, but the properties of the operations we perform on them. This is very common in both physics and mathematics. The vector product, the matrix commutator, and Poisson brackets are very different operations, yet they share common properties and exhibit analogous behavior. If we equip a linear vector space with the scalar product, we obtain a *unitary space*, and if it is complete – meaning every convergent sequence converges to some element of the space – we obtain a *Hilbert space*. If we equip a linear vector space with the Lie operation, we obtain a *Lie algebra*. Both structures have excellent applications in physics and mathematics. Today, it is hard to imagine quantum theory without Hilbert spaces. And the Lie algebra of the vector product is the basis for describing all rotational motions; matrix or operator commutators have wide applications in quantum theory, and Poisson brackets allow us to find the time evolution of variables in classical mechanics.

* * *

Structure coefficients of a Lie algebra

In a linear vector space, we can express elements as linear combinations of other elements. Ideally, we choose a so-called basis – the maximal set of independent elements. There must be the right number of basis elements. If there are too few, it is not the maximum number of independent vectors; if there are too many, the elements are necessarily dependent (one of them can be written as a combination of the others). An ideal basis is orthonormal, i.e., the elements are mutually orthogonal and have unit magnitude. If we expand the elements of the space into the corresponding basis, we can write:

$$[\mathbf{f},\mathbf{g}] = [f_k\,\mathbf{e}_k\,, g_l\,\mathbf{e}_l\,] = f_k g_l\,[\mathbf{e}_k,\mathbf{e}_l\,]. \tag{3.229}$$

To determine the Lie operation, it suffices to know the result of the operation only for the basis elements. It is clear that the result of the operation $[\mathbf{e}_k, \mathbf{e}_l\,]$ is an element of the space, and we can therefore expand it again into the basis $\{\mathbf{e}_m\}$. However, the expansion coefficients (coordinates) c^m will depend on which two basis elements we are performing the Lie operation on:

$$[\mathbf{e}_k,\mathbf{e}_l] = c_{kl}^m\,\mathbf{e}_m\,. \tag{3.230}$$

The quantities c_{kl}^m are called the *structure coefficients* of the Lie algebra. The result of the Lie operation can now be written as

$$[\mathbf{f},\mathbf{g}] = c_{kl}^m\,x_k y_l\,\mathbf{e}_m\,. \tag{3.231}$$

The Lie algebra is determined by the values of these coefficients. The antisymmetry of the Lie operation (3.225) implies the antisymmetry of the structure coefficients

$$c_{kl}^m = -c_{lk}^m\,. \tag{3.232}$$

◖ **Example 3.24: Ordered triples**

For ordered triples, a basis can be selected

$$\mathbf{e}_1 = (1,0,0)\,; \qquad \mathbf{e}_2 = (0,1,0)\,; \qquad \mathbf{e}_3 = (0,0,1)\,.$$

Now we will determine the structural coefficients from the vector product (Lie operation), which we must perform for all combinations of the basis elements:

$$[\mathbf{e}_1,\mathbf{e}_2] = \mathbf{e}_1\times\mathbf{e}_2 = \mathbf{e}_3\,, \qquad [\mathbf{e}_2,\mathbf{e}_3] = \mathbf{e}_2\times\mathbf{e}_3 = \mathbf{e}_1\,, \qquad [\mathbf{e}_3,\mathbf{e}_1] = \mathbf{e}_3\times\mathbf{e}_1 = \mathbf{e}_2\,.$$

The other combinations are zero. The nonzero structural coefficients are therefore

$$c_{12}^3 = c_{23}^1 = c_{31}^2 = 1 \quad ; \quad c_{21}^3 = c_{32}^1 = c_{13}^2 = -1.$$ ◗

◖ **Example 3.25: Square matrices**

In 2×2 complex matrices, a basis can be chosen

$$\boldsymbol{\sigma}_0 = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}; \quad \boldsymbol{\sigma}_1 = \frac{1}{2}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}; \quad \boldsymbol{\sigma}_2 = \frac{1}{2}\begin{pmatrix}0 & -\mathrm{i}\\ \mathrm{i} & 0\end{pmatrix}; \quad \boldsymbol{\sigma}_3 = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

The matrix $\boldsymbol{\sigma}_0$ is the identity matrix (up to the normalization constant); $\boldsymbol{\sigma}_k$ $k = 1,2,3$ are Pauli matrices, which in quantum theory act as spin operators. We can easily calculate

$$[\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2] = \boldsymbol{\sigma}_1\boldsymbol{\sigma}_2 - \boldsymbol{\sigma}_2\boldsymbol{\sigma}_1 = \mathrm{i}\boldsymbol{\sigma}_3\,,$$
$$[\boldsymbol{\sigma}_2,\boldsymbol{\sigma}_3] = \boldsymbol{\sigma}_2\boldsymbol{\sigma}_3 - \boldsymbol{\sigma}_3\boldsymbol{\sigma}_2 = \mathrm{i}\boldsymbol{\sigma}_1\,,$$
$$[\boldsymbol{\sigma}_3,\boldsymbol{\sigma}_1] = \boldsymbol{\sigma}_3\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_1\boldsymbol{\sigma}_3 = \mathrm{i}\boldsymbol{\sigma}_2\,,$$
$$[\boldsymbol{\sigma}_0,\boldsymbol{\sigma}_1] = [\boldsymbol{\sigma}_0,\boldsymbol{\sigma}_2] = [\boldsymbol{\sigma}_0,\boldsymbol{\sigma}_3] = 0\,.$$

The identity matrix commutes with every matrix. The nonzero structural coefficients are

$$c_{12}^{3} = c_{23}^{1} = c_{31}^{2} = \mathrm{i}\,; \qquad c_{21}^{3} = c_{32}^{1} = c_{13}^{2} = -\mathrm{i}\,.$$ ◗

Other properties of some Lie algebras

For matrices (and other objects for which a multiplication operation is defined), there is another important relation that holds for the Lie algebra of commutators:

$$[\mathbb{A}\mathbb{B},\mathbb{C}] = \mathbb{A}[\mathbb{B},\mathbb{C}] + [\mathbb{A},\mathbb{C}]\mathbb{B}\,, \tag{3.233}$$

$$[\mathbb{A},\mathbb{B}\mathbb{C}] = \mathbb{B}[\mathbb{A},\mathbb{C}] + [\mathbb{A},\mathbb{B}]\mathbb{C}\,. \tag{3.234}$$

Proof:

$$\mathbb{A}[\mathbb{B},\mathbb{C}] + [\mathbb{A},\mathbb{C}]\mathbb{B} = \mathbb{A}(\mathbb{B}\mathbb{C}-\mathbb{C}\mathbb{B}) + (\mathbb{A}\mathbb{C}-\mathbb{C}\mathbb{A})\mathbb{B} =$$
$$= \mathbb{A}\mathbb{B}\mathbb{C} - \mathbb{A}\mathbb{C}\mathbb{B} + \mathbb{A}\mathbb{C}\mathbb{B} - \mathbb{C}\mathbb{A}\mathbb{B} = \mathbb{A}\mathbb{B}\mathbb{C} - \mathbb{C}\mathbb{A}\mathbb{B} = [\mathbb{A}\mathbb{B},\mathbb{C}]\,.$$

We can prove the second relation in a similar way. Using these relations, we can define the Lie operation for matrix powers as well, for example:

$$[\mathbb{A}^2,\mathbb{B}] = [\mathbb{A}\mathbb{A},\mathbb{B}] = \mathbb{A}[\mathbb{A},\mathbb{B}] + [\mathbb{A},\mathbb{B}]\mathbb{A}\,.$$

Similarly, using the basic operation $[\mathbb{A},\mathbb{B}]$ and the relations (3.233) and (3.234), one can determine step by step the result of the commutation relation for arbitrary matrix powers $[\mathbb{A}^k,\mathbb{B}^l]$.

3.3.6 Tensors and Metrics

Suppose we have a linear vector space with a basis $\{\mathbf{e}_k\}$. We can expand the vector $\mathbf{A}$ in this basis into the expression

$$\mathbf{A} = \sum_{k=1}^{N} A^k \mathbf{e}_k = A^k \mathbf{e}_k \tag{3.235}$$

We refer to the numbers A^k as the components (coordinates, expansion coefficients) of a vector, and the objects $\mathbf{e}_k$ as basis elements. Different index positions indicate that vector components transform differently than basis elements. We will continue to use the *summation convention*, but summation will always be performed over one lower index (which transforms as basis elements) and one upper index (which transforms as vector components). A pair of identical upper and lower indices is automatically summed; these are called dummy indices. The positions of the free indices (over which no summation is performed) must always remain the same on both sides of the equality. Let's move from one basis to another, a tilded basis:

$$\{\mathbf{e}_k\} \;\rightarrow\; \{\tilde{\mathbf{e}}_k\}\,. \tag{3.236}$$

A vector **A** is an object whose representation cannot depend on the choice of basis, i.e.

$$\mathbf{A} = \tilde{A}^k \tilde{\mathbf{e}}_k = A^k \mathbf{e}_k \,. \tag{3.237}$$

The components of vectors will be transformed between two bases using a matrix **S**:

$$\tilde{A}^k = S^k{}_l A^l \,. \tag{3.238}$$

Note that the summation is taken over the dummy index l (one is at the top and the other at the bottom). The free index k is at the top on both sides of the equality. Even with matrices, we must distinguish between upper and lower indices. Let us denote the transformation matrix of the basis elements by **U**:

$$\tilde{\mathbf{e}}_k = U^l{}_k \mathbf{e}_l \,. \tag{3.239}$$

Verify that this is the only case in which the summation is performed across one upper and one lower index, the free index k has the same position on both sides, and the transformation matrix **U**, like the matrix **S**, has its first index at the top and its second at the bottom. Let us now determine the relationship between the two transformation matrices **S** and **U**. We start from the expression of vector **A** in the new basis (3.239):

$$\mathbf{A} = \tilde{A}^k \tilde{\mathbf{e}}_k = S^k{}_l A^l \, U^n{}_k \mathbf{e}_n = U^n{}_k S^k{}_l A^l \mathbf{e}_n \,.$$

It is clear that in the new basis, the result must be $A^l \mathbf{e}_l$ or $A^n \mathbf{e}_n$, as you wish. However, this can be achieved in only one way: the following must hold in the last expression

$$U^n{}_k S^k{}_l = \delta^n{}_l \,, \tag{3.240}$$

where we denoted $\delta^n{}_l$ as the Kronecker delta. In matrix notation, this condition states

$$\mathbf{U} \cdot \mathbf{S} = \mathbf{1} \,. \tag{3.241}$$

It is clear that the matrices **U** and **S** are inverses of each other. This is evident directly from the decomposition of vector **A** (3.237) into both bases. For the result to be the same, the components of the vectors (upper indices) must be transformed "inversely" to the elements of the basis (lower indices). Only in this way will the combinations (3.237) yield a result independent of the choice of basis. We will call the upper indices *contravariant*. These indices are transformed in the same way as the components of the vector (via matrix **S**). We will call the lower indices *covariant*. These indices are transformed in the same way as the basis elements (via matrix **U**). There may be more than one index; e.g., from the components of two vectors, we can construct the expression

$$T^{kl} \equiv A^k B^l \,; \qquad \tilde{T}^{kl} \equiv S^k{}_o S^l{}_p T^{op} \,, \tag{3.242}$$

which is transformed as the product of vector components. Using T^{kl}, we can again create an object independent of the coordinate system, known as a second-order tensor:

$$\overleftrightarrow{\mathbf{T}} \equiv T^{kl} \mathbf{e}_k \otimes \mathbf{e}_l \,. \tag{3.243}$$

We sometimes use a double arrow to denote an object with two indices. The symbol $\mathbf{e}_k \otimes \mathbf{e}_l$ is called the tensor (diadic) product; it is an ordered pair of basis elements. We interpret the expression $\mathbf{A} \otimes \mathbf{B}$ as an object whose components form the matrix $A^k B^l$:

► $$\mathbf{A} \otimes \mathbf{B} = A^k B^l \mathbf{e}_k \otimes \mathbf{e}_l \,. \tag{3.244}$$

Scalar product, raising and lowering indices

Suppose that a scalar product $\mathbf{A}\cdot\mathbf{B}$ is defined on our linear vector space, satisfying the basic properties of a scalar product. If we express both vectors in the basis, we obtain

$$\mathbf{A}\cdot\mathbf{B} = A^k B^l\ \mathbf{e}_k \cdot \mathbf{e}_l = g_{kl} A^k B^l , \tag{3.245}$$

where we have denoted

$$g_{kl} \equiv \mathbf{e}_k \cdot \mathbf{e}_l \tag{3.246}$$

the so-called metric coefficients (metric). We see that we can determine the result of the scalar product of any two vectors if we know the metric coefficients, i.e., the results of the scalar products of all the elements of the basis with each other.

Let's denote the inverse matrix of the metric

$$g^{kl} \equiv \left(g_{kl}\right)^{-1} ; \qquad g^{kl} g_{lm} = \delta^k{}_m . \tag{3.247}$$

Let's now introduce auxiliary (dual) objects

$$\mathbf{e}^k \equiv g^{kl}\mathbf{e}_l ; \qquad A_k \equiv g_{kl} A^l . \tag{3.248}$$

These are not actual basis elements or vector components, but formal linear combinations defined by the metric. It always holds that an upper index denotes a transformation of the vector components, and a lower index denotes a transformation of the basis elements. Using the metric, we can thus freely lower or raise the indices; we only need to follow the rule that we sum over one upper and one lower index (this ensures the invariance of the sum with respect to the basis transformation). Free indices always retain their position. Let's consider an example:

$$g_{lo} T^{klm} = T^k{}_o{}^m .$$

We lowered the middle index using a metric. The scalar product can now be written in several ways:

$$\mathbf{A}\cdot\mathbf{B} = g_{kl} A^k B^l = A^k B_k ,$$

where we lowered the second index using the metric. However, we could also have lowered the first index:

$$\mathbf{A}\cdot\mathbf{B} = g_{kl} A^k B^l = A_l B^l = A_k B^k .$$

Therefore, the following applies

► $$\mathbf{A}\cdot\mathbf{B} = g_{kl} A^k B^l = A^k B_k = A_k B^k . \tag{3.249}$$

The contravariant (upper) component is the actual component of the vector, while the covariant (lower) component contains the metric. We can also interpret the definition of the inverse metric (3.247) as lowering or raising indices:

$$\begin{aligned} g^{kl} g_{lm} &= \delta^k{}_m ; \\ g^{kl} g_{lm} &= g^k{}_m ; \end{aligned} \quad \Rightarrow \quad g^k{}_m = \delta^k{}_m . \tag{3.250}$$

The metric and Kronecker delta are thus a single object. If both indices are at the bottom, these are metric coefficients. If both indices are at the top, this is the inverse matrix of the metric coefficients, and if the indices are mixed, this is the Kronecker delta, i.e., the elements of the identity matrix. The metric is thus nothing more than the identity matrix with appropriately shifted indices. Using tensor notation, we can write

$$\mathbf{1} = \delta^k{}_l \mathbf{e}_k \otimes \mathbf{e}^l = g_{kl}\mathbf{e}^k \otimes \mathbf{e}^l = g^{kl}\mathbf{e}_k \otimes \mathbf{e}_l \,. \tag{3.251}$$

Four-vectors, Minkowski metric

In special relativity, we call any quadruple that transforms under the Lorentz transformation a four-vector. The basic quadruplets include the *event* (time and space coordinates of the event), *four-momentum* (energy and momentum), *wave four-vector* (angular frequency and wave vector), *four-potential of the electromagnetic field* (scalar and vector potential), *four-current* (source terms of Maxwell's equations – charge density and charge flux), or four-gradient. In the SI system, we must ensure that all four components have the same dimension. The simplest way to do this is by multiplying or dividing the time component by the universal constant c (the speed of light in a vacuum):

► $$\begin{aligned} x^\mu &\equiv \begin{pmatrix} ct \\ \mathbf{x} \end{pmatrix}; \qquad p^\mu \equiv \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix}; \qquad k^\mu \equiv \begin{pmatrix} \omega/c \\ \mathbf{k} \end{pmatrix}; \\ A^\mu &\equiv \begin{pmatrix} \phi/c \\ \mathbf{A} \end{pmatrix}; \qquad j^\mu \equiv \begin{pmatrix} \rho c \\ \mathbf{j} \end{pmatrix}; \qquad \partial_\mu \equiv \begin{pmatrix} \partial/\partial ct \\ \partial/\partial \mathbf{x} \end{pmatrix}. \end{aligned} \tag{3.252}$$

Note 1: We will use Greek indices to denote four-vectors (the index 0 corresponds to the time component, and the indices 1, 2, and 3 to the spatial components).

Note 2: In the case of the four-gradient, it is a covariant (lower) index, because

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu},$$

So the actual components of the vectors are in the denominator; if we write the index in the numerator, it must be in the opposite position, because the transformation matrix becomes the inverse!

Note 3: The metric in special relativity is called the Minkowski metric. It is diagonal and has a negative sign in the time component. The same applies to the inverse matrix (the metric with upper indices). The metric with mixed indices is the identity matrix, i.e., its elements are the Kronecker delta:

► $$\begin{aligned} g_{\mu\nu} &= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}; \qquad g^{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}; \\ g^\mu{}_\nu &= \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}; \qquad g_\mu{}^\nu = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}. \end{aligned} \tag{3.253}$$

For simplicity, the Minkowski metric is often written as $g_{\mu\nu}$ = diag(–1, 1, 1, 1); it is sometimes denoted by the symbol $\eta_{\mu\nu}$. Using this metric, we can now easily determine the covariant components of ordinary four-vectors and the contravariant component of the four-gradient:

► $$x_\mu \equiv \begin{pmatrix} -ct \\ \mathbf{x} \end{pmatrix}; \qquad p_\mu \equiv \begin{pmatrix} -E/c \\ \mathbf{p} \end{pmatrix}; \qquad k_\mu \equiv \begin{pmatrix} -\omega/c \\ \mathbf{k} \end{pmatrix};$$
$$A_\mu \equiv \begin{pmatrix} -\phi/c \\ \mathbf{A} \end{pmatrix}; \qquad j_\mu \equiv \begin{pmatrix} -\rho c \\ \mathbf{j} \end{pmatrix}; \qquad \partial^\mu \equiv \begin{pmatrix} -\partial/\partial ct \\ \partial/\partial \mathbf{x} \end{pmatrix}. \tag{3.254}$$

As an example of working with indices, let's look at some typical scalar products, starting with a wave vector and an event:

$$k \cdot x = k^\mu x_\mu = k^0(-x_0) + k^1 x_1 + k^2 x_2 + k^3 x_3 = -\omega t + \mathbf{k}\cdot\mathbf{x}.$$

On the left is a four-vector product; the last term on the right is a standard product in $\mathcal{R}^3$. Similarly, let's determine the results of the other examples

$$\mathrm{d}s^2 \equiv \mathrm{d}x_\mu \mathrm{d}x^\mu = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 ;$$

$$j \cdot A \equiv j_\mu A^\mu = -\rho\phi + \mathbf{j}\cdot\mathbf{A} ;$$

$$\frac{\partial\rho}{\partial t} + \operatorname{div}\mathbf{j} = 0 \qquad \Leftrightarrow \qquad \partial_\mu j^\mu = 0 ;$$

$$\Box f = 0 \qquad \Leftrightarrow \qquad \partial_\mu \partial^\mu f = 0 .$$

A shorthand notation is often used in which the derivative is written after the comma. The indices before the comma are the actual indices, while those after the comma are the derivatives:

$$\frac{\partial A^\mu}{\partial x^\nu} \equiv \partial_\nu A^\mu \equiv A^\mu{}_{,\nu} .$$

This is actually the most concise notation for a derivative, making it immediately clear how the derivative is derived. Let's look at some more examples:

$$\frac{\partial\varphi}{\partial x_\mu} \equiv \partial^\mu \varphi \equiv \varphi_,{}^\mu ;$$

$$\frac{\partial\varphi}{\partial x^\mu} \equiv \partial_\mu \varphi \equiv \varphi_{,\mu} ;$$

$$\frac{\partial^2 T^\alpha{}_\beta}{\partial x^\mu \partial x_\nu} \equiv \partial_\mu \partial^\nu T^\alpha{}_\beta \equiv T^\alpha{}_{\beta,\mu}{}^\nu ;$$

$$\Box f \equiv \partial_\mu \partial^\mu f \equiv f_{,\mu}{}^\mu .$$

3.4 Dirac Notation and Operators in Quantum Theory

In this chapter, we will examine the most important mathematical concepts required in quantum theory. For the simplicity of the presentation, all discussions are conducted for the case where the eigenvalues of the operators are distinct and form a countable set. More general cases involving multiple eigenvalues and a continuous spectrum are discussed only briefly in this text. In Section 3.3.1, we extended the concept of a vector to objects more general than ordered triples and introduced the linear vector space. In Section 3.3.2, we added the scalar product to the operations of vector scalar multiplication and addition and introduced the unitary space as a prototype of spaces with a scalar product. In this chapter, we will consistently use Dirac notation, which was introduced by Paul Adrien Maurice Dirac for quantum theory. Elements of linear vector spaces are denoted by $|\,\mathbf{f}\,\rangle$, $|\,\mathbf{x}\,\rangle$, $|\,\mathbf{a}\,\rangle$, and scalar products $\langle\,\mathbf{f}\,|\,\mathbf{g}\,\rangle$, $\langle\,\mathbf{x}\,|\,\mathbf{y}\,\rangle$, $\langle\,\mathbf{a}\,|\,\mathbf{b}\,\rangle$, etc.

3.4.1 Hilbert Spaces

Let us recall the structures of some spaces with a scalar product that we introduced in Section 3.3.2:

$$\mathcal{R}^3 \qquad \langle \mathbf{f}\,|\,\mathbf{g}\rangle \equiv f_1 g_1 + f_2 g_2 + f_3 g_3; \qquad f_k, g_k \in \mathcal{R},\; k = 1, 2, 3 \tag{3.255}$$

$$\mathcal{C}^N \qquad \langle \mathbf{f}\,|\,\mathbf{g}\rangle = \sum_{k=1}^{N} f_k^* g_k\,; \qquad f_k, g_k \in \mathcal{C},\; k = 1, \ldots N \tag{3.256}$$

$$\ell^2 \qquad \langle \mathbf{f}\,|\,\mathbf{g}\rangle = \sum_{k=1}^{\infty} f_k^* g_k\,; \qquad \{f_k, g_k\}_{k=1}^{\infty} \in \mathcal{C} \tag{3.257}$$

$$\mathcal{L}^2 \qquad \langle \mathbf{f}\,|\,\mathbf{g}\rangle \equiv \int_{-\infty}^{+\infty} f^*(x)\, g(x)\, \mathrm{d}x\,; \qquad f(x),\, g(x) : \mathcal{R} \to \mathcal{C} \tag{3.258}$$

In spaces of ℓ^2, or $\mathcal{L}^2$, we need to include only elements with $\|\,\mathbf{f}\,\| < \infty$; that is, we require that sequences be squared-summable and functions be squared-integrable. Unitary spaces allow us to determine the magnitude of an element, find angles between elements, project an element in a certain direction, expand elements into various bases, etc. However, if we find a sequence of elements that converges in some sense in a unitary space, it may happen that the limit of this sequence is not part of the unitary space. The simplest example is the open interval (0, 1), for whose elements (numbers from this interval) we can define the operations of extension, composition, and scalar product in the space $\mathcal{R}^N$ for $N = 1$. The sequence of elements $|\,f_k\,\rangle = 1/k$ clearly converges to zero as $k \to \infty$, but zero is not part of the unitary space we have defined. If we were to take a closed interval, we would avoid this difficulty. Therefore, Hilbert spaces are generally used instead of unitary spaces.

A Hilbert space is a complete unitary space. Completeness means that any sequence that converges in some sense always converges to an element that is part of that space. Simply put, these are spaces that contain their limit, i.e., the limit is part of them. A *separable Hilbert space* is a Hilbert space with a countable basis.

Note 1: By adding the operation [,] from a linear vector space, we obtain a Lie algebra; by adding the operation $\langle \; | \; \rangle$, we obtain a unitary space.

Note 2: The notation system was developed by P. A. M. Dirac. It is also known as bracket notation or brackets.

$\langle \; | \; \rangle$ "bracket";

$\langle \; |$ "bra" (can be precisely defined, the indicated scalar product operation);

$| \; \rangle$ "ket" (vector from $\mathcal{V}$).

Note 3: For complex N-tuples, $|\,\mathbf{f}\,\rangle$ can be interpreted as a column matrix, and $\langle\,\mathbf{f}\,|$ as a transposed complex conjugate matrix:

$$|\,\mathbf{f}\,\rangle = \begin{pmatrix} f_1 \\ \vdots \\ f_N \end{pmatrix}; \qquad \langle\,\mathbf{f}\,| = \begin{pmatrix} f_1^* & \cdots & f_N^* \end{pmatrix}.$$

Then the scalar product

$$\langle\,\mathbf{f}\,|\,\mathbf{g}\,\rangle = \begin{pmatrix} f_1^* & \cdots & f_N^* \end{pmatrix} \cdot \begin{pmatrix} g_1 \\ \vdots \\ g_N \end{pmatrix} = f_k^* g_k$$

is then defined using matrix multiplication. For spaces other than N-tuples, it is not necessary for our purposes to interpret the individual components of the scalar product a $\langle\,\mathbf{f}\,|\,\mathbf{g}\,\rangle$ in any particular way.

Note 4: For $\mathcal{L}^2$, this can be understood as

$$\langle\,\mathbf{f}\,| = \int_{-\infty}^{+\infty} f^*(x) \;\cdots\; \mathrm{d}x$$

i.e., as an implied scalar product operation. One simply needs to specify the appropriate function on which the operation acts. A similar situation arises in differentiation when we write d/dx.

Note 5: Dirac notation has greatly simplified most of the notation used in quantum theory. Mathematicians were initially skeptical of it, but eventually came to accept it. Today, we cannot imagine describing the processes occurring in the microworld without this notation. In particular, the notation for projection operators, the completeness theorem, and the spectral expansion theorem would be extremely confusing and cumbersome without it.

Linear span

Let $\mathcal{H}$ be a Hilbert space and $|\,\mathbf{f}\,\rangle$ its nonzero element. We will call the set of elements $\{|\,\mathbf{g}\,\rangle;\ |\,\mathbf{g}\,\rangle = \alpha\,|\,\mathbf{f}\,\rangle;\ \alpha \in\ \mathcal{C} = \backslash\{0\},\ |\,\mathbf{f}\,\rangle, |\,\mathbf{g}\,\rangle \in\ \mathcal{H}\}$ a span generated by $|\,\mathbf{f}\,\rangle$.

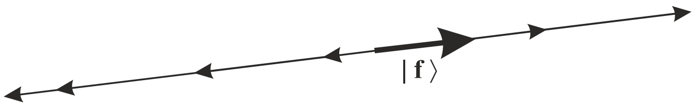


Fig. 3.20: Linear span of vector $|\,\mathbf{f}\,\rangle$

3.4.2 Operators

By an operator, we mean the mapping $\hat{\mathbf{A}}$: $\mathcal{H} \to \mathcal{H}$ that assigns an element $|\,\mathbf{f}\,\rangle$ from the space $\mathcal{V}$ to an element $|\,\mathbf{g}\,\rangle$ of that space:

$$\hat{\mathbf{A}}|\,\mathbf{f}\,\rangle = |\,\mathbf{g}\,\rangle\,.$$

The standard terminology (*preimage, image, domain, range*…) remains in use.

◖ **Example 3.26: Space $\mathcal{R}^3$**

An operator on $\mathcal{R}^3$ can be any 3×3 matrix, for example

$$\hat{\mathbf{A}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}; \qquad |\,\mathbf{f}\,\rangle = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \qquad \Rightarrow$$

$$\hat{\mathbf{A}}|\,\mathbf{f}\,\rangle = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = |\,\mathbf{g}\,\rangle\,, \qquad \text{generally}$$

$$\hat{\mathbf{A}}|\,\mathbf{f}\,\rangle = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_3 \\ f_2 + f_3 \end{pmatrix}.$$

◗

◖ **Example 3.27: Space $\mathcal{L}^2\,(-\infty\,,\infty)$**

We can consider the derivative of a function to be a typical operator in the function space. However, we must always choose functions that are integrable with respect to the square and verify that they retain this property even after differentiation:

$$\hat{\mathbf{D}} \equiv \frac{\mathrm{d}}{\mathrm{d}x}; \qquad |\,\mathbf{f}\,\rangle = x\,\mathrm{e}^{-x} \qquad \Rightarrow$$

$$\hat{\mathbf{D}}|\,\mathbf{f}\,\rangle = \frac{\mathrm{d}}{\mathrm{d}x}\left(x\,\mathrm{e}^{-x}\right) = (1-x)\,\mathrm{e}^{-x} = |\,\mathbf{g}\,\rangle\,.$$ ◗

Unit operator

The unit operator is defined by the rule

► $$\hat{\mathbf{1}}|\,\mathbf{f}\,\rangle \;\equiv\; |\,\mathbf{f}\,\rangle\,. \tag{3.259}$$

For N-tuples, the identity operator is the diagonal matrix with ones on the diagonal (the identity matrix) – check it out!

Square of an operator

We can define the square of an operator if the range of the operator is a subset of its domain; in that case, we can write

$$\hat{\mathbf{A}}^2\,|\,\mathbf{f}\,\rangle \;\equiv\; \hat{\mathbf{A}}\Big(\,\hat{\mathbf{A}}\,|\,\mathbf{f}\,\rangle\Big). \tag{3.260}$$

◖ **Example 3.28: Square of the derivative**

According to the previous definition, the second power of a derivative is the second derivative of the given function:

$$\hat{\mathbf{D}} \;\equiv\; \frac{\mathrm{d}}{\mathrm{d}x}\,; \qquad |\,\mathbf{f}\,\rangle = \mathrm{e}^{-x^2} \qquad \Rightarrow$$

$$\hat{\mathbf{D}}^2\,|\,\mathbf{f}\,\rangle \;\equiv\; \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{-x^2}\right) \;=\; \frac{\mathrm{d}}{\mathrm{d}x}\left(-2x\,\mathrm{e}^{-x^2}\right) \;=\; (-2+4x^2)\,\mathrm{e}^{-x^2}\,.$$ ◗

Power of an operator

We define the power of an operator by induction (assuming knowledge of the square and that the range is a subset of the domain of the operator):

► $$\hat{\mathbf{A}}^n\,|\,\mathbf{f}\,\rangle \;\equiv\; \hat{\mathbf{A}}\Big(\,\hat{\mathbf{A}}^{n-1}\,|\,\mathbf{f}\,\rangle\Big). \tag{3.261}$$

Function of an operator

Let $f(x)$ be an analytic function with the Taylor series

► $$f(x) \;=\; \sum_{k=0}^{\infty} c_k x^k\,. \tag{3.262}$$

Then we can define a function of the operator using a formula

$$f(\hat{\mathbf{A}}) \;=\; \sum_{k=0}^{\infty} c_k \hat{\mathbf{A}}^k\,. \tag{3.263}$$

You can find the expansion of some important functions at the end of Section 3.10.5:

◖ **Example 3.29: Exponential of a matrix**

Problem: A matrix operator is defined on the space C^2 $\hat{\mathbf{A}} = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}$. Find $\exp(\hat{\mathbf{A}})$.

Solution: First, we will determine the individual powers of the given matrix (this is one of the Pauli matrices, which are used in quantum theory as spin operators):

$$\hat{\mathbf{A}}^2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \hat{\mathbf{1}},$$

$$\hat{\mathbf{A}}^3 = \hat{\mathbf{A}} \cdot \hat{\mathbf{A}}^2 = \hat{\mathbf{A}} \cdot \hat{\mathbf{1}} = \hat{\mathbf{A}},$$

$$\hat{\mathbf{A}}^4 = \hat{\mathbf{A}} \cdot \hat{\mathbf{A}}^3 = \hat{\mathbf{A}} \cdot \hat{\mathbf{A}} = \hat{\mathbf{A}}^2 = \hat{\mathbf{1}},$$

$$\vdots$$

$$\hat{\mathbf{A}}^{2n-1} = \hat{\mathbf{A}}, \qquad \hat{\mathbf{A}}^{2n} = \hat{\mathbf{1}}, \qquad n = 1, 2, \ldots.$$

Now we can easily find the desired function of the matrix:

$$\exp(\hat{\mathbf{A}}) = 1 + \hat{\mathbf{A}} + \frac{\hat{\mathbf{A}}^2}{2!} + \frac{\hat{\mathbf{A}}^3}{3!} + \frac{\hat{\mathbf{A}}^4}{4!} + \cdots =$$

$$= \left(1 + \frac{1}{2!} + \frac{1}{4!} + \frac{1}{6!} + \cdots\right) \hat{\mathbf{1}} + \left(1 + \frac{1}{3!} + \frac{1}{5!} + \frac{1}{7!} + \cdots\right) \hat{\mathbf{A}} =$$

$$= \mathrm{ch}(1)\, \hat{\mathbf{1}} + \mathrm{sh}(1)\, \hat{\mathbf{A}} = \begin{pmatrix} \mathrm{ch}\,1 & -\mathrm{i}\,\mathrm{sh}\,1 \\ \mathrm{i}\,\mathrm{sh}\,1 & \mathrm{ch}\,1 \end{pmatrix}.$$

This is how expressions such as sin(d/dx) ones make sense. Later, we will learn how to find the function of an operator using the spectral expansion of the operator; see Equation (3.296). This is a more efficient method than the Taylor series expansion. ◗

Inverse operator

We call an operator $\hat{\mathbf{A}}^{-1}$ the inverse of $\hat{\mathbf{A}}$ if the following holds for it:

► $$\hat{\mathbf{A}} \cdot \hat{\mathbf{A}}^{-1} = \hat{\mathbf{A}}^{-1} \cdot \hat{\mathbf{A}} = \hat{\mathbf{1}}. \tag{3.264}$$

For a given operator $\hat{\mathbf{A}}$, finding the inverse operator can be quite difficult; sometimes the inverse operator does not exist at all.

Adjoint operator

We call an operator $\hat{\mathbf{A}}^{\dagger}$ the adjoint of $\hat{\mathbf{A}}$ if the following holds:

► $$\langle\, \mathbf{f} \,|\, \hat{\mathbf{A}}\mathbf{g} \,\rangle = \langle\, \hat{\mathbf{A}}^{\dagger}\mathbf{f} \,|\, \mathbf{g} \,\rangle. \tag{3.265}$$

Applying the original operator to the right side of the scalar product yields the same result as applying its adjoint operator to the left side. An adjoint operator does not always exist.

◖ **Example 3.30:** Inverse and adjoint operators for a 2×2 matrix

Prove that the inverse and adjoint operators of the given matrix have the forms:

$$\hat{\mathbf{A}} = \begin{pmatrix} 1 & 2\mathrm{i} \\ 0 & \mathrm{i} \end{pmatrix}; \qquad \hat{\mathbf{A}}^{-1} = \begin{pmatrix} 1 & -2 \\ 0 & -\mathrm{i} \end{pmatrix}; \qquad \hat{\mathbf{A}}^{\dagger} = \begin{pmatrix} 1 & 0 \\ -2\mathrm{i} & -\mathrm{i} \end{pmatrix}.$$

Simply substitute into equations (3.264) and (3.265):

$$\hat{\mathbf{A}}\cdot\hat{\mathbf{A}}^{-1} = \begin{pmatrix} 1 & 2\mathrm{i} \\ 0 & \mathrm{i} \end{pmatrix}\cdot\begin{pmatrix} 1 & -2 \\ 0 & -\mathrm{i} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \hat{\mathbf{1}},$$

$$\hat{\mathbf{A}}^{-1}\cdot\hat{\mathbf{A}} = \begin{pmatrix} 1 & -2 \\ 0 & -\mathrm{i} \end{pmatrix}\cdot\begin{pmatrix} 1 & 2\mathrm{i} \\ 0 & \mathrm{i} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \hat{\mathbf{1}};$$

$$\hat{\mathbf{A}}\,|\,\mathbf{g}\,\rangle = \begin{pmatrix} 1 & 2\mathrm{i} \\ 0 & \mathrm{i} \end{pmatrix}\cdot\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} g_1 + 2\mathrm{i}\,g_2 \\ \mathrm{i}\,g_2 \end{pmatrix},$$

$$\hat{\mathbf{A}}^{\dagger}\,|\,\mathbf{f}\,\rangle = \begin{pmatrix} 1 & 0 \\ -2\mathrm{i} & -\mathrm{i} \end{pmatrix}\cdot\begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} f_1 \\ -2\mathrm{i}\,f_1 - \mathrm{i}\,f_2 \end{pmatrix} \quad\Rightarrow$$

$$\langle\,\mathbf{f}\,|\,\hat{\mathbf{A}}\mathbf{g}\,\rangle = \begin{pmatrix} f_1^* & f_2^* \end{pmatrix}\cdot\begin{pmatrix} g_1 + 2\mathrm{i}\,g_2 \\ \mathrm{i}\,g_2 \end{pmatrix} = f_1^* g_1 + 2\mathrm{i}\,f_1^* g_2 + \mathrm{i}\,f_2^* g_2\,,$$

$$\langle\,\hat{\mathbf{A}}^{\dagger}\,\mathbf{f}\,|\,\mathbf{g}\,\rangle = \begin{pmatrix} f_1^* & 2\mathrm{i}\,f_1^* + \mathrm{i}\,f_2^* \end{pmatrix}\cdot\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = f_1^* g_1 + 2\mathrm{i}\,f_1^* g_2 + \mathrm{i}\,f_2^* g_2\,.$$

◗

Note: Finding the adjoint of a matrix is very easy; simply take the complex conjugate of the original matrix and then transpose it (flip it across the diagonal), i.e.

$$\hat{\mathbf{A}}^{\dagger} = (\hat{\mathbf{A}}^{*})^{T}\,.$$

Let us now state some very useful relationships for calculating the inverse and adjoint operators for the product of two operators:

$$(\hat{\mathbf{A}}\hat{\mathbf{B}})^{-1} = \hat{\mathbf{B}}^{-1}\,\hat{\mathbf{A}}^{-1}\,, \tag{3.266}$$

$$(\hat{\mathbf{A}}\hat{\mathbf{B}})^{\dagger} = \hat{\mathbf{B}}^{\dagger}\hat{\mathbf{A}}^{\dagger}\,. \tag{3.267}$$

Their proof follows directly from the definitions of the inverse and adjoint operators:

$$(\hat{\mathbf{A}}\hat{\mathbf{B}})^{-1}\cdot\hat{\mathbf{A}}\hat{\mathbf{B}} = \hat{\mathbf{1}} \qquad /\quad \hat{\mathbf{B}}^{-1}\quad \text{from right}\,,$$

$$(\hat{\mathbf{A}}\hat{\mathbf{B}})^{-1}\cdot\hat{\mathbf{A}} = \hat{\mathbf{B}}^{-1} \qquad /\quad \hat{\mathbf{A}}^{-1}\quad \text{from right}\,,$$

$$(\hat{\mathbf{A}}\hat{\mathbf{B}})^{-1} = \hat{\mathbf{B}}^{-1}\,\hat{\mathbf{A}}^{-1}\,.$$

We proceed analogously for the adjoint operator:

$$\langle\,(\hat{\mathbf{A}}\hat{\mathbf{B}})^{\dagger}\mathbf{f}\,|\,\mathbf{g}\,\rangle = \langle\,\mathbf{f}\,|\,\hat{\mathbf{A}}\hat{\mathbf{B}}\mathbf{g}\,\rangle = \langle\,\hat{\mathbf{A}}^{\dagger}\mathbf{f}\,|\hat{\mathbf{B}}\mathbf{g}\,\rangle = \langle\,\hat{\mathbf{B}}^{\dagger}\hat{\mathbf{A}}^{\dagger}\,\mathbf{f}\,|\,\mathbf{g}\,\rangle\,.$$

Commutativity of operators

For operators, $\hat{\mathbf{A}}\hat{\mathbf{B}} \neq \hat{\mathbf{B}}\hat{\mathbf{A}}$ holds in general. We say that the operators do not commute. We can assess the degree of non-commutativity using what is known as a *commutator*

► $$[\hat{\mathbf{A}},\hat{\mathbf{B}}] \equiv \hat{\mathbf{A}}\hat{\mathbf{B}} - \hat{\mathbf{B}}\,\hat{\mathbf{A}}\,. \tag{3.268}$$

If the commutator of operators $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ is zero, the operators commute; if it is nonzero, they do not commute. The result of the commutator is again an operator. Let us list the most important properties of commutators (try to prove them)

$$1)\quad [\hat{\mathbf{A}},\hat{\mathbf{B}}] = -[\hat{\mathbf{B}},\hat{\mathbf{A}}],$$
$$2)\quad [\hat{\mathbf{A}}+\hat{\mathbf{B}},\hat{\mathbf{C}}] = [\hat{\mathbf{A}},\hat{\mathbf{C}}]+[\hat{\mathbf{B}},\hat{\mathbf{C}}],$$
$$3)\quad [\alpha\cdot\hat{\mathbf{A}},\hat{\mathbf{B}}] = \alpha\cdot[\hat{\mathbf{A}},\hat{\mathbf{B}}],$$
$$4)\quad [\hat{\mathbf{A}},[\hat{\mathbf{B}},\hat{\mathbf{C}}]]+[\hat{\mathbf{B}},[\hat{\mathbf{C}},\hat{\mathbf{A}}]]+[\hat{\mathbf{C}},[\hat{\mathbf{A}},\hat{\mathbf{B}}]] = 0.$$

But these are precisely the defining properties of the Lie algebra (3.225) up to (3.228). The commutators form a Lie algebra on the space of operators.

◖ **Example 3.31: Commutation relation [d/dx, x]**

Consider two operators on $\mathcal{L}^2$: $\hat{\mathbf{D}} = \mathrm{d}/\mathrm{d}x$ and $\hat{\mathbf{X}} = x$; they act on the element $|x^5\rangle$ like this

$$\hat{\mathbf{D}}|x^5\rangle = \frac{\mathrm{d}}{\mathrm{d}x}x^5 = 5x^4, \qquad \hat{\mathbf{X}}|x^5\rangle = x\cdot x^5 = x^6.$$

Let's determine their commutator

$$[\hat{\mathbf{D}},\hat{\mathbf{X}}]\,|\,\mathbf{f}\,\rangle = (\hat{\mathbf{D}}\hat{\mathbf{X}} - \hat{\mathbf{X}}\hat{\mathbf{D}})\,|\,\mathbf{f}\,\rangle = \left(\frac{\mathrm{d}}{\mathrm{d}x}x - x\frac{\mathrm{d}}{\mathrm{d}x}\right)f(x) = \frac{\mathrm{d}}{\mathrm{d}x}\big(x\,f(x)\big) - x\frac{\mathrm{d}}{\mathrm{d}x}f(x) =$$
$$= f(x) + x\,f'(x) - x\,f'(x) = f(x) = |\,\mathbf{f}\,\rangle \quad \Rightarrow$$
$$[\hat{\mathbf{D}},\hat{\mathbf{X}}]\,|\,\mathbf{f}\,\rangle = |\,\mathbf{f}\,\rangle = \quad \text{pro}\ \forall\ |\,\mathbf{f}>\in\mathcal{L}^2 \quad \Rightarrow$$
$$[\hat{\mathbf{D}},\hat{\mathbf{X}}] = \hat{\mathbf{1}}.$$

In the same way, we can determine other commutative relations.

◗

We will focus only on *linear operators*, i.e., operators with a linear response:

$$\hat{\mathbf{A}}(\alpha|\,\mathbf{f}\,\rangle + \beta|\,\mathbf{g}\,\rangle) = \alpha\hat{\mathbf{A}}|\,\mathbf{f}\,\rangle + \beta\hat{\mathbf{A}}|\,\mathbf{g}\,\rangle.$$

All of the operators discussed so far have been linear. In quantum theory, we encounter primarily two types of linear operators: *unitary* operators and *Hermitian* operators. Let us now define these operators.

Unitary operators

Definition: A unitary operator preserves the scalar product, i.e.,

$$\blacktriangleright \qquad \langle\,\mathbf{f}\,|\,\mathbf{g}\,\rangle = \langle\,\hat{\mathbf{U}}\mathbf{f}\,|\,\hat{\mathbf{U}}\mathbf{g}\,\rangle. \qquad (3.269)$$

The scalar product remains unchanged before and after the application of a unitary operator.

Theorem: For unitary operators, the adjoint and inverse operators are identical, i.e.,

$$\blacktriangleright \qquad \hat{\mathbf{U}}^\dagger = \hat{\mathbf{U}}^{-1}. \qquad (3.270)$$

Proof: From the definition of the adjoint operator, we know that

$$\langle \hat{\mathbf{U}}\mathbf{f} \,|\, \hat{\mathbf{U}}\mathbf{g} \rangle = \langle \hat{\mathbf{U}}^\dagger \hat{\mathbf{U}}\mathbf{f} \,|\, \mathbf{g} \rangle .$$

To preserve the scalar product (the definition of a unitary operator), it is necessary that

$$\hat{\mathbf{U}}^\dagger \hat{\mathbf{U}} = \hat{\mathbf{1}} ,$$

but by the definition of the inverse operator, this precisely means that

$$\hat{\mathbf{U}}^\dagger = \hat{\mathbf{U}}^{-1} .$$

◖ **Example 3.32: Operator $\hat{\mathbf{U}} \equiv \mathrm{e}^{\mathrm{i}x}$**

We will show that the operator $\hat{\mathbf{U}} \equiv \mathrm{e}^{\mathrm{i}x}$ is unitary on the space $\mathcal{L}^2$:

$$|\,\mathbf{f}\,\rangle = f(x) , \qquad \hat{\mathbf{U}}|\,\mathbf{f}\,\rangle = \mathrm{e}^{\mathrm{i}x} f(x) ,$$

$$|\,\mathbf{g}\,\rangle = g(x) , \qquad \hat{\mathbf{U}}|\,\mathbf{g}\,\rangle = \mathrm{e}^{\mathrm{i}x} g(x) ,$$

$$\langle \hat{\mathbf{U}}\mathbf{f} \,|\, \hat{\mathbf{U}}\mathbf{g} \rangle = \int \left(\mathrm{e}^{\mathrm{i}x} f(x)\right)^* \mathrm{e}^{\mathrm{i}x} g(x)\,\mathrm{d}x =$$

$$= \int \mathrm{e}^{-\mathrm{i}x} f^*(x)\,\mathrm{e}^{\mathrm{i}x} g(x)\,\mathrm{d}x =$$

$$= \int f^*(x)\,g(x)\,\mathrm{d}x = \langle\, \mathbf{f} \,|\, \mathbf{g} \,\rangle .$$

◗

Hermitian operators

Definition: A Hermitian operator acts identically on both parts of a scalar product, i.e.,

$$\langle \hat{\mathbf{A}}\mathbf{f} \,|\, \mathbf{g} \rangle = \langle \mathbf{f} \,|\, \hat{\mathbf{A}}\mathbf{g} \rangle . \tag{3.271}$$

Theorem: For a Hermitian operator, the adjoint operator is equal to the original operator; it is *self-adjoint*:

$$\hat{\mathbf{A}}^\dagger = \hat{\mathbf{A}} . \tag{3.272}$$

Proof: This follows immediately from the definition of the adjoint operator.

Note: In pure mathematics, the definitions of self-adjoint and Hermitian operators differ slightly in terms of the domain requirements; for our purposes, we will not distinguish between self-adjoint and Hermitian operators. Since a Hermitian operator acts identically on both parts of the scalar product, it is often written as

$$\langle \hat{\mathbf{A}}\mathbf{f} \,|\, \mathbf{g} \rangle = \langle \mathbf{f} \,|\, \hat{\mathbf{A}}\mathbf{g} \rangle = \langle \mathbf{f} \,|\, \hat{\mathbf{A}} \,|\, \mathbf{g} \rangle .$$

The central position of $\hat{\mathbf{A}}$ implies that we can apply the operator to the left or right at our discretion. This structure is sometimes called the *Dirac sandwich*.

◖ **Example 3.33: Operator $\hat{\mathsf{B}} \equiv \mathrm{i}\,\mathrm{d}/\mathrm{d}x$**
We will show that the operator $\hat{\mathsf{B}} = \mathrm{i}\,\mathrm{d}/\mathrm{d}x$ on the space $\mathcal{L}^2(-\infty, \infty)$ is Hermitian:

$$\langle \hat{\mathsf{B}}\mathbf{f} \,|\, \mathbf{g} \rangle = \int_{-\infty}^{\infty} \left(\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x} f(x) \right)^* g(x)\,\mathrm{d}x =$$

$$= -\mathrm{i} \int_{-\infty}^{\infty} \frac{\mathrm{d}f^*(x)}{\mathrm{d}x} g(x)\,\mathrm{d}x \overset{\text{p.partes}}{=}$$

$$= -\mathrm{i}\left[f^*(x)\,g(x) \right]_{-\infty}^{\infty} + \mathrm{i} \int_{-\infty}^{\infty} f^*(x) \frac{\mathrm{d}g(x)}{\mathrm{d}x}\,\mathrm{d}x =$$

$$\int_{-\infty}^{\infty} f^*(x) \left(\mathrm{i}\frac{\mathrm{d}g(x)}{\mathrm{d}x} \right) \mathrm{d}x = \langle \mathbf{f} \,|\, \hat{\mathsf{B}}\mathbf{g} \rangle .$$

The expression in square brackets is zero, since functions from $\mathcal{L}^2(-\infty, \infty)$ are integrable with respect to the square on $(-\infty, \infty)$, and therefore the following must hold

$$\lim_{x\to\pm\infty} f(x) = \lim_{x\to\pm\infty} g(x) = 0 \quad \text{pro } \forall\, f, g \in \mathcal{L}^2 .$$

The derivative operator $\hat{\mathsf{D}} = \mathrm{d}/\mathrm{d}x$ itself is not Hermitian; if we were to perform integration by parts, the sign would change and the following would hold:

$$\langle \hat{\mathsf{D}}\mathbf{f} \,|\, \mathbf{g} \rangle = -\langle \mathbf{f} \,|\, \hat{\mathsf{D}}\mathbf{g} \rangle .$$ ◗

3.4.3 Projection Operators

The goal of this section is to learn how to find the projections of vectors onto a given linear span. This is a task of paramount importance, not only in quantum theory. Expansions into various types of series (such as Fourier series) are nothing more than the search for the projections of a given function onto the vectors of some basis that represent a linear span in space.

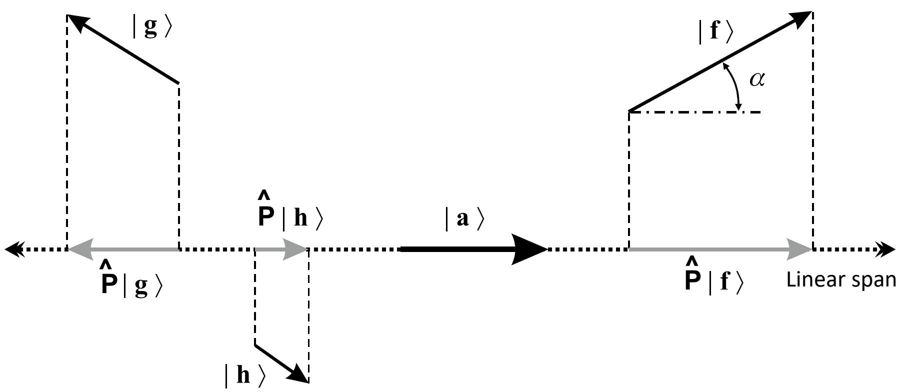


Fig. 3.21: Projection of vectors onto a specified direction

From the entire beam, it is sufficient to choose a single vector that fully describes the beam. We choose this "representative" unit vector, i.e., such that $\langle \mathbf{a}|\mathbf{a}\rangle = 1$, i.e. $\|\mathbf{a}\| = 1$. This is easy to visualize in $\mathcal{R}^2$. In the figure, we see the unit vector $|\mathbf{a}\rangle$ representing the linear span and the vectors $|\mathbf{f}\rangle$, $|\mathbf{g}\rangle$, $|\mathbf{h}\rangle$, which we will project onto this span.

The projection of any vector $|\mathbf{f}\rangle$ has magnitude $\|\mathbf{f}\|\cdot\cos\alpha$ and direction $|\mathbf{a}\rangle / \|\mathbf{a}\|$. The sign of the cosine function in the magnitude of the projection determines whether the projected vector points in the direction of $|\mathbf{a}\rangle$ or in the opposite direction. The projection can be written as the product of its magnitude and direction:

$$\hat{\mathsf{P}}|\mathbf{f}\rangle = \|\mathbf{f}\|\cdot\cos\alpha\frac{|\mathbf{a}\rangle}{\|\mathbf{a}\|} = \|\mathbf{f}\|\cdot\frac{\langle\mathbf{a}|\mathbf{f}\rangle}{\|\mathbf{a}\|\cdot\|\mathbf{f}\|}\cdot\frac{|\mathbf{a}\rangle}{\|\mathbf{a}\|} = \frac{\langle\mathbf{a}|\mathbf{f}\rangle}{\|\mathbf{a}\|\cdot\|\mathbf{a}\|}|\mathbf{a}\rangle = \frac{\langle\mathbf{a}|\mathbf{f}\rangle}{\langle\mathbf{a}|\mathbf{a}\rangle}|\mathbf{a}\rangle .$$

Note that when rearranging expressions in Dirac notation, we can freely move numbers (vector norms and scalar products). The expression for the projection consists of a coefficient that determines the magnitude of the projection and the vector $|\mathbf{a}\rangle$. If we formally write the coefficient after the vector, we obtain the operator form:

$$\hat{\mathsf{P}}|\mathbf{f}\rangle = |\mathbf{a}\rangle\frac{\langle\mathbf{a}|\mathbf{f}\rangle}{\langle\mathbf{a}|\mathbf{a}\rangle} = \frac{|\mathbf{a}\rangle\langle\mathbf{a}|\mathbf{f}\rangle}{\langle\mathbf{a}|\mathbf{a}\rangle} .$$

Let us denote

► $$\hat{\mathsf{P}} \equiv \frac{|\mathbf{a}\rangle\langle\mathbf{a}|}{\langle\mathbf{a}|\mathbf{a}\rangle} . \tag{3.273}$$

This expression is called a projection operator. It has no meaning on its own; it represents a scalar product operation that must be performed. Only by applying the projection operator to a vector $|\mathbf{f}\rangle$ we obtain a meaningful expression – the projection of the vector given by the coefficient $\langle\mathbf{a}|\mathbf{f}\rangle / \langle\mathbf{a}|\mathbf{a}\rangle$ and the direction of the vector $|\mathbf{a}\rangle$. The situation is similar to the operator d/dx; this, too, is merely an indicated derivative that must be performed on a specific function. If the vector $|\mathbf{a}\rangle$ is the unit vector, the expressions simplify further:

$$\begin{aligned}\hat{\mathsf{P}} &\equiv |\mathbf{a}\rangle\langle\mathbf{a}| ,\\ \hat{\mathsf{P}}|\mathbf{f}\rangle &\equiv |\mathbf{a}\rangle\langle\mathbf{a}|\mathbf{f}\rangle .\end{aligned} \tag{3.274}$$

The projection coefficient is $\langle\mathbf{a}|\mathbf{f}\rangle$, and the direction is $|\mathbf{a}\rangle$.

◖ Example 3.34: Projection in a plane

Find the projection of the vector $|\mathbf{f}\rangle|$ onto the vectors $|\mathbf{a}\rangle$ and $|\mathbf{b}\rangle$. The vectors are given as follows:

$$|\mathbf{f}\rangle = \begin{pmatrix}1\\3\end{pmatrix}; \qquad |\mathbf{a}\rangle = \begin{pmatrix}1\\1\end{pmatrix}; \qquad |\mathbf{b}\rangle = \begin{pmatrix}-1\\+1\end{pmatrix}.$$

Solution: First, we find the projection operators:

$$\hat{\mathsf{P}}_a \equiv \frac{|\mathbf{a}\rangle\langle\mathbf{a}|}{\langle\mathbf{a}|\mathbf{a}\rangle} = \frac{\begin{pmatrix}1\\1\end{pmatrix}\cdot\begin{pmatrix}1 & 1\end{pmatrix}}{\begin{pmatrix}1 & 1\end{pmatrix}\cdot\begin{pmatrix}1\\1\end{pmatrix}} = \frac{\begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}}{2} = \begin{pmatrix}1/2 & 1/2\\1/2 & 1/2\end{pmatrix},$$

$$\hat{\mathbf{P}}_b \equiv \frac{|\,\mathbf{b}\,\rangle\langle\,\mathbf{b}\,|}{\langle\,\mathbf{b}\,|\,\mathbf{b}\,\rangle} = \frac{\begin{pmatrix}-1\\+1\end{pmatrix}\cdot\begin{pmatrix}-1 & +1\end{pmatrix}}{\begin{pmatrix}-1 & +1\end{pmatrix}\cdot\begin{pmatrix}-1\\+1\end{pmatrix}} = \frac{\begin{pmatrix}+1 & -1\\-1 & +1\end{pmatrix}}{2} = \begin{pmatrix}1/2 & -1/2\\-1/2 & 1/2\end{pmatrix}.$$

Now we can easily find the projections we're looking for:

$$\hat{\mathbf{P}}_a\,|\,\mathbf{f}\,\rangle = \begin{pmatrix}1/2 & 1/2\\1/2 & 1/2\end{pmatrix}\cdot\begin{pmatrix}1\\3\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix},$$

$$\hat{\mathbf{P}}_b\,|\,\mathbf{f}\,\rangle = \begin{pmatrix}1/2 & -1/2\\-1/2 & 1/2\end{pmatrix}\cdot\begin{pmatrix}1\\3\end{pmatrix} = \begin{pmatrix}-1\\+1\end{pmatrix}.$$

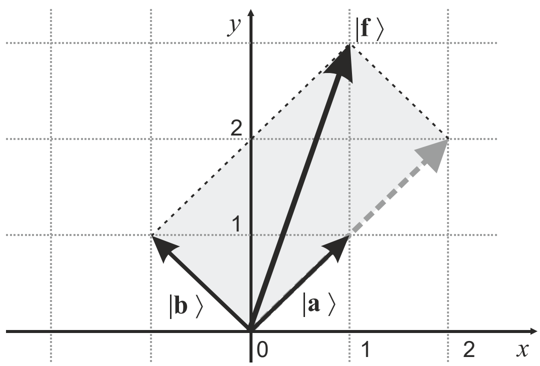


Fig. 3.22: The projection of the vector $|\,\mathbf{f}\,\rangle$ onto two perpendicular directions ◗

Let's use this example to demonstrate that some simple and useful relationships hold true. The calculations are so straightforward that we'll just list the results here:

$$\begin{aligned}
&1. \quad \hat{\mathbf{P}}_a^\dagger = \hat{\mathbf{P}}_a\,; \qquad \hat{\mathbf{P}}_b^\dagger = \hat{\mathbf{P}}_b\,.\\
&2. \quad \hat{\mathbf{P}}_a^2 = \hat{\mathbf{P}}_a\,; \qquad \hat{\mathbf{P}}_b^2 = \hat{\mathbf{P}}_b\,.\\
&3. \quad \hat{\mathbf{P}}_a + \hat{\mathbf{P}}_b = \hat{\mathbf{1}}.
\end{aligned} \tag{3.275}$$

The first property means that projection operators are Hermitian. For matrices, the implication is straightforward: a matrix remains unchanged after being transposed about the main diagonal and then complex conjugated. The second relation also has a very simple meaning: A projection performed twice in succession (the square of the operator) is identical to a projection performed once. Both properties are characteristic for projection operators and are generally considered the definition of a projection operator:

Projection operator

The projection operator is a linear operator that satisfies

► $$\begin{aligned}\hat{\mathbf{P}}^2 &= \hat{\mathbf{P}}\,;\\ \hat{\mathbf{P}}^\dagger &= \hat{\mathbf{P}}\,.\end{aligned} \tag{3.276}$$

It is easy to show that both properties hold for definition (3.273); for example, for the first property we have:

$$\hat{\mathsf{P}}^2 = \frac{|\,\mathbf{a}\,\rangle\langle\,\mathbf{a}\,|}{\langle\,\mathbf{a}\,|\,\mathbf{a}\,\rangle}\cdot\frac{|\,\mathbf{a}\,\rangle\langle\,\mathbf{a}\,|}{\langle\,\mathbf{a}\,|\,\mathbf{a}\,\rangle} = \frac{|\,\mathbf{a}\,\rangle\langle\,\mathbf{a}\,|\,\mathbf{a}\,\rangle\langle\,\mathbf{a}\,|}{\langle\,\mathbf{a}\,|\,\mathbf{a}\,\rangle\langle\,\mathbf{a}\,|\,\mathbf{a}\,\rangle} = \frac{|\,\mathbf{a}\,\rangle\langle\,\mathbf{a}\,|}{\langle\,\mathbf{a}\,|\,\mathbf{a}\,\rangle} = \hat{\mathsf{P}}.$$

In the middle expression, we have reduced the scalar product in the numerator (in the middle) by one of the scalar products in the denominator. These are simple complex numbers that can be factored out of the expressions and reduced.
The meaning of the third relation (3.275) is easy to understand: The vectors$|\,\mathbf{a}\,\rangle$, $|\,\mathbf{b}\,\rangle$ are mutually perpendicular and form an orthogonal basis in the plane. The projections onto these vectors are nothing more than the decomposition of the original vector into this basis. Try adding both projections together. You will get the original vector. The third relation is precisely the mathematical expression of the fact that the sum of all projections yields the original vector:

$$\hat{\mathsf{P}}_a + \hat{\mathsf{P}}_b = \hat{\mathbf{1}} \qquad \Rightarrow \qquad \hat{\mathsf{P}}_a\,|\,\mathbf{f}\,\rangle + \hat{\mathsf{P}}_b\,|\,\mathbf{f}\,\rangle = |\,\mathbf{f}\,\rangle. \tag{3.277}$$

This property is called the *completeness relation*. If a given basis is complete (i.e., no vector is missing), then the sum of all projection operators equals the identity operator. This means that the sum of all projections of any vector yields the original vector.

3.4.4 Expanding an Element to the Base

Let us consider a countable basis (the maximal set of linearly independent vectors) $\{|\mathbf{e}_k\rangle\}$ in a unitary space. By appropriately combining the individual elements, we can always ensure that the basis vectors are mutually orthogonal and have unit magnitude. We will denote such elements simply by their ordinal number: $|\,k\,\rangle$. A basis consisting of vectors $|\,k\,\rangle$ has two fundamental properties:

► $$\langle\,k\,|\,l\,\rangle = \delta_{kl}\,, \tag{3.278}$$

► $$\sum_k |\,k\,\rangle\langle\,k\,| = \hat{\mathbf{1}}. \tag{3.279}$$

Property (3.278) expresses orthogonality. The scalar product of two different vectors is zero (perpendicularity), and that of two identical vectors is one (unit magnitude). Property (3.279) is the completeness relation. The sum of all projections yields the identity operator, i.e., the sum of projections of any vector yields the original vector. In the completeness relation, it is possible to use Einstein's summation convention. In the case of nonseparable spaces with an infinite basis, both relations have the form:

$$\langle\,x\,|\,y\,\rangle = \delta(y-x)\,, \tag{3.280}$$

$$\int |\,x\,\rangle\langle\,x\,|\mathrm{d}x = \hat{\mathbf{1}}. \tag{3.281}$$

In the orthonormality relation, the Kronecker symbol is replaced by the Dirac delta function, and in the completeness relation, the summation is replaced by integration.

Expanding an element to its basis is simple in Dirac notation. All you need to do is insert the completeness relation in the form of the identity operator before the element:

$$|\,\mathbf{f}\,\rangle \;=\; \hat{\mathbf{1}}|\,\mathbf{f}\,\rangle \;=\; \sum_k |\,k\,\rangle\langle\,k\,|\,\mathbf{f}\,\rangle \;=\; \sum_k c_k\,|\,k\,\rangle \quad, \qquad c_k \equiv \langle\,k\,|\,\mathbf{f}\,\rangle\,. \tag{3.282}$$

Expansion coefficients are given by the scalar product of the vector being expanded and a basis vector.

◖ **Example 3.35: Fourier series**

Consider the space $\mathcal{L}^2(0,\,2\pi)$ of periodic functions that are square-integrable. Due to the requirement $f(0)=f(2\pi)$, the set of functions forms a complete basis in this space (the proof can be found in standard mathematics textbooks):

$$|\,f_k\,\rangle = \mathrm{e}^{\mathrm{i}\,kx} \quad , \qquad k = 0, \pm 1, \pm 2, \ldots . \tag{3.283}$$

Although these functions are perpendicular to each other, they are not unit functions:

$$\langle\, f_k \,|\, f_l \,\rangle = \int_0^{2\pi} \mathrm{e}^{-\mathrm{i}\,k\,x}\mathrm{e}^{\mathrm{i}\,l\,x}\,\mathrm{d}x \;=\; \int_0^{2\pi} \mathrm{e}^{\mathrm{i}\,(l-k)\,x}\,\mathrm{d}x \;=\; \begin{cases} 0 & \text{pro } l \neq k \\ 2\pi & \text{pro } l = k \end{cases} \;=\; 2\pi\,\delta_{k\,l}\,.$$

The fact that the scalar product is zero for $l \neq k$ follows from the periodicity of the trigonometric functions on the interval $\langle\,0,\,2\pi\,\rangle$. If we divide the basis elements by their magnitudes, we obtain an orthonormal basis

$$|\,k\,\rangle = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{\mathrm{i}\,kx} \quad , \qquad k = 0, \pm 1, \pm 2, \ldots \quad , \tag{3.284}$$

for which the orthonormality relation (3.278) and the completeness relation (3.279) hold. The expansion of any function from our space is then

$$|\,\mathbf{f}\,\rangle \;=\; \sum_k c_k\,|\,k\,\rangle\,, \qquad c_k \equiv \langle\,k\,|\,\mathbf{f}\,\rangle\,. \tag{3.285}$$

Written in standard notation we have

$$f(x) \;=\; \frac{1}{\sqrt{2\pi}}\sum_k c_k\,\mathrm{e}^{\mathrm{i}\,kx}\,, \qquad c_k = \frac{1}{\sqrt{2\pi}}\int_0^{2\pi} \mathrm{e}^{-\mathrm{i}\,kx}\,f(x)\,\mathrm{d}x\,. \tag{3.286}$$

These are well-known relations for Fourier series. The scalar product is complex conjugate in the left-hand argument, which is why there is a minus sign in c_k.

◗

Representation

In this basis, we can easily rewrite the operator equation $\hat{\mathbf{A}}\,|\,\mathbf{f}\,\rangle = |\,\mathbf{g}\,\rangle$. We insert the identity operator in front of the vector.

$$\hat{\mathbf{A}}\hat{\mathbf{1}}|\,\mathbf{f}\,\rangle = |\,\mathbf{g}\,\rangle \qquad \Rightarrow$$

$$\sum_l \hat{\mathbf{A}}\,|\,l\,\rangle\langle\,l|\,\mathbf{f}\,\rangle = |\,\mathbf{g}\,\rangle \qquad /\langle\,k\,| \quad \text{zleva} \quad \Rightarrow$$

$$\sum_l \langle\,k|\,\hat{\mathbf{A}}\,|\,l\,\rangle\langle\,l|\,\mathbf{f}\,\rangle = \langle\,k\,|\,\mathbf{g}\,\rangle\,. \tag{3.287}$$

However, the resulting expression is nothing more than matrix multiplication

$$\begin{pmatrix} A_{11} & \cdots & A_{1n} & \cdots \\ \vdots & \ddots & & \\ A_{n1} & & A_{nn} & \\ \vdots & & & \ddots \end{pmatrix} \cdot \begin{pmatrix} f_1 \\ \vdots \\ f_n \\ \vdots \end{pmatrix} = \begin{pmatrix} g_1 \\ \vdots \\ g_n \\ \vdots \end{pmatrix}, \qquad \begin{array}{l} \hat{\mathbf{A}} \to A_{kl} \equiv \langle k | \hat{\mathbf{A}} | l \rangle \\ |\mathbf{f}\rangle \to f_n \equiv \langle n | \mathbf{f} \rangle \\ |\mathbf{g}\rangle \to g_n \equiv \langle n | \mathbf{g} \rangle \end{array}$$

In short, we have

$$\sum_l A_{kl} f_l = g_k . \tag{3.288}$$

If we assign a square matrix to the operator

$$\hat{\mathbf{A}} \to A_{kl} \equiv \langle k | \hat{\mathbf{A}} | l \rangle \tag{3.289}$$

and a column matrix for the vector

$$|\mathbf{f}\rangle \to f_l \equiv \langle l | \mathbf{f} \rangle , \tag{3.290}$$

we can treat operator expressions as ordinary matrix multiplication. We say that we have chosen a *representation* of the given space. In fact, this is nothing more than the choice of a specific basis. If the basis has an infinite but countable number of elements, the vectors will correspond to sequences and the operators to infinite matrices. We see that in any separable Hilbert space, there exists a unique mapping of elements to the space of sequences ℓ^2 (an isomorphism). In the case of non-separable spaces with an uncountable basis, we obtain a similar relation:

$$\int \langle x | \hat{\mathbf{A}} | y \rangle \langle y | \mathbf{f} \rangle \, \mathrm{d}y = \langle x | \mathbf{g} \rangle , \tag{3.291}$$

which is nothing more than an integral transform

$$\begin{gathered} \int A(x,y) f(y) \, \mathrm{d}y = g(x) ; \\ A(x,y) \equiv \langle x | \hat{\mathbf{A}} | y \rangle , \\ f(y) \equiv \langle y | \mathbf{f} \rangle , \\ g(x) \equiv \langle x | \mathbf{g} \rangle \end{gathered} \tag{3.292}$$

with the kernel $A(x, y)$. The variables x and y act as continuous indices.

Transition between bases

If we have two sets of basis vectors $\{|k\rangle\}$ and $\{|k'\rangle\}$, there is a simple relation between the expansion coefficients, which we can derive by applying the completeness relation:

$$\begin{gathered} f'_k \equiv \underset{\uparrow}{\langle k' | \mathbf{f} \rangle} = \sum_k \langle k' | k \rangle \langle k | \mathbf{f} \rangle = \sum_k S_{k'k} f_k , \\ S_{k'k} \equiv \langle k' | k \rangle . \end{gathered} \tag{3.293}$$

Matrix $\mathbf{S}$ is called the transition matrix.

3.4.5 Spectral Theory

In operator theory, one of the fundamental tasks is to find directions in which the action of a given operator manifests itself as a complex stretching:

$$\hat{\mathbf{A}}\,|\,\mathbf{f}\,\rangle = \lambda|\mathbf{f}\,\rangle\,; \qquad \lambda \in C\,. \tag{3.294}$$

A vector $|\,\mathbf{f}\,\rangle$ is called an *eigenvector* (characteristic vector) of the operator $\hat{\mathbf{A}}$, and the scaling factor λ is called an *eigenvalue* (characteristic value). For example, for the inertia tensor, the eigenvectors lie along the axes in which the body does not "wobble" during rotation. For linear operators, every multiple of an eigenvector is also an eigenvector with the same eigenvalue. This constitutes an entire span in Hilbert space, or an *eigenvector direction*. There may be a whole series of such eigenvector directions and eigenvalues for linear operators; their maximum number is equal to the *dimension of the space* (the number of basis elements). For separable spaces, we can therefore formulate the problem of finding eigenvalues and eigenvectors using the following equations:

$$\hat{\mathbf{A}}\,|\,\mathbf{f}_k\,\rangle = \lambda_k|\mathbf{f}_k\,\rangle\,; \qquad k = 1, 2, \ldots; \quad \lambda_k \in C\,. \tag{3.295}$$

The set of all eigenvalues $\{\lambda_1, \ldots, \lambda_n, \ldots\}$ is called the *spectrum of the operator* $\hat{\mathbf{A}}$. If we find the spectrum of an operator and its eigenvectors, we can relatively easily solve equations involving that operator. For example, using eigenvalues and eigenvectors, we can solve systems of ordinary linear differential equations (see Chapter 1.5).

Eigenvalues and eigenvectors of the Hermitian operator

Theorem: A Hermitian operator has real eigenvalues, and the eigenvectors corresponding to two distinct eigenvalues are orthogonal to each other.

Proof: Calculating the scalar product, we use the Hermitian property and apply the operator to both the left and right sides of the scalar product. The result is the same:

$$\langle\,\mathbf{f}_k\,|\,\hat{\mathbf{A}}\,\mathbf{f}_k\,\rangle \;=\; \begin{cases} \langle\,\mathbf{f}_k\,|\,\lambda_k\,\mathbf{f}_k\,\rangle = \lambda_k\langle\,\mathbf{f}_k\,|\,\mathbf{f}_k\,\rangle \\ \langle\,\hat{\mathbf{A}}\,\mathbf{f}_k\,|\,\mathbf{f}_k\,\rangle = \lambda_k^*\langle\,\mathbf{f}_k\,|\,\mathbf{f}_k\,\rangle \end{cases} \quad\Rightarrow\quad \lambda_k = \lambda_k^* \quad\Rightarrow\quad \lambda_k \in \mathcal{R}\,.$$

The eigenvalues are therefore real. To find the eigenvalue of the lhs of the scalar product, we will use the first part of the proof:

$$\langle\,\mathbf{f}_k\,|\,\hat{\mathbf{A}}\,\mathbf{f}_l\,\rangle \;=\; \begin{cases} \langle\,\mathbf{f}_k\,|\,\lambda_l\,\mathbf{f}_l\,\rangle = \lambda_l\langle\,\mathbf{f}_k\,|\,\mathbf{f}_l\,\rangle \\ \langle\,\hat{\mathbf{A}}\,\mathbf{f}_k\,|\,\mathbf{f}_l\,\rangle = \lambda_k^*\langle\,\mathbf{f}_k\,|\,\mathbf{f}_l\,\rangle = \lambda_k\langle\,\mathbf{f}_k\,|\,\mathbf{f}_l\,\rangle \end{cases}$$

$$\Downarrow$$

$$\lambda_l\langle\,\mathbf{f}_k\,|\,\mathbf{f}_l\,\rangle = \lambda_k\langle\,\mathbf{f}_k\,|\,\mathbf{f}_l\,\rangle\,,$$

$$\Downarrow$$

$$(\lambda_l - \lambda_k)\langle\,\mathbf{f}_k\,|\,\mathbf{f}_l\,\rangle = 0\,,$$

$$\Downarrow$$

$$\langle\,\mathbf{f}_k\,|\,\mathbf{f}_l\,\rangle = 0 \text{ pro } \lambda_k \neq \lambda_l.$$

Eigenvalues and Eigenvectors of a Unitary Operator

Theorem: The eigenvalues of a unitary operator lie on the complex unit circle, and the eigenvectors corresponding to two distinct eigenvalues are orthogonal to each other.

Proof: We begin our proof with the definition of a unitary operator:

$$\langle \mathbf{f}_k \,|\, \mathbf{f}_k \rangle = \langle \hat{\mathbf{U}}\,\mathbf{f}_k \,|\, \hat{\mathbf{U}}\,\mathbf{f}_k \rangle = \lambda_k^* \lambda_k \langle \mathbf{f}_k \,|\, \mathbf{f}_k \rangle \,.$$

By comparing the first and last expressions, it is clear that

$$\lambda_k \lambda_k^* = 1 \qquad \Rightarrow \qquad |\lambda_k| = 1.$$

It remains to prove that the vectors are orthogonal. To do this, we will need a lemma:

Lemma: If there exists an inverse operator to $\hat{\mathbf{A}}$, it has eigenvalues $1/\lambda_k$.:

Apparently, the following is true

$$\hat{\mathbf{A}}|\psi\rangle = \lambda|\psi\rangle \qquad \Rightarrow \qquad \hat{\mathbf{A}}^{-1}\hat{\mathbf{A}}|\psi\rangle = \lambda\lambda^{(-1)}|\psi\rangle \qquad \Rightarrow$$

$$|\psi\rangle = \lambda\lambda^{(-1)}|\psi\rangle \qquad \Rightarrow$$

$$\lambda^{(-1)} = \frac{1}{\lambda}.$$

■

We will now proceed in the same way as we did for the Hermitian operator. On the left-hand side of the scalar product, we will again use the first part of the proof:

$$\langle \mathbf{f}_k \,|\, \hat{\mathbf{U}}\,\mathbf{f}_l \rangle \;=\; \begin{cases} \langle \mathbf{f}_k \,|\, \lambda_l\, \mathbf{f}_l \rangle = \lambda_l \langle \mathbf{f}_k \,|\, \mathbf{f}_l \rangle \\ \langle \hat{\mathbf{U}}^{-1}\, \mathbf{f}_k \,|\, \mathbf{f}_l \rangle = \dfrac{1}{\lambda_k^*} \langle \mathbf{f}_k \,|\, \mathbf{f}_l \rangle = \lambda_k \langle \mathbf{f}_k \,|\, \mathbf{f}_l \rangle \end{cases}$$

$$\Downarrow$$

$$\lambda_l \langle \mathbf{f}_k \,|\, \mathbf{f}_l \rangle = \lambda_k \langle \mathbf{f}_k \,|\, \mathbf{f}_l \rangle \,,$$

$$\Downarrow$$

$$(\lambda_l - \lambda_k)\langle \mathbf{f}_k \,|\, \mathbf{f}_l \rangle = 0 \,,$$

$$\Downarrow$$

$$\langle \mathbf{f}_k \,|\, \mathbf{f}_l \rangle = 0 \quad \text{for } \lambda_k \neq \lambda_l \,.$$

Note 1: In quantum theory, the real eigenvalues of Hermitian operators are used as possible outcomes of a measurement of the dynamic variable corresponding to the operator $\hat{\mathbf{A}}$. The orthogonality of the eigenvectors is also useful. A suitable Hermitian operator can "generate" a convenient orthonormal basis in Hilbert space through its eigenvectors.

Note 2: In quantum theory, unitary operators are used to describe the time evolution of an object's state. Their properties give rise, e.g., to Ehrenfest theorems on the relationship between classical and quantum physics, as well as the virial theorem, whose classical analogy led to the discovery of dark matter in the universe.

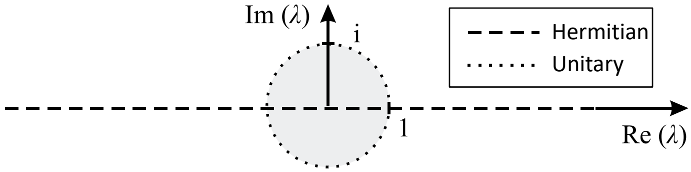


Fig. 3.23: Eigenvalues of the Hermitian and Unitary Operators

◖ **Example 3.36: Eigenvalues and eigenvectors**

Determine the eigenvalues and eigenvectors of the matrix $\hat{\mathbf{A}}$ from Example 3.29, p. 298)

$$\hat{\mathbf{A}} = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}.$$

For this matrix, we will formulate the eigenvalue and eigenvector problem:

$$\hat{\mathbf{A}} \,|\, \mathbf{f} \,\rangle = \lambda \,|\, \mathbf{f} \,\rangle$$

$$\Downarrow$$

$$\begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \cdot \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \lambda \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$$

$$\Downarrow$$

$$\begin{pmatrix} -\lambda & -\mathrm{i} \\ \mathrm{i} & -\lambda \end{pmatrix} \cdot \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The equation has a non-trivial solution only if the determinant is zero. This condition yields two possible values for the eigenvalue λ. For each of these, we can then easily determine the corresponding eigenvector. Note! The condition on the determinant makes the equations for the components of the eigenvector dependent. However, this is fine; the solution to the equations must have one free parameter so that it represents a full linear span in space. For the eigenvectors, we can find the normalized eigenvectors and the corresponding projection operators. The result is:

$$\lambda_1 = -1, \qquad |\, \mathbf{f}_1 \,\rangle = c \begin{pmatrix} +\mathrm{i} \\ +1 \end{pmatrix}, \qquad |\, 1 \,\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} +\mathrm{i} \\ +1 \end{pmatrix}, \qquad \hat{\mathbf{P}}_1 = |\, 1 \,\rangle\langle\, 1 \,| = \frac{1}{2} \begin{pmatrix} 1 & +\mathrm{i} \\ -\mathrm{i} & 1 \end{pmatrix},$$

$$\lambda_2 = +1, \qquad |\, \mathbf{f}_2 \,\rangle = c \begin{pmatrix} -\mathrm{i} \\ +1 \end{pmatrix}, \qquad |\, 2 \,\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} -\mathrm{i} \\ +1 \end{pmatrix}, \qquad \hat{\mathbf{P}}_2 = |\, 2 \,\rangle\langle\, 2 \,| = \frac{1}{2} \begin{pmatrix} 1 & -\mathrm{i} \\ +\mathrm{i} & 1 \end{pmatrix}.$$

Note that the matrix **A** was Hermitian (a matrix reflected across the diagonal and complex-conjugate is identical to the original matrix). Therefore, it has real eigenvalues, and its eigenvectors form an orthonormal basis:

$$\langle\, 1 \,|\, 1 \,\rangle = \langle\, 2 | 2 \,\rangle = 1, \qquad \langle\, 1 \,|\, 2 \,\rangle = \langle\, 2 | 1 \,\rangle = 0.$$

This set of vectors is complete; it forms a basis in the space of complex pairs:

$$|\, 1 \,\rangle\langle\, 1 \,| + |\, 2 \,\rangle\langle\, 2 \,| = \hat{\mathbf{P}}_1 + \hat{\mathbf{P}}_2 = \hat{\mathbf{1}}.$$ ◗

Spectral expansion (decomposition) theorem

Theorem: Let $\hat{\mathbf{A}}$ be a linear operator with a set of eigenvectors that forms a complete orthonormal basis in a Hilbert space. Then, for the analytic function of the operator defined by the Taylor series (3.263), we can write

► $$f(\hat{\mathbf{A}}) = \sum_k f(\lambda_k)|\,k\,\rangle\langle\,k\,| = \sum_k f(\lambda_k)\hat{\mathbf{P}}_k \,. \tag{3.296}$$

Proof: First, let's consider the effect of the operator powers on eigenvectors:

$$\hat{\mathbf{A}}\,|\,\mathbf{f}_k\,\rangle = \lambda_k|\mathbf{f}_k\,\rangle \quad ,$$

$$\hat{\mathbf{A}}^2\,|\,\mathbf{f}_k\,\rangle = \lambda_k\hat{\mathbf{A}}|\mathbf{f}_k\,\rangle = \lambda_k^2\,|\mathbf{f}_k\,\rangle\,,$$

$$\vdots$$

$$\hat{\mathbf{A}}^n\,|\,\mathbf{f}_k\,\rangle = \lambda_k^n\,|\mathbf{f}_k\,\rangle\,,$$

$$\Downarrow$$

$$f(\hat{\mathbf{A}})\,|\,\mathbf{f}_k\,\rangle = \sum_n c_n\hat{\mathbf{A}}^n\,|\,\mathbf{f}_k\,\rangle = \sum_n c_n\lambda_k^n\,|\,\mathbf{f}_k\,\rangle = f(\lambda_k)\,|\,\mathbf{f}_k\,\rangle\,.$$

In the next step of the proof, we will apply the operator to a general vector. However, we will first decompose it into a basis consisting of eigenvectors, where we know the action from the last equality (we substitute the sum of all projectors at the arrow):

$$f(\hat{\mathbf{A}})\underset{\uparrow}{|}\,\mathbf{f}\,\rangle = \sum_k f(\hat{\mathbf{A}})|\,k\,\rangle\langle\,k\,|\,\mathbf{f}\,\rangle = \sum_k f(\lambda_k)|\,k\,\rangle\langle\,k\,|\,\mathbf{f}\,\rangle.$$

But that is precisely the equality we wanted to prove. If we omit from the expressions any vector $|\,\mathbf{f}\,\rangle$ on which the operators act, we obtain the spectral expansion theorem.

Note 1: If an operator has a multiple eigenvalue of order N, this is not a problem. The eigenvectors corresponding to the multiple eigenvalue form a complete subspace $\mathcal{P}$ of dimension N, and it is possible to choose N independent, orthogonal eigenvectors corresponding to this multiple eigenvalue.

Note 2: If the space is non-separable, then under certain additional assumptions, it will be possible to modify the spectral expansion theorem to the following form:

$$f(\hat{\mathbf{A}}) = \int f(\lambda_x)\,|\,x\,\rangle\langle\,x\,|\,\mathrm{d}x\,.$$

Note 3: It is often much simpler to use the spectral expansion theorem instead of the Taylor expansion to express the operator's function. The corresponding series runs only over the operator's eigenvalues.

Note 4: If we know the spectrum of an operator and its eigenvectors, we can easily write any function of the operator and thus solve equations in which this function of the operator appears. In particular, the inverse operator is given by the formula

$$\hat{\mathbf{A}}^{-1} = \sum_k \frac{1}{\lambda_k} | k \rangle\langle k | = \sum_k \frac{1}{\lambda_k} \hat{\mathbf{P}}_k \ .$$

We see that, in addition to the conditions of the theorem, the non-zero nature of all eigenvalues is necessary for its existence.

◖ **Example 3.37: Finding the exponential of a matrix**

A matrix operator is defined on the space C^2

$$\hat{\mathbf{A}} = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}.$$

We are to find the matrix exp($\hat{\mathbf{A}}$). We have already solved this problem using the Taylor series expansion in Example 3.29. From Example 3.36, we know the eigenvalues, eigenvectors, and projection operators of this matrix operator. Therefore, from the spectral expansion theorem, we can write:

$$\mathrm{e}^{\hat{\mathbf{A}}} = \mathrm{e}^{\lambda_1} \hat{\mathbf{P}}_1 + \mathrm{e}^{\lambda_2} \hat{\mathbf{P}}_2 = \mathrm{e}^{-1} \frac{1}{2} \begin{pmatrix} 1 & +\mathrm{i} \\ -\mathrm{i} & 1 \end{pmatrix} + \mathrm{e}^{+1} \frac{1}{2} \begin{pmatrix} 1 & -\mathrm{i} \\ +\mathrm{i} & 1 \end{pmatrix} = \begin{pmatrix} \mathrm{ch}\,1 & -\mathrm{i}\,\mathrm{sh}\,1 \\ \mathrm{i}\,\mathrm{sh}\,1 & \mathrm{ch}\,1 \end{pmatrix}.$$

Unlike Taylor's series, this series now has only two terms.

◗

◖ **Example 3.38: Solving the heat diffusion equation**

Let us determine the time evolution of the temperature of a rod of length L, both ends of which are kept at zero temperature. The initial temperature of the rod is given by the function $T_0(x)$. The task is therefore to find the solution to the heat diffusion equation

$$\frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial x^2} \quad , \qquad T = T(t, x)$$

with initial and boundary conditions

$$T(t_0, x) = T_0(x) \,, \qquad T(t, 0) = T(t, L) = 0 \,.$$

Let us now reformulate the problem in Dirac notation. First, let us define the Hilbert space

$$\mathcal{H} = \left\{ f(x) : \quad f \in L^2(0, L) \quad \wedge \quad f(0) = f(L) = 0 \right\}.$$

This is the space of functions defined on the interval <0, L>, which are periodic and square-integrable. The boundary condition of the original equation is moved into the definition of the space. If the temperature of the rod at both ends were nonzero, we could shift the origin of the temperature scale. This has no effect on the form of the equation due to the derivatives in the heat diffusion equation. The requirement of zero temperature at both ends of the rod therefore does not affect from the generality of the solution. The problem now takes the form:

$$\frac{\mathrm{d}|T\rangle}{\mathrm{d}t} = -\kappa\hat{\mathbf{A}}|T\rangle, \qquad |T\rangle \in \mathcal{H}, \qquad \hat{\mathbf{A}} \equiv -\frac{\mathrm{d}^2}{\mathrm{d}x^2}. \tag{3.297}$$

From Example 3.33, we know that the operator $\hat{\mathbf{B}} = \mathrm{i}\,\mathrm{d}/\mathrm{d}x$ is Hermitian (it has real eigenvalues and an orthogonal set of eigenvectors). Therefore, the square of this operator, $\hat{\mathbf{A}} = \hat{\mathbf{B}}^2$, is also Hermitian. The minus sign is not essential here; it merely ensures that the eigenvalues of the operator $\hat{\mathbf{A}}$ are non-negative. We can write the solution to the problem formally immediately (!):

$$|T(t)\rangle = \mathrm{e}^{-\kappa\hat{\mathbf{A}}(t-t_0)}|T(t_0)\rangle. \tag{3.298}$$

Indeed: if we substitute the initial time, the solution satisfies the initial condition. If we differentiate the solution with respect to time, we find that the solution (3.298) satisfies the initial equation (3.297):

$$\frac{\mathrm{d}}{\mathrm{d}t}|T(t)\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{-\kappa\hat{\mathbf{A}}(t-t_0)}|T(t_0)\rangle = -\kappa\hat{\mathbf{A}}\,\mathrm{e}^{-\kappa\hat{\mathbf{A}}(t-t_0)}|T(t_0)\rangle = -\kappa\hat{\mathbf{A}}|T(t)\rangle.$$

The only problem is that the solution we found (3.298) involves the operator $\hat{\mathbf{A}}$. To determine it, we need to know the spectrum and eigenvectors of the operator $\hat{\mathbf{A}}$ on the space $\mathcal{H}$. So let's first solve the problem

$$\hat{\mathbf{A}}|f\rangle = \lambda|f\rangle, \qquad |f\rangle \in \mathcal{H} \qquad \Rightarrow$$
$$f'' + \lambda f = 0, \qquad f(0) = f(L) = 0.$$

The solution to this ordinary linear differential equation with boundary conditions is:

$$\lambda_k = \frac{\pi^2 k^2}{L^2},$$
$$|f_k\rangle = c\sin\left(\frac{k\pi}{L}x\right),$$
$$|k\rangle = \sqrt{\frac{2}{L}}\sin\left(\frac{k\pi}{L}x\right), \qquad k = 1, 2, \ldots.$$

The eigenvalues are real ($\hat{\mathbf{A}}$ is the Hermitian operator), the eigenvectors are mutually orthogonal, and they form a natural basis in $\mathcal{H}$. Writing the solution to (3.298) is now simply a matter of applying the spectral expansion theorem:

$$|T(t)\rangle = \mathrm{e}^{-\kappa\hat{\mathbf{A}}(t-t_0)}|T(t_0)\rangle =$$
$$= \sum_{k=1}^{\infty}\mathrm{e}^{-\lambda_k\kappa(t-t_0)}\,|k\rangle\langle k|T(t_0)\rangle =$$
$$= \sum_{k=1}^{\infty}c_k\,\mathrm{e}^{-\frac{\pi^2k^2\kappa}{L^2}(t-t_0)}\,|k\rangle;$$

$$|\,k\,\rangle \equiv \sqrt{\frac{2}{L}}\sin\left(\frac{k\pi}{L}x\right), \qquad c_k \equiv <k\,|\,T(t_0)\,\rangle\,. \tag{3.299}$$

Of course, we can also write the solution in standard form, without using Dirac notation:

$$\begin{aligned} T(t,x) &= \sqrt{\frac{2}{L}}\sum_{k=1}^{\infty} c_k\,\mathrm{e}^{-\frac{\pi^2 k^2 \kappa}{L^2}(t-t_0)}\sin\left(\frac{k\pi}{L}x\right); \\ c_k &\equiv \sqrt{\frac{2}{L}}\int_0^L \sin\left(\frac{k\pi}{L}x\right)T_0(x)\,\mathrm{d}x\,. \end{aligned} \tag{3.300}$$

For $t = t_0$, we have the Fourier series for the initial condition. For $t \neq t_0$, this is nothing more than a series expansion into individual Fourier modes. Each Fourier component decays exponentially with time. We can see that the spectral expansion theorem can also be useful when solving partial differential equations.

◗

3.5 From Gradient to Helicity

Derivation is a very effective tool for determining the properties of functions. In physics, we most often differentiate scalar or vector fields. A scalar field assigns a single value to a position, such as density, temperature, or pressure, while a vector field assigns a triple of values, such as a velocity field, an electric field, or a magnetic field:

$$\begin{aligned} f(\mathbf{x}): &\qquad \mathcal{R}^3 \to \mathcal{R}\,, \\ \mathbf{K}(\mathbf{x}): &\qquad \mathcal{R}^3 \to \mathcal{R}^3\,. \end{aligned} \tag{3.301}$$

If the field varies with time, a time coordinate is added, and the mapping is from $\mathcal{R}^4$. In relativity, we use four-vectors, so ultimately this can be a mapping $\mathcal{R}^4 \to \mathcal{R}^4$. In wave theory, complex functions are usually used, in which case it is a mapping $\mathcal{R}^4 \to \mathcal{C}^4$. We often use the symbol ψ as a placeholder for any field, whether scalar, vector, or otherwise. When differentiating with respect to spatial coordinates, the indicated operation can often be used

► $$\nabla \equiv \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right). \tag{3.302}$$

There are many different ways to write this operation; we most often use the symbol ∇, but there are other options as well:

$$\nabla \equiv \text{grad} \equiv \frac{\partial}{\partial \mathbf{x}} \equiv \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \equiv \left(\partial_x, \partial_y, \partial_z \right) \equiv \left(\partial_1, \partial_2, \partial_3 \right) \equiv \partial_k\,. \tag{3.303}$$

All these notations are simply shorthand for the same operation, which is called the gradient. We denote it with the symbol of an inverted delta and call it "*nabla*." The name was introduced by the Scottish mathematical physicist Peter Guthrie Tait (1831–1901) based on the triangular shape of an Assyrian harp from the 7th century BCE. Assyria was located in northern Mesopotamia. The word *nabla* (Nbl) comes from Aramaic, which adapted it from the Hebrew *Nev(b)el*. However, the Sumerians were already familiar with the same instrument as early as 3100 BCE. James Clerk Maxwell coined the name "*slope*" for this operator, derived from the English word meaning gradient or inclination. However, Tait's proposal prevailed. A scalar function can be acted upon in only one way, whereas for vector fields there are multiple possibilities:

► $$\begin{array}{lll} \nabla f: & \partial_k f & \text{gradient} \\ \nabla \cdot \mathbf{K}: & \partial_k K_k & \text{divergence} \\ \nabla \times \mathbf{K}: & \varepsilon_{klm} \partial_l K_m & \text{rotation (curl)} \\ \nabla \otimes \mathbf{K}: & \partial_k K_l & \text{tensor gradient} \end{array} \tag{3.304}$$

In the following text, we will take a closer look at the various differential operations (gradient, divergence, curl, and helicity). We will explore not only their mathematical foundations but also their practical applications in physics.

3.5.1 Gradient

Let's now consider the gradient of a scalar function. Gradient can be useful in calculations such as directional derivatives, Taylor series expansions, convective derivatives, normal vectors to isosurfaces, and the relationship between force and potential energy. Of course, there are other areas in which the gradient plays an important role.

Directional derivative

When calculating partial derivatives, we always approach the function along the axis. Let us now assume that we approach it in a general direction and define the derivative in the direction s using the relation

$$\frac{\partial f}{\partial \mathbf{s}} \equiv \lim_{h\to 0} \frac{f(\mathbf{x}+h\mathbf{s})-f(\mathbf{x})}{h\,||\,\mathbf{s}\,||}. \tag{3.305}$$

We can now easily calculate this limit using Taylor's expansion:

$$\frac{\partial f}{\partial \mathbf{s}} \equiv \lim_{h\to 0} \frac{f(\mathbf{x})+\dfrac{\partial f}{\partial x_k}hs_k+\dfrac{1}{2!}\dfrac{\partial^2 f}{\partial x_k \partial x_l}hs_k hs_l+\cdots-f(\mathbf{x})}{h\,||\,\mathbf{s}\,||} = \frac{\partial f}{\partial x_k}\frac{s_k}{||\,\mathbf{s}\,||}.$$

The first and last terms are subtracted. In all other terms, except for the second, the quantity h remains, which approaches zero as the limit. If we introduce the unit vector $\boldsymbol{\sigma} = \mathbf{s}/||\mathbf{s}||$, we have a useful relationship for the derivative in that direction

► $$\frac{\partial f}{\partial \mathbf{s}} = (\boldsymbol{\sigma}\cdot\nabla)f; \qquad \boldsymbol{\sigma} \equiv \frac{\mathbf{s}}{||\,\mathbf{s}\,||}. \tag{3.306}$$

◖ **Example 3.39: Differentiate the scalar function in the direction s = (1, 2)**.
The solution is straightforward; simply use the formula: (3.306):

$$\mathbf{s} = (1, 2); \qquad \boldsymbol{\sigma} = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right);$$

$$\boldsymbol{\sigma}\cdot\nabla = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)\cdot(\partial_x, \partial_y) = \frac{\partial_x}{\sqrt{5}} + 2\frac{\partial_y}{\sqrt{5}};$$

$$\frac{\partial f}{\partial \mathbf{s}} = (\boldsymbol{\sigma}\cdot\nabla)f = \frac{1}{\sqrt{5}}\frac{\partial f}{\partial x} + \frac{2}{\sqrt{5}}\frac{\partial f}{\partial y}.$$

◗

Taylor expansion

Using the gradient, it is very easy to write the first non-trivial term of the Taylor expansion of a scalar or vector field in three dimensions:

$$\psi(\mathbf{x}+\mathbf{h}) = \psi(\mathbf{x}) + \frac{\partial \psi}{\partial x_k}h_k + \cdots = \psi(\mathbf{x}) + h_k\frac{\partial \psi}{\partial x_k} + \cdots$$

We can now easily write the relationship in its final form using the gradient operation:

► $$\psi(\mathbf{x}+\mathbf{h}) = \psi(\mathbf{x}) + (\mathbf{h}\cdot\nabla)\psi + \cdots \qquad (3.307)$$

This relationship is elegant and is very commonly used in physics.

Convective derivative

Suppose we have a field that varies in both time and space, and our task is to find its total time derivative:

$$\frac{\mathrm{d}\psi(t,\mathbf{x})}{\mathrm{d}t} = \frac{\partial\psi}{\partial t} + \frac{\partial\psi}{\partial x_k}\frac{\mathrm{d}x_k}{\mathrm{d}t} = \frac{\partial\psi}{\partial t} + \frac{\partial\psi}{\partial x_k}u_k \,,$$

where we have denoted the velocity field $\mathbf{u}(t, \mathbf{x})$. Using the gradient, we have:

► $$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \frac{\partial\psi}{\partial t} + (\mathbf{u}\cdot\nabla)\psi\,; \;\text{ resp. }\quad \frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + (\mathbf{u}\cdot\nabla)\,. \qquad (3.308)$$

The first term is the partial derivative with respect to time, which describes local changes. The second term is called the *convective* derivative and describes changes caused by flow. A typical application is, e.g., the equation of motion for a fluid element:

$$\Delta m\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \Delta\mathbf{F}\,.$$

We normalize the equation to a unit volume and expand the total time derivative according to equation (3.308):

► $$\rho\frac{\partial\mathbf{u}}{\partial t} + \rho(\mathbf{u}\cdot\nabla)\mathbf{u} = \mathbf{f}\,, \qquad (3.309)$$

where $\mathbf{f}$ is the volume density of the forces acting on the fluid.

Normal to the isosurface

The general hyperplane in N dimensions can be implicitly written as

$$\phi(x_1,\cdots,x_N) = \text{const}\,. \qquad (3.310)$$

An example is the equation of the sphere surface in three dimensions: $x^2 + y^2 + z^2 = R^2$. In N dimensions, equation (3.310) represents an $N-1$-dimensional set. In two dimensions, it is a curve; in three dimensions, it is a surface. In the same way, we can write the equation of constant surfaces of a certain quantity, known as isosurfaces – e.g., isotherms for temperature, isobars for pressure, or isophotes for light intensity. We simply substitute the appropriate quantity for ϕ. Let us now differentiate equation (3.310):

$$\frac{\partial\phi}{\partial x_k}\mathrm{d}x_k = 0\,; \qquad \Rightarrow \qquad (\nabla\phi\,|\,\mathrm{d}\mathbf{l}) = 0 \qquad \Rightarrow \qquad \nabla\phi \perp \mathrm{d}\mathbf{l}. \qquad (3.311)$$

The gradient of the function ϕ is therefore perpendicular to any surface vector drawn from a given point. Thus, the gradient is perpendicular to the isosurface defined by the equation (3.310). Since the derivative is the tangent slope of the function, the gradient points perpendicularly from a given isosurface toward other isosurfaces with higher values. The normal to the isosurface can therefore be found very easily:

► $$\mathbf{n} = \nabla\phi\,; \qquad \boldsymbol{\nu} = \frac{\nabla\phi}{\|\nabla\phi\|}\,. \tag{3.312}$$

The first option is a general normal vector; the second is a normal vector normalized to one, i.e., the normal vector **ν** has a magnitude of one. For a closed surface, we always choose the normal vector so that it points outward from the surface.

◖ **Example 3.40: Find the perpendiculars to the parabola $y = x^2$ at various points**

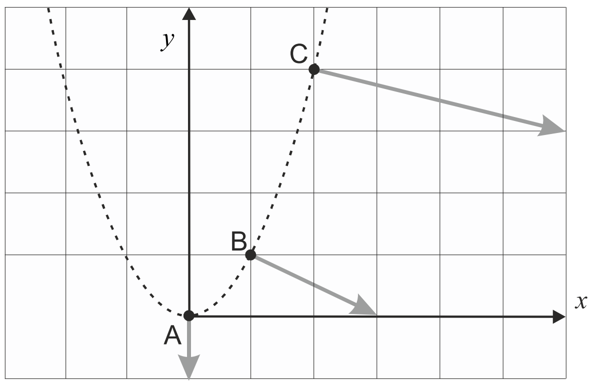


Fig. 3.24: Normal vectors to a parabola

First, we rewrite the equation in its implicit form

$$\phi(x, y) = x^2 - y = 0 \tag{3.313}$$

Now we can easily find the gradient that is perpendicular to the hypersurface (in this case, a parabola):

$$\mathbf{n} = \nabla\phi = (2x, -1)\,. \tag{3.314}$$

If we substitute various points from the figure, we get

$$\begin{aligned} \mathbf{n}_A &= (0, -1)\,; \\ \mathbf{n}_B &= (2, -1)\,; \\ \mathbf{n}_C &= (4, -1)\,. \end{aligned} \tag{3.315}$$

◗

Force and energy

Suppose we have a potential field in which work is done at the expense of potential energy

$$\mathrm{d}A = -\mathrm{d}W_\mathrm{p}\,. \tag{3.316}$$

We will now express the work as force multiplied by the distance and the cosine of the angle between them, and then expand the differential on the right-hand side:

$$F\,\mathrm{d}s\,\cos\alpha = -\frac{\partial W_\mathrm{p}}{\partial x_k}\mathrm{d}x_k\,,$$

$$\mathbf{F} \cdot \mathrm{d}\mathbf{s} = -\nabla W_\mathrm{p} \cdot \mathrm{d}\mathbf{s} ,$$

$$\left(\mathbf{F} \mid \mathrm{d}\mathbf{s}\right) = \left(-\nabla W_\mathrm{p} \mid \mathrm{d}\mathbf{s}\right).$$

Since the above relationship holds for any path element d**s**, it must hold that

► $$\mathbf{F} = -\nabla W_\mathrm{p} . \tag{3.317}$$

This equation states that the force always points toward the minimum of the potential energy. For a conservative field, therefore, a single scalar quantity – potential energy – is sufficient to describe it. We can always obtain the corresponding force from equation (3.317). For example, for the gravitational field of a body of mass M located at the origin of the coordinates and acting on a body of mass $m << M$, we have:

$$W_\mathrm{p} = -G\frac{mM}{r} . \tag{3.318}$$

It is now easy to identify the individual components of force:

$$\begin{aligned} F_k &= -\frac{\partial W_\mathrm{p}}{x_k} = \frac{\partial}{x_k}\left[G\frac{mM}{r}\right] = GmM\frac{\partial}{r}\left(\frac{1}{r}\right)\frac{\partial r}{\partial x_k} = \\ &= GmM\left(-\frac{1}{r^2}\right)\frac{\partial\sqrt{x_1^2 + x_2^2 + x_3^2}}{\partial x_k} = -G\frac{mM}{r^2}\frac{x_k}{r} . \end{aligned} \tag{3.319}$$

We have indeed obtained the correct values for the force components, including the direction vector:

$$\mathbf{F} = -G\frac{mM}{r^2}\frac{\mathbf{r}}{r} . \tag{3.320}$$

We can easily find the magnitude of the force:

$$F = \sqrt{\mathbf{F}\cdot\mathbf{F}} = \sqrt{F_x^2 + F_y^2 + F_z^2} = \sqrt{\frac{G^2 m^2 M^2}{r^6}\left(x^2 + y^2 + z^2\right)} = G\frac{mM}{r^2} . \tag{3.321}$$

3.5.2 Divergence

A very important task is to determine whether a vector field originates at a given point (e.g., electric field emerges from a positive charge), or whether it passes through that point or vanishes at it (e.g., an electric field in a negative charge). To consider this problem, it will be useful to introduce the concept of the flux of a vector field:

► $$\mathrm{d}\phi \equiv \mathbf{K}\cdot\mathrm{d}\mathbf{S} ; \qquad \mathrm{d}\mathbf{S} = \boldsymbol{\nu}\,\mathrm{d}S . \tag{3.322}$$

An elementary surface is characterized by a vector whose magnitude is equal to the area of the surface and whose direction is the normal to the surface. The unit of the flux is equal to the unit of the vector field multiplied by the square of the meter:

$$[\mathrm{d}\phi] = [\mathbf{K}]\mathrm{m}^2. \tag{3.323}$$

The unit of magnetic flux density is Tm^2, the unit of electric flux density is $\mathrm{Vm}^{-1}\mathrm{m}^2$ = Vm, the unit of velocity flux is $\mathrm{ms}^{-1}\mathrm{m}^2 = \mathrm{m}^3\mathrm{s}^{-1}$, and so on.

Note: The flux of a vector field refers to intensive vector quantities (those that do not depend on the amount of substance), such as electric, magnetic, or velocity fields. Do not confuse this flux with the flux of an additive (proportional to the amount of substance) quantity A, which is defined as the amount of the quantity flowing through a unit area per unit time (mass flux, energy flux, charge flux, momentum flux, entropy flux, etc.), i.e., $[\mathbf{j}] = [A]/(\mathrm{m}^2\mathrm{s})$. For example, the charge flux has the unit $\mathrm{Cm}^{-2}\mathrm{s}^{-1}$.

From definition (3.322), it is clear that the flux of the field in the direction of the normal is maximal, zero perpendicular to the normal, and in the case of oblique flux, only the projection of the field vector onto the direction of the normal is used:

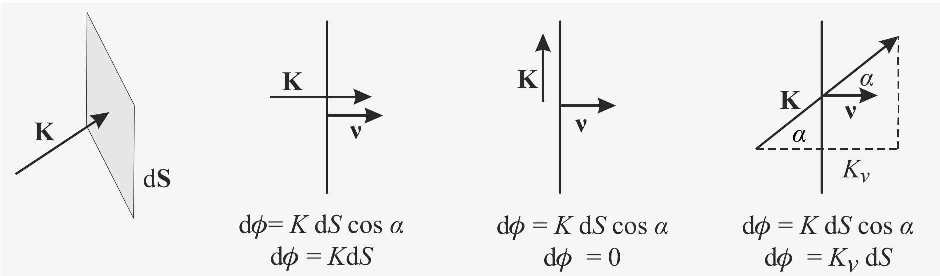


Fig. 3.25: Defining the flux of a vector field through a surface

Let us determine the total of the field through a rectangular prism that we create around the point. Later, we will shrink the prism down to the given point. The flux of the field will have six components (corresponding to the six faces of the prism). We will label the point of the prism closest to the origin (x, y, z), the farthest point (x+Δx, y+Δy, z+Δz), and our point, i.e., the center of the prism ($\tilde{x}$, $\tilde{y}$, $\tilde{z}$). The total flux of the field **K** through the surface of the prism will be (we always locate the field vector at the center of the face currently being calculated, with the normal vector pointing outward):

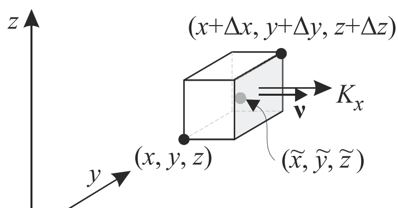


Fig. 3.26: The flux of a vector field through a surface

$$\Delta\phi = \sum_{a=1}^{6} \Delta\phi_a \, .$$

We will label the faces in the following order: right, left, back, front, top, and bottom. The faces of the cuboid have edges of length Δx, Δy, and Δz:

$$\begin{aligned}\Delta\phi = {} & K_x(x+\Delta x, \tilde{y}, \tilde{z})\,\Delta y\,\Delta z - K_x(x, \tilde{y}, \tilde{z})\,\Delta y\,\Delta z + \\ & + K_y(\tilde{x}, y+\Delta y, \tilde{z})\,\Delta z\,\Delta x - K_y(\tilde{x}, y, \tilde{z})\,\Delta z\,\Delta x + \\ & + K_z(\tilde{x}, \tilde{y}, z+\Delta z)\,\Delta x\,\Delta y - K_z(\tilde{x}, \tilde{y}, z)\,\Delta x\,\Delta y \, .\end{aligned}$$

Now we'll take out the volume of the elementary cuboid from all the members:

$$\Delta\phi = \begin{bmatrix} \dfrac{K_x(x+\Delta x,\tilde{y},\tilde{z}) - K_x(x,\tilde{y},\tilde{z})}{\Delta x} + \\ +\dfrac{K_y(\tilde{x},y+\Delta y,\tilde{z}) - K_y(\tilde{x},y,\tilde{z})}{\Delta y} + \\ +\dfrac{K_z(\tilde{x},\tilde{y},z+\Delta z) - K_z(\tilde{x},\tilde{y},z)}{\Delta z} \end{bmatrix} \Delta x\,\Delta y\,\Delta z\,.$$

If we now take the limit as Δx, Δy, and $\Delta z \to 0$, the sum of the partial derivatives will appear in the square brackets

$$\mathrm{d}\phi = \left(\frac{\partial K_x}{\partial x} + \frac{\partial K_y}{\partial y} + \frac{\partial K_z}{\partial z}\right)\mathrm{d}V. \tag{3.324}$$

We call the expression in parentheses the divergence of the field. It is a single number, which we denote by

► $$\operatorname{div}\mathbf{K} \equiv \nabla\cdot\mathbf{K} = \partial_l K_l = \frac{\partial K_x}{\partial x} + \frac{\partial K_y}{\partial y} + \frac{\partial K_z}{\partial z}. \tag{3.325}$$

The flux through an infinitesimal rectangular prism led around our chosen point is

► $$\mathrm{d}\phi \equiv \sum_a \mathbf{K}_a \cdot \mathrm{d}\mathbf{S}_a = \operatorname{div}\mathbf{K}\;\mathrm{d}V. \tag{3.326}$$

It is clear that the operation we are looking for is the divergence. If it is positive, the field emerges from the point; if it is negative, the field sinks toward the point; and if it is zero, the field passes through the point:

► $$\operatorname{div}\mathbf{K}\begin{cases} >0: & \text{Field emerges} \\ =0: & \text{Field passes through} \\ <0: & \text{Field sinks} \end{cases} \tag{3.327}$$

If we have a finite set Ω, we can easily calculate the flux of the field across its boundary $\partial\Omega$. We fill the region Ω completely with many cuboids. The flux always cancels out at their adjacent faces, because the outer normals of the adjacent cuboids will point in opposite directions. The only non-zero flux will be at the boundary of the region where there are no more adjacent prisms or other shapes. Therefore, the following will hold

$$\phi \equiv \oiint_{S=\partial\Omega} \mathbf{K}\cdot\mathrm{d}\mathbf{S} = \iiint_{V=\Omega} \operatorname{div}\mathbf{K}\;\mathrm{d}V. \tag{3.328}$$

The left-hand side shows the sum of the flux through each small surface, and the right-hand side shows the result. The derived relationship is called *Gauss theorem* and is useful in physics for finding various relationships and for converting surface and volume integrals. Gauss's theorem is most commonly written in the form

► $$\oiint_{\partial\Omega} \mathbf{K}\cdot\mathrm{d}\mathbf{S} = \iiint_{\Omega} \operatorname{div}\mathbf{K}\;\mathrm{d}V. \tag{3.329}$$

Let's now look at how to calculate divergence on four two-dimensional fields

$$\mathbf{A} = (-y, +x, 0)\,;$$
$$\mathbf{B} = (+x, +y, 0)\,;$$
$$\mathbf{C} = \alpha\left(\frac{x}{r^3}, \frac{y}{r^3}, \frac{z}{r^3}\right); \qquad (3.330)$$
$$\mathbf{D} = (+y, +x, 0)$$

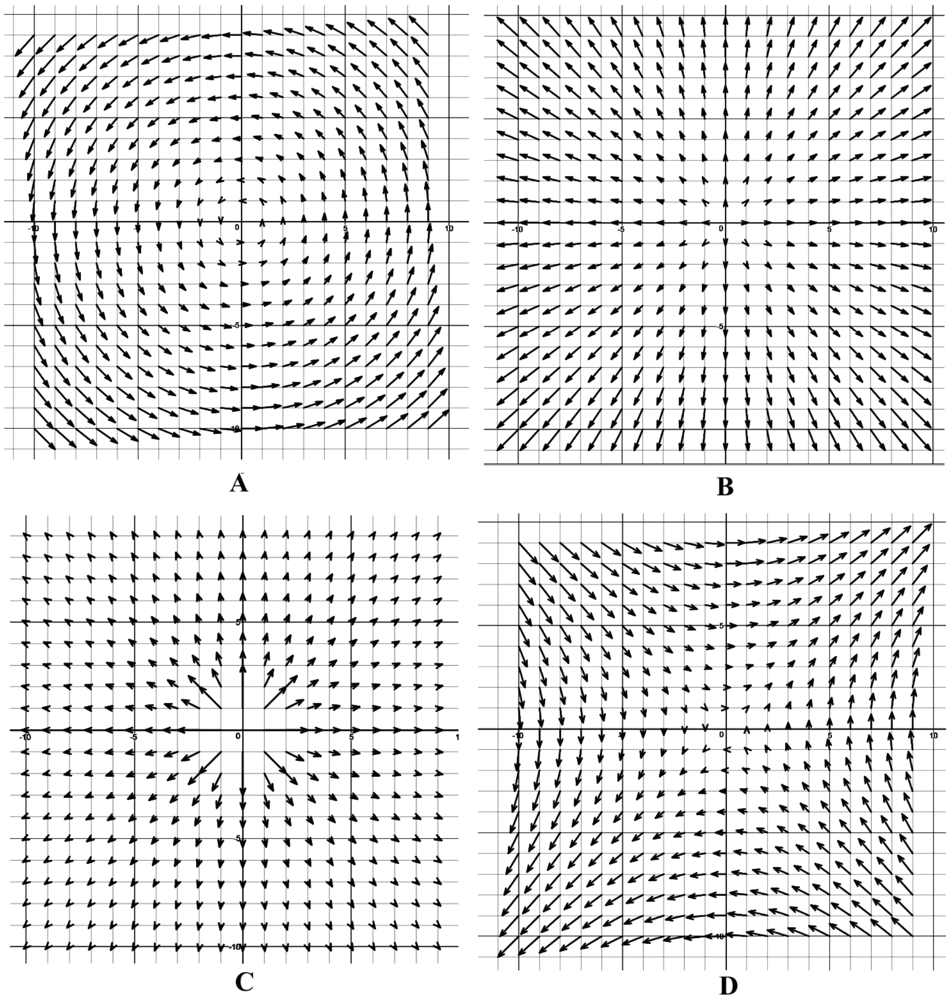


Fig. 3.27: Examples of the four fields **A**, **B**, **C**, **D**

If we plot the components of the first field, we obtain a typical vortex around the origin; if we plot the second field, radial vectors emerge that continue to increase in magnitude (proportionally to the distance from the origin). Such a field is non-physical and must continue to emerge with increasing distance. Every point in space is its source. The third field is the electric field intensity of a point charge; the field decreases from the origin, and its source is only at the origin. The fourth field clearly has no source and does not form vortices. Let us now calculate the divergences of these fields:

$$\operatorname{div}\mathbf{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} = 0 + 0 = 0 \,. \tag{3.331}$$

$$\operatorname{div}\mathbf{B} = \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} = 1 + 1 = 2 \,. \tag{3.332}$$

$$\operatorname{div}\mathbf{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} = 0 + 0 = 0 \,. \tag{3.333}$$

The first field creates a vortex but has no source; in the second field, a radial component of the field is generated at every point in space. The fourth field has neither sources nor vortices. For the third field, the calculation is more complex but straightforward:

$$\begin{aligned} \operatorname{div}\mathbf{C} &= \partial_k C_k = \alpha\, \partial_k \left(\frac{x_k}{r^3} \right) = \alpha\, \frac{(\partial_k x_k) r^3 - x_k \partial_k \left(r^3 \right)}{r^6} = \\ &= \alpha\, \frac{3r^3 - x_k \dfrac{\partial r^3}{\partial r} \dfrac{\partial r}{\partial x_k}}{r^6} = \alpha\, \frac{3r^3 - x_k 3r^2 \dfrac{x_k}{r}}{r^6} = \alpha\, \frac{3r^3 - 3r^3}{r^6} = \begin{cases} 0\,; & r \neq 0\,, \\ \infty\,; & r = 0\,. \end{cases} \end{aligned} \tag{3.334}$$

We have already calculated the derivative $\partial r/\partial x_k$ in Eq. (3.319); the result is x_k/r. The source of the field is only at the origin (a point charge); everywhere else, the divergence is zero, and the field only passes through these points. The infinite value of the divergence is related to the fact that we consider the charge to be point-like; therefore, the charge density $\rho_Q = \Delta Q/\Delta V$ is infinite at the origin. This is analogous to the concept of a mass point. These issues are formally addressed by the theory of distributions (generalized functions), which we will explore in Chapter 3.8.

Coulomb law

Maxwell equations contain two terms involving divergence:

$$\operatorname{div}\mathbf{B} = 0\,, \tag{3.335}$$

$$\operatorname{div}\mathbf{D} = \rho_Q\,. \tag{3.336}$$

The first equation states that the magnetic field only passes through each point in space; there are no sources of it anywhere. Magnetic monopoles do not exist. According to some theories, magnetic monopoles were so diluted during the inflationary phase of the universe that so few remained in the region we can observe (we cannot see the entire universe) that we cannot detect them. The second equation describes how electric fields emerge in regions with a positive charge and vanish in regions with a negative charge. Let's imagine we have just a single isolated charge. We choose coordinates with the origin at the charge and draw a spherical surface around the charge. We integrate equation (3.336) over the volume of the resulting sphere as follows:

$$\iiint\limits_V \operatorname{div}\mathbf{D}\, \mathrm{d}V = \iiint\limits_V \rho_Q\, \mathrm{d}V\,. \tag{3.337}$$

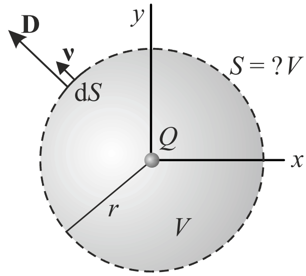


Fig. 3.28: Volume and area of integration

Using Gauss theorem, we convert the integral on the left to a surface integral. The integral on the right gives the total charge of the enclosed particle:

$$\oiint_{S=\partial V} \mathbf{D}\cdot \mathrm{d}\mathbf{S} = Q. \tag{3.338}$$

The electric induction vector and the surface element vector have the same direction; the angle between them is zero, and the cosine of their dot product is equal to one:

$$\oiint_{S=\partial V} D\,\mathrm{d}S = Q. \tag{3.339}$$

The magnitude of D is the same over the entire surface of the sphere, so we can factor it out of the integral. The remaining integral gives the total surface area of the sphere:

$$D\,4\pi r^2 = Q. \tag{3.340}$$

If we convert the electric induction to the electric field intensity, we obtain the resulting integral relation, which is nothing other than Coulomb law:

$$E = \frac{Q}{4\pi\varepsilon_0 r^2}. \tag{3.341}$$

3.5.3 Rotation (Curl)

Let's now move on to another task. We will test whether our vector field creates a vortex around a selected point. If you imagine a vortex drawn on a piece of paper, you will only see it from a direction perpendicular to the paper. If you look at it from the plane of the paper, you won't be able to detect it. To detect a vortex, three independent views from three different directions are required. Therefore, the vortex test will be of a vector nature. A useful tool for our testing will be the *circulation* element of the field:

$$\mathrm{d}C \equiv \mathbf{K}\cdot \mathrm{d}\mathbf{l}, \tag{3.342}$$

where **K** is the field and **dl** is an element of the curve. If the field flows in the direction of the curve, the circulation will be $K\mathrm{d}l$; if opposite, the circulation will be $-K\mathrm{d}l$. If the field flows perpendicular to the curve, the circulation will be zero.

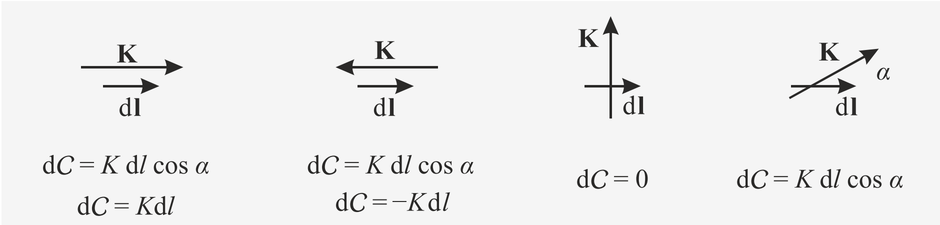


Fig. 3.29: Field circulation element

Let's now imagine a point in space around which we draw a generally oriented rectangle and project both the point and the rectangle onto all three coordinate planes. In the (x, y) plane, the situation will look like this:

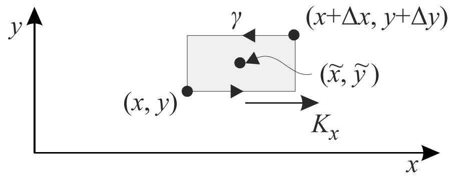


Fig. 3.30: Calculation of the circulation of a vector field around a rectangle

Now we find the line integral of the field around this rectangle along the positively oriented curve γ formed by the edges of the rectangle (the points labels are in the Figure). We write the contributions in the following order: bottom, right, top, and left edges; we take the field at the midpoint of the edge. The edges have lengths Δx, Δy:

$$\Delta C = \sum_{a=1}^{4} \Delta C_a = K_x(\tilde{x}, y)\,\Delta x + K_y(x+\Delta x, \tilde{y})\,\Delta y - K_x(\tilde{x}, y+\Delta y)\,\Delta x - K_y(x, \tilde{y})\,\Delta y =$$

$$= \left[\frac{K_y(x+\Delta x, \tilde{y}) - K_y(x, \tilde{y})}{\Delta x} - \frac{K_x(\tilde{x}, y+\Delta y) - K_x(\tilde{x}, y}{\Delta y}\right]\Delta x\,\Delta y\,.$$

Now we take the limit as Δx and Δy approach 0, and we obtain (we denote the surface element with a normal in the direction of the z-axis as $\mathrm{d}S_z \equiv \mathrm{d}x\,\mathrm{d}y$):

$$\mathrm{d}C = \left[\frac{\partial K_y}{\partial x} - \frac{\partial K_x}{\partial y}\right]\mathrm{d}S_z\,. \tag{3.343}$$

If we superimpose all three projections, we obtain a general rectangle, and the circulation around it will be (the other terms can be obtained by cyclic permutation)

$$\mathrm{d}C = \left[\frac{\partial K_z}{\partial y} - \frac{\partial K_y}{\partial z}\right]\mathrm{d}S_x + \left[\frac{\partial K_x}{\partial z} - \frac{\partial K_z}{\partial x}\right]\mathrm{d}S_y + \left[\frac{\partial K_y}{\partial x} - \frac{\partial K_x}{\partial y}\right]\mathrm{d}S_z\,. \tag{3.344}$$

It is clear that the terms in square brackets are the components of the vector product of the gradient and our field; therefore, for an element of the field's circulation, we have:

► $$\mathrm{d}C \equiv \sum_a \mathbf{K}_a \cdot \mathrm{d}\mathbf{l}_a = (\nabla \times \mathbf{K}) \cdot \mathrm{d}\mathbf{S}\,. \tag{3.345}$$

The non-zero circulation and the presence of a vortex are determined by the vector product of the gradient and the field, which we call the field's rotation:

► $$\operatorname{rot}\mathbf{K} \equiv \nabla \times \mathbf{K}\,. \qquad (3.346)$$

The individual components of the rotation correspond to our view of the vortex along the coordinate axes. If all components of the rotation are zero, the field does not form a vortex around the point. If any component is nonzero, there is a vortex around the point:

► $$\operatorname{rot}\mathbf{K}\begin{cases} = (0,0,0): & \text{pole netvoří vír,} \\ \neq (0,0,0): & \text{pole tvoří vír.} \end{cases} \qquad (3.347)$$

Now let's determine the rotation of the fields **A**, **B**, **C**, and **D** from the equation (3.330):

$$\operatorname{rot}\mathbf{A} = (0,0,2)\,; \quad \operatorname{rot}\mathbf{B} = \operatorname{rot}\mathbf{C} = \operatorname{rot}\mathbf{D} = (0,0,0)\,. \qquad (3.348)$$

Only the first field has a non-zero rotation; the vortex is visible only when viewed along the z-axis. This is a non-physical vortex whose intensity increases with distance from the center (at every point in space, there is a sort of blower that amplifies the vortex). Therefore, the rotation is non-zero at all points in space.

Let us now consider a finite area S bounded by a closed curve γ. Our task is to calculate the circulation of the field along this curve. We will fill the area bounded by the curve with many small rectangles that are adjacent to one another. Each rectangle represents, mathematically, a positively oriented closed curve. At the shared edges, the circulation of the field will always be zero, since the edges are oppositely oriented. The only non-zero circulation will be at the edges adjacent to the boundary of our region. By taking the limit from (3.345), we have

► $$\oint_{\gamma=\partial S} \mathbf{K}\cdot \mathrm{d}\mathbf{l} = \iint_S (\operatorname{rot}\mathbf{K})\cdot \mathrm{d}\mathbf{S}\,, \qquad (3.349)$$

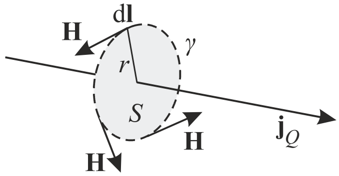
which is Stokes theorem, which converts a surface integral into a line integral.

Ampère law

One of Maxwell's equations takes the following form for time-independent fields:

$$\operatorname{rot}\mathbf{H} = \mathbf{j}_Q\,, \qquad (3.350)$$

where $\mathbf{j}_Q$ is the current density (charge flux, i.e., the amount of charge passing through a unit area per unit time; the unit is $\mathrm{A/m^2}$). The meaning of the equation is clear: a magnetic field vortex forms around conductors carrying current. Let's draw a circle around a point on the conductor; we'll denote the area enclosed by this circle as S:



$$\int_a^b f' g\,\mathrm{d}x + \int_a^b f g'\,\mathrm{d}x = [f g]_a^b$$

Fig. 3.31: Area and curve of integration

We will integrate the equation (3.350) over this area:

$$\iint_S (\operatorname{rot}\mathbf{H})\cdot \mathrm{d}\mathbf{S} = \iint_S \mathbf{j}_Q \cdot \mathrm{d}\mathbf{S} \,. \tag{3.351}$$

The integral on the right gives the total current flowing through the surface, i.e., the current flowing through the conductor. We convert the integral on the left into an integral over the boundary using Stokes' theorem:

$$\oint_{\gamma=\partial S} \mathbf{H}\cdot \mathrm{d}\mathbf{l} = I \,. \tag{3.352}$$

Both the field and the curve element have the same direction, so the scalar product is equal to the product of the magnitudes of both vectors, i.e.,

$$\oint_{\gamma=\partial S} H\,\mathrm{d}l = I \,. \tag{3.353}$$

Field H has the same magnitude along the entire circumference of the circle, so we factor it out of the integral. This gives the circumference of the circle:

$$H\,2\pi r = I \,. \tag{3.354}$$

After converting to magnetic induction, we obtain Ampère law in its well-known form

$$B = \frac{\mu_0 I}{2\pi r} \,. \tag{3.355}$$

A field generated by a linear object decreases with distance as $1/r$. Fields generated by point sources (such as Coulomb's law) decrease with distance as $1/r^2$. Maxwell's equation (3.350) is the differential form of Ampère law.

3.5.4 Helicity

In nature, we often observe structures twisted into helices. Whether it be DNA molecules or simply a plasma discharge that had enough time to twist into a helix after a while. Plasma theory shows (see [2]) that the helical structure of a magnetic field has minimum energy. Therefore, for example, in a tokamak, the poloidal field generated by the flowing current is not sufficient; a toroidal field generated by auxiliary coils must also be present. The resulting field has a helical shape, is in a state of minimum energy, and is therefore relatively stable. In mathematics, the concept of helicity is introduced for similarly structured fields. The helicity density of a vector field **K** is defined as

► $$\mathcal{H}(t,\mathbf{x}) \equiv \mathbf{K}\cdot \operatorname{rot}\mathbf{K} \,. \tag{3.356}$$

We then define total helicity as the integral

$$G(t) = \iiint_V \mathcal{H}(t,\mathbf{x})\,\mathrm{d}V \,. \tag{3.357}$$

Helicity is a scalar quantity that characterizes the helicity of a vector field. It is zero for all fields satisfying the no-vortex condition (rot **K** = 0) and also for all planar vortices with circular or spiral streamlines. Fields with non-zero helicity must form spatial helices with a non-zero pitch α; the helicity is proportional to sin α. An example of the introduction of helicity density can be the velocity field

$$\mathcal{H}_u = \mathbf{u} \cdot \operatorname{rot} \mathbf{u} = \mathbf{u} \cdot \boldsymbol{\omega} . \tag{3.358}$$

We call the quantity $\boldsymbol{\omega} \equiv \operatorname{rot} \mathbf{u}$ the vorticity of the velocity field. When describing the helicity of a magnetic field, the magnetic potential $\mathbf{A}$ is used, which is related to the field itself by the equation $\mathbf{B} = \operatorname{rot} \mathbf{A}$:

$$\mathcal{H}_A = \mathbf{A} \cdot \operatorname{rot} \mathbf{A} = \mathbf{A} \cdot \mathbf{B} . \tag{3.359}$$

We will deal with the helicity of a magnetic field in detail only in the third volume of this textbook, see [2], devoted to plasma physics, where the concept of helicity plays a very important role.

3.6 Multidimensional Integrals

In many situations, it is necessary to integrate a quantity – usually a scalar or vector – along a curve, over a surface, or over the volume of a body. To this aim, we introduce an element (an infinitesimally small part) of the given set. A line element can be either scalar or vector, and a surface element can be either as well, but a volume element exists only as a scalar; volume, as we understand it in three dimensions, has no direction.

3.6.1 Line Integral

First, let's consider the curve, which we'll denote by the symbol γ. We'll start by parameterizing the curve; a single parameter, u, is sufficient for this:

$$\gamma: \begin{cases} x = x(u)\,; \\ y = y(u)\,; \\ z = z(u)\,. \end{cases} \tag{3.360}$$

We can write this in abbreviated form

► $$\gamma:\ \mathbf{r} = \mathbf{r}(u)\,. \tag{3.361}$$

The parameter u ranges from u_1 to u_2. The point $\mathbf{r}(u_1)$ is the starting point of the curve, and $\mathbf{r}(u_2)$ is the endpoint. We will now divide the curve into a large number of segments and refine this division until we arrive at an infinitesimal element of the curve

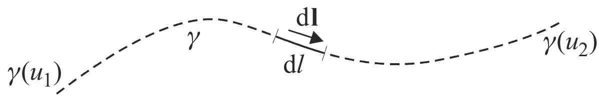


Fig. 3.32: Integration curve

► $$\mathrm{d}\mathbf{l} = (\mathrm{d}x, \mathrm{d}y, \mathrm{d}z) = (x'\mathrm{d}u, y'\mathrm{d}u, z'\mathrm{d}u)\,. \tag{3.362}$$

A curve element is vector pointing in the direction of the curve; therefore, we call it a *vector element*. We can also define a scalar element as the magnitude of the element:

► $$\mathrm{d}l \equiv \sqrt{\mathrm{d}\mathbf{l}\cdot\mathrm{d}\mathbf{l}} = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2} = \sqrt{x'^2 + y'^2 + z'^2}\,\mathrm{d}u\,, \tag{3.363}$$

where the prime denotes the derivative with respect to the parameter u. Using these elements, we can introduce two types of line integrals – the first-kind line integral sums a scalar field over a scalar element along the curve, while the second-kind line integral sums the projection of a vector field in the direction of the curve (represented by the scalar product) along the curve:

► $$I_1 = \int_\gamma f(\mathbf{r})\,\mathrm{d}l, \tag{3.364}$$

► $$I_2 = \int_\gamma \mathbf{K}(\mathbf{r})\cdot\mathrm{d}\mathbf{l}\,. \tag{3.365}$$

We can use integrals of the first kind, for example, to determine the mass of a fiber with a varying cross-section or to determine the length of a curve (in this case, $f = 1$). Integrals of the second kind are used to calculate mechanical work (the field will be the applied force), electric potential (the field will be the electric field intensity), and other quantities. Integrals of both kinds can be easily converted to a standard Riemann integral: We substitute the parametric representation of the curve into both fields (scalar or vector) and express the corresponding element in terms of the differentials of the curve's parametric representation. This converts the entire integration into an integral over the curve's parameter, i.e., a standard one-dimensional Riemann integral:

$$I_1 = \int_\gamma f(\mathbf{r})\,\mathrm{d}l = \int_\gamma f(\mathbf{r})\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2} = \int_{u_1}^{u_2} f\left(\mathbf{r}(u)\right)\sqrt{x'^2 + y'^2 + z'^2}\,\mathrm{d}u\,. \tag{3.366}$$

If we choose x directly as the parameter, we have

$$I_1 = \int_\gamma f(\mathbf{r})\,\mathrm{d}l == \int_{x_1}^{x_2} f\left(\mathbf{r}(x)\right)\sqrt{1 + y'^2 + z'^2}\,\mathrm{d}x\,. \tag{3.367}$$

For the length of a plane curve $y(x)$, the relationship simplifies to the formula

$$l = \int_{x_1}^{x_2} \sqrt{1 + y'^2}\,\mathrm{d}x\,. \tag{3.368}$$

For a line integral of the second kind, we proceed in a similar manner

$$\begin{aligned} I_2 &= \int_\gamma \mathbf{K}(\mathbf{r})\cdot \mathrm{d}\mathbf{l} = \int_\gamma K_x \mathrm{d}x + K_y \mathrm{d}y + K_z \mathrm{d}z = \\ &= \int_{u_1}^{u_2} \left[K_x\left(\mathbf{r}(u)\right)x' + K_y\left(\mathbf{r}(u)\right)y' + K_z\left(\mathbf{r}(u)\right)z'\right]\mathrm{d}u = \\ &= \int_{u_1}^{u_2} \mathbf{K}\left(\mathbf{r}(u)\right)\cdot \mathbf{r}'\,\mathrm{d}u\,. \end{aligned} \tag{3.369}$$

◖ Example 3.41: The length of the sine wave from 0 to 2π

We will parameterize the curve $y = \sin(x)$ using the variable x, i.e., we will identify x with the curve's parameter u. For the integration, we will use equation (3.368):

$$\begin{gathered} y(x) = \sin(x)\,; \\ l = \int_0^{2\pi} \sqrt{1 + y'^2}\,\mathrm{d}x = \int_0^{2\pi} \sqrt{1 + \cos^2(x)}\,\mathrm{d}x\,. \end{gathered} \tag{3.370}$$

Integrals of this type are rarely solvable analytically; in most cases, it is necessary to use numerical methods. This case leads to what is known as an elliptic integral. However, there is a simple way to quickly find the result of the integral. On the Wolfram Alpha website, enter the integration using the string "integrate sqrt(1+cos(x)^2) dx from 0 to

2pi". The program will attempt to calculate the integral analytically; if that fails, it will use a numerical method. The result is 7.6404.

◗

◖ Example 3.42: Length of the cable on the Golden Gate Bridge in San Francisco

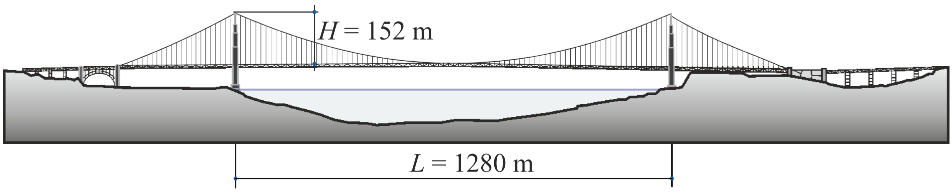


Fig. 3.33: The chain bridge in San Francisco

We will calculate the length of the cable between the supports. Let's assume that the cable is approximately shaped like a catenary ch(x). We will take the span between the columns, L = 1280 meters, as the unit of length and shift the catenary so that the origin of the coordinates is at the lowest point; that is, the rough shape will be given by the function ch(x/L) − 1. Now we will stretch the catenary to the bridge's parameters:

$$y(x) = A\left[\operatorname{ch}(x/L) - 1\right] \tag{3.371}$$

It follows from the condition $y(\pm L/2) = H$ that

$$A = \frac{H}{\operatorname{ch}(1/2) - 1} = 1190{,}98 \text{ m} . \tag{3.372}$$

Now we can easily determine the length of the cable:

$$l = \int_{-L/2}^{L/2} \sqrt{1 + y'^2}\, \mathrm{d}x = \int_{-L/2}^{L/2} \sqrt{1 + (A/L)^2 \operatorname{sh}^2(x/L)}\, \mathrm{d}x . \tag{3.373}$$

Let's make the substitution $\xi = x/L$:

$$l = L \int_{-1/2}^{1/2} \sqrt{1 + (A/L)^2 \operatorname{sh}^2(\xi)}\, \mathrm{d}\xi . \tag{3.374}$$

The integration string is " 1280*integrate sqrt(1+0.8672 sinh(x)^2) dx from -0.5 to 0.5", which yields 1,327 meters. This refers only to the sections of cable between the pylons.

◗

◖ Example 3.43: Work done by a spatial oscillator field

Let's assume the course of potential energy

$$W_P = \frac{1}{2} k r^2 = \frac{1}{2} k \left(x^2 + y^2\right) . \tag{3.375}$$

We can easily determine the force acting on it: $\mathbf{F} = -\nabla W_\mathrm{P}$:

$$F_x = -kx , \qquad F_y = -ky . \tag{3.376}$$

In any direction, the system behaves like a harmonic oscillator. No matter where we displace the body, a restoring force will act on it toward the origin, and this force will be greater the farther the body is from the origin. Our task will be to calculate the work done when moving the body between points $A = (a, 0)$ a $B = (0, b)$:

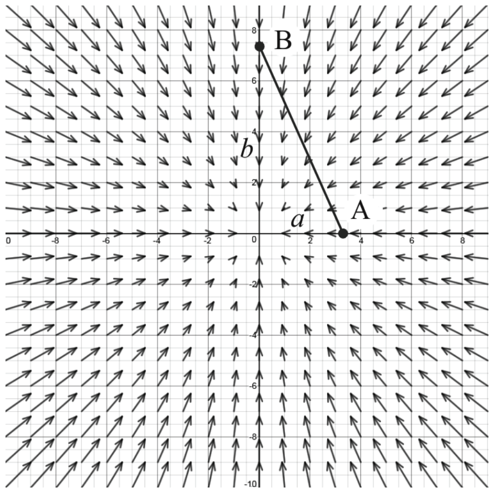


Fig. 3.34: Spatial oscillator

First, we find the parametric equation of the curve and its vector element:

$$X = A + (B - A)\,t\,; \qquad t \in \langle 0, 1 \rangle\,; \quad A = (a, 0); \quad B = (0, b)\,. \tag{3.377}$$

$$\begin{aligned} x &= a - at\,, \qquad & \mathrm{d}x - a\,\mathrm{d}t\,, \\ y &= bt, \qquad & \mathrm{d}y = b\,\mathrm{d}t. \end{aligned} \tag{3.378}$$

The role of the curve's parameter here is played by t, i.e., $u = t$. Now the calculation of the work done is straightforward; we just need to substitute the relations for all fields and differentials from (3.378):

$$\begin{aligned} \Delta A &= \int_\gamma \mathbf{F} \cdot \mathrm{d}\mathbf{l} = \int_\gamma F_x \mathrm{d}x + F_y \mathrm{d}y = \int_\gamma -kx\,\mathrm{d}x - ky\,\mathrm{d}y = \\ &= \int_0^1 -k(a - at)(-a\,\mathrm{d}t) - (kbt)(b\,\mathrm{d}t) = \int_0^1 \left[+ka^2 - ka^2 t + kb^2 t \right] \mathrm{d}t = \\ &= \frac{1}{2} k a^2 - \frac{1}{2} k b^2\,. \end{aligned} \tag{3.379}$$

Another solution: Since potential energy appeared in the problem statement, we know that the field is conservative. We can then determine the work done from the difference in potential energy:

$$\Delta A = -\Delta W_\mathrm{P} = W_\mathrm{P}(A) - W_\mathrm{P}(B) = \frac{1}{2} k a^2 - \frac{1}{2} k b^2. \tag{3.380}$$

◗

3.6.2 Surface and Volume Integrals

Surface integrals

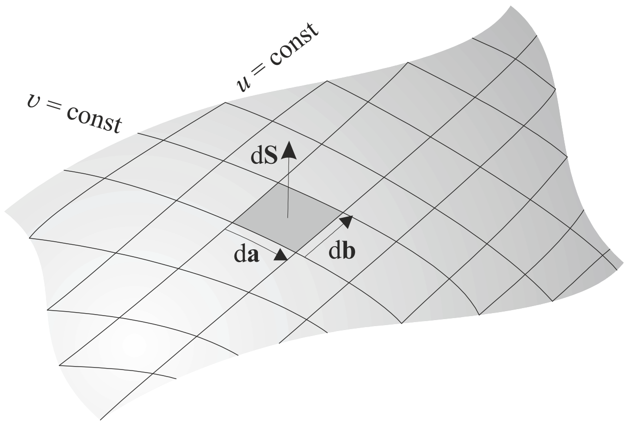


Fig. 3.34: Surface element for calculating a surface integral

We describe the surface with parameters, *u* and *v*, which form a coordinate grid on it:

$$\sigma : \begin{cases} x = x(u,v)\ ; \\ y = y(u,v)\ ; \\ z = z(u,v)\ . \end{cases} \tag{3.381}$$

For brevity, we can write

► $$\sigma : \quad \mathbf{r} = \mathbf{r}(u,v). \tag{3.382}$$

The surface element shown in the figure will have edges

$$\mathrm{d}\mathbf{a} \cong \mathbf{r}(u+\mathrm{d}u,v) - \mathbf{r}(u,v) \cong \frac{\partial \mathbf{r}}{\partial u}\mathrm{d}u\ ; \qquad \mathrm{d}\mathbf{b} \cong \mathbf{r}(u,v+\mathrm{d}v) - \mathbf{r}(u,v) \cong \frac{\partial \mathbf{r}}{\partial v}\mathrm{d}v\ . \tag{3.383}$$

The vector element of the surface is given by the vector product, and the scalar element is its magnitude:

► $$\mathrm{d}\mathbf{S} = \left(\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right)\mathrm{d}u\,\mathrm{d}v\ ; \tag{3.384}$$

► $$\mathrm{d}S = \left\|\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}\right\|\mathrm{d}u\,\mathrm{d}v\ ; \tag{3.385}$$

Now, similarly to line integrals, we introduce the first-kind surface integral (the sum of a scalar field over a scalar element) and the second-kind surface integral (the flux of a vector field through the surface):

► $$I_1 \equiv \iint_\sigma f\,\mathrm{d}S\,, \tag{3.386}$$

► $$I_2 \equiv \iint_\sigma \mathbf{K}\cdot\mathrm{d}\mathbf{S} \tag{3.387}$$

We use the first kind to calculate charge, mass, or area, and the second kind to calculate the flux of a field through a surface.

◖ **Example 3.44: Circle and surface integral**. Let's parameterize the circle $x^2+y^2 \le R^2$ using polar coordinates:

$$\begin{aligned} x &= r\cos\varphi\,,\\ y &= r\sin\varphi\,,\\ z &= 0. \end{aligned} \tag{3.388}$$

The parameters take the values

$$\begin{aligned} u &= r \in <0, R>;\\ v &= \varphi \in <0, 2\pi). \end{aligned} \tag{3.389}$$

Let us now determine the vector surface element:

$$\mathrm{d}\mathbf{S} = \frac{\partial \mathbf{r}}{\partial r}\times\frac{\partial \mathbf{r}}{\partial \varphi}\,\mathrm{d}r\,\mathrm{d}\varphi = (\cos\varphi, \sin\varphi, 0)\times(-r\sin\varphi, r\cos\varphi, 0)\,\mathrm{d}r\,\mathrm{d}\varphi \qquad \Rightarrow$$

$$\mathrm{d}\mathbf{S} = (0, 0, r\ \mathrm{d}r\,\mathrm{d}\varphi) \tag{3.390}$$

The vector element points only along the z-axis; thus, as expected, it is perpendicular to the plane (x, y) in which the circle lies. The scalar element will be its magnitude, i.e.,

$$\mathrm{d}S = r\ \mathrm{d}r\,\mathrm{d}\varphi \tag{3.391}$$

Now we can easily determine the area of the circle

$$S = \iint_\sigma \mathrm{d}S = \int_{\varphi=0}^{\varphi=2\pi}\int_{r=0}^{r=R} r\ \mathrm{d}r\,\mathrm{d}\varphi = \frac{R^2}{2}2\pi = \pi R^2\,. \tag{3.392}$$

Let us now find the flux of the field

$$\mathbf{K} = \left(\alpha y, \beta z, \gamma x^2\right) \tag{3.393}$$

through this circle. The calculation is straightforward:

$$\begin{aligned} I_2 &= \iint_\sigma \mathbf{K}\cdot\mathrm{d}\mathbf{S} = \iint_\sigma K_x \mathrm{d}S_x + K_y \mathrm{d}S_y + K_z \mathrm{d}S_z = \\ &= \iint_\sigma \left(0 + 0 + \gamma x^2 \mathrm{d}S_z\right) = \int_{\varphi=0}^{\varphi=2\pi}\int_{r=0}^{r=R}\left(\gamma r^2\cos^2\varphi\right)\left(r\,\mathrm{d}r\,\mathrm{d}\varphi\right) = \\ &= \gamma\frac{R^4}{4}\frac{2\pi}{2} = \gamma\pi\frac{R^4}{4}. \end{aligned} \tag{3.394}$$

◗

Volume integral

It remains to introduce the volume integral, which has a single element, namely a scalar. We describe the volume set Ω using three parameters::

$$\Omega: \begin{cases} x = x(u,v,w)\,; \\ y = y(u,v,w)\,; \\ z = z(u,v,w)\,. \end{cases} \tag{3.395}$$

For brevity, we can write

► $$\Omega: \quad \mathbf{r} = \mathbf{r}(u,v,w)\,. \tag{3.396}$$

We define the volume element as shown in the figure

$\mathrm{d}V = \mathrm{dS\;dc}\cos\alpha$
$\mathrm{d}V = |\mathrm{d}\mathbf{S}\cdot\mathrm{d}\mathbf{c}|$
$\mathrm{d}V = |(\mathrm{d}\mathbf{a}\times\mathrm{d}\mathbf{b})\cdot\mathrm{d}\mathbf{c}|$

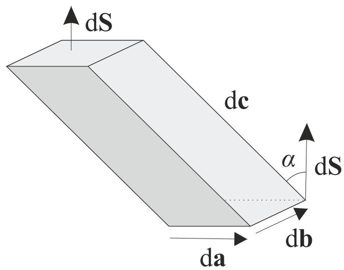


Fig. 3.35: Volume element

► $$\mathrm{d}V = \left|\left(\frac{\partial \mathbf{r}}{\partial u}\times\frac{\partial \mathbf{r}}{\partial v}\right)\cdot\frac{\partial \mathbf{r}}{\partial w}\right|\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w; \tag{3.397}$$

And we will understand the volume integral as

► $$I = \iiint_V f(\mathbf{r})\,\mathrm{d}V. \tag{3.398}$$

Basic types of integrals:

► $\int_\gamma f\,\mathrm{d}l\,;$ $\quad \gamma : \mathbf{r} = \mathbf{r}(u)\,;$ $\quad \mathrm{d}l = \sqrt{x'^2 + y'^2 + z'^2}\,\mathrm{d}u\,;$ (3.399)

► $\int_\gamma \mathbf{K}\cdot\mathrm{d}\mathbf{l}\,;$ $\quad \gamma : \mathbf{r} = \mathbf{r}(u)\,;$ $\quad \mathrm{d}\mathbf{l} = (x'\mathrm{d}u, y'\mathrm{d}u, z'\mathrm{d}u);$ (3.400)

► $\iint_\sigma f\,\mathrm{d}S$ $\quad \sigma : \mathbf{r} = \mathbf{r}(u,v)\,;$ $\quad \mathrm{d}S = \left\|\frac{\partial \mathbf{r}}{\partial u}\times\frac{\partial \mathbf{r}}{\partial v}\right\|\mathrm{d}u\,\mathrm{d}v\,;$ (3.401)

► $\iint_\sigma \mathbf{K}\cdot\mathrm{d}\mathbf{S}$ $\quad \sigma : \mathbf{r} = \mathbf{r}(u,v)\,;$ $\quad \mathrm{d}\mathbf{S} = \left(\frac{\partial \mathbf{r}}{\partial u}\times\frac{\partial \mathbf{r}}{\partial v}\right)\mathrm{d}u\,\mathrm{d}v\,;$ (3.402)

► $\iiint_\Omega f\,\mathrm{d}V$ $\quad \Omega : \mathbf{r} = \mathbf{r}(u,v,w)\,;$ $\quad \mathrm{d}V = \left|\left(\frac{\partial \mathbf{r}}{\partial u}\times\frac{\partial \mathbf{r}}{\partial v}\right)\cdot\frac{\partial \mathbf{r}}{\partial w}\right|\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}w;$ (3.403)

3.6.3 Integration by Parts in *N* Dimensions

Let's first review the situation in one dimension. For the derivative of the product of two functions, we have

$$f'g + fg' = (fg)' . \tag{3.404}$$

By integrating this relation, we immediately obtain

$$\int_a^b f'g\,\mathrm{d}x + \int_a^b fg'\,\mathrm{d}x = [fg]_a^b, \qquad \text{or} \tag{3.405}$$

$$\int_a^b f'g\,\mathrm{d}x + \int_a^b fg'\,\mathrm{d}x = +f(b)g(b) - f(a)g(a) \tag{3.406}$$

The signs on the right-hand side can be interpreted as one-dimensional outer normals to the interval <*a*, *b*>, that is,

$$\int_a^b f'g\,\mathrm{d}x + \int_a^b fg'\,\mathrm{d}x = \nu_b f(b)g(b) + \nu_a f(a)g(a) . \tag{3.407}$$

Both expressions on the right are actually integrals over the boundary of the set Ω = <*a*, *b*>, which consists of only two points. We can therefore formally rewrite it as

$$\int_\Omega f'g\,\mathrm{d}^1x + \int_\Omega fg'\,\mathrm{d}^1x = \int_{\partial\Omega} \nu fg\,\mathrm{d}^0x . \tag{3.408}$$

The index above the differential determines the dimension of the integration. The integration on the right is formal; it takes place over only two points, so it is a summation. In *N* dimensions, similar relations hold:

$$\frac{\partial f}{\partial x_k} g + f\frac{\partial g}{\partial x_k} = \frac{\partial}{\partial x_k}(fg) , \tag{3.409}$$

► $$\int_\Omega \left(\frac{\partial f}{\partial x_k} g + f\frac{\partial g}{\partial x_k} \right) \mathrm{d}^N x = \int_{\partial\Omega} fg\nu_k\,\mathrm{d}^{N-1}x . \tag{3.410}$$

The last expression is the integration by parts theorem in *N* dimensions. It allows to transfer the derivative from one function to another. For $g = 1$, the relation yields

► $$\int_\Omega \left(\frac{\partial f}{\partial x_k} \right) \mathrm{d}^N x = \int_{\partial\Omega} f\nu_k\,\mathrm{d}^{N-1}x . \tag{3.411}$$

Let us now derive Gauss's theorem from this relation as a special case of integration by parts in three dimensions:

$$\iiint_V \frac{\partial f}{\partial x_k}\,\mathrm{d}V = \oiint_{S=\partial V} f\nu_k\,\mathrm{d}S . \tag{3.412}$$

Now, for the function f, we take the component of a vector field, i.e., $f = K_k$, and sum over the index k:

$$\iiint_V \frac{\partial K_k}{\partial x_k}\,\mathrm{d}V = \oiint_{S=\partial V} K_k \nu_k \,\mathrm{d}S. \tag{3.413}$$

The above relation can easily be rewritten as

$$\iiint_V (\operatorname{div}\mathbf{K})\,\mathrm{d}V = \oiint_{S=\partial V} (\mathbf{K}\cdot\boldsymbol{\nu})\,\mathrm{d}S. \tag{3.414}$$

Introducing the vector element $\mathrm{d}\mathbf{S} = \boldsymbol{\nu}\,\mathrm{d}S$, we have Gauss theorem in its standard form

$$\iiint_V (\operatorname{div}\mathbf{K})\,\mathrm{d}V = \oiint_{S=\partial V} \mathbf{K}\cdot\mathrm{d}\mathbf{S}. \tag{3.415}$$

3.6.4 Exterior Algebra

A unified approach to various types of integrals can be obtained by introducing the so-called exterior algebra, which is based on the antisymmetric product of differentials, e.g.

$$\mathrm{d}x \wedge \mathrm{d}y = -\mathrm{d}y \wedge \mathrm{d}x\,. \tag{3.416}$$

This operation mirrors the properties of the vector product of basis vectors; we regard the differentials themselves as elements of a linear vector space, i.e., we can “stretch” and “add” them. It is clear that the newly introduced product of two identical differentials must yield zero, for example

$$\mathrm{d}x \wedge \mathrm{d}x = 0\,. \tag{3.417}$$

◖ **Example 3.45: An illustration of a two-element exterior algebra**

On the two elements $\mathbf{e}_1$ and $\mathbf{e}_2$, we can introduce an external algebra by adding the identity element $\mathbf{e}_0$ (which does not change the element when multiplied) and the result of the external product $\mathbf{e}_{12}$. We can simply derive the multiplication of all elements, e.g.

$$\begin{gathered}\mathbf{e}_2 \wedge \mathbf{e}_1 = -\mathbf{e}_1 \wedge \mathbf{e}_2 = \mathbf{e}_{12}\;; \\ \mathbf{e}_{12} \wedge \mathbf{e}_1 = (\mathbf{e}_1 \wedge \mathbf{e}_2) \wedge \mathbf{e}_1 = -(\mathbf{e}_2 \wedge \mathbf{e}_1) \wedge \mathbf{e}_1 = -\mathbf{e}_2 \wedge (\mathbf{e}_1 \wedge \mathbf{e}_1) = 0\end{gathered} \tag{3.418}$$

and so on. The result of multiplying the individual elements can be written in a table

$\wedge$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_{12}$
$\mathbf{e}_0$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_{12}$
$\mathbf{e}_1$	$\mathbf{e}_1$	0	$\mathbf{e}_{12}$	0
$\mathbf{e}_2$	$\mathbf{e}_2$	$-\mathbf{e}_{12}$	0	0
$\mathbf{e}_{12}$	$\mathbf{e}_{12}$	0	0	0

The algebra being introduced will be a linear span of these elements. ◗

◖ **Example 3.46: Use of the exterior algebra in integrals**

The general form of the relation (3.411) can be understood as

$$\int_{\Omega} \mathrm{d}F = \int_{\partial\Omega} F \,; \qquad \text{spec. for one dimension} \int_{<a,b>} \mathrm{d}F = \left[F\right]_a^b . \tag{3.419}$$

The integral on the right hand side refers to values on the boundary of the region; therefore, we do not write the differential sign here, just as in one dimension, where the boundary consists of two points. However, the expression F itself may contain differentials. For Stokes' theorem, F on the right hand side is given by the relation (we calculate the circulation of the vector $\mathbf{K}$ along γ)

$$F = \mathbf{K}\cdot \mathrm{d}\mathbf{l} = K_x \mathrm{d}x + K_y \mathrm{d}y + K_z \mathrm{d}z \,. \tag{3.420}$$

If we define the differential using the exterior product, we have for the left-hand side:

$$\mathrm{d}F = \mathrm{d}K_x \wedge \mathrm{d}x + \mathrm{d}K_y \wedge \mathrm{d}y + \mathrm{d}K_z \wedge \mathrm{d}z = \left(\frac{\partial K_x}{\partial x}\mathrm{d}x + \frac{\partial K_x}{\partial y}\mathrm{d}y + \frac{\partial K_x}{\partial z}\mathrm{d}z\right)\wedge \mathrm{d}x + \cdots =$$

$$= \left(\frac{\partial K_z}{\partial y} - \frac{\partial K_y}{\partial z}\right)\mathrm{d}y\wedge \mathrm{d}z + \left(\frac{\partial K_x}{\partial z} - \frac{\partial K_z}{\partial x}\right)\mathrm{d}z\wedge \mathrm{d}x + \left(\frac{\partial K_y}{\partial x} - \frac{\partial K_x}{\partial y}\right)\mathrm{d}x\wedge \mathrm{d}y =$$

$$= (\operatorname{rot}\mathbf{K})_x \mathrm{d}S_x + (\operatorname{rot}\mathbf{K})_y \mathrm{d}S_y + (\operatorname{rot}\mathbf{K})_z \mathrm{d}S_z = \operatorname{rot}\mathbf{K}\cdot \mathrm{d}\mathbf{S} \,.$$

It is clear that Stokes' theorem is a special case of the relation (3.419)

$$\int_S \operatorname{rot}\mathbf{K}\cdot \mathrm{d}\mathbf{S} = \int_{\partial S} \mathbf{K}\cdot \mathrm{d}\mathbf{l} \,. \tag{3.421}$$

The same procedure can be followed for the other integrals. ◗

3.6.5 Measure and Metric

Suppose we have a metric g_{kl} on a given N-dimensional space, i.e., the distance in generalized coordinates q_k is given by the relation

► $$\mathrm{d}l^2 = g_{kl}\, \mathrm{d}q_k\, \mathrm{d}q_l \,. \tag{3.422}$$

Then a measure element (line element, surface element, volume element, etc.) can generally be written as (the reader can find the proof in the specialized literature)

► $$\mathrm{d}\mu = \sqrt{\left|\det g\right|}\, \mathrm{d}q_1 \ldots \mathrm{d}q_N \,. \tag{3.423}$$

All integrals of the first kind can then be written uniformly in the form

► $$\int_{\Omega} f\, \mathrm{d}\mu = \int_{\Omega} f \sqrt{\left|\det g\right|}\, \mathrm{d}q_1 \ldots \mathrm{d}q_N \,. \tag{3.424}$$

◖ **Example 3.47: Determine the volume of a sphere in three dimensions.**
For the spherical metric, we have:

$$\mathrm{d}l^2 = \mathrm{d}r^2 + r^2 \sin^2\theta \, \mathrm{d}\varphi^2 + r^2 \mathrm{d}\theta^2 , \tag{3.425}$$

so the metric tensor is

$$g_{kl} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 \sin^2\theta & 0 \\ 0 & 0 & r^2 \end{pmatrix} . \tag{3.426}$$

For the volume of the sphere, we have

$$V = \int \mathrm{d}\mu = \int \sqrt{\det(g)} \, \mathrm{d}r \, \mathrm{d}\varphi \mathrm{d}\theta = \int r^2 \sin\theta \, \mathrm{d}r \, \mathrm{d}\varphi \mathrm{d}\theta . \tag{3.427}$$

Substituting the limits, we obtain

$$V = \left(\int_0^R r^2 \, \mathrm{d}r \right) \cdot \left(\int_0^{2\pi} \mathrm{d}\varphi \right) \cdot \left(\int_0^{\pi} \sin\theta \, \mathrm{d}\theta \right) = \frac{4}{3} \pi R^3 , \tag{3.428}$$

which is the well-known formula for the volume of a sphere.

◗

3.7 Some Special Functions

Oscillations and waves are the most natural phenomena we observe. Oscillations are described by a parabolic potential energy, which leads to the equation of motion

$$\ddot{\psi} + \omega^2 \psi = 0 \,. \tag{3.429}$$

The variable ψ is a placeholder for any oscillating quantity. The dot denotes the time derivative. The solution to this equation is a linear combination of sine and cosine functions: $A \cos \omega t + B \sin \omega t$. If there were a negative sign in front of the second term, the solution would be a combination of hyperbolic sine and cosine functions. Such solutions also have their place in describing nature. We can write a similar equation for an instantaneous snapshot of one-dimensional spatial waves (without time dependence):

$$\psi'' + k^2 \psi = 0 \,. \tag{3.430}$$

The comma denotes the spatial derivative. The solution to the equation is again a linear combination of sines and cosines: $A \cos kx + B \sin kx$. If the sign of the second term of the equation is negative, the solution again leads to hyperbolic sines and cosines. In three spatial dimensions, the natural generalization of the equation is

$$\Delta \psi + k^2 \psi = 0. \tag{3.431}$$

The solution depends on the region in which the waves propagate. Waves in a cylindrical region lead to Bessel functions, while waves on a sphere lead to spherical functions. Both classes of functions are extremely important in physics, and therefore we will at least briefly familiarize ourselves with them. The equation (3.431) is called the Helmholtz equation – it is an equation for the eigenfunctions of the Laplace operator.

3.7.1 Bessel Functions

Let us now find the natural mode of a simple cylinder. To do this, we will rewrite Helmholtz's equation (3.431) in cylindrical coordinates (r, φ, z).

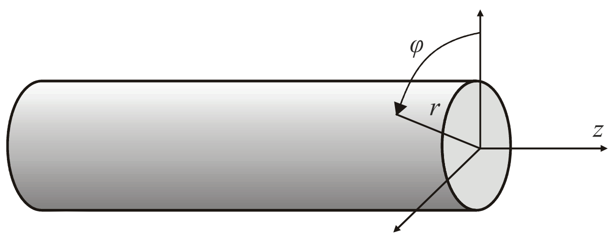


Fig. 3.36: Cylindrical coordinates

Along the z-axis, the solution is unbounded and reduces to ordinary sine and cosine functions; therefore, we are not interested in it and will attempt to solve equation (3.431) in a cross-section, i.e., for the variables (r, φ), where the solution is deformed by the cylindrical geometry from traditional sine and cosine functions into other functions:

$$\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial \psi}{\partial r}+\frac{1}{r^2}\frac{\partial^2\psi}{\partial\varphi^2}+k^2\psi=0\,;\qquad \psi=\psi(r,\varphi). \tag{3.432}$$

In the azimuthal direction, the solution must be periodic, i.e., it must repeat every 2π. We can expand periodic functions into sines and cosines using Fourier series. For a single mode of the solution, we will therefore have

$$\psi=\psi(r,\varphi)=f(r)\,\mathrm{e}^{\mathrm{i}m\varphi}\,. \tag{3.433}$$

From the requirement of periodicity, it follows that

$$\mathrm{e}^{\mathrm{i}m\varphi}=\mathrm{e}^{\mathrm{i}m(\varphi+2\pi)}\qquad\Rightarrow\qquad \mathrm{e}^{2\pi\mathrm{i}m}=1\qquad\Rightarrow\qquad m=0,\pm1,\pm2,\dots \tag{3.434}$$

We call the number m the wave mode. After substituting the assumed form of the solution (3.433) into the Helmholtz equation (3.432) and differentiating with respect to the azimuthal angle, we obtain the equation for the radial part of the solution $f(r)$:

$$\frac{\mathrm{d}^2 f}{\mathrm{d}r^2}+\frac{1}{r}\frac{\mathrm{d}f}{\mathrm{d}r}+k^2 f-\frac{m^2}{r^2}f=0\,. \tag{3.435}$$

The second and fourth terms are caused by the cylindrical geometry. If we remove them, we would obtain a wave equation like (3.430). Let us multiply the equation by r^2

$$r^2\frac{\mathrm{d}^2 f}{\mathrm{d}r^2}+r\frac{\mathrm{d}f}{\mathrm{d}r}+\left(k^2r^2-m^2\right)f=0\,. \tag{3.436}$$

and perform the substitution

$$x=kr \tag{3.437}$$

► $$x^2\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}+x\frac{\mathrm{d}f}{\mathrm{d}x}+\left(x^2-m^2\right)f=0\,. \tag{3.438}$$

Do not forget that the variable x has nothing to do with the Cartesian variable x; it is a rescaled radial distance. The given equation is called Bessel's equation, and its solution is analogous to the cosines and sines in Cartesian geometry. The equation must be solved using an expansion into infinite series. The general solution has the form:

► $$f(x)=c_1J_m(x)+c_2Y_m(x)\,. \tag{3.439}$$

We call the functions J_m Bessel functions of the first kind, and they have finite values at the origin. We call the functions Y_m Bessel functions of the second kind, and they are singular at the origin ($\to -\infty$), which is a consequence of the singular terms in (3.435). The function $J_0(x)$ corresponds to the cosine function from Cartesian coordinates, and the function $J_1(x)$ corresponds to the sine. A similar relationship holds between them:

$$\frac{\mathrm{d}J_0(x)}{\mathrm{d}x}=-J_1(x)\,. \tag{3.440}$$

Functions of the first kind can be easily written using a series; for the second kind, an integral expression is clearer:

$$J_m(x)=\sum_{k=0}^{\infty}\frac{(-1)^k}{k!(k+m)!}\left(\frac{x}{2}\right)^{2k+m}, \tag{3.441}$$

$$Y_m(x) = \frac{1}{\pi}\int_0^{\pi} \sin\left(x\sin\theta - n\theta\right)\mathrm{d}\theta - \frac{1}{\pi}\int_0^{\infty}\left[\mathrm{e}^{mt} + (-1)^m\,\mathrm{e}^{-mt}\right]\mathrm{e}^{-x\,\mathrm{sh}\,t}\,\mathrm{d}t\,. \qquad (3.442)$$

Analogy of the hyperbolic solution

Bessel functions can also be defined without difficulty for a non-integer index m (the factorials in the defining series are replaced by Γ-functions) or for a complex argument. Among complex arguments, the most important is the purely imaginary argument ($x \rightarrow \mathrm{i}x$). In this case, the equation (3.438) becomes

► $$x^2\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} + x\frac{\mathrm{d}f}{\mathrm{d}x} - (x^2 + m^2)\,f = 0\,, \qquad (3.443)$$

whose solution is analogous to the hyperbolic functions

► $$f(x) = c_1 I_m(x) + c_2 K_m(x)\,, \qquad (3.444)$$

where I_m and K_m are called hyperbolic or modified Bessel functions. The I_m functions have finite values at the origin and diverge at infinity (analogous to exp[x]); they are easily defined using a series. Conversely, the K_m functions diverge at the origin and approach zero at infinity (analogous to exp[$-x$]), and their integral representation is simpler:

$$I_m(x) = \sum_{k=0}^{\infty}\frac{1}{k!(k+m)!}\left(\frac{x}{2}\right)^{2k+m}\;; \qquad (3.445)$$

$$K_m(x) = \frac{\sqrt{\pi}}{(m-1/2)!}\left(\frac{x}{2}\right)^m\int_1^{\infty}\mathrm{e}^{-tx}\left(t^2-1\right)^{m-1/2}\mathrm{d}t\,. \qquad (3.446)$$

Asymptotic relationships

Asymptotic relationships near the origin ($x \ll 1$) can be useful:

$$J_m(x) \approx \frac{1}{m!}\left(\frac{x}{2}\right)^m, \qquad Y_m(x) \approx \frac{(m-1)!}{\pi}\left(\frac{x}{2}\right)^{-m}\;;\quad m>0\,,$$

$$I_m(x) \approx \frac{1}{m!}\left(\frac{x}{2}\right)^m, \qquad K_m(x) \approx \frac{(m-1)!}{\pi}\left(\frac{x}{2}\right)^{-m}\;;\quad m>0\,. \qquad (3.447)$$

Sometimes asymptotic relationships at infinity ($x \gg 1$) are also useful:

$$J_m(x) \approx \sqrt{\frac{2}{\pi x}}\cos\left(x - \frac{m\pi}{2} - \frac{\pi}{4}\right), \qquad Y_m(x) \approx \sqrt{\frac{2}{\pi x}}\sin\left(x - \frac{m\pi}{2} - \frac{\pi}{4}\right),$$

$$I_m(x) \approx \frac{1}{\sqrt{2\pi x}}\mathrm{e}^x\,, \qquad K_m(x) \approx \sqrt{\frac{\pi}{2x}}\;\mathrm{e}^{-x}\,. \qquad (3.448)$$

To plot Bessel functions, we can use the website www.wolframalpha.com; the commands are intuitive. For example, "besselj[0,x] from x=0 to x=10" plots the function J_0.

The first argument in parentheses (no matter of the type of parentheses) is the number m. The individual Bessel functions are called besselj, bessely, besseli, and besselk.

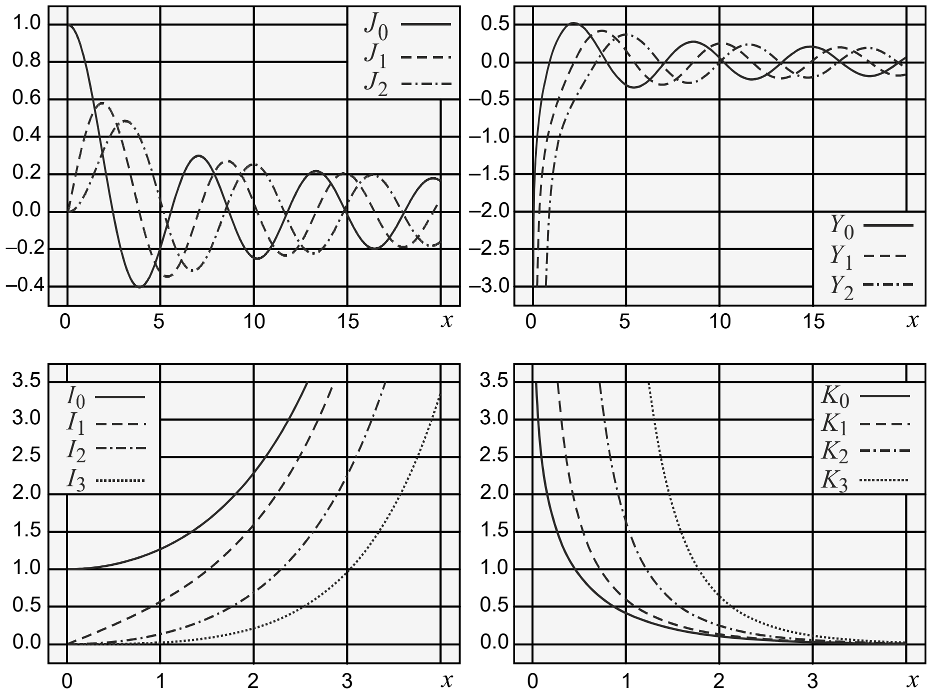


Fig. 3.36: Graphs of the individual Bessel functions

3.7.2 Spherical Harmonics

Spherical functions are obtained from the Helmholtz equation (the eigenvalue problem of the Laplace operator) under spherical symmetry for the cross-section $r = \text{const}$, i.e., on the general sphere. Spherical harmonics form the basis for the spherically symmetric potential in quantum theory; see Section 2.5. They are suitable for expanding the angular parts of functions in spherical coordinates. Any data on the sphere can be expanded into spherical harmonics; for example, they are used in helioseismology (the study of waves in the Sun) or in the frequency analysis of the cosmic microwave background. The initial equation again has the form (3.431), i.e., in spherical coordinates

$$\frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial \psi}{\partial r} + \frac{1}{r^2}\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{r^2}\frac{1}{\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi^2} + k^2\psi = 0\,, \qquad (3.449)$$

where r is the distance from the center, θ is the angle from the polar axis, and φ is the azimuthal angle. For a fixed r, only the angular part of the Laplace operator remains; we denote the eigenvalue by the usual symbol λ:

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2\psi}{\partial\varphi^2} = \lambda\psi; \qquad \lambda \equiv -r^2k^2\,. \qquad (3.450)$$

The solutions for the eigenvalues of the Laplace operator are spherical functions

$$Y_{lm}(\varphi,\theta) \equiv \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{\mathrm{i}\,m\varphi}\,P_{lm}(\cos\theta)\,; \tag{3.451}$$

$$P_{lm}(x) \equiv \frac{(1-x^2)^{m/2}}{2^l\;l!}\frac{\mathrm{d}^{l+m}}{\mathrm{d}x^{l+m}}(x^2-1)^l\;;\quad l=0,1,2,\ldots;\quad |m|\le l\,;\quad m=0,\pm1,\ldots \tag{3.452}$$

The eigenvalues are indexed by two discrete numbers l and m (on the sphere, we have two degrees of freedom). In quantum theory, these numbers correspond to the principal and magnetic numbers. The polynomials P_{lm} are called *associated Legendre polynomials*. For $m = 0$, they are called *Legendre polynomials*:

$$P_l(x) \equiv \frac{1}{2^l\;l!}\frac{\mathrm{d}^l}{\mathrm{d}x^l}(x^2-1)^l\;;\quad l=0,1,2,\;\ldots \tag{3.453}$$

Another way to write a Legendre polynomial is in the complex plane using a line integral along a curve that circles the origin counterclockwise; z, t are complex:

$$P_l(z) = \frac{1}{2\pi\,\mathrm{i}}\oint_\gamma\left(1-2tz+t^2\right)^{-1/2}t^{-l-1}\,\mathrm{d}t\,. \tag{3.454}$$

There are some useful relationships for Legendre polynomials, for example

$$\int_{-1}^{+1}P_l(x)\,\mathrm{d}x = 2\delta_{0l}\;;\qquad \int_{-1}^{+1}x\,P_l(x)\,\mathrm{d}x = \frac{2}{3}\delta_{1l}\,. \tag{3.455}$$

In the first case, therefore, only the integral of $P_0(x)$ is nonzero; in the second case, only the integral of $P_1(x)$ is nonzero. This is evident from the relationship (3.453). Legendre polynomials can be used in multipole expansions, e.g. for the inverse of the distance

$$\frac{1}{|\,\mathbf{r}-\mathbf{r}'\,|} = \sum_{l=0}^{\infty}\frac{\min^l(r,r')}{\max^{l+1}(r,r')}P_l(\cos\theta)\,, \tag{3.456}$$

where θ is the angle between the position vectors of the observer $\mathbf{r}$ and the source $\mathbf{r}'$.

* * *

If the wave is also expanded in the radial direction, we need three numbers to describe the wave: m describes the modes in the azimuthal direction, l in the polar direction, and n in the radial direction. For $l = 0$, the solution on the sphere is constant and no waves develop in the angular degrees of freedom. For $l > 1$, the typical wave dimension is

$$\Delta\theta = \frac{\pi}{l}\,;\qquad l>1\,. \tag{3.457}$$

The aim of this chapter was to provide a brief introduction to the existence of Bessel and spherical functions, not the details of their calculation. On wolphramalpha.com, the notation for spherical functions is sphericalharmonic(l,m), where you substitute specific numbers for l and m. Finally, take a look at the waves on the sphere for some l and m. The (+) regions mean bulging outward, and the (–) regions indicate inward curvature.

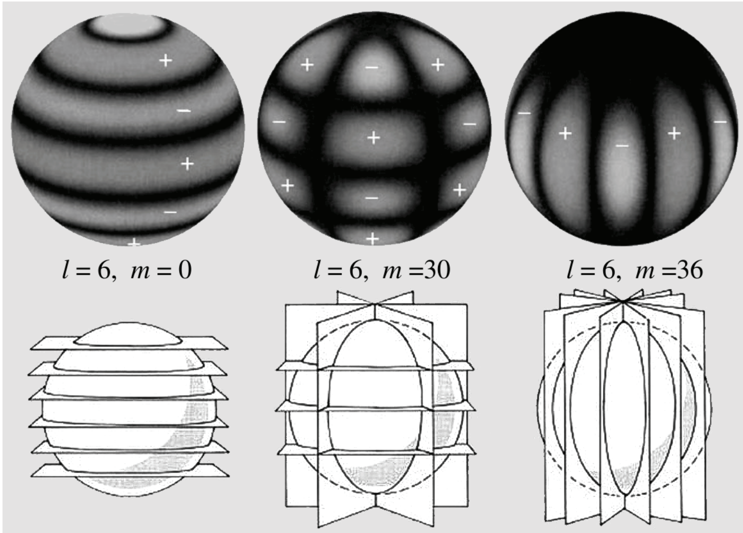


Fig. 3.37: Spherical functions. Robert Noyes, Sky Publishing Corporation.

3.7.3 Error Function and Chandrasekhar Function

Error function

In statistical physics, many processes follow a Gaussian probability density

$$f(x) \equiv \frac{2}{\sqrt{\pi}} e^{-x^2}; \qquad x \geq 0. \tag{3.458}$$

The sum of the probability densities from zero up to a certain value is called the cumulative distribution function. In the case of the Gaussian distribution, this is the so-called error function

$$\phi(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} \, d\xi . \tag{3.459}$$

Its value can also be defined for negative x. The error function can be used to express the first Rosenbluth potential H for the Maxwell distribution (see Part 3 of this textbook).

Chandrasekhar function

A function very closely related to the error function is the Chandrasekhar function. It is useful in describing collisions of charged particles and appears in the friction term of the Fokker-Planck equation for the Maxwell distribution. This function is also used to describe the entry of a charged particle into the so-called runaway regime, where collisions are negligible and the particle is accelerated by an electric field to very high energies. This function is given by the relation

$$\psi(x) \equiv \frac{\dfrac{2}{\sqrt{\pi}} \displaystyle\int_0^x \xi^2 \, \mathrm{e}^{-\xi^2} \, \mathrm{d}\xi}{x^2} . \tag{3.460}$$

There is a simple relationship between the Chandrasekhar and the error functions:

$$\psi(x) = \frac{\phi - x\phi'}{2x^2} . \tag{3.461}$$

Using the fact that $\phi' = 2 \exp[-x^2]/\pi^{1/2}$, we can easily reverse the relationship:

$$\phi = 2x^2\psi + \frac{2x}{\sqrt{\pi}} \mathrm{e}^{-x^2} . \tag{3.462}$$

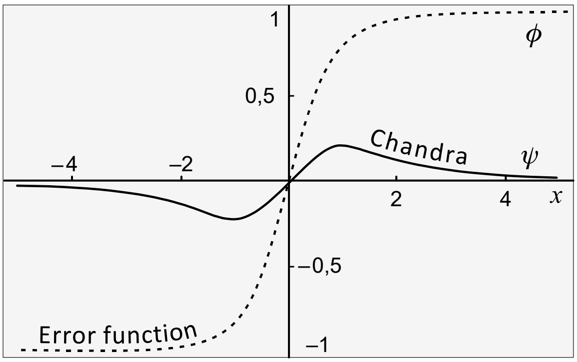


Fig. 3.38: The Chandrasekhar function (ψ) and the error function (ϕ)

The extreme of the Chandrasekhar function can be found numerically:

$$x_0 \approx 0{,}968 ; \qquad \psi(x_0) \approx 0{,}214 . \tag{3.463}$$

The limiting behavior of the Chandrasekhar function is

$$\psi(x) = \frac{2}{3\sqrt{\pi}} x ; \qquad x \ll 1 , \tag{3.464}$$

$$\psi(x) = \frac{1}{2x^2} ; \qquad x \gg 1 . \tag{3.465}$$

The Chandrasekhar function can be replaced by a simpler rational function

$$G(x) = \frac{2x}{3\pi^{1/2} + 4|x|^3} , \tag{3.466}$$

which has the same limiting behavior as the Chandrasekhar function and whose values do not differ by more than ten percent over the entire range.

3.8 Generalized Functions

In physics, we very often encounter the need to describe a point charge or a mass point. We imagine the charge or mass of a particle as localized at a single point, which leads to the problem of infinite charge or mass density at that point. The solution is to introduce so-called generalized functions, in particular the Dirac distribution. Let's illustrate the problem using a linear charge density localized at the point $x = 0$:

$$\tau(x) = \begin{cases} 0\,; & x \neq 0 \\ \neq 0 & x = 0. \end{cases} \tag{3.467}$$

However, the integral over the density must yield the total charge Q:

$$\int_{-\infty}^{+\infty} \tau(x)\,\mathrm{d}x = Q. \tag{3.468}$$

It is clear that the charge density is not a "normal" function. It has a nonzero value at a single point, and yet the integral over it should yield a finite number. However, such functions do not exist; we can introduce them as limits of sequences of functions, and their meaning lies only in the scalar product with another, so-called test function.

3.8.1 Dirac Distribution

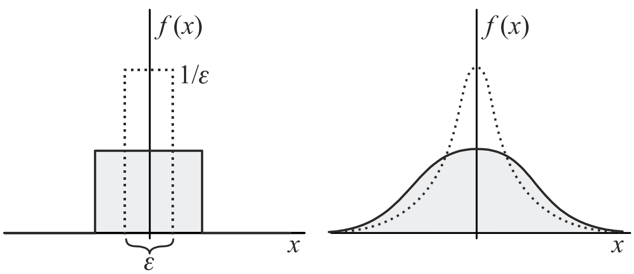


Fig. 3.39: A sequence of rectangles and hills

Sequence of rectangles

Let's introduce rectangular functions

$$f_\varepsilon(x) \equiv \begin{cases} 1/\varepsilon, & x \in \,<-\varepsilon/2, \varepsilon/2>\,; \\ 0, & x \notin \,<-\varepsilon/2, \varepsilon/2>. \end{cases} \tag{3.469}$$

All rectangles have the same area, equal to one, and the functions f_ε have some quite interesting properties:

$$\int_{-\infty}^{+\infty} f_\varepsilon(x)\,\mathrm{d}x = 1\,; \qquad f_\varepsilon(0) = \frac{1}{\varepsilon}\,; \qquad \lim_{\varepsilon \to 0} f_\varepsilon(x) = \begin{cases} \infty & \text{pro } x = 0 \\ 0 & \text{pro } x \neq 0 \end{cases}. \tag{3.470}$$

We can formally define the Dirac distribution as the limit of these rectangular functions

► $$\delta(x) = \lim_{\varepsilon \to 0} f_\varepsilon(x) = \begin{cases} \infty & \text{pro } x = 0 \\ 0 & \text{pro } x \neq 0 \end{cases}; \qquad \int_{-\infty}^{+\infty} \delta(x)\,\mathrm{d}x = 1. \tag{3.471}$$

It is clear that this cannot be a real function, because the integral of a function cannot be affected by the value at a single point. We are talking about a so-called generalized function, and we will explain its true meaning later.

Sequence of peaks (Cauchy-Lorentz distributions)

The rectangles from the previous example are not smooth functions. However, this is not an insurmountable problem; instead of rectangles, we can use functions that are continuous along with all their derivatives according to the relation

$$f_\varepsilon(x) \equiv \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + x^2}. \tag{3.472}$$

The area under these functions is equal to one for every ε, because

$$\int_{-\infty}^{+\infty} f_\varepsilon(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + x^2}\,\mathrm{d}x = \frac{1}{\pi}\left[\operatorname{atg}\frac{x}{\varepsilon}\right]_{-\infty}^{+\infty} = \frac{1}{\pi}\left(\frac{\pi}{2} + \frac{\pi}{2}\right) = 1. \tag{3.473}$$

For small ε, the “hills” narrow while their height increases:

$$\int_{-\infty}^{+\infty} f_\varepsilon(x)\,\mathrm{d}x = 1; \qquad f_\varepsilon(0) = \frac{1}{\pi\varepsilon};$$
$$\lim_{\varepsilon \to 0} f_\varepsilon(x) = \begin{cases} \infty & \text{pro } x = 0 \\ 0 & \text{pro } x \neq 0 \end{cases}. \tag{3.474}$$

Now we can define the Dirac distribution as the limit of these continuous functions:

$$\delta(x) = \lim_{\varepsilon \to 0} f_\varepsilon(x). \tag{3.475}$$

The Cauchy–Lorentz distribution, from which we have now constructed the Dirac distribution, describes the shape of spectral lines in spectroscopy or the resonance curve in the theory of forced oscillations. It is named after the French mathematician Augustin Cauchy (1759–1857) and the Dutch physicist Hendrik Lorentz (1853–1928).

The Dirac distribution (or *generalized function*) does not possess the properties of ordinary functions. Although its value is nonzero at a single point, the integral of this distribution yields a nonzero value. This follows from the limit-based nature of the definition of this distribution. Its fundamental properties include:

► $$\int_{-\infty}^{+\infty} \delta(x) f(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} \delta(x) f(0)\,\mathrm{d}x = f(0) \int_{-\infty}^{+\infty} \delta(x)\,\mathrm{d}x = f(0). \tag{3.476}$$

The reason is obvious. The distribution δ is zero everywhere except at a point $x = 0$. The result of the integral can only be affected by the value of the function at this point. However, we can factor $f(0)$ out of the integral and obtain as a result $f(0)$.

3.8.2 Tempered Distributions

The Dirac distribution can be understood as a very simple mapping that assigns a value to a function at the origin (mapping assigning a number to a function is the *functional*).

$$\hat{T}_\delta f(x) \equiv f(0)\,; \qquad \text{resp. } f(x) \overset{T_\delta}{\to} f(0)\,. \tag{3.477}$$

Such a definition has no issues with infinities at the origin or with an integral in which the "function" has a single value, and it is entirely correct. In general, a distribution can be understood as a functional given by the scalar product

$$\hat{T}_g f(x) \equiv \langle g \,|\, f \rangle\,; \tag{3.478}$$

The scalar product acts on any function f from the so-called space of test functions. The function g is fixed; it defines this mapping and is called a *tempered distribution*. The better the properties of the functions from the test space (for example, they converge sufficiently quickly to zero at the boundaries of the domain), the worse the properties the function g defining the mapping may have. The space of test functions can be, for example, the *Schwartz* (*Sobolev*) space $\mathcal{S}$ of functions that satisfy:

1) Functions are infinitely differentiable, i.e., belong to class C^∞;
2) Functions decrease at $\pm\infty$ faster than any power of $1/|x|^k$.

We will continue to denote test functions by ψ, i.e.

► $$\hat{T}_g \psi(x) \equiv \langle g \,|\, \psi \rangle\,; \qquad \psi \in \mathscr{S}\,. \tag{3.479}$$

Solutions to integral equations are often sought "in the sense of a scalar product." For example, instead of the Laplace-Poisson equation

$$\Delta\phi = f\,,$$

we solve the equation

$$\langle \Delta\phi | \psi \rangle = \langle f \,|\, \psi \rangle\,,$$

where ϕ is the searched solution and ψ is an arbitrary function from the space of test functions. These solutions are called solutions *in the sense of distributions* or *weak solutions*. Their class is much richer than the class of solutions to the original equation. The solutions found may have a "wilder" character and are closer to physical reality. The outstanding Soviet and Russian mathematician Olga Alexandrovna Ladyzhenskaya (1922–2004) was one of the first to study them.

Sequence of Dirichlet kernels

We can also introduce the Dirac distribution using a simple function

$$f(x) \equiv \frac{\sin x}{x}\,; \qquad f(0) \to 1\,; \qquad \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \pi\,. \tag{3.480}$$

Let's define a sequence

$$f_k(x) = \frac{k}{\pi}\frac{\sin kx}{kx}, \tag{3.481}$$

that has simple properties

$$\int_{-\infty}^{+\infty} f_k(x)\,\mathrm{d}x = 1; \qquad f_k(0) = \frac{k}{\pi}; \qquad \lim_{k\to\infty} f_k(0) = \infty. \tag{3.482}$$

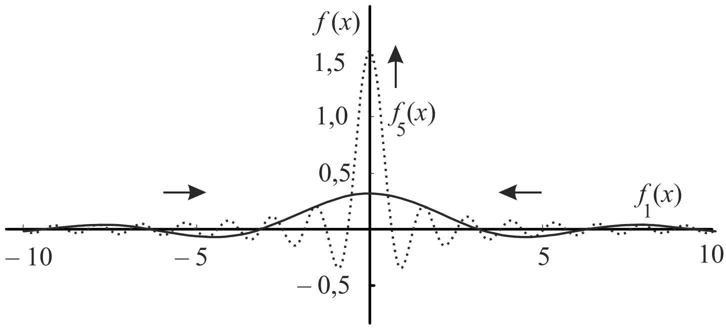


Fig. 3.40: A sequence of Dirichlet kernels

Unfortunately, it is no longer true that, as $k \to \infty$, the function values for $x \neq 0$ approach zero, as was the case with the sequences of rectangles or "hills" for small epsilon. For example, for $x = \pi/2$, $f_k(x) \sim \sin(k\pi/2) = \{0, 1, 0, -1, 0, 1, \ldots\}$ and the limit does not exist at all. However, on every interval that does not contain zero, the integral of this rapidly oscillating function as $k \to \infty$ will be zero, and this sequence of functions is again suitable for realizing the Dirac distribution (the integral of each is equal to one, and at zero, as $k \to \infty$, the sequence diverges to infinity. We can therefore define

► $$\delta(x) = \lim_{k\to\infty} \frac{k}{\pi}\frac{\sin kx}{kx}. \tag{3.483}$$

However, we understand this limit as the limit of a sequence of distributions (when integrating over a test function, the non-existence of certain limits will not be significant). Note that the functions $f_k(x)$ are known from the proof of the Fourier series theorem and are called the Dirichlet kernel. It is named after the famous German mathematician Johann Peter Gustav Lejeune Dirichlet (1805–1859).

Dirac distribution as the Fourier image of the unit function

Let us first compute the following integral:

$$\int_{-k}^{+k} \mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}k = \left[\frac{1}{\mathrm{i}x}\mathrm{e}^{\mathrm{i}kx}\right]_{-k}^{+k} = \frac{\mathrm{e}^{\mathrm{i}kx} - \mathrm{e}^{-\mathrm{i}kx}}{\mathrm{i}x} = 2k\frac{\sin kx}{kx}. \tag{3.484}$$

The integral, up to the coefficient $\pi/2$, gives the Dirichlet kernel. For the Dirac distribution, we can therefore also write

$$\delta(x) = \lim_{k\to\infty} \frac{1}{2\pi}\int_{-k}^{+k} \mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}k = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}k\,.$$

We interpret the integral over improper limits precisely in the sense as specified. The Dirac distribution is thus proportional to the Fourier image of the identity function:

► $$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}k\,. \tag{3.485}$$

3.8.3 Convolution and Fourier Transform

Convolution

We have already shown that the direct generalization of matrix multiplication on the space of functions is convolution. The two operations differ only in whether the index is discrete or continuous and whether we use summation or integration. For a mapping given by matrix multiplication, the following holds (we assign element g to element f)

$$\mathbf{g} = \mathbf{A}\cdot\mathbf{f}\,;$$
$$g_k = \sum_l A_{kl} f_l\,. \tag{3.486}$$

The identity mapping is given by the identity matrix, whose elements form the Kronecker symbol (it has ones on the diagonal and zeros off the diagonal):

$$\mathbf{1}\cdot\mathbf{f} = \mathbf{f}\,;$$
$$\sum_l \delta_{kl} f_l = f_k\,. \tag{3.487}$$

In the case of a function space, instead of a matrix, the mapping involves a function of two variables $A(x, y)$, which we call the convolution kernel:

$$A * f = g\,;$$
$$\int_\Omega A(x,y) f(y)\,\mathrm{d}y = g(x)\,. \tag{3.488}$$

The integral (3.488) is called a *convolution*. Convolution is the analogue of matrix multiplication in function spaces. The continuous variables x and y take on the role of indices. The *convolution kernel* takes on the role of the matrix. Various integral transforms (Laplace, Fourier, Abel, etc.) are special cases of convolutions. The kernel of the identity operator is the Dirac distribution (it is nonzero only for $x = y$):

$$\int \delta(x-y)\, f(y)\,\mathrm{d}y = f(x)\,.$$

The Dirac distribution takes on the role of the Kronecker delta and represents the identity operator on the space of functions. The Dirac distribution for convolution can be expressed by relation (3.485):

$$\delta(x-y)=\frac{1}{2\pi}\int_k \mathrm{e}^{\mathrm{i}k(x-y)}\,\mathrm{d}k\,. \qquad (3.489)$$

In N dimensions, the relationship is similar:

$$\delta(\mathbf{x}-\mathbf{y})=\frac{1}{(2\pi)^N}\int_k \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})}\,\mathrm{d}^N\mathbf{k}\,. \qquad (3.490)$$

Fourier transform

The Fourier transform can be understood as a convolution with the kernel exp[i$\mathbf{k}\cdot\mathbf{x}$] or as a series expansion of a function into plane waves:

► $$F(\mathbf{x})\equiv\int A(\mathbf{k})\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\,\mathrm{d}^N\mathbf{k}\,; \qquad (3.491)$$

The expansion coefficients, or amplitudes $A(\mathbf{k})$, determine how strongly each wave is represented in the spectrum. In some situations, a relationship is useful that tells us how to determine the coefficients of the Fourier expansion of the function $F(\mathbf{x})$:

► $$A(\mathbf{k})=\frac{1}{(2\pi)^N}\int F(\mathbf{x})\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\,\mathrm{d}^N\mathbf{x}\,. \qquad (3.492)$$

Let us prove that the amplitudes determined in this way lead to the correct expansion of the function $F(\mathbf{x})$. To do this, we substitute (3.492) into the right-hand side of (3.491):

$$\int A(\mathbf{k})\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\,\mathrm{d}^N\mathbf{k}=\int\left(\frac{1}{(2\pi)^N}\int F(\mathbf{x}')\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}'}\,\mathrm{d}^N\mathbf{x}'\right)\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\,\mathrm{d}^N\mathbf{k}=$$

$$=\int F(\mathbf{x})\left(\frac{1}{(2\pi)^N}\int \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}\mathrm{d}^N\mathbf{k}\right)\mathrm{d}^N\mathbf{x}'=$$

$$\int F(\mathbf{x}')\delta(\mathbf{x}-\mathbf{x}')\,\mathrm{d}^N\mathbf{x}'=F(\mathbf{x})\,,$$

which is what we wanted to prove – after substituting the amplitudes from (3.492) into the rhs of (3.491), we obtain the original function. We usually interpret the expansion (3.491) as a direct transformation from **k**-space to **x**-space, and relation (3.492) for the coefficients $A(\mathbf{k})$ as an inverse transformation from **x**-space to **k**-space:

$$F(\mathbf{x})\equiv\mathscr{F}\big(A(\mathbf{k})\big)\,;\qquad A(\mathbf{k})\equiv\mathscr{F}^{-1}\big(F(\mathbf{x})\big). \qquad (3.493)$$

Theorem: There exists a simple relation for the Fourier transform of a convolution:

► $$\mathscr{F}(f*g)=\mathscr{F}(f)\cdot\mathscr{F}(g)\,. \qquad (3.494)$$

Proof:

$$\mathscr{F}(f*g)=\int \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\left(\int f(\mathbf{x}-\mathbf{y})g(\mathbf{y})\,\mathrm{d}^N\mathbf{y}\right)\mathrm{d}^N\mathbf{x}=$$

$$=\int\left(\int \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}f(\mathbf{x}-\mathbf{y})g(\mathbf{y})\,\mathrm{d}^N\mathbf{x}\right)\mathrm{d}^N\mathbf{y}=\quad\Big|\ \text{subst.: }\ \mathbf{x}-\mathbf{y}\equiv\mathbf{z}$$

$$=\int\left(\int \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{y}+\mathbf{z})}f(\mathbf{z})g(\mathbf{y})\,\mathrm{d}^N\mathbf{x}\right)\mathrm{d}^N\mathbf{y}=$$

$$= \left(\int e^{i\mathbf{k}\cdot\mathbf{z}} f(\mathbf{z})\,d^N\mathbf{z}\right)\cdot\left(\int e^{i\mathbf{k}\cdot\mathbf{y}} g(\mathbf{y})\,d^N\mathbf{y}\right) =$$

$$= \mathscr{F}(f)\cdot\mathscr{F}(g).$$

■

Note: The product of the coefficients for the direct and inverse transforms must equal $1/(2\pi)^N$. It is sometimes useful to choose the coefficients symmetrically:

$$F(\mathbf{x}) \equiv \frac{1}{(2\pi)^{N/2}}\int A(\mathbf{k})\,e^{i\mathbf{k}\cdot\mathbf{x}}\,d^N\mathbf{k}\,; \tag{3.495}$$

$$A(\mathbf{k}) = \frac{1}{(2\pi)^{N/2}}\int F(\mathbf{x})\,e^{-i\mathbf{k}\cdot\mathbf{x}}\,d^N\mathbf{x}\,. \tag{3.496}$$

In such a case, however, the simple convolution relation (3.494) will not hold

3.8.4 Green Function

Let us now consider a special case – an equation with a linear operator and a nonzero right-hand side in the space $\mathcal{L}^2$

$$\hat{L}\phi = f. \tag{3.497}$$

First, let us find the solution for a unit impulse on the right-hand side (it will be represented by the Dirac distribution)::

$$\hat{L}\,G(\mathbf{x}) = \delta(\mathbf{x}) \tag{3.498}$$

This solution is called the *Green function*. The general solution to the equation (3.497) is the convolution of the Green's function and the right-hand side of the equation

► $$\phi(\mathbf{x}) = G * f = \int G(\mathbf{x}-\mathbf{y})f(\mathbf{y})\,d^N\mathbf{y}\,. \tag{3.499}$$

The proof is very simple. We will show that by applying the operator $\hat{L}$ to the found solution, we obtain the right-hand side of the original equation:

$$\hat{L}\phi(\mathbf{x}) = \int \hat{L}\,G(\mathbf{x}-\mathbf{y})f(\mathbf{y})\,d^N\mathbf{y} = \int \delta(\mathbf{x}-\mathbf{y})f(\mathbf{y})\,d^N\mathbf{y} = f(\mathbf{x})\,.$$

The Green function is very useful in solving partial differential equations. Any boundary conditions are included in the Green function; naturally, the Green function must automatically satisfy these conditions. If the partial differential equation also has a time variable, i.e., $\phi = \phi(t, \mathbf{x})$, we must specify the initial condition $\phi_0(\mathbf{x}) = \phi(t = 0, \mathbf{x})$, the boundary conditions, and the right-hand side $f(t, \mathbf{x})$. The general solution is then the sum of two terms: the spatial convolution of the Green's function with the initial condition and the spacetime convolution of the Green's function with the right-hand side:

► $$\phi(t,\mathbf{x}) = G \underset{(\mathbf{x})}{*} \phi_0 + G \underset{(t,\mathbf{x})}{*} f, \tag{3.500}$$

$$\phi(t,\mathbf{x}) = \int G(t,\mathbf{x}-\mathbf{y})\,\phi_0(\mathbf{y})\,\mathrm{d}^N\mathbf{y} + \int G(t-\tau,\mathbf{x}-\mathbf{y})\,f(\tau,\mathbf{y})\,\mathrm{d}\tau\,\mathrm{d}^N\mathbf{y}\,. \tag{3.501}$$

In the final section of this chapter, we will demonstrate the use of the Green function in calculating the general solution to the diffusion equation in unbounded 3D space.

General solution of the diffusion equation

The diffusion equation is one of the most important equations in physics. It leads to a number of problems: the diffusion of molecules (such as perfume), heat conduction, magnetic field diffusion, and the Schrödinger time equation in quantum theory has a similar form. Let us start with the form:

► $$\frac{\partial \phi}{\partial t} - D\,\Delta\phi = f(t,\mathbf{x})\,, \tag{3.502}$$

where $\phi(t, \mathbf{x})$ is the quantity of interest (it can be either scalar or vector), D is the diffusion coefficient, $f(t, \mathbf{x})$ is the right-hand side describing the sources of the quantity (e.g., perfume atomizers), and the initial distribution of the quantity ϕ is

$$\phi(0, \mathbf{x}) = \phi_0(\mathbf{x})\,. \tag{3.503}$$

Consider this diffusion equation in an unbounded domain.

Theorem: The general solution to the diffusion equation in N dimensions can be written as the spatial convolution of the Green function with the initial condition and the space-time convolution with the right-hand side:

$$\phi = G \underset{(\mathbf{x})}{*} \phi_0 \quad + \quad G \underset{(t,\mathbf{x})}{*} \mathbf{f}\,, \tag{3.504}$$

or

$$\phi(t,\mathbf{x}) = \int G(t,\mathbf{x}-\boldsymbol{\xi})\,\phi_0(\boldsymbol{\xi})\,\mathrm{d}^N\boldsymbol{\xi} \;+\; \int G(t-\tau,\mathbf{x}-\boldsymbol{\xi})\,\mathbf{f}(\tau,\boldsymbol{\xi})\,\mathrm{d}\tau\,\mathrm{d}^N\boldsymbol{\xi}\,, \tag{3.505}$$

where the Green function G is given by the relation

$$G(t,\mathbf{x}) = \frac{1}{(4\pi D t)^{N/2}} \exp\left[-\frac{|\mathbf{x}|^2}{4Dt}\right]. \tag{3.506}$$

Proof: We perform a Fourier transform on the spatial part of the equation:

$$\frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}t} + D k^2 \tilde{\phi} = \tilde{f}\,.$$

We multiply all terms by the exponential $\exp[Dk^2 t]$ and simplify:

$$\frac{\mathrm{d}\tilde{\phi}}{\mathrm{d}t}\mathrm{e}^{Dk^2 t} + Dk^2\,\mathrm{e}^{Dk^2 t}\,\tilde{\phi} = \tilde{f}\,\mathrm{e}^{Dk^2 t} \qquad \Rightarrow$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{e}^{Dk^2 t}\,\tilde{\phi}\right) = \tilde{f}\,\mathrm{e}^{Dk^2 t}\,.$$

We integrate both sides over time and simplify again

$$\tilde{\phi}\,\mathrm{e}^{Dk^2 t} - \tilde{\phi}_0 = \int_0^t \tilde{f}\,\mathrm{e}^{Dk^2\tau}\,\mathrm{d}\tau \qquad \Rightarrow$$

$$\tilde{\phi} = \tilde{\phi}_0\,\mathrm{e}^{-Dk^2 t} + \int_0^t \tilde{f}(\tau,\mathbf{k})\,\mathrm{e}^{-Dk^2(t-\tau)}\,\mathrm{d}t' .$$

If we denote

$$\exp[-Dk^2 t] \equiv \tilde{G}\,, \tag{3.507}$$

we obtain the following relation using equation (3.494):

$$\tilde{\phi} = \tilde{\phi}_0 \tilde{G} + \int_0^t \tilde{f}(t',\mathbf{k})\tilde{G}(t-\tau,\mathbf{k})\,\mathrm{d}\tau \quad \Rightarrow$$

$$\mathscr{F}(\phi) = \mathscr{F}(\phi_0 \underset{(\mathbf{x})}{*} G) + \mathscr{F}(\phi \underset{(t,\mathbf{x})}{*} G) \quad \Rightarrow$$

$$\phi = \phi_0 \underset{(\mathbf{x})}{*} G \;+\; \phi \underset{(t,\mathbf{x})}{*} G \quad \Rightarrow$$

$$\phi = G \underset{(\mathbf{x})}{*} \phi_0 \;+\; G \underset{(t,\mathbf{x})}{*} \phi .$$

Now all that remains is to determine the Green function G from equation (3.507):

$$\tilde{G} \equiv \exp[-Dk^2 t] \quad \Rightarrow$$

$$G(t,\mathbf{x}) = \mathscr{F}^{-1}(\tilde{G}) = \frac{1}{(2\pi)^N}\int \mathrm{e}^{-Dk^2 t}\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\,\mathrm{d}^N\mathbf{k} \quad \Rightarrow$$

$$G(t,\mathbf{x}) = \frac{1}{(2\pi)^N}\mathrm{e}^{-|\mathbf{x}|^2/4Dt}\int \mathrm{e}^{-Dt(\mathbf{k}+\mathrm{i}\mathbf{x}/2Dt)^2}\,\mathrm{d}^N\mathbf{k} \quad \Rightarrow$$

$$G(t,\mathbf{x}) = \frac{1}{(2\pi)^N}\mathrm{e}^{-|\mathbf{x}|^2/4Dt}\left(\int_{-\infty}^{+\infty} \mathrm{e}^{-Dt\xi^2}\,\mathrm{d}\xi\right)^N \quad \Rightarrow$$

$$G(t,\mathbf{x}) = \frac{1}{(2\pi)^N}\mathrm{e}^{-|\mathbf{x}|^2/4Dt}\left(\frac{\pi}{Dt}\right)^{N/2} \quad \Rightarrow$$

$$G(t,\mathbf{x}) = \frac{1}{(4\pi Dt)^{N/2}}\exp\left[-\frac{|\mathbf{x}|^2}{4Dt}\right].$$

■

In the follow-up textbook [2], we will examine magnetic field diffusion in greater detail. We obtain the general relations (3.504) and (3.506) by direct calculation. The Green function describes the time evolution of the initial Dirac delta function. In this case, it leads to a gradually dissipating Gaussian packet.

3.9 Pfaff Differential Forms

You surely remember the concept of the small increment of a multivariable function, or the differential (see Chapter 3.1). For example, for the function

$$f(x, y) = x^2 + y^2 \tag{3.508}$$

is the first-order differential

$$\mathrm{d}f = 2x\,\mathrm{d}x + 2y\,\mathrm{d}y\,. \tag{3.509}$$

Let's now turn the problem around. Imagine we write an expression similar to the one on the right-hand side of equation (3.509) and ask whether there exists a function for which the expression would be the first differential. For example

$$\mathrm{d}\omega_1 = 2x\,\mathrm{d}x + 2y\,\mathrm{d}y\,, \tag{3.510}$$

$$\mathrm{d}\omega_2 = 2y\,\mathrm{d}x + xy\,\mathrm{d}y\,. \tag{3.511}$$

Such a function exists for the first expression – it is the function (3.508) – while we will never find such a function for the second expression. In general, we call expressions of this type Pfaffian differential forms and write them in the form

$$\mathrm{d}\omega = a_1(x_1,\ldots,x_N)\,\mathrm{d}x_1 + \cdots + a_N(x_1,\ldots,x_N)\,\mathrm{d}x_N\,. \tag{3.512}$$

Using the more concise summation convention, the expression takes a simpler form

► $$\mathrm{d}\omega = a_k(\mathbf{x})\,\mathrm{d}x_k\,. \tag{3.513}$$

The question, then, is: when is a Pfaff form in the form of a total differential of a function? The answer is very interesting. All differential forms fall into two broad categories. The first is not in the form of a total differential of a function, and this type has no "nice" properties. The second type is in the form of a total differential of some function; it has many very elegant properties and is very easy to work with. That is why mathematicians and physicists always prefer differential forms that are in the form of a total differential. Let us now formulate the so-called *five-equivalences theorem*:

3.9.1 Five-Equivalences Theorem

Let the differential form $\mathrm{d}\omega = a_k \mathrm{d}x_k$ have coefficients whose derivatives are continuous up to and including the second order. Then the following statements are equivalent:

1) There exists a function $f(x_1,\ldots, x_N)$ such that the form is its first differential, i.e., the coefficients of the form are the partial derivatives of this function:

$$a_k = \frac{\partial f}{\partial x_k}\,. \tag{3.514}$$

2) There exists a function ϕ such that the line integral between two points is simply the difference between the final and initial values of this function (we call this the potential of the differential form):

$$\int_A^B a_k \mathrm{d}x_k = \phi(B) - \phi(A)\,. \tag{3.515}$$

3) The line integral between two points does not depend on the path of integration:

$$\int_\gamma a_k \mathrm{d}x_k \quad \text{does not depend on the curve } \gamma\,. \tag{3.516}$$

4) The line integral along any closed curve of differential form is zero:

$$\oint a_k \mathrm{d}x_k = 0\,. \tag{3.517}$$

5) The form coefficients satisfy the following relations:

$$\frac{\partial a_k}{\partial x_l} = \frac{\partial a_l}{\partial x_k} \quad \text{pro } \forall\, k,l. \tag{3.518}$$

Note 1: If we have a differential form, either all the properties listed in the five-equivalence theorem hold for it, or none of them do. There is nothing in between. Therefore, differential forms are divided into two disjoint groups.

Note 2: To prove the theorem, it would suffice to prove only the individual implications in the "circle": $1 \Rightarrow 2 \Rightarrow \cdots \Rightarrow 5$. This makes it possible to derive any statement from any other. We will not present the entire proof here; we will limit ourselves to just a few parts.

Note 3: The fifth statement is essentially a guide to identifying "correct" differential forms, i.e., forms in the shape of a total differential. If we verify that property 5) holds, then all the other properties follow.

Note 4: In physics, we would say that the coefficients of a differential form represent a conservative field, such as the gravitational field. The line integral of the gravitational force represents the mechanical work done. This work does not depend on the path between two points; it is zero along a closed curve. There is potential energy, and the work done is the difference in potential energy between the end and starting points. The last condition is also easy to interpret. If we move both terms to the left-hand side, we get

► $$\frac{\partial a_k}{\partial x_l} - \frac{\partial a_l}{\partial x_k} = 0 \quad \text{for } \forall\, k,l\,.$$

It is simply a condition that the field's rotation is zero; mathematically speaking, it is therefore a non-vortex field.

Let us now outline a proof of some of the implications of the five equivalences theorem:

Implication 1 ⇒ 2

Let's try to find, in phase space $(x_1, \ldots, x_N)$, the integral of a differential form in the form of a total differential between two points A and B:

$$\int_A^B \mathrm{d}\omega = \int_A^B a_k \,\mathrm{d}x_k \overset{(1)}{=} \int_A^B \frac{\partial f}{\partial x_k} \mathrm{d}x_k = \int_A^B \mathrm{d}f = f(B) - f(A)\,.$$

If a differential form is in the form of a total differential, then the coefficients are given by the partial derivatives of the function *f*, and the resulting integral is simply the difference between the values of *f* at the starting and ending points. The potential of the differential form is thus the function *f* itself.

Implication 2 ⇒ 3

If the value of the integral depends only on the initial and final values of the function *f*, then it does not depend on the path of integration. It makes no difference whether we integrate along curve γ_1, γ_2, or γ_3 in Figure 3.41.

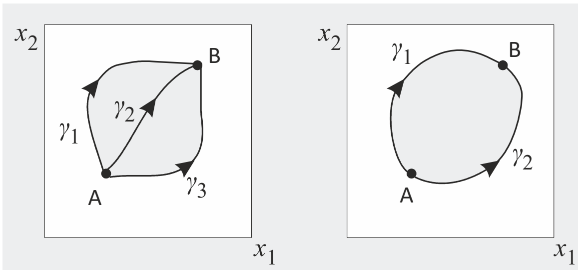


Fig. 3.41: Paths of integration

Implication 3 ⇒ 4

We can think of the closed curve on the right side of the figure as the sum of two separate curves. However, we must pay attention to the orientation of the curve, which changes the sign of the line integral:

$$\oint \mathrm{d}\omega = \int_{\gamma_1} \mathrm{d}\omega + \int_{-\gamma_2} \mathrm{d}\omega = \int_{\gamma_1} \mathrm{d}\omega - \int_{\gamma_2} \mathrm{d}\omega \overset{(3)}{=} 0 .$$

Implication 1 ⇒ 5:

The proof is based on the commutativity of the second derivatives of a function *f*:

$$\frac{\partial a_k}{\partial x_l} \overset{(1)}{=} \frac{\partial}{\partial x_l}\left(\frac{\partial f}{\partial x_k}\right) = \frac{\partial}{\partial x_k}\left(\frac{\partial f}{\partial x_l}\right) = \frac{\partial a_l}{\partial x_k} .$$

◖ Example 3.48: Differential form 11

Determine whether the form dω is in the shape of a total differential:

$$\mathrm{d}\omega = 2xy\,\mathrm{d}x + x^2\mathrm{d}y .$$

From property (3.518) of the five-equivalence theorem, we find “cross” derivations

$$a_x = 2xy ; \qquad a_y = x^2 \qquad \Rightarrow \qquad \frac{\partial a_x}{\partial y} = 2x ; \qquad \frac{\partial a_y}{\partial x} = 2x .$$

Both derivatives are equal, and the differential form is therefore in the form of a total differential. You can easily verify that this is a total differential of the function

$$f(x,y) = x^2 y\,.$$

◗

◖ Example 3.49: Differential form 12

Determine whether the form dω is in the shape of a total differential:

$$\mathrm{d}\omega = \frac{x}{y}\,\mathrm{d}x + \mathrm{d}y\,.$$

For "cross" derivatives, we have:

$$a_x = \frac{x}{y}; \qquad a_y = 1 \qquad \Rightarrow \qquad \frac{\partial a_x}{\partial y} = -\frac{x}{y^2}; \qquad \frac{\partial a_y}{\partial x} = 0\,.$$

The derivatives are not equal, and therefore the differential form is not in the form of a total differential, and there is no function f such that the form is its first differential.

◗

However, not all differential forms that are not in the "correct" form need to be discarded. Some of them can be easily "corrected." The form from Example 3.49 can be modified by multiplying it by the function y:

$$\mathrm{d}\sigma = y\,\mathrm{d}\omega = y\left(\frac{x}{y}\,\mathrm{d}x + \mathrm{d}y\right) = x\,\mathrm{d}x + y\,\mathrm{d}y\,.$$

The new differential form clearly has potential and is the differential of the function $(x^2 + y^2)/2$. If we find that the differential form is not in the form of a total differential, we can try to find the so-called integration factor μ so that the new form

$$\mathrm{d}\sigma = \mu(x_1,\ldots,x_N)\,\mathrm{d}\omega \tag{3.519}$$

is in the form of a total differential. Unfortunately, this is not always possible; in particular, for differential forms with many variables, finding an integrating factor is extremely difficult. However, for differential forms with up to three variables, an integrating factor always exists, as follows from the next theorem.

3.9.2 Theorem on the Existence of an Integrating Factor

Theorem: For $n = 1, 2$ the Pfaff's differential form always has an integrating factor.

Proof: Let us examine whether the new form $\mathrm{d}\sigma = \mu(x_1, \ldots, x_N)\, a_k(x_1, \ldots, x_N)\,\mathrm{d}x_k$ is in the form of a total differential. We are therefore looking for a function f such that the new form is a total differential, i.e.,

$$\partial f/\partial x_k = \mu\, a_k\,.$$

This will be possible if the cross-derivatives of the coefficients are equal:

$$\frac{\partial \mu a_k}{\partial x_l} = \frac{\partial \mu a_l}{\partial x_k} \qquad \text{for } \forall\, k, l.$$

The integration factor must be determined from these equations. For the problem to be solvable, their total number must be less than the dimension of the phase space, N:

$$\binom{N}{2} < N \qquad \Rightarrow \qquad \frac{N(N-1)}{2} < N \qquad \Rightarrow \qquad N \le 2\,.$$

The integration factor therefore always exists for one and two dimensions. In three dimensions, we can write

$$\frac{\partial \mu a_k}{\partial x_l} = \frac{\partial \mu a_l}{\partial x_k}; \qquad k, l, m = 1, 2, 3 \qquad \Rightarrow$$

$$\frac{\partial \mu a_k}{\partial x_l} - \frac{\partial \mu a_l}{\partial x_k} = 0 \qquad \Rightarrow \qquad \varepsilon_{mkl}\left(\frac{\partial \mu a_k}{\partial x_l}\right) = 0 \qquad \Rightarrow$$

$$a_m \varepsilon_{mkl}\left(\frac{\partial \mu a_k}{\partial x_l}\right) = 0 \qquad \Rightarrow \qquad a_m a_k \varepsilon_{mkl}\left(\frac{\partial \mu}{\partial x_l}\right) + \mu a_m \varepsilon_{mkl}\left(\frac{\partial a_k}{\partial x_l}\right) = 0\,.$$

The first term is automatically zero (due to the reduction of the symmetric and anti-symmetric terms in the indices m and k). The second term gives the condition

$$a_m \varepsilon_{mkl}\left(\frac{\partial a_k}{\partial x_l}\right) = 0 \qquad \Rightarrow \qquad \mathbf{a} \cdot \operatorname{rot} \mathbf{a} = 0\,.$$

For differential forms in three dimensions, a necessary condition for the existence of an integrating factor is that the relation **a** rot **a** = 0 holds (the helicity of the field formed by the coefficients of the differential form is zero, i.e., the field does not have a helical structure). For more than three variables, the existence of an integrating factor is generally not guaranteed in any way. There are also more sophisticated theorems that, under certain conditions, allow for the existence of an integrating factor in higher dimensions (for example, Caratheodory principle), but these go beyond the scope of this textbook.

■

3.10 Important Relationships

3.10.1 Conic Sections

Ellipse

An ellipse is a set of points in a plane that have the same sum of distances from two fixed points, known as foci.

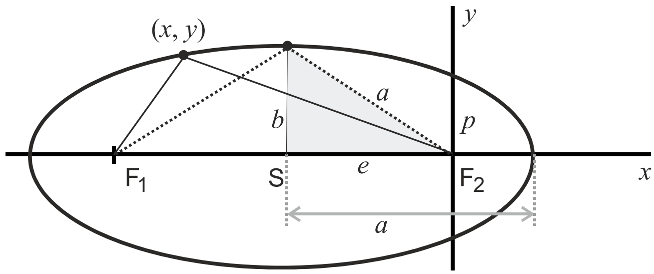


Fig. 3.42: Ellipse

In the Cartesian coordinate system, the equation of an ellipse has the form (the y-axis passes through one of the foci)

► $$\left(\frac{x+e}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1\,. \qquad (3.520)$$

The quantity e is called the eccentricity, a is the major axis, and b is the minor axis. There is a simple relationship between these parameters (see figure)

► $$a^2 = e^2 + b^2\,. \qquad (3.521)$$

Let's rewrite this equation in polar coordinates

$$x = r\cos\varphi\,; \qquad y = r\sin\varphi\,. \qquad (3.522)$$

Substituting (3.522) into (3.520) gives a quadratic equation for r, which has the solution

$$r = \frac{-b^2 e\cos\varphi \pm b^2\sqrt{e^2\cos^2\varphi + b^2\cos^2\varphi + a^2\sin^2\varphi}}{b^2\cos^2\varphi + a^2\sin^2\varphi}\,.$$

A solution with a negative value is unacceptable, as it would result in a negative radial distance. We eliminate the minor semi-axis b from the equation using equation (3.521):

$$r = \frac{-(a^2-e^2)e\cos\varphi + (a^2-e^2)a}{a^2 - e^2\cos^2\varphi} = \frac{(a^2-e^2)(a - e\cos\varphi)}{(a-e\cos\varphi)(a+e\cos\varphi)} =$$

$$= \frac{a^2-e^2}{a+e\cos\varphi} = \frac{a[1-(e/a)^2]}{1+(e/a)\cos\varphi}\,.$$

The resulting equation of the ellipse is therefore

► $$r = \frac{p}{1+\varepsilon\cos\varphi}\,; \qquad \varepsilon \equiv e/a = \sqrt{1-(b/a)^2}\,; \qquad p \equiv a\left(1-\varepsilon^2\right). \tag{3.523}$$

For an ellipse, $\varepsilon < 1$ and $p > 0$. The quantity ε is called the *numerical eccentricity* and is a dimensionless parameter characterizing the elongation of the ellipse. We obtain the area of the ellipse by rescaling the coordinate axes so that the semi-axes are of equal length and the ellipse becomes a circle:

► $$S = \pi ab\,. \tag{3.524}$$

Hyperbola

A hyperbola is a set of points in a plane that have the same difference in distance from two fixed points, known as foci.

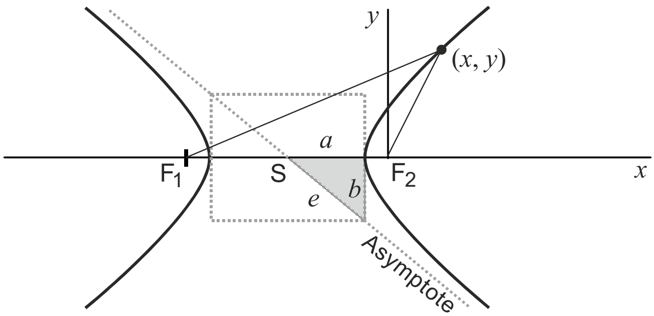


Fig. 3.43: Hyperbola

In the Cartesian coordinate system, the equation of a hyperbola has the form (the y-axis passes through one of the hyperbola's foci)

► $$\left(\frac{x+e}{a}\right)^2 - \left(\frac{y}{b}\right)^2 = 1\,. \tag{3.525}$$

The quantity e is called the eccentricity, a is the major axis, and b is the minor axis. There is a simple relationship between these parameters (see figure)

► $$a^2 + b^2 = e^2\,. \tag{3.526}$$

Using a procedure identical to that for the ellipse, we derive the equation of the hyperbola in polar coordinates. The resulting equation of the hyperbola is

► $$r = \frac{p}{1+\varepsilon\cos\varphi}\,;$$
$$\varepsilon \equiv e/a = \sqrt{1+(b/a)^2}\,; \qquad p \equiv a\left(1-\varepsilon^2\right). \tag{3.527}$$

For a hyperbola, $\varepsilon > 1$and $p < 0$. The quantity ε is called the numerical eccentricity and is a dimensionless parameter that characterizes the shape of the hyperbola. If we require that $p > 0$, i.e., we define $p = |a(1–\varepsilon^2)|$, the denominator will contain –1 instead of +1.

Parabola

A parabola is the set of all points in a plane that are equidistant from the focus and the directrix.

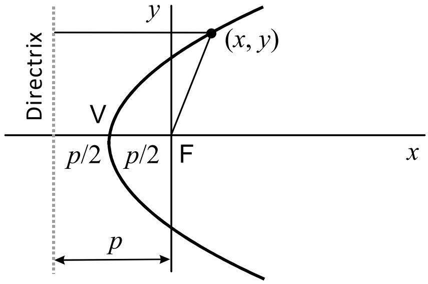


Fig. 3.44: Parabola

In the Cartesian coordinate system, the equation of a parabola has the form (the y-axis passes through the focus)

► $$y^2 = 2p\left(x + \frac{p}{2}\right). \tag{3.528}$$

Using a procedure similar to that for the ellipse and the hyperbola, we derive the equation of the parabola in polar coordinates. The resulting equation of the parabola is

► $$r = \frac{p}{1+\varepsilon\cos\varphi}; \qquad \varepsilon = \pm 1. \tag{3.529}$$

The "+" sign applies to a parabola symmetric about the vertical y-axis, and the "–" sign applies to a parabola symmetric about the horizontal x-axis. The equations of all conic sections therefore have the same form:

► $$r = \frac{p}{1+\varepsilon\cos\varphi}; \qquad \begin{cases} |\varepsilon| < 1 & \text{Ellipse} \\ |\varepsilon| > 1 & \text{Hyperbola} \\ |\varepsilon| = 1 & \text{Parabola} \end{cases} \tag{3.530}$$

3.10.2 Trigonometry

Simple definitions

$$\operatorname{tg} x = \frac{\sin x}{\cos x}, \tag{3.531}$$

$$\operatorname{ctg} x = \frac{\cos x}{\sin x}, \tag{3.532}$$

$$\operatorname{cosec} x = \frac{1}{\sin x}, \tag{3.533}$$

$$\sec x = \frac{1}{\cos x}, \tag{3.534}$$

$$\operatorname{ch} x = \frac{\exp(x) + \exp(-x)}{2} , \tag{3.535}$$

$$\operatorname{sh} x = \frac{\exp(x) - \exp(-x)}{2} . \tag{3.536}$$

$$\cos x = \frac{\exp(\mathrm{i}\, x) + \exp(-\mathrm{i}\, x)}{2} , \tag{3.537}$$

$$\sin x = \frac{\exp(\mathrm{i}\, x) - \exp(-\mathrm{i}\, x)}{2\mathrm{i}} . \tag{3.538}$$

Summation formulas

$$\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y , \tag{3.539}$$

$$\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y , \tag{3.540}$$

$$\operatorname{tg}(x \pm y) = \frac{\operatorname{tg} x \pm \operatorname{tg} y}{1 \mp \operatorname{tg} x \operatorname{tg} y} , \tag{3.541}$$

$$\operatorname{ctg}(x \pm y) = \frac{\pm \operatorname{ctg} x \operatorname{ctg} y - 1}{\operatorname{ctg} x \mp \operatorname{ctg} y} , \tag{3.542}$$

$$\sin x + \sin y = 2 \sin \tfrac{x+y}{2} \cos \tfrac{x-y}{2} , \tag{3.543}$$

$$\sin x - \sin y = 2 \cos \tfrac{x+y}{2} \sin \tfrac{x-y}{2} , \tag{3.544}$$

$$\cos x + \cos y = 2 \cos \tfrac{x+y}{2} \cos \tfrac{x-y}{2} , \tag{3.545}$$

$$\cos x - \cos y = -2 \sin \tfrac{x+y}{2} \sin \tfrac{x-y}{2} . \tag{3.546}$$

Double angle

$$\sin 2x = 2 \sin x \cos x , \tag{3.547}$$

$$\cos 2x = \cos^2 x - \sin^2 x , \tag{3.548}$$

$$\operatorname{tg} 2x = \frac{2 \operatorname{tg} x}{1 - \operatorname{tg}^2 x} , \tag{3.549}$$

$$\operatorname{ctg} 2x = \frac{\operatorname{ctg}^2 x - 1}{2 \operatorname{ctg} x} . \tag{3.550}$$

Half angle

$$\sin \frac{x}{2} = \pm \sqrt{\frac{1 - \cos x}{2}} , \tag{3.551}$$

$$\cos\frac{x}{2} = \pm\sqrt{\frac{1+\cos x}{2}}\,, \qquad \text{(sign based on quadrant)} \tag{3.552}$$

$$\mathrm{tg}\frac{x}{2} = \frac{1-\cos x}{\sin x} = \frac{\sin x}{1+\cos x}\,, \tag{3.553}$$

$$\mathrm{ctg}\frac{x}{2} = \frac{\sin x}{1-\cos x} = \frac{1+\cos x}{\sin x}\,, \tag{3.554}$$

Squares

$$\cos^2 x + \sin^2 x = 1\,, \tag{3.555}$$

$$\cos^2 x - \sin^2 x = \cos 2x\,, \tag{3.556}$$

$$\mathrm{ch}^2 x - \mathrm{sh}^2 x = 1\,, \tag{3.557}$$

$$\cos^2 x = \frac{1}{1+\mathrm{tg}^2 x}\,, \tag{3.558}$$

$$\sin^2 x = \frac{1}{1+\mathrm{ctg}^2 x}\,, \tag{3.559}$$

$$\cos^2 x = \frac{1+\cos 2x}{2}\,, \tag{3.560}$$

$$\sin^2 x = \frac{1-\cos 2x}{2}\,. \tag{3.561}$$

Conversion to tg(x/2) and cotg(x/2)

$$\sin x = \frac{2\,\mathrm{tg}(x/2)}{1+\mathrm{tg}^2(x/2)} = \frac{2\,\mathrm{ctg}(x/2)}{1+\mathrm{ctg}^2(x/2)}\,, \tag{3.562}$$

$$\cos x = \frac{1-\mathrm{tg}^2(x/2)}{1+\mathrm{tg}^2(x/2)} = \frac{\mathrm{ctg}^2(x/2)-1}{\mathrm{ctg}^2(x/2)+1}\,. \tag{3.563}$$

Shifts by $\pi/2$

$$\sin\left(x-\tfrac{\pi}{2}\right) = -\cos x\,, \tag{3.564}$$

$$\cos\left(x-\tfrac{\pi}{2}\right) = \sin x\,, \tag{3.565}$$

$$\mathrm{tg}\left(x-\tfrac{\pi}{2}\right) = -\mathrm{ctg}\, x\,, \tag{3.566}$$

$$\mathrm{ctg}\left(x-\tfrac{\pi}{2}\right) = -\mathrm{tg}\, x\,. \tag{3.567}$$

3.10.3 Operators in Curvilinear Coordinates

In many applications, Cartesian coordinates are not sufficient; instead, we need to use a curvilinear coordinate system (coordinate surfaces are not planes). We will denote Cartesian coordinates by **r** or **x**:

$$\mathbf{r} = (x, y, z)\,; \qquad \mathbf{x} = (x_1, x_2, x_3)\,.$$

Which notation we use depends solely on whether we need to number the axes or not. Both notations are equivalent. Similarly, in polar coordinates, we will use one of the following options

$$\mathbf{r} = (u, v, w); \qquad \mathbf{q} = (q_1, q_2, q_3)\,.$$

All four notations describe the coordinates of the same point. A vector can be understood as the difference between the coordinates of two nearby points (in the Cartesian system, these points do not have to be close together, but in the curvilinear coordinate system, they must be). We can then express the coordinate transformation as follows:

$$A_k = \mathrm{d}x_k = \frac{\partial x_k}{\partial q_l}\mathrm{d}q_l = T_{kl}\tilde{A}_l\,;$$

$$\tilde{A}_k = \mathrm{d}q_k = \frac{\partial q_k}{\partial x_l}\mathrm{d}x_l = U_{kl}A_l\,.$$

The untilded components of a vector are Cartesian, while tilded are curvilinear. We transform from one set to the other using a linear transformation, i.e., the new components are a linear combination of the old ones. The transformation matrices **T** and **U** are called the Jacobi matrices of the corresponding transformations:

$$T_{kl} \equiv \frac{\partial x_k}{\partial q_l}; \qquad U_{kl} \equiv \frac{\partial q_k}{\partial x_l}.$$

The two transformations are inverses of each other, i.e., the product of the two transformation matrices yields the identity matrix:

$$\left(\mathbf{T}\cdot\mathbf{U}\right)_{kl} = \frac{\partial x_k}{\partial q_n}\frac{\partial q_n}{\partial x_l} = \frac{\partial x_k}{\partial x_l} = \delta_{kl}\,.$$

Surfaces in curvilinear coordinates (in 2D coordinate lines) are defined by isosurfaces

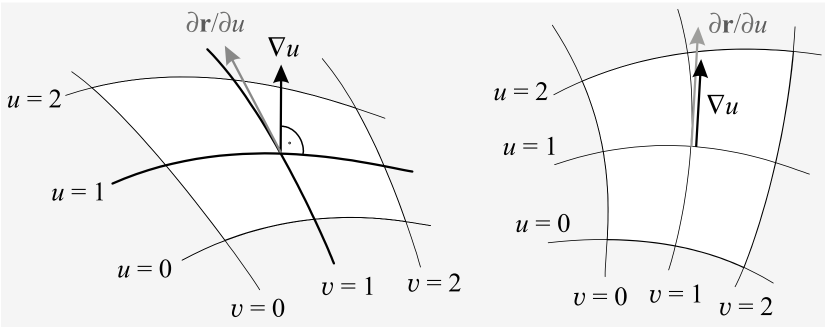
$$q_k = \mathrm{const}\,.$$

Curvilinear coordinates and Lamé coefficients

Let us now draw a coordinate grid with two curvilinear coordinates u and v. The normal to the coordinate surface u = const is given by the gradient $\mathbf{n} = \partial u/\partial \mathbf{r}$, and is shown in the figure at the selected point by a red vector. Conversely, the tangent from a given point along the second coordinate line is given by the vector $\boldsymbol{\tau} = \partial \mathbf{r}/\partial u$:

► $$\mathbf{n} = \nabla u = \frac{\partial u}{\partial \mathbf{r}} = \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}\right); \qquad \boldsymbol{\tau} = \frac{\partial \mathbf{r}}{\partial u} = \left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u}\right).$$

The components of both vectors are formed by elements of the Jacobian matrices, and their orientations are illustrated in the following figure. The first (black) is a normal to the coordinate plane, and the second (gray) is a tangent to the coordinate line.



Neither of these vectors is normalized, and they point in different directions because the coordinate system on the left is not orthogonal. Orthogonality can always be achieved by choosing an appropriate combination of coordinates (the coordinate planes and lines intersect at right angles). Commonly used cylindrical and spherical coordinates are orthogonal, therefore we will limit ourselves to orthogonal coordinates in which both the normal to the coordinate plane and the tangent to the coordinate line have the same direction (on the right in the figure) and point in the direction of the unit vector $\mathbf{e}_u$. Since the scalar product of the vectors $\mathbf{n}$ and $\boldsymbol{\tau}$ is clearly equal to one, we can write:

► $$\boldsymbol{\tau} = h_u \mathbf{e}_u \; ; \qquad \mathbf{n} = \frac{1}{h_u}\mathbf{e}_u \, .$$

The scaling parameter h_u is called the *Lamé coefficient* and expresses how many times longer the tangent vector is than the unit vector. If we multiply the unit vector by the Lamé coefficient, we obtain the vector $\boldsymbol{\tau}$; if we divide it by this coefficient, we obtain the vector $\mathbf{n}$ (both now point in the same direction). The coefficients are named after the French mathematician Gabriel Lamé. In the general case, the Lamé coefficients are given by the relations (we do not use summation notation in the rest of this chapter)

$$\boldsymbol{\tau}_k = \partial \mathbf{r}/\partial q_k = h_k \mathbf{e}_k \, .$$

From here, we immediately obtain the relationships for the individual Lamé coefficients

► $$h_k = \sqrt{\left(\frac{\partial x}{\partial q_k}\right)^2 + \left(\frac{\partial y}{\partial q_k}\right)^2 + \left(\frac{\partial y}{\partial q_k}\right)^2} \, .$$

If we have the relationships for the transformation from curvilinear to Cartesian coordinates, we can determine the Lamé coefficients. However, it is often more convenient to determine them geometrically, i.e., from the distance between two nearby points:

$$\mathrm{d}\mathbf{l} = \frac{\partial \mathbf{r}}{\partial q_1}\mathrm{d}q_1 + \frac{\partial \mathbf{r}}{\partial q_2}\mathrm{d}q_2 + \frac{\partial \mathbf{r}}{\partial q_3}\mathrm{d}q_3 = h_1\mathbf{e}_1\mathrm{d}q_1 + h_2\mathbf{e}_2\mathrm{d}q_2 + h_3\mathbf{e}_3\mathrm{d}q_3 \, .$$

For the distance element, we then have

$$\blacktriangleright \qquad \mathrm{d}l = \sqrt{h_1^2 \mathrm{d}{q_1}^2 + h_2^2 \mathrm{d}{q_2}^2 + h_3^2 \mathrm{d}{q_3}^2}\,. \qquad (3.568)$$

It is clear that this is a generalization of the Pythagorean theorem; under the square root are the squares of the individual legs (in 3D), and due to the curvilinear nature of the coordinates, the Lamé scaling parameters appear as coefficients. It is precisely this relationship that allows us to determine the Lamé coefficients geometrically. If we move infinitesimally in the direction of the k-th coordinate, the distance changes as

$$\mathrm{d}l_k = h_k\, \mathrm{d}q_k\,. \qquad (3.568')$$

We will therefore gradually move the point along each coordinate line, determine the corresponding distance, and "read off" the corresponding Lamé coefficient.

Cartesian coordinates (*x*, *y*, *z*)

In Cartesian coordinates, the coordinate planes are mutually perpendicular, and the shifts in the direction of the individual coordinate axes are

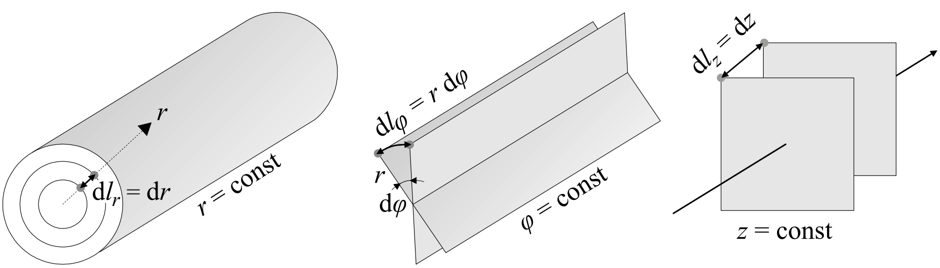
$$\blacktriangleright \qquad \begin{aligned} \mathrm{d}l_x &= \mathrm{d}x & &\Rightarrow & h_x &= 1 \\ \mathrm{d}l_y &= \mathrm{d}y & &\Rightarrow & h_y &= 1 \\ \mathrm{d}l_z &= \mathrm{d}z & &\Rightarrow & h_z &= 1 \end{aligned}$$

With these coefficients, the Pythagorean theorem (3.568) takes on its standard form: the distance between two points (in 3D, the diagonal of a rectangular prism) is the sum of the squares of the differences in their distances (the legs of the prism).

Cylindrical coordinates (*r*, *φ*, *z*)

Cylindrical coordinates have a preferred axis. The distance from this axis is denoted by r and is the first coordinate. Surfaces with a constant r coordinate are concentric cylindrical surfaces. The second coordinate is the azimuthal angle measured around the axis from some origin. We denote it by φ and it is the second coordinate. Surfaces of constant azimuthal angle form a fan of planes through which the axis passes. The last coordinate is the distance measured along the axis; we denote it by z. Surfaces of constant coordinate z form planes perpendicular to the axis. The system of all three surfaces is orthogonal, i.e., each is perpendicular to every other. The individual shifts and corresponding Lamé coefficients are (see figure):

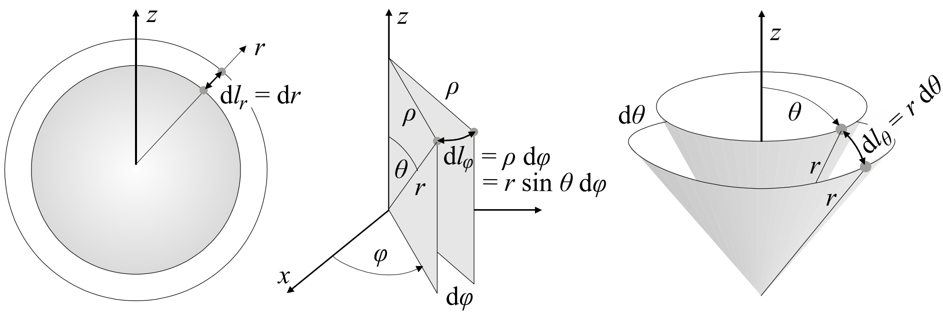
$$\blacktriangleright \qquad \begin{aligned} \mathrm{d}l_r &= \mathrm{d}r & &\Rightarrow & h_r &= 1 \\ \mathrm{d}l_\varphi &= r\,\mathrm{d}\varphi & &\Rightarrow & h_\varphi &= r \\ \mathrm{d}l_z &= \mathrm{d}z & &\Rightarrow & h_z &= 1 \end{aligned}$$



Spherical coordinates (r, φ, θ)

Spherical coordinates use a preferred axis with an origin through which the base coordinate plane passes. Typically, the z-axis is chosen as the axis and the xy-plane as the base plane. The first coordinate is the distance of the point from the origin, r. Coordinate surfaces with a constant r form the surfaces of concentric spheres. The second coordinate is the azimuthal angle φ measured in the base coordinate plane from the x-axis. Coordinate surfaces with a constant azimuth form a fan of planes with a common line (axis). The third coordinate is the deviation from the z-axis, denoted by θ. Coordinate surfaces form open cones with their vertices at the origin. The system of all three coordinate surfaces is orthogonal. The coordinates consist of a radial distance and two angles. The individual shifts and corresponding Lamé coefficients are (see figure):

$$\begin{aligned} \mathrm{d}l_r &= \mathrm{d}r & &\Rightarrow & h_r &= 1 \\ \mathrm{d}l_\varphi &= r\sin\theta\,\mathrm{d}\varphi & &\Rightarrow & h_\varphi &= r\sin\theta \\ \mathrm{d}l_\theta &= r\,\mathrm{d}\theta & &\Rightarrow & h_\theta &= r \end{aligned}$$



Gradient

Expressing the gradient in curvilinear coordinates is more or less straightforward:

$$\nabla f = \frac{\partial f}{\partial \mathbf{r}} = \sum_k \frac{\partial f}{\partial q_k}\frac{\partial q_k}{\partial \mathbf{r}} = \sum_k \frac{\partial f}{\partial q_k}\frac{1}{h_k}\mathbf{e}_k \qquad \Rightarrow$$

$$\nabla f = \left(\frac{1}{h_1}\frac{\partial}{\partial q_1}, \frac{1}{h_2}\frac{\partial}{\partial q_2}, \frac{1}{h_3}\frac{\partial}{\partial q_3}\right). \tag{3.569}$$

In Cartesian, cylindrical, and spherical coordinates, the gradient is given by:

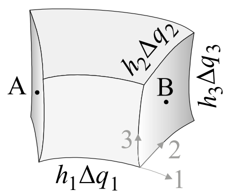
$$\begin{aligned} &\text{Cartesian} & \nabla f &= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right); \\ &\text{Cylindrical} & \nabla f &= \left(\frac{\partial f}{\partial r}, \frac{\partial f}{r\partial \varphi}, \frac{\partial f}{\partial z}\right); \\ &\text{Spherical} & \nabla f &= \left(\frac{\partial f}{\partial r}, \frac{\partial f}{r\sin\theta\,\partial \varphi}, \frac{\partial f}{r\,\partial \theta}\right). \end{aligned} \tag{3.569'}$$

Divergence

In Cartesian coordinates, we introduced divergence as a test for the sources of fields. We integrated the flux of the field over a small prism, which we then reduced to zero in the limit. A byproduct of the calculation was Gauss theorem

$$\oiint_{\partial\Omega} \mathbf{K}\cdot \mathrm{d}\mathbf{S} = \iiint_{\Omega} \operatorname{div}\mathbf{K}\, \mathrm{d}V,$$

which is stated independently of the choice of coordinate system. Let us now apply it in curvilinear coordinates to a small curvilinear prism $\mathcal{K}$, whose faces are coordinate surfaces.



Later, we will again reduce this prism to zero, i.e., to the point at which we calculate the divergence:

$$\operatorname{div}\mathbf{K} = \lim_{\Delta V\to 0}\frac{1}{\Delta V}\oiint_{\partial\mathcal{K}} \mathbf{K}\cdot \mathrm{d}\mathbf{S}\,.$$

This definition of divergence mirrors the definition in a Cartesian coordinate system, but is independent of the choice of coordinates. Divergence represents the flux density of the field through an elementary cuboid. Given the limit as the volume approaches zero, we can replace the field flux with the sum of the fluxes through the individual faces of the cuboid:

$$\operatorname{div}\mathbf{K} = \lim_{\Delta V\to 0}\frac{1}{\Delta V}\sum_{k=1}^{6}\mathbf{K}\cdot\Delta\mathbf{S}_k\,.$$

The calculation is straightforward, so we will demonstrate it only for the pair of walls perpendicular to the first axis (right and left). On both surfaces, only the K_1 component of the field is included in the flux; on the right wall, the flux is positive (the outward normal points in the direction of the K_1 component), and on the left wall, it is negative:

$$\operatorname{div}\mathbf{K} = \lim_{\Delta V\to 0}\frac{1}{h_1\Delta q_1 h_2\Delta q_2 h_3\Delta q_3}\begin{bmatrix} K_1(\mathrm{B})h_2(\mathrm{B})\Delta q_2\;\; h_3(\mathrm{B})\Delta q_3 - \\ -K_1(\mathrm{A})h_2(\mathrm{A})\Delta q_2\;\; h_3(\mathrm{A})\Delta q_3 + \cdots \end{bmatrix}.$$

$$\operatorname{div}\mathbf{K} = \lim_{\Delta V\to 0}\frac{1}{h_1h_2h_3}\left[\frac{K_1(\mathrm{B})h_2(\mathrm{B})h_3(\mathrm{B}) - K_1(\mathrm{A})h_2(\mathrm{A})\;\; h_3(\mathrm{A})}{\Delta q_1} + \cdots\right]$$

After taking the limits, the fractions become partial derivatives. We can either calculate the other contributions in a similar manner or derive them from a cyclic index change:

► $$\operatorname{div}\mathbf{K} = \frac{1}{h_1h_2h_3}\left[\frac{\partial h_2h_3K_1}{\partial q_1} + \frac{\partial h_3h_1K_2}{\partial q_2} + \frac{\partial h_1h_2K_3}{\partial q_3} + \right] \tag{3.570}$$

This is the general formula for divergence. For our three cases of divergence, we have

Cartesian $\operatorname{div}\mathbf{K} = \frac{\partial K_x}{\partial x} + \frac{\partial K_y}{\partial y} + \frac{\partial K_z}{\partial z}$;

► Cylindrical $\operatorname{div}\mathbf{K} = \frac{1}{r}\frac{\partial (rK_r)}{\partial r} + \frac{1}{r}\frac{\partial K_\varphi}{\partial \varphi} + \frac{\partial K_z}{\partial z}$; (3.570′)

Spherical $\operatorname{div}\mathbf{K} = \frac{1}{r^2}\frac{\partial (r^2 K_r)}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial K_\varphi}{\partial \varphi} + \frac{1}{r\sin\theta}\frac{\partial (\sin\theta\, K_\theta)}{\partial \theta}$.

Laplace operator for scalar function

The Laplace operator can be expressed as

$$\Delta f = \operatorname{div}\nabla f\ .$$

This expression does not depend on the choice of coordinates, and the Laplace operator can therefore be constructed from the formulas for the gradient and the divergence:

► $$\Delta f = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial K_1}{\partial q_1}\right) + \frac{\partial}{\partial q_2}\left(\frac{h_3 h_1}{h_2}\frac{\partial K_2}{\partial q_2}\right) + \frac{\partial}{\partial q_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial K_3}{\partial q_3}\right)\right]. \quad (3.571)$$

For the coordinate systems we are considering, the Laplace operator is

Cartesian $\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}$;

► Cylindrical $\Delta f = \frac{1}{r}\frac{\partial}{\partial r} r \frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \varphi^2} + \frac{\partial^2 f}{\partial z^2}$; (3.571′)

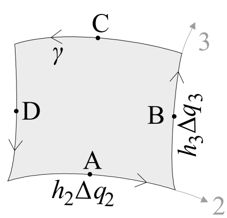
Spherical $\Delta f = \frac{1}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial f}{\partial r} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 f}{\partial \varphi^2} + \frac{1}{r^2\sin^2\theta}\frac{\partial}{\partial \theta}\sin\theta\frac{\partial f}{\partial \theta}$.

Rotation

For the last of the basic operators, we will proceed in a similar manner. In Cartesian coordinates, we introduced divergence as a test for the vorticity of fields. We integrated the circulation of the field over the boundary of a small rectangle, which we then reduced to zero in the limit. A byproduct of this calculation was Stokes theorem

$$\oint_{\gamma=\partial S} \mathbf{K}\cdot \mathrm{d}\mathbf{l} = \iint_S (\operatorname{rot}\mathbf{K})\cdot \mathrm{d}\mathbf{S},$$

which is expressed independently of the choice of coordinate system. Let us now apply it in curvilinear coordinates to a small curvilinear rectangle O, whose edges are formed by coordinate lines. Later, we will again reduce this rectangle to zero in the limit, i.e., to the point at which we calculate the rotation:



The result is a definition of the coordinate independent k^{th} component of rotation:

$$\left(\operatorname{rot}\mathbf{K}\right)_k = \lim_{\Delta S_k \to 0} \frac{1}{\Delta S_k} \oint_{\gamma=\partial O} \mathbf{K}\cdot \mathrm{d}\mathbf{l}\,.$$

Rotation corresponds to the surface density of the field's circulation through an elementary rectangle. We will calculate the first component; for the others, we will use a cyclic permutation. Given the limit as the volume approaches zero, we can replace the field flux with the sum of the fluxes through the individual walls of the rectangle:

$$\left(\operatorname{rot}\mathbf{K}\right)_1 = \lim_{\Delta S_1 \to 0} \frac{1}{\Delta S_1} \sum_{k=1}^{4} \mathbf{K}\cdot \Delta\mathbf{l}_k\,.$$

The calculation is straightforward. The integration curve is formed by the boundaries of a rectangle, which must be oriented in the mathematically positive direction (counter-clockwise). At the respective edges of the rectangle, only the component oriented in the direction of the edge is included in the circulation, and it is positive if it points in the direction of the integration curve and negative if it points in the opposite direction:

$$\left(\operatorname{rot}\mathbf{K}\right)_1 = \lim_{\Delta S_1 \to 0} \frac{1}{h_2\Delta q_2 h_3 \Delta q_3}\begin{bmatrix} K_2(\mathrm{A})h_2(\mathrm{A})\Delta q_2 + K_3(\mathrm{B})h_3(\mathrm{B})\Delta q_3 - \\ -K_2(\mathrm{C})h_2(\mathrm{C})\Delta q_2 - K_3(\mathrm{D})h_3(\mathrm{D})\Delta q_3 \end{bmatrix} =$$

$$= \lim_{\Delta S_1 \to 0} \frac{1}{h_2 h_3}\begin{bmatrix} \dfrac{K_3(\mathrm{B})h_3(\mathrm{B}) - K_3(\mathrm{D})h_3(\mathrm{D})}{\Delta q_2} - \\ - \dfrac{K_2(\mathrm{C})h_2(\mathrm{C}) - (K_2(\mathrm{A})h_2(\mathrm{A})}{\Delta q_3} \end{bmatrix} = \frac{1}{h_2 h_3}\left[\frac{\partial h_3 K_3}{\partial q_2} - \frac{\partial h_2 K_2}{\partial q_3}\right].$$

The general relationship for all three components of rotation will therefore be:

$$\left(\operatorname{rot}\mathbf{K}\right) = \frac{1}{h_1 h_2 h_3}\left(h_1\left[\frac{\partial h_3 K_3}{\partial q_2} - \frac{\partial h_2 K_2}{\partial q_3}\right], h_2\left[\frac{\partial h_1 K_1}{\partial q_3} - \frac{\partial h_3 K_3}{\partial q_1}\right], h_3\left[\frac{\partial h_2 K_2}{\partial q_1} - \frac{\partial h_1 K_1}{\partial q_2}\right]\right);$$

In Cartesian, cylindrical, and spherical coordinates for rotation, we have

Cartesian $$\operatorname{rot}\mathbf{K} = \left(\frac{\partial K_z}{\partial y} - \frac{\partial K_y}{\partial z}, \frac{\partial K_x}{\partial z} - \frac{\partial K_z}{\partial x}, \frac{\partial K_y}{\partial x} - \frac{\partial K_x}{\partial y}\right);$$

► Cylindrical $$\operatorname{rot}\mathbf{K} = \left(\frac{1}{r}\frac{\partial K_z}{\partial \varphi} - \frac{\partial K_\varphi}{\partial z}, \frac{\partial K_r}{\partial z} - \frac{\partial K_z}{\partial r}, \frac{1}{r}\frac{\partial}{\partial r}(rK_\varphi) - \frac{1}{r}\frac{\partial K_r}{\partial \varphi}\right); \quad (3.572)$$

Spherical $$\operatorname{rot}\mathbf{K} = \begin{pmatrix} \dfrac{1}{r\sin\theta}\dfrac{\partial}{\partial\theta}(\sin\theta\, K_\varphi) - \dfrac{1}{r\sin\theta}\dfrac{\partial K_\theta}{\partial\varphi}, \\ \dfrac{1}{r\sin\theta}\dfrac{\partial K_r}{\partial\varphi} - \dfrac{1}{r}\dfrac{\partial}{\partial r}(rK_\varphi), \\ \dfrac{1}{r}\dfrac{\partial}{\partial r}(rK_\theta) - \dfrac{1}{r}\dfrac{\partial K_r}{\partial\theta} \end{pmatrix}.$$

3.10.4 Some Integrals and Series

In relationships, it is referred to as $n! = n(n-1)\ldots 1$; $n!! = n(n-2)(n-4)\ldots 1$.

$$\int_0^\infty x^n \, e^{-ax} \, dx = \frac{n!}{a^{n+1}}; \quad a > 0; \quad n = 1, 2, \ldots \tag{3.573}$$

$$\int_0^\infty x^{2n} \, e^{-ax^2} \, dx = \frac{(2n-1)!!\sqrt{\pi}}{2^{n+1} a^{(2n+1)/2}}; \quad a > 0; \quad n = 1, 2, \ldots \tag{3.574}$$

$$\int_0^\infty x^{2n+1} \, e^{-ax^2} \, dx = \frac{n!}{2a^{n+1}}; \quad a > 0; \quad n = 0, 1, 2, \ldots \tag{3.575}$$

$$\int_{-\infty}^{\infty} e^{-ax^2} \, dx = \sqrt{\frac{\pi}{a}}; \qquad a > 0 \qquad \text{(Gauss integral)} \tag{3.576}$$

$$\int_0^\infty e^{-ax^2} \, dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}; \qquad a > 0 \tag{3.577}$$

$$\int_{-\infty}^{\infty} e^{-ax^2 + bx} \, dx = \sqrt{\frac{\pi}{a}} \, e^{-b^2/4a}; \qquad a > 0 \tag{3.578}$$

$$\int \frac{1}{\sqrt{a^2 - x^2}} \, dx = -\arccos\left(\frac{x}{a}\right) \tag{3.579}$$

$$\int \frac{1}{\sqrt{a^2 + x^2}} \, dx = \operatorname{ash}\left(\frac{x}{a}\right) \tag{3.580}$$

$$\int \frac{x}{\sqrt{a^2 + x^2}} \, dx = \sqrt{a^2 + x^2} \tag{3.581}$$

$$\int_0^\infty \frac{x^3}{e^x + 1} \, dx \cong 5,6822; \tag{3.582}$$

$$\int_0^\infty \frac{x^3}{e^x - 1} \, dx \cong \frac{\pi^4}{15}; \tag{3.583}$$

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1 - q}; \; |\, q \,| < 1 \quad \text{(Sum of a geometric series)} \tag{3.584}$$

$$V_{2N} = \pi^N R^{2N} / N! \qquad \text{(Sphere volume in even number of dimensions)} \tag{3.585}$$

Calculation of the Gauss integral

The Gauss integral can be easily determined using a simple trick in which we convert this integral into an integral over an infinite plane in polar coordinates. The individual steps are straightforward, so we will not comment on them:

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-ax^2}\,\mathrm{d}x = \sqrt{\left(\int_{-\infty}^{+\infty} \mathrm{e}^{-ax^2}\,\mathrm{d}x\right)\left(\int_{-\infty}^{+\infty} \mathrm{e}^{-ax^2}\,\mathrm{d}x\right)} =$$

$$= \sqrt{\left(\int_{-\infty}^{+\infty} \mathrm{e}^{-ax^2}\,\mathrm{d}x\right)\left(\int_{-\infty}^{+\infty} \mathrm{e}^{-ay^2}\,\mathrm{d}y\right)} = \sqrt{\iint_{\mathcal{R}\times\mathcal{R}} \mathrm{e}^{-a(x^2+y^2)}\,\mathrm{d}x\,\mathrm{d}y} =$$

$$= \sqrt{\int_0^{\infty}\int_0^{2\pi} \mathrm{e}^{-ar^2}\, r\,\mathrm{d}\varphi\,\mathrm{d}r} = \sqrt{2\pi\int_0^{\infty} \mathrm{e}^{-ar^2}\, r\,\mathrm{d}r}$$

We can easily calculate the last integral using the substitution $\xi = \alpha r^2$, and the result is

$$\int_{-\infty}^{+\infty} \mathrm{e}^{-ax^2}\,\mathrm{d}x = \sqrt{\frac{\pi}{a}}\,. \tag{3.586}$$

We can easily convert the result into an integral (3.577) ranging only from zero to infinity. By repeatedly differentiating equation (3.577) with respect to the parameter a, we obtain equations (3.574) for even powers. The procedure is the same for odd powers. We can easily compute the integral (3.575) for $n = 0$. We obtain higher powers again by differentiating with respect to the parameter a.

Calculation of the integral in the Stefan-Boltzmann law

First, we rewrite the integral as follows:

$$\int_0^{\infty} \frac{x^3}{\mathrm{e}^x - 1}\,\mathrm{d}x = \int_0^{\infty} x^3\,\mathrm{e}^{-x}\left(\frac{1}{1-\mathrm{e}^{-x}}\right)\mathrm{d}x\,.$$

The expression in round brackets can be interpreted as the sum of a geometric series with a quotient of $q = \mathrm{e}^{-x}$, i.e.

$$\int_0^{\infty} \frac{x^3}{\mathrm{e}^x - 1}\,\mathrm{d}x = \int_0^{\infty}\left(x^3\,\mathrm{e}^{-x}\sum_{n=0}^{\infty}\mathrm{e}^{-nx}\right)\mathrm{d}x = \int_0^{\infty}\left(x^3\sum_{n=0}^{\infty}\mathrm{e}^{-(n+1)x}\right)\mathrm{d}x =$$

$$= \int_0^{\infty}\left(x^3\sum_{n=1}^{\infty}\mathrm{e}^{-nx}\right)\mathrm{d}x = \sum_{n=1}^{\infty}\int_0^{\infty} x^3\,\mathrm{e}^{-nx}\,\mathrm{d}x\,.$$

The last integral is easy to solve. We perform integration by parts three times in a row, thereby gradually reducing the power of x, and finally perform a simple integration of the exponential. The result is

$$\int_0^{\infty} \frac{x^3}{\mathrm{e}^x - 1}\,\mathrm{d}x = \sum_{n=1}^{\infty}\frac{6}{n^4}\,. \tag{3.587}$$

All that remains is to determine the sum of the Riemann series on the right-hand side of the equation. First, we will determine the sum of the simpler series involving squares, and only then the series involving fourth powers. We will therefore find the sums step by step

$$S_2 = \sum_{n=1}^{\infty} \frac{1}{n^2}; \qquad S_4 = \sum_{n=1}^{\infty} \frac{1}{n^4}. \tag{3.588}$$

We can easily determine both sums from the Fourier series of the functions x^2 and x^4 on the interval $<-\pi, +\pi>$:

$$\begin{aligned} x^2 &= \frac{\pi^2}{3} + \sum_{n=1}^{\infty} (-1)^n \frac{4}{n^2} \cos(nx); \\ x^4 &= \frac{\pi^4}{5} + \sum_{n=1}^{\infty} (-1)^n \left(\frac{8\pi^2}{n^2} - \frac{48}{n^4} \right) \cos(nx). \end{aligned} \tag{3.589}$$

Substituting $x = \pi$, we obtain the value S_2 from the first expansion and the value S_4 from the second (we will need the value S_2 for the calculation):

$$S_2 = \frac{\pi^2}{6}; \qquad S_4 = \frac{\pi^4}{90}. \tag{3.590}$$

The following holds for the integral in question:

$$\int_0^{\infty} \frac{x^3}{e^x - 1} \, dx = \sum_{n=1}^{\infty} \frac{6}{n^4} = 6\, S_4 = 6 \frac{\pi^4}{90} = \frac{\pi^4}{15}.$$

The result of the integration is

$$\int_0^{\infty} \frac{x^3}{e^x - 1} \, dx = \frac{\pi^4}{15}. \tag{3.591}$$

3.10.5 Expansion of Certain Functions

$$\exp x \equiv 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots \tag{3.592}$$

$$\operatorname{ch} x \equiv 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \frac{x^8}{8!} + \cdots \tag{3.593}$$

$$\cos x \equiv 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} \pm \cdots \tag{3.594}$$

$$\operatorname{sh} x \equiv x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \frac{x^9}{9!} \pm \cdots \tag{3.595}$$

$$\sin x \equiv x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} - \frac{x^9}{9!} \pm \cdots \tag{3.596}$$

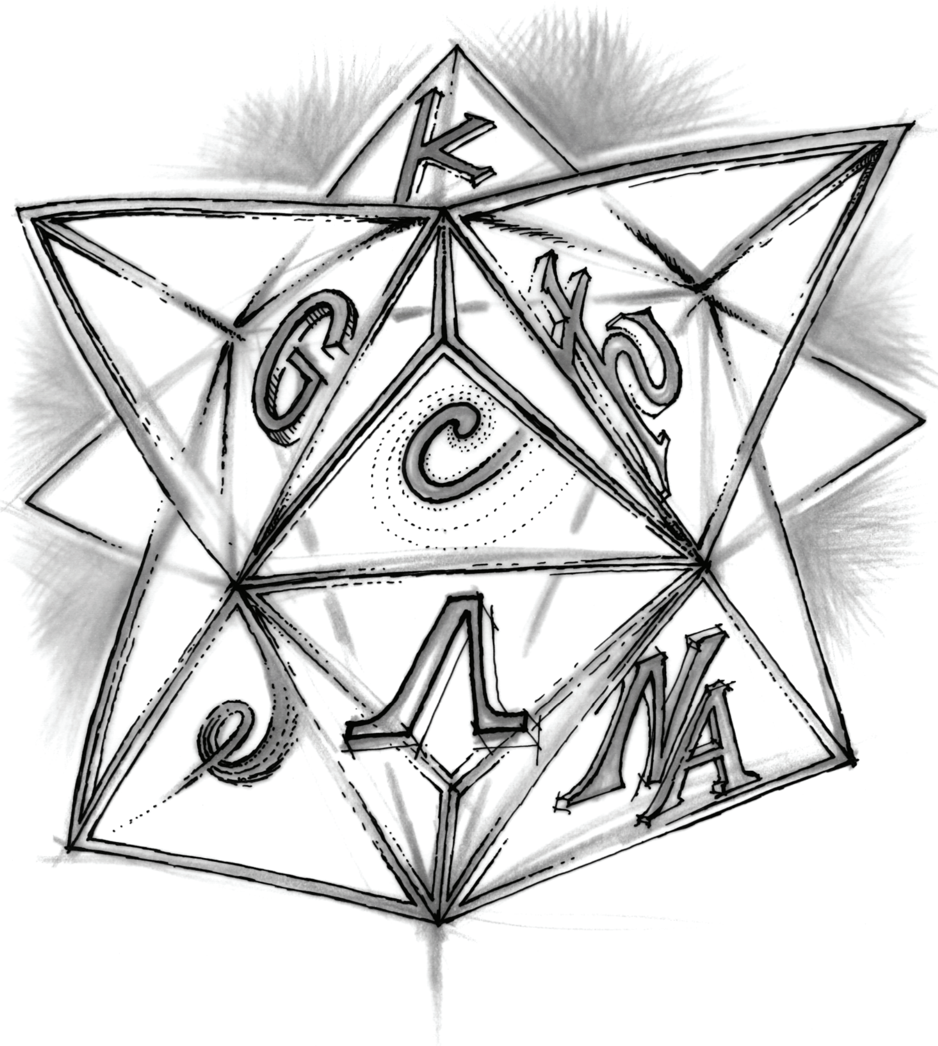
$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \quad \text{(Geometric series, } x < 1\text{)} \tag{3.597}$$

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} \mp \cdots \tag{3.598}$$

$$\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{x^3}{16} - \frac{5x^4}{128} + \frac{7x^5}{256} - \cdots \tag{3.599}$$

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \sum_{l=0}^{\infty} \frac{\min^l(r,r')}{\max^{l+1}(r,r')} P_l(\cos\theta)\,; \quad P_l \text{ je Legendre polynomial} \tag{3.560}$$

List of Symbols



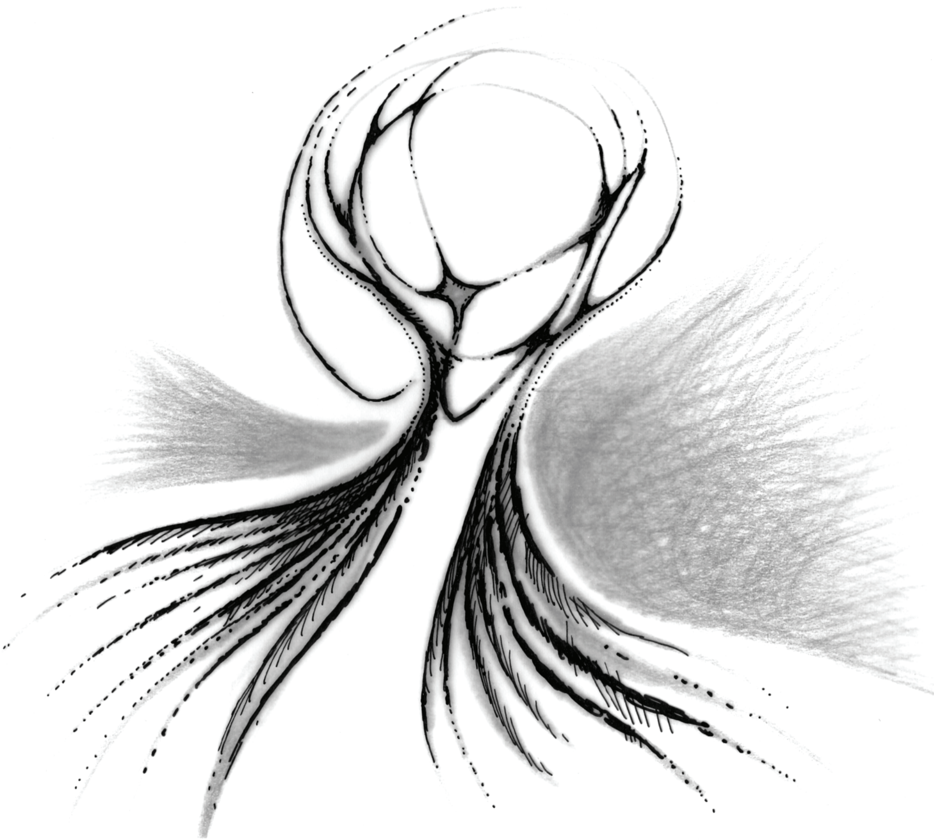
$\hat{\mathbf{1}}$	unit matrix
a	major axis of an ellipse
ash	inverse (area) hyperbolic sine
atg	arctangent
ath	inverse (area) hyperbolic tangent
$\mathbf{a}_k$	crystal lattice vectors
$\hat{\mathbf{a}}$	annihilation operator
$\hat{\mathbf{a}}^\dagger$	creation operator
A	amplitude dynamical variable mechanical work
A^μ	four-potential of a field
$\mathbf{A}$	vector potential displacement vector
A	matrix stability matrix
$\hat{\mathbf{A}}$	operator corresponding to A
$\mathcal{A}$	atractor
$\mathscr{A}$	probability amplitude
Ai	Airi funcion
$\mathbf{b}$	angular momentum
$\mathbf{b}_k$	reciprocal lattice vectors
b	semi-major axis of an ellipse angular momentum
$\hat{\mathbf{b}}$	annihilation operator
$\hat{\mathbf{b}}^\dagger$	creation operator
$\mathbf{B}$	magnetic induction
Bi	Airi function
c	velicity velocity of light
ch	hyperbolic cosine
cos	cosine
cosec	cosecant
ctg	cotangent
c_{kl}^m	structural coefficients of a Lie algebra
$\hat{\mathbf{c}}$	projection of annih. operator
$\hat{\mathbf{c}}^\dagger$	projection of creat. operator
$\mathscr{C}$	capacity
C	cisrculation (field) limit cycle complex numbers
C^3	space of complex triples
C^N	space of complex N-tuples
det	determinant
div	divergence
$\mathrm{d}\mu$	measure element
$\mathrm{d}\omega$	differential form
D	diffusion coefficient
$\mathbf{D}$	magnetic field induction
$\hat{\mathbf{D}}$	derivative operator
$\mathcal{D}$	densely covered set
e	rulet number
e	elementary charge excentricity
$\mathbf{e}_k$	unit vector
exp	exponential function
E, $\mathscr{E}$	energy
$\mathbf{E}$	elektric fielt intensity
f	frequency function number of degrees of freedom
F, $\mathbf{F}$	force
F_k	force components
$\mathscr{F}$	Fourier transform
$F^{\mu\nu}$	electromagnetic field tensor
$\boldsymbol{g}$	gravity acceleration
g	magnitude of gravity acceleration degree of degeneration
$g^{\mu\nu}$	metric tensor
G	Chandrasekhar funtion approximation gravitational constant Green function helicity
$\mathbf{G}$	displacement in k-space
h	Planck constant increment damping in the well
$\hbar$	reduced Planck constant
$\mathbf{H}$	magnetic field intensity
H	Hamilton function height
H_n	Hermite polynomial
$\mathscr{H}$	energy density
$\mathcal{H}$	Hilbert space
$\mathcal{H}$	helicity density
$\hat{\mathbf{H}}$	Hamilton operator
I	electric current intensity

	isospin
	moment of inertia
$\mathbb{I}$	unit matrix
I_m	Modified Bessel function of the first kind
Im	imaginary part
I_3	isospin projection
$\mathbf{j}_Q$	current density
	charge flux
j^μ	four-current
J	adiabatic invariant
	coupling constant
J_m	Bessel function of the first kind
$\mathcal{J}$	invariant set
k	rate of reaction
	oscillation stiffness
	wave vector magnitude
$\mathbf{k}$	wave vector
k^μ	wave four-vector
K	integration constant
$\mathscr{K}$	annulus
K_m	modified Bessel function of the second kind
l	azimuthal quantum number
	distance
ln	natural logarithm
ℓ^2	space of sequences
L	Lagrange function
	angular momentum
	well width
	distance
$\mathbf{L}$	angular momentum vector
$\hat{\mathbf{L}}$	angular momentum operator
$\mathcal{L}^2$	space of functions
L_k	Lagrange points
$\mathscr{L}$	Lagrange function density
	inductance
	Laplace transform
m	mass
	magnetic quantum number
m_S	magnetic spin number
$\mathbf{M}$	magnetization vector
M	mass
	magnetization
$\mathbb{M}$	generator of rotations
$\mathcal{M}_O$	open set
$\mathcal{M}_U$	closed set
n	concentration
	principal quantum number
$\mathbf{n}$	normal vector
N	number of particles
	number of parameters
$\hat{N}$	particle number operator
$\hat{\mathcal{N}}$	particle density operator
$\mathbf{p}$	momentum vector
p	momentum
	cone section parameter
	pressure
$\mathbf{p}_E$	electric dipole moment
$\mathbf{p}_M$	magnetic dipole moment
p^μ	four-momentum
$\mathbf{P}$	polarization vector
P	momentum
	polarization
	photon polarization
	probability
P_l	Legendre polynomial
P_{lm}	associated Legendre polynomial
$\mathscr{P}$	field momentum density
$\hat{\mathbf{P}}$	projection operator
	particle exhange operator
$\mathbf{q}$, q_k	generalized coordinates
Q	charge
	heat
	generalized coordinate
r	radial distance
	radius
$\mathbf{r}$	position vector
rot	rotation
R	reflectivity
	number of bonds
	Rayleigh function
$\mathscr{R}$	electric resistance
$\mathcal{R}$	real numbers
$\mathcal{R}^3$	the space of real triples
$\mathcal{R}^N$	the space of real N-tuples
R_L	Larmor radius
$\mathbb{R}$	rotational matrix
$\mathbb{R}_{\text{inf}}$	infinitesimal rotation matrix
Re	real part
Res	residue

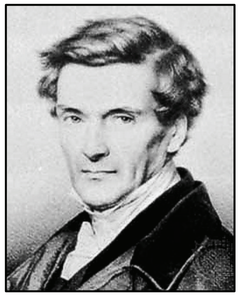
s	spin
	distance
$\mathbf{s}$	direction vector
sec	secant
sh	hyperbolic sine
sin	sine
S	action
	scattering matrix
	strangeness
	transformation matrix
S.	scalar part of a quaternion
$\hat{\mathbf{S}}$	spin operator
$\mathbf{S}$, S	area (surface)
$\mathcal{S}$	strange atractor
$\mathscr{S}$	Schwartz (Sobolev) space
S_{kl}	change-of-basis matrix
t	time
T	absolute temperature
	kinetic energy
	transmission coefficient
	period
	temperature
$\hat{\mathbf{T}}$	kinetic energy operator
$\hat{T}_g$	tempered distribution
$T^{\mu\nu}$	energy-momentum tensor
u	rapidity
	internal energy density
$\mathbf{u}$	velocity field
U	transformation matrix
	internal energy
U_ε	ε-neighborhood
$\hat{\mathbf{U}}$	time-evoluton operator
$\mathbf{v}$, v	velocity
V	mixing matrix
	volume
	potential energy
V	vector part of quaternion
	kvaternionu
V.P.	Value Principal
	(Cauchy principal value)
$\mathscr{V}$	potential energy density
	generating function
$\mathcal{V}$	linear vector space
$\hat{\mathbf{V}}$	potential energy operator
$\hat{\mathscr{V}}$	virial operator
w	probability
W	energy
x_k	coordinates
$\mathbf{x}$	position vector
x^μ	event
$\mathcal{X}$	chaotic set
y	coordinate
Y	hypercharge
Y_m	Bessel function
	of the second kind
Y_{lm}	spherical harmonic
z	complex variable
	coordinate
Z	degree of ionization
α	coefficient
	constant
	angle
α^k	Dirac matrices
β	constant
	relativistic coefficient v/c
	Dirac matrix
γ	integration path
	Lorentz factor
γ^k, γ^5	Dirac matrices
Γ^k	base matrices
δ	Dirac distribution
	Kronecker symbol
	equation parameter
	angle (wave phase)
Δ	finite increment
	Laplace operator
ε	energy of one state
	left hand side of the equation
	Levi-Civita tensor
	numerical eccentricity
	permittivity
	control parameter
$\boldsymbol{\eta}$	eigenvector
θ	deviation from the z-axis
κ	electrical susceptibility
	constant in the KG equation
	thermal conductivity
	coefficient
λ	eigenvalue
	Lyapunov exponent
	parameter
	wavelength
	latitude

$\mathbf{\Lambda}$	Lorentz matrix
μ	dimensionless mass
	integration factor
	permeability
	reduced mass
ν	frequency
	radial quantum number
	neutrino wave function
$\mathbf{\nu}$	normal vector
$\boldsymbol{\xi}$, ξ	phase variable
	dimensionless parameter
	auxiliary variable
π	canonical conjugate
ρ	density
	distance
ρ_Q	charge density
σ	spin
	Pauli matrix
$\boldsymbol{\sigma}$	unit direction vector
τ	linear density
$\Sigma^{\alpha\beta}$	basis matrices
φ	phase
	field
	angle
	system state
$\hat{\varphi}$	field operator
ϕ	phase
	phase space
	error function
	scalar potential
	field flux
χ	magnetic susceptibility
ψ	Chandrasekhar function
	system state
	wave function
ω	angular frequency
ω_c	cyclotron frequency
Ω	spacetime region
∇	gradient
$\square$	D'Alembert operator

List of Personalities



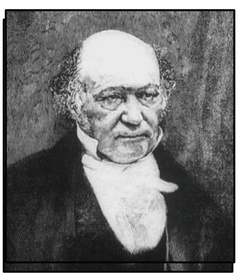
Theoretical Mechanics



Coriolis, Gustave Gaspard (1792–1843), a French mathematician and physicist, worked in the fields of mathematical analysis, mechanics, and hydraulics. He became famous for his calculations of the forces acting in rotating systems. He also studied friction. He was the first to use the term "mechanical work" in connection with the action of forces on bodies and applied the correct formula for kinetic energy. In one of his works, he also explored the theory of billiard ball collisions. In 1829, he became a professor of mechanics at the École Centrale Paris. His name is one of the 72 names engraved on the Eiffel Tower.

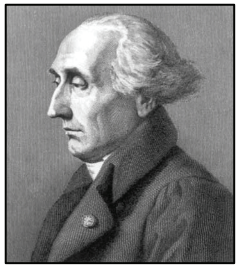
Euler, Leonhard (1707–1783), Swiss mathematician and astronomer, a student of Johann Bernoulli. He worked at the Academy in St. Petersburg and at the Academy of Sciences in Berlin. He had a phenomenal memory and once settled a dispute between two students whose results of a complex calculation differed by the fiftieth decimal place by simply calculating the answer in his head. In 1735, Euler lost sight in his right eye, and in 1766, in his left. Nevertheless, he continued to publish his results, which he dictated. Euler was the most prolific mathematician of all time (even though he had 13 children). During his lifetime, he published over 800 works. He was awarded the prize of the Paris Academy twelve times. When asked to explain his astonishing productivity, he replied: "*It seems that my pen is more intelligent than I am.*"

Independently of Lagrange, he discovered the necessary conditions for minimizing a functional in the calculus of variations. In physics, these equations are known as Lagrange equations of motion. He also worked in theoretical astronomy. He investigated the problem of the motion of three or more bodies and proved that no analytical solution exists. He theoretically analyzed the motion of the Moon (a problem ultimately solved by Laplace) and the perturbations of the orbits of Jupiter and Saturn. Euler theoretically derived a method for correcting chromatic aberration in refracting telescopes (practically demonstrated by the English lawyer and mathematician Chester Moor Hall in 1733).

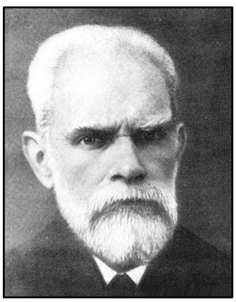


Hamilton, William Rowan (1805–1865), a prominent Irish mathematician. Under the guidance of his uncle, a linguist, he learned to speak fourteen languages. At the age of seventeen, he discovered an error in Laplace's treatise *Celestial Mechanics*. He predicted conical refraction in biaxial crystals, which was soon experimentally confirmed by Humphrey Lloyd. Hamilton also extended the principle of least energy, described by Maupertuis, to Hamilton principle, a fundamental variational principle in theoretical mechanics that leads to Lagrange equations. In differential calculus, the Hamilton operator is named after him; in physics, there are Hamilton equations of motion, the Hamilton function, and the Hamilton–Jacobi equation. He spent the last third of his life under the influence of alcohol.

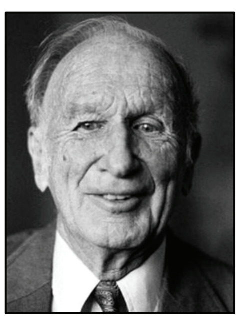
Hopf, Eberhard Frederich Ferdinand (1902–1983), a German mathematician and astronomer who was born in Austro-Hungarian monarchy (in Salzburg). He is the founder of ergodic theory and bifurcation theory (the branching of solutions to differential equations). He worked on partial differential equations, integral equations, fluid mechanics, and differential geometry. He discovered the maximum principle in the theory of elliptic differential equations. He studied at the University of Berlin, where he habilitated in 1929. From 1931, he worked at MIT in the United States. In 1936, he returned to Germany, working at universities in Leipzig and Munich. From 1949 until his death, he worked at Indiana University in Bloomington. In the theory of differential equations, Hopf bifurcation is named after him.



Lagrange, Joseph (1736–1813), a French mathematician and theoretical physicist, was one of the most prominent scientific figures of the eighteenth century. Lagrange succeeded Euler as director of the Berlin Academy. His 1788 work *Mécanique Analytique* was a comprehensive treatment of mechanics from a mathematical perspective. Lagrange became a co-founder of the calculus of variations (in mathematics, he solved problems similar to those of Euler). He understood mechanical problems as the search for an optimal trajectory based on an integral principle. He discovered the necessary conditions for the existence of a solution, which are the equations of motion. Today, these equations are called Lagrange equations and form the foundation of theoretical mechanics. Lagrange points – a set of five equilibrium points in the vicinity of two mutually orbiting bodies – are also named after him. One of his sayings goes: "*I have always observed that people's demands are inversely proportional to what they truly deserve. This is one of the-foundations of morality.*"

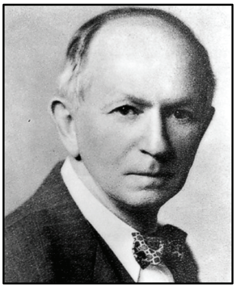


Lyapunov, Alexander Mikhailovich, (1857–1918), a Russian mathematician whose fundamental work focused on differential equations, potential theory, solution stability, and probability theory. Lyapunov stability – the stability of solutions to differential equations with respect to perturbations of the initial conditions – is named in his honor. In the field of physics, he studied the stability conditions of rotating liquids. He studied at the University of St. Petersburg. Among his teachers was Chebyshev. He completed his studies in 1880. In 1880, he received a gold medal for his work on hydrostatics. In 1895, he became a head of the Department of Mechanics at the University of Kharkiv. In 1902, he returned to St. Petersburg. In 1917, he moved to Odessa with his seriously ill wife. In 1918, his wife died of tuberculosis. Lyapunov wanted to end his life and shot himself in the head. He died three days later from his injuries.



Lorenz, Edward Norton (1917–2008), American mathematician and meteorologist, co-founder of the theory of deterministic chaos. He discovered the strange attractor and was the first to use the term "butterfly effect" to describe instability. He studied mathematics at Dartmouth College in New Hampshire and at

Harvard. During World War II, he provided weather forecasts for the military. After the war, he studied meteorology at MIT, where he became a professor. He developed a mathematical model of the movement of air masses in the atmosphere. He is the recipient of numerous awards and medals. The Lorenz strange attractor is named after him.



Lotka, Alfred James (1880–1949), an American mathematician and chemist. He is best known for applying physical methods to biology, particularly in his work on population dynamics and energetics. He proposed an equation describing the change in the number of individuals in a system composed of predators and prey. The Italian mathematician Vitto Volterra independently derived this equation. Evolutionary equations of this type are today called Lotka–Volterra equations. The equations are also used in other systems of two competing groups. Lotka was born in Lviv, in what is now Ukraine (then part of Austria-Hungary). His parents were American. He studied in Birmingham, Leipzig, and at Cornell University in the United States.

Newton Isaac (1642–1727), is considered one of the most significant scientists in human history. He was an English physicist, mathematician, astronomer, philosopher, and theologian. Newton laid the foundations of classical mechanics in his three laws of motion (the law of inertia, the law of force, and the law of action and reaction). The law of force became the very first mathematical tool for predicting the trajectory of bodies. To solve the equation of motion, Newton developed the foundations of differential and integral calculus; independently of him differential calculus discovered Gottfried Leibniz. Newton proposed a force formula for gravitation. It applies to the motion of bodies both on Earth and in space. Furthermore, in mechanics, Newton dealt with the laws of conservation of momentum and angular momentum. Newton conception of mechanics employs absolute space and time existing independently of bodies. Newton published the complete foundations of mechanics in 1687 in the Principia (*Philosophiæ Naturalis Principia Mathematica*), the publication of which was sponsored by Edmond Halley, who became famous for predicting the return of Halley Comet based on Newton law of gravitation.

Isaac Newton (1642–1727) Newton’s interests, however, were not limited to mechanics. He also studied optics. He constructed a reflecting telescope with an eyepiece positioned perpendicular to the instrument’s optical axis. Using a prism, he split light into its individual colors. Unlike Huygens, he conceived of light as a stream of particles. Today we know that both were right. We call this phenomenon, which was only clarified on the basis of quantum mechanics, wave-particle duality.

In mathematics, Newton introduced integral and differential calculus, generalized the binomial theorem, and worked on the numerical solution of transcendental and differential equations. Newton also devoted much of his time to alchemy and had his own laboratory. Most of his writings are devoted to religious issues. The following are named after Newton: Newton laws of motion, Newton method, the Newtonian telescope, the unit of force known as the Newton, and craters on Mars and the Moon.

Noether, Emmy (1882–1935), an outstanding German-American mathematician who demonstrated that every symmetry in nature is closely linked to a conservation law. A conserved quantity is directly defined by the given symmetry. She worked in Erlangen and Göttingen with figures such as Felix Klein and David Hilbert. Her work in the field of invariant theory contributed to the final form of the general theory of relativity formulated by Albert Einstein in 1916. Emmy Noether was likely the first woman ever to hold an academic degree, as habilitation had previously been granted only to men.

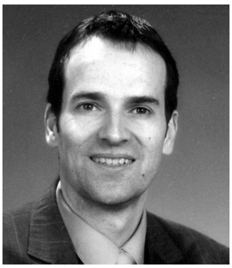
Pol, Balthasar (1889–1959), full name Balthasar van der Pol, Dutch physicist. Van der Pol studied physics at the University of Utrecht, where he earned his PhD in 1920. He conducted experiments on the propagation of electromagnetic waves. In the theoretical field, he studied electrical circuits, worked on mathematical physics and the theory of differential equations. In 1935, he was awarded the IEEE Medal for his work. The van der Pol oscillator and the asteroid 10443 are named after him.

Rayleigh, John William Strutt (1842–1919), an English baron who studied physics, acoustics, and optics, particularly the propagation of waves in fluids. Poor health prevented him from completing his studies at two schools (Eton and Harrow). In 1857, he began four years of private study. In 1861, he entered Trinity College, Cambridge. He completed his studies in 1865. He devoted to Maxwell theory of electromagnetism, both experimentally and theoretically. In 1878, he published the two-volume work *The Theory of Sound*, which became a cornerstone of acoustic literature. He derived an equation describing the dependence of light scattering in the atmosphere on wavelength, thereby becoming the first to explain the blue color of the sky. Like many others, he also attempted to derive the law of blackbody radiation. His relationship (Rayleigh law) correctly describes the radiation intensity for long wavelengths. For short wavelengths, the intensity diverges. It was not until Max Planck that the law was successfully derived for the entire spectrum. He received the Nobel Prize in 1904 for isolating inert atmospheric argon. He competed with William Ramsay for credit for this discovery, but Ramsay demonstrably began his work only after the publication of Rayleigh's results. Ramsay received the Nobel Prize in Chemistry for his long-term research on the properties of argon in the same year. In theoretical mechanics, Rayleigh introduced the dissipation function describing losses caused by the conversion of energy into heat. This function is called the Rayleigh dissipation function.

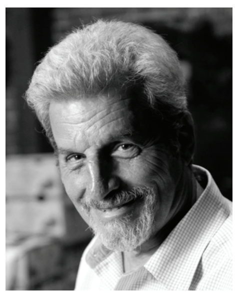
Tonti, Enzo (1935–2021), an Italian theoretical physicist, was born in Milan, where he completed his studies in mathematics and physics in 1961. He then worked at the Polytechnic University of Milan. In 1975, he became a professor at the University of Milan. Since 1976, he has worked at the Faculty of Engineering in Trieste. He has focused on the mathematical structure of physical theories. Tonti was fascinated by the analogies between various physical theories. He is best known for his work in the field of variational formulations of physical theories. He discovered the conditions that a system of differential equations must satisfy in order to be formulated variationally (*Tonti conditions*).

Volterra, Vito (1860–1940), an Italian mathematician and physicist who contributed to the application of mathematics to the biological and social sciences. Independently of Alfred Lotka, he formulated evolutionary equations for two competing groups. He also worked on integral equations. He studied at the University of Pisa and became a professor at the University of Turin in 1892. In the period of World War II, he refused to participate in the practices of the Nazi leader Benito Mussolini. The Lotka–Volterra evolutionary equations are named after him.

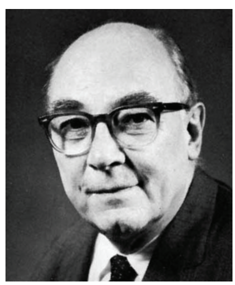
Quantum Theory



Abele, Hartmut, a German-Austrian quantum physicist. He graduated from the University of Heidelberg. He gained a deeper understanding of quantum physics and neutrons during a research stay at the Laue-Langevin Institute in Grenoble, France. He defended his doctoral thesis on neutrino oscillations in Heidelberg and then worked at Yale University in the US. Upon his return, he became a professor at the universities of Heidelberg and Munich. Since 2008, he has been working at the University of Vienna. His team set up a cold neutron experiment (2011) in which the quantum states of a neutron in a gravitational field were observed. This was the first time that manifestations of gravity had ever been detected for an elementary particle. Abele thus paved the way for the study of gravity at the microscopic level.



Aharonov, Yakir (1932), Israeli quantum physicist. His work focuses on non-local phenomena, quantum field theory and its interpretations. In 1959, together with David Bohm, he proposed a thought experiment in which an electron beam is influenced by a region of zero magnetic field in which there is a non-zero vector potential. The Aharonov–Bohm effect was experimentally demonstrated in 1986 by the Japanese physicist Akira Tomonura. Classical electrodynamics is incomplete; only quantum theory describes phenomena in which electromagnetic potentials change the phase of the wave function. Potentials are thus not a mathematical tool, but a real physical entity. In 1988, Aharonov published the concept of so-called weak measurement, which does not affect quantum state of the measured object. Aharonov studied in Haifa, Israel, and later worked at the University of Bristol in England, where he earned his PhD. under the supervision of David Bohm. In 1998, he received the Wolf Prize.

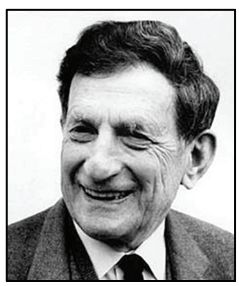


Anderson, Carl David (1905–1991), an American physicist who, together with Victor Hess of Austria, received the Nobel Prize in 1936 for the discovery of the positron (a positively charged electron), the first known particle of antimatter. Anderson earned his PhD in 1930 at the Caltech in Pasadena, where he worked with physicist Robert Millican. Starting in 1927, they studied X-ray photoelectrons (which are emitted during collisions of atoms with high-energy photons); in 1930, they began investigating cosmic and gamma radiation. Anderson photographed tracks of secondary cosmic-ray showers in a cloud chamber, and while studying the photographs, he discovered numerous tracks whose shape suggested they could have been produced by positively charged particles. In 1932, he stated that the tracks were caused by positrons, positively charged particles with the same mass as electrons. This claim was verified in 1933 by the British physicist P. Blackett and his Italian colleague G. Occhialini. The existence of the positron as the antiparticle of the electron was predicted in 1928 by P. Dirac.

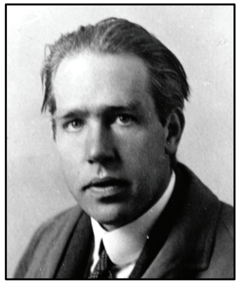
In 1936, Anderson discovered the muon (a heavy electron), a particle 207 times heavier than an electron. At first, he thought he had found the meson predicted by Japanese physicist Hideki Yukawa, but it turned out that the muon interacts only weakly and is actually a heavier version of the electron. The actual meson was then discovered in 1947 by British physicist Cecil Powell and named the pi meson or pion.

Bell, John Stewart (1928–1990), an Irish physicist who formulated Bell inequalities, which made it possible to rule out the classical interpretation of quantum mechanics (the hidden-parameter interpretation). In 1948, he earned a bachelor's degree in experimental physics from the University of Belfast and received his Ph.D. in 1956 from the University of Birmingham. He then worked at the Harwell Laboratory and, after several years, moved to CERN. There he began to focus intensively on theoretical physics, particle physics, and quantum theory. Based on inequalities he proposed, the classical interpretation of quantum theory – according to which the probabilistic nature of measurements is caused by a lack of knowledge of all the system's parameters – was definitively ruled out in 1983.

Binnig, Gerd Karl (1947), German physicist and co-recipient of the 1986 Nobel Prize for the invention of the scanning tunneling microscope. Binnig was born in Frankfurt and studied at Goethe University, where he earned his PhD in 1978. Then he joined an IBM research laboratory near Zurich, Switzerland, and together with Heinrich Rohrer, they began working on the problem of how to image the fine details of material structures. They came up with the idea of a probe moving around the surface and emitting electrons. A small probe with a sharp tip moves, and electrons tunnel between the sample and the tip. Minute changes (even at the atomic scale) in the distance between the probe and the surface are recorded. By sampling the surface appropriately, a detailed three-dimensional image is obtained. The new microscope is capable of detecting individual atoms and, with the help of the tip, even manipulating them.



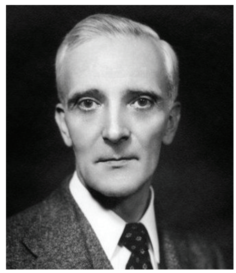
Bohm, David (1917–1992), an American-British theoretical physicist. He worked in the fields of plasma physics, nuclear physics, and quantum theory. During World War II, he worked on the Manhattan Project. The following phenomena are named after him: anomalous Bohm diffusion in plasma and the Aharon-Bohm effect, in which a beam of electrons changes its phase due to a non-zero vector potential even in a region where the magnetic field is zero. He studied at Pennsylvania State University, then at Caltech and the University of California, Berkeley. He worked in Israel, Brazil, and England. He discovered the Aharonov–Bohm effect together with his student Aharon while staying in Bristol, England, where Aharonov was his student.



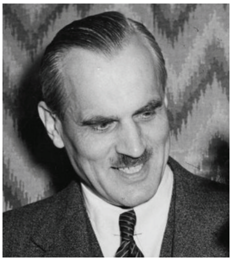
Bohr, Niels (1885–1962), a Danish physicist who proposed the first successful model of the atom in 1913. If an electron is in a specific orbit (circumference corresponds to an integer multiple of the electron's wavelength), it does not emit radiation. When an electron jumps between two energy levels, a corresponding energy quantum is emitted. The electron's angular momentum is quantized, with Planck constant as the fundamental quantum. This model applies to hydrogen but does not explain the laws of the microworld. Later, the same results were derived from quantum theory. Bohr is the author of the correspondence principle – for large quantum numbers, quantum formulas must transition into classical ones. He is also a proponent of the probabilistic (so-called Copenhagen) interpretation of quantum theory. In 1922, he received the Nobel Prize in Physics.

Bose, Satyendra Nath (1854–1948), see [1], Statistical physics.

Brillouin, Léon (1889–1969), a French-American physicist. In 1926, he contributed to the development of the WKB approximation in quantum theory. This was an approximate method for solving the Schrödinger equation. In solid-state physics, he discovered the Brillouin zones, which reflect the periodicity of the crystal lattice in *k*-space. Brillouin studied at the *École Normale Supérieure* in France; in 1928, he became professor at the Sorbonne, and later a professor at the Collège de France. During World War II, he emigrated to the United States, becoming a professor at the University of Wisconsin (1941) and later at Harvard (1946). In his later years, he was a professor at Columbia University. He is the author of more than 200 papers.



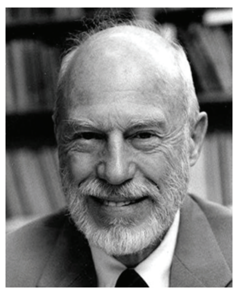
Broglie, Louis de (1892–1987), a French physicist who proposed the principle of particle-wave duality in 1923, which states that objects in the microworld can behave as particles in some situations and as waves in others. Broglie expressed the wave properties of particles with the equation: $\lambda = h/mv$, where λ is the wavelength, h is Planck constant, m is the mass of the particle, and v is its velocity. De Broglie came from a French noble family and was a duke. His great-grandfather was executed by guillotine during the French Revolution. During World War I, he worked on problems related to radio communication and became interested in science. He served at the top of the Eiffel Tower. He presented his hypothesis on the wave properties of particles as part of his doctoral thesis in 1923. The idea was so groundbreaking that the committee hesitated to award him his doctorate. Just six years later, in 1929, he received the Nobel Prize in Physics for this work.



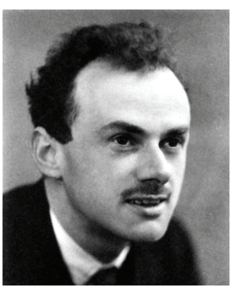
Compton, Arthur (1892–1962), American physicist, winner of the 1927 Nobel Prize for his research on the scattering of photons by free electrons. Compton studied X-rays. In 1922, he discovered the angular dependence of the change in wavelength of a high-energy photon upon scattering off electrons. Compton discovery confirmed that electromagnetic radiation has both wave and particle nature, and became a key experiment in quantum mechanics. This scattering is called the Compton effect. When a photon gains energy from an electron, we refer to the so-called inverse Compton effect. This is one of the fundamental mechanisms by which photons in the universe can acquire high energy. Compton also conducted research on cosmic rays, reflection, polarization, and the spectral properties of X-rays.

Compton was born in Wooster, Ohio. He graduated from Wooster College and Princeton. In 1923, he became a professor of physics at the University of Chicago. He became director of the laboratory where, in 1942, Enrico Fermi commissioned the world's first nuclear reactor. Compton participated in the construction of the first atomic bomb. From 1945 to 1953, he served as president of the University of Washington.

Davisson, Clinton (1881–1958), an American experimental physicist who, in 1937, was awarded the Nobel Prize in Physics jointly with George Thomson. Both scientists independently discovered that electrons can diffract just like light waves, thereby confirming Louis de Broglie hypothesis that electrons should behave as both waves and particles. In 1927, Davisson and Lester Germer investigated an electron beam reflected off a metal crystal. They found that it exhibited diffraction patterns similar to those of X-rays and other electromagnetic waves. This discovery confirmed the quantum mechanical understanding of wave-particle duality and enabled the study of the nuclear, atomic, and molecular structures of matter. Davisson earned his doctorate at Princeton University and spent most of his career at Bell Telephone Laboratories. He first studied electron emission from metals at elevated temperatures and later helped develop the electron microscope.

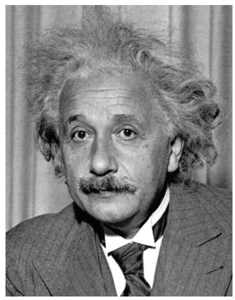


DeWitt, Bryce (1923–2004), an American theoretical physicist who worked on quantum gravity and numerical simulations. He suprted Everett many-worlds interpretation of quantum theory. The Wheeler–DeWitt wave function, which describes the universe as a whole, is named after him. He is a recipient of the Dirac and Einstein prizes. He studied theoretical physics at Harvard, where he earned his PhD in 1950 under the supervision of Julian Schwinger, a co-recipient of the Nobel Prize for the development of quantum electrodynamics. He worked at the Institute for Advanced Study in Princeton, the University of North Carolina at Chapel Hill, and the University of Texas.



Dirac, Paul Adrien Maurice (1902–1984), an English physicist and one of the leading figures in 20th-century quantum theory. He laid the foundations of quantum electrodynamics. In 1928, he derived the famous Dirac equation – a relativistic equation for the electron. He correctly interpreted states with negative energy as antiparticles and predicted the existence of the positron. The positron was discovered by Carl Anderson in 1932. Independently of Fermi, he derived the statistical distribution for particles with half-integer spin (the Fermi–Dirac distribution). He is the author of Dirac notation in quantum theory. He demonstrated the equivalence of Schrödinger and Heisenberg approaches to quantum mechanics. He is the author of the second quantization method, which enables the transition from quantum particle theory to quantum field theory. He predicted vacuum polarization and the non-trivial dynamic properties of the vacuum. He predicted the existence of the magnetic monopole. He is the author of the many-particle formalism in quantum theory. For his work he was awarded the Nobel Prize in Physics in 1933. In later years, he explored hypothetical variability of fundamental constants (gravitational, speed of light, etc.). Throughout his life, he was an advocate of the principle of simplicity in physical equations. He was one of the first to realize that symmetries in nature are a primary principle in the formulation of physical equations.

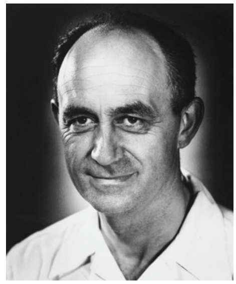
Ehrenfest, Paul (1880–1933), see [1], Statistical Physics.



Einstein, Albert (1879–1955), Einstein, Albert (1879–1955), a German-American physicist, author of the special and general theories of relativity, a scientist who elucidated the photoelectric effect and Brownian motion, and who, together with Bose, derived the statistical distribution of particles with integer spin. He published the special theory of relativity in 1905. In it, he reconciled classical mechanics with Maxwell electrodynamics, from which followed the independence of the speed of light on the motion of the source. The special theory of relativity brought with it length contraction, time dilation, and the realization that time and space are not absolute.

In 1905, Einstein also explained the photoelectric effect. He posited that light consists of particles (photons) whose energy is equal to $\hbar\omega$. The ability to eject an electron from a metal using light is determined by the frequency of individual photons, not their number. That same year, Einstein also provided an explanation for Brownian motion. In 1916, Einstein published a new theory of gravity – general relativity. He describes gravity as curved spacetime. The bodies themselves contribute to the curvature of spacetime and move within it along the straightest possible paths – geodesics. Albert Einstein was awarded the Nobel Prize in Physics for 1921; paradoxically, however, it was for his explanation of the photoelectric effect and not for general relativity.

Everett, Hugh (1930–1982), an American physicist who proposed the many-worlds interpretation of quantum theory, according to which, at the moment of measurement, one of the possible outcomes is realized in our universe, while the other are realized in parallel universes. This interpretation of quantum theory was not accepted by many physicists and caused Everett considerable trouble. He studied at The Catholic University of America and at Princeton. In addition to quantum theory, he also worked on mathematical optimization of various problems, mathematical modeling, statistical analysis, and game theory.



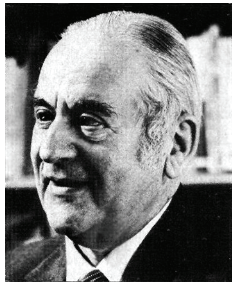
Fermi, Enrico (1901–1954), an Italian-American physicist who focused primarily on quantum theory and particle physics. He named the small neutral particle produced during beta decay the *neutrino* (from the Italian "neutro" and "ino"). Particles with half-integer spin are named *fermions* in his honor. These are particles that satisfy the Pauli exclusion principle. These particles follow a statistical distribution known as the Fermi–Dirac distribution. In 1942, Enrico Fermi constructed and launched the world's first nuclear reactor beneath the University of Chicago stadium. It was built from graphite bricks, which also served as a moderator. In 1943, he founded the Argonne National Laboratory. Enrico Fermi also studied the acceleration of cosmic rays and proposed the statistical acceleration of charged particles as they bounce off magnetic mirrors. Today, we call this mechanism the *Fermi mechanism*. In 1938, he was awarded the Nobel Prize in Physics for the discovery of artificial radioactive elements produced by bombarding nuclei with neutrons. An X-ray observatory launched into space in 2008 is also named after Enrico Fermi.

Fock, Vladimir Alexandrovich (1898–1974), a Soviet physicist who focused primarily on quantum mechanics and quantum electrodynamics. He graduated from St. Petersburg University. Fock states with many particles are named after him.

Gerlach, Walter (1889–1979), Gerlach, Walter (1889–1979), German physicist and co-discoverer of spin in the famous Stern–Gerlach experiment. Gerlach graduated from Eberhard Karls University in Frankfurt am Main. During World War I, he served in the German army. He contributed to the development of wireless telegraphy. In 1922, he conducted an experiment with Stern that led to the discovery of spin. A beam of silver atoms passed through a non-uniform magnetic field and struck a glass disk. It split into two beams, a phenomenon soon explained as a consequence of the existence of spin—the intrinsic rotational and magnetic moment of particles. In 1925, Gerlach became a professor at the University of Tübingen. He succeeded Paschen in this position. In 1929, he became a professor at the University of Munich, where he succeeded Wien. He remained in this position until May 1945, when he was imprisoned by the Allied forces. He allegedly participated in the development of German weapons. In 1946, he returned to Germany, worked at the University of Bonn as a visiting professor, and from 1948 served as a professor at the University of Munich. He was simultaneously head of the Department of Physics and rector of the university (1948–1951). He held many positions in German science.

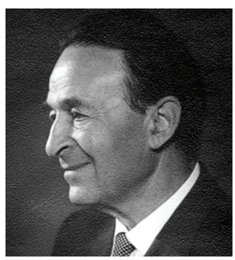
Germer Lester (1896–1971), an American experimental physicist who, together with Clinton Davisson, demonstrated the wave properties of the electron, thereby confirming the wave-particle duality proposed by Louis de Broglie. The famous Germer-Davisson experiment was pivotal to the development of the electron microscope. Germer graduated from Columbia University. During World War I, he served as a fighter pilot in the U.S. Army. He later joined Bell Telephone Laboratories. In addition to physics, his other passion was until his death mountaineering.

Gordon, Walter (1893–1939), German theoretical physicist. He spent his childhood in Switzerland and his later years in Sweden (due to the political situation in Germany). He studied at the University of Berlin and earned his doctorate in 1921 under the supervision of Max Planck. In 1922, he became an assistant to Max von Laue. From 1926, he worked in Hamburg, where he became a professor in 1930. From 1933, he lived in Stockholm, Sweden. He worked in theoretical physics. In 1927, together with Oskar Klein, they proposed a relativistic variant of the Schrödinger equation, known as the Klein–Gordon equation. They originally assumed that this was the correct equation for the electron. It turned out that their equation describes quantum-relativistic particles with spin 0, whereas Dirac equation from 1928 is suitable for particles with spin ½.



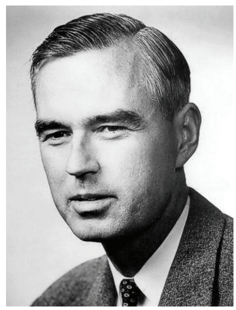
Goudsmit, Samuel Abraham (1902–1978), a Dutch-American physicist who, together with Uhlenbeck, interpreted the results of the Stern-Gerlach experiment as existence of spin. He also studied line spectra. He studied physics at Leiden University (student of Paul Ehrenfest) and earned his PhD in 1927. From 1927 to 1946, he served as a professor at the University of Michigan. During World War II, he worked at MIT. He was involved in the Manhattan Project (atomic bomb development). He collaborated closely with Werner Heisenberg and Otto Hahn. He was also interested in archaeology and Egyptology.

Heisenberg, Werner (1901–1976), a German theorist who studied the fundamental principles of quantum theory. He is the author of matrix quantum mechanics (1925). This is a different method for calculating quantum states than Schrödinger wave mechanics. Heisenberg derived the famous uncertainty principle, according to which it is impossible to simultaneously measure the position and momentum of an object. Measuring one quantity disturbs the result of measuring the other. He was awarded the Nobel Prize in Physics in 1932 for laying the foundations of quantum theory. Heisenberg studied theoretical physics at the University of Munich. He earned his PhD under Sommerfeld in 1923 and became an assistant to Max Born in Göttingen. He worked for three years in Copenhagen with Niels Bohr, where they collaborated on Copenhagen interpretation of quantum theory. He proposed a model of ferromagnets with two phase transitions. From 1927 to 1941, he was a professor of physics in Leipzig; from 1942 to 1945, he was director of the Max Planck Institute in Berlin; and from 1946 director of the Max Planck Institute in Copenhagen.



Klein, Oskar Benjamin (1894–1977), Swedish theoretical physicist. He earned his PhD in 1921 from Stockholm University. He worked at the University of Michigan (USA), in Leiden, and at Lund University. He collaborated with Niels Bohr and Paul Ehrenfest. He is a co-author of the Kaluza–Klein model, which was the first attempt to unify electricity and magnetism with gravity by adding fifth dimension. Today, a similar approach is used in string theory (M-theory). He is also a co-creator of the Klein–Gordon equation from 1927. Originally, Klein and Gordon derived this equation as a relativistic equation for the electron, but it turned out that it is the correct quantum relativistic equation for particles with zero spin. Together with Alfvén, he held the view that events in the universe are predominantly influenced by electromagnetic interaction. Klein paradox is also named after him: the solution to the Dirac equation implies that a relativistic massless particle is not exponentially damped when passing through a potential barrier. The phenomenon has been experimentally verified (e.g., the motion of an electron with zero effective mass in graphene). Together with Yoshio Nishina, he derived in 1929 a formula for the effective cross section of Thomson scattering (photon on an electron) in the lowest order of quantum electrodynamics (the Klein–Nishina formula). In 1959, he was awarded the Planck Medal.

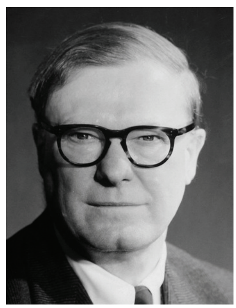
Kronig, Ralph (1904–1995), a German-American physicist who made significant contributions to quantum mechanics. He was a co-discoverer of particle spin and conducted research on X-ray absorption spectroscopy and the quantum behavior of periodic structures. The Kronig–Penney model is named after him; it describes the formation of allowed and forbidden bands in a particle's spectrum in periodic potential. Furthermore, the Coster–Kronig transition is named after him, in which an electron is emitted when jumps to a different energy level. Kronig studied in Dresden, Germany, and later at Columbia University in the US, where he earned PhD in 1925. Among European scientists, he was most influenced by Paul Ehrenfest



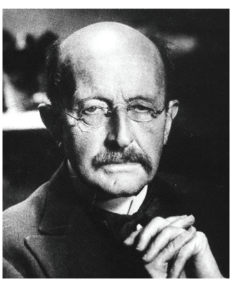
Lamb, Willis Eugene (1913–2008), an American physicist and co-recipient of the 1955 Nobel Prize in Physics, which he shared with Polykarp Kusch for experimental work leading to the refinement of quantum electrodynamics. Lamb joined Columbia University in New York in 1938 and worked at the renowned Radiation Laboratory at MIT during World War II. In 1947, Lamb detected deviations from the hyperfine structure of spectral lines predicted by quantum electrodynamics. These deviations were caused by non-trivial dynamic properties of the vacuum, particularly the presence of virtual electron-positron pairs in the vacuum. From 1951 to 1956, he was a professor of physics at Stanford University in California, where he developed microwave techniques for measuring the hyperfine structures of helium spectral lines. Until 1962, he was a professor of theoretical physics at the University of Oxford. In the same year, he was appointed a professor at Yale University. In 1974, he became a professor of physics and optical sciences at the University of Arizona.

Neumann, John von (1903–1957), a Hungarian-American mathematician who, independently of Dirac, demonstrated in 1944 that Schrödinger and Heisenberg quantum mechanics are equivalent. In 1932, he proposed a controversial interpretation of quantum mechanics, according to which the outcome of a measurement is influenced by the observer's consciousness. In 1944, he developed game theory. He became a pioneer of digital computers, designing the basic architecture of the computer. He was deeply involved in numerical mathematics; at the end of World War II, he contributed to the construction of the first atomic bomb. The von Neumann computer architecture, von Neumann algebra in quantum theory, and von Neumann cellular automata are named after him. Von Neumann studied chemistry at the University of Berlin until 1923, when he moved to the Swiss Federal Institute of Technology in 1926 and became a chemical engineer. He earned his PhD at the University of Budapest in mathematics, specifically in set theory. At the age of twenty, he published a definition of natural numbers, as we use them today. which can be rearranged into the standard equation for oscillations in the variable x.

Pauli, Wolfgang (1900–1958), an Austrian-German-American physicist, formulated the Pauli exclusion principle in 1925, which states that two fermions cannot occupy the same quantum state. This principle is responsible for the distinct properties of different atoms and for the chemical properties of substances. He made a significant contribution to the development of quantum mechanics. The Pauli equation, the first quantum equation to include spin, is named after him. In the 1930s, he predicted the existence of the neutrino. For his work, particularly for the discovery of the exclusion principle, he was awarded the Nobel Prize in Physics in 1945. Pauli was born in Vienna; his godfather was Ernst Mach. Pauli's paternal grandparents came from a Jewish family in Prague. He published his first scientific article on general relativity at the age of 18. He studied in Munich under Sommerfeld, where he earned his PhD in 1921 based on a thesis on the hydrogen molecule quantum properties. Pauli spent a year at the University of Göttingen (under Max Born). He also worked at the Institute of Theoretical Physics in Copenhagen (now the Niels Bohr Institute), at the University of Hamburg, and in Zurich, Switzerland. In 1931, he was a visiting professor at the University of Michigan, and in 1935 at Princeton. By 1939, Pauli moved to USA, where he worked as a professor of theoretical physics at Princeton. After World War II, he became a U.S.

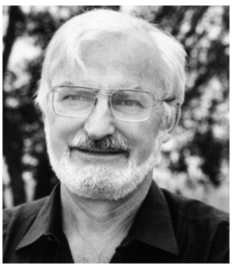


Penney, William (1909–1991), an English mathematician and theoretical physicist, one of the founders of British nuclear research. He studied at Imperial College, earned his master's degree at the University of Wisconsin in USA, and received his doctorate from Trinity College, Cambridge. From 1967 to 1973, he served as rector of the Imperial College. He was one of 20 British physicists who worked in the Manhattan Project. Penney calculated the destructive effects of the shock wave generated by the explosion. In 1945, he was a member of the committee that selected the cities of Hiroshima and Nagasaki for the attack. After the war, he was involved in the design of the British atomic bomb and oversaw the development of the British hydrogen bomb. He served as director and chairman of various organizations involved in atomic energy. In 1967, he was knighted and became a baron. He is a co-author of the Kronig–Penney model of particle interaction with a periodic potential.

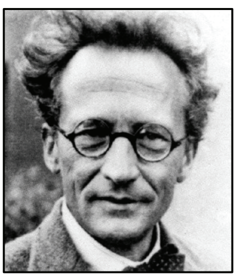


Planck, Max (1858–1947), a German physicist who solved the radiation of a perfectly black body, assuming that energy is quantized. He introduced this assumption purely for mathematical reasons, so that the equations would be solvable. He did not have much confidence in the physical interpretation. In 1918, he received the Nobel Prize for his quantum theory, which was successfully tested by Einstein in the photoelectric effect and by Bohr in the first atomic model. Planck was deeply involved in thermodynamics; one of the possible formulations of the second law of thermodynamics is named after him. Planck was a critic of the probabilistic interpretation of entropy. In 1900, he first used the universal gas constant and Avogadro number. The so-called Planck scales are named after Planck – typical units of mass, length, time, etc., derived from combinations of fundamental constants. Planck's name is also borne by the largest network of scientific institutes in Germany (Max Planck Institute), a crater on the Moon, and a European probe studying cosmic background radiation.

Pontecorvo, Bruno (1913–1993), an Italian-Russian nuclear physicist. In the first half of his life, he worked in Italy, became Enrico Fermi's assistant, and performed experiments with slow neutrons that led to the discovery of the nuclear chain reaction. He predicted neutrino oscillations, and the mixing matrix of neutrino mass states is named after him. In 1948, he obtained British citizenship. In 1950, under strange circumstances, he emigrated to the Soviet Union, where he worked in Dubna until his death. In accordance with his wishes, half of his ashes are interred in Rome and half in Dubna. Since 1995, the prestigious Pontecorvo Prize is awarded for achievements in nuclear and particle physics research.



Rohrer, Heinrich (1933), a Swiss physicist and co-recipient of the 1986 Nobel Prize in Physics, which was awarded to him for the invention of the scanning tunneling microscope. Rohrer studied at the Swiss Federal Institute of Technology, where he earned his PhD in 1960. In 1963, he joined the IBM research laboratory near Zurich. There, together with Gerd Binnig, he set about constructing a device that later enabled them to reveal the microscopic structure of the surfaces of the materials under investigation and to observe individual atoms. The new microscope utilizes electron tunneling between the probe tip and the surface of the sample being examined. The scanning tunneling microscope is used to manipulate individual atoms, study biological samples, analyze industrial materials (such as superconductors), or test miniature electrical circuits.



Schrödinger, Erwin (1887–1961), an Austrian physicist who, in 1926, developed wave mechanics as one of the possible formulations of quantum mechanics. From the so-called Schrödinger equation, it is possible to determine the wave function, which represents the amplitude of the probability of a particle's occurrence, and its square represents the probability density. For his work, he was awarded the Nobel Prize in Physics in 1933. Schrödinger studied at the University of Vienna. After World War I, he began working at the University of Zurich. From 1927, he worked at the University of Berlin at the invitation of Max Planck. Due to the persecution of Jews, he left the university in 1933 and spent the next seven years traveling through Austria, Great Britain, Belgium, and Italy, changing jobs many times. It was not until 1940 that he settled for the next 15 years in Ireland at the Dublin Institute for Advanced Studies. In 1956, Schrödinger retired and returned to Vienna, Austria.

Stern, Otto (1888–1969), a German-born scientist and recipient of the 1943 Nobel Prize in Physics for his research on molecular bonds as a tool for studying molecular characteristics and for measuring the magnetic moment of the proton. Stern's early scientific work consisted of theoretical studies in statistical physics. In 1914, he became a lecturer in theoretical physics at the University of Frankfurt, and in 1923, a professor of physical chemistry at the University of Hamburg. It was here, in the early 1920s, that he and Walter Gerlach presented their historic experiment with molecular beams. A collimated beam of silver atoms passed through a non-uniform magnetic field and struck a glass disk. It split into two beams, a phenomenon that was soon explained as a consequence of the existence of spin – the intrinsic rotational and magnetic moment of particles.

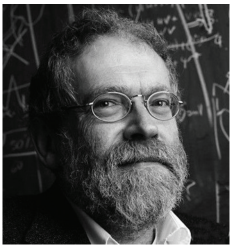
In 1933, Stern measured the magnetic moment of the proton and pointed out its inconsistency with existing theory. In 1933, when the Nazis came to power, Stern leaved Germany. He went to the United States, where he became a professor of physics at the Carnegie Institute of Technology in Pittsburgh until his retirement in 1945.

Tonomura, Akira (1942–2012), an outstanding Japanese quantum physicist and the inventor of the electron holographic microscope, which can record not only the intensity of an electron beam but also its phase. Tonomura has been a long-time employee of Hitachi's research and development laboratories. He studied at the University of Tokyo and earned his Ph.D. from Gakushuin University. He recorded the first electron hologram as early as 1968. Together with colleagues, using electron holography, he observed the Aharonov–Bohm effect – the phase shift of electrons passing through a region with a zero magnetic field but a non-zero potential. In the 1990s, he developed a method for observing magnetic tubes and vortices in superconductors. In 2000, he constructed a holographic microscope with a resolution of 49.5 pm. He is the recipient of numerous international awards and medals.

Uhlenbeck, George Eugene (1900–1988), a Dutch-American physicist who, together with Goudsmit, showed that the deflection of a beam of silver atoms in an external magnetic field (the Stern–Gerlach experiment) is caused by the existence of an additional quantum number, spin. Uhlenbeck studied chemical engineering in Delft and then physics and mathematics in Leiden, where he earned his bachelor's degree in 1920 and his master's degree in 1923. From 1925, he worked in Leiden as Ehrenfest's assistant. There he met Goudsmith, with whom he co-discovered spin. Uhlenbeck was a longtime friend of Enrico Fermi. In 1938, he was a visiting professor at Columbia University. In 1939, he became a professor of theoretical physics at the University of Ann Arbor.

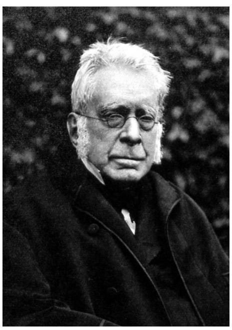
During World War II, he led a theoretical group at the Radiation Laboratory in Cambridge (USA). After the war, he returned to Ann Arbor. From 1960 until his retirement, he worked at the Rockefeller Institute for Medical Research in New York.

Wigner, Eugene (1902–1995), a Hungarian-born American physicist who, together with Hans Jensen and Maria Mayer, was awarded the 1963 Nobel Prize for his contributions to atomic physics. Wigner earned his Ph.D. from the Technical University of Berlin in 1925. He worked in Berlin and Göttingen, then moved to United States, where he he spent most of his academic career at Princeton. He formulated the law of parity. He demonstrated that the nuclear force has a short range and is independent of charge. In 1936, he worked on the theory of neutron absorption, which proved useful for nuclear reactors. After World War II, he assisted Enrico Fermi in building the first nuclear reactor. The following are named after Wigner: the Wigner probability distribution, Wigner theorem, the Wigner effect, the Wigner–Eckart theorem, and others.

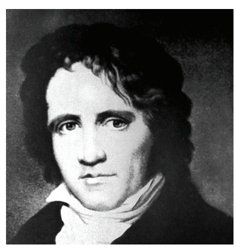


Zeilinger, Anthon (1945), an Austrian quantum physicist and pioneer of quantum information theory. He was the first to achieve quantum teleportation of photons. He prepared experiments exploring the boundary between the quantum and the macroscopic world. He studies interference phenomena phenomena in neutrons, atoms, and large molecules, entangled quantum states, quantum cryptography, and teleportation. He demonstrated that quantum communication is possible even via satellites. He has been affiliated with many universities worldwide, including Oxford, MIT, Humboldt University, and others. He is currently a professor at the University of Vienna. He was awarded the 2022 Nobel Prize in Physics for his discoveries.

Mathematics



Airy, George Biddell (1801–1892), English mathematician and Royal Astronomer (1835 to 1881). He modernized the Greenwich Observatory. Had he not ignored John Couch Adams's calculations, he might discover Neptune. He improved the orbital elements of Venus and the Moon, mathematically described the rainbow, and calculated the density of the Earth from the oscillations of a pendulum placed in a deep mine. In mathematics, the Airy functions are named after him. He studied at Cambridge, where he became a professor after three years and later director of the Cambridge Observatory. In 1834, he served as chairman of the International Commission on Weights and Measures.



Bessel, Friedrich Wilhelm (1784–1846), German mathematician and astronomer, director of the observatory in Königsberg. He measured the positions of 50,000 stars. He was the first to measure parallax (star 61 Cygni). He used parallax to calculate the distances of nearby stars. In 1844, based on changes in the position of Sirius, he predicted the existence of its small companion, Sirius B. This was the first application of the law of gravity outside the Solar System. This companion was discovered in the Clark brothers' optical workshop in 1862 during a test of a 45-centimeter-diameter lens. Bessel studied the mathematical functions introduced by Daniel Bernoulli, which are now known as Bessel functions. The asteroid Bessel is named after him.

Cauchy, Augustin Louis (1789–1857), a French mathematician who authored 789 papers. Only Leonhard Euler and Arthur Cayley surpassed him. He brought precision and rigor to mathematics. He coined the term "determinant." He systematized his studies and soon thereafter defined the concepts of limit, continuity, and convergence. Independently of Jean-le-Rond d'Alembert, Cauchy founded complex analysis, and together with Bernhard Riemann, they derived the important Cauchy–Riemann conditions for the existence of a derivative in complex analysis.

Dirichlet, Johann Peter Gustav Lejeune (1805–1859), a Belgian mathematician who lived and worked primarily in France and later in Germany. After Gauss's death, he gained his position at Göttingen. He worked on solving Fermat theorem (non-existence of a solution to the equation $x^n + y^n = z^n$ in the field of integers) for $n = 5$ and 14. He studied polynomial equations and was involved in number theory. In mechanics, he studied the potentials of equilibrium systems. He sought solutions to Laplace equation with fixed boundary conditions (now called Dirichlet conditions). He also studied the convergence of trigonometric series used to solve partial differential equations. A function called the Dirichlet kernel is key to proving convergence.

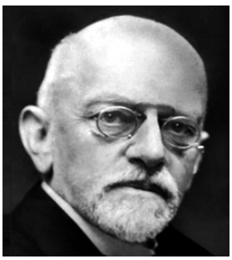
Euler, Leonhard (1707–1803), see section Theoretical Mechanics

Fourier, Jean-Baptiste Joseph de (1768–1830), French physicist and mathematician. Together with the Danish physicist H. Oersted, he constructed a voltage source similar to a Voltaic pile using bismuth and antimony plates. In 1822 he mathematically formulated the theory of heat conduction, thereby contributing to the development of steam engines. He laid the foundation for the Fourier method for solving partial differential equations. He demonstrated how powerful Fourier series are in mathematical physics and mathematical analysis, as they can be used to approximate periodic functions over a finite interval. For non-periodic functions, he introduced the Fourier transform. He also worked on statistics and probability theory. Fourier's work is an example of how the demands of physics have driven significant progress in mathematics.

Gauss, Karl Fridrich (1777–1855), a German mathematician sometimes referred to as the "prince of mathematicians." He was a child prodigy; at the age of three, he pointed out an error in his father's wage calculation and told him the correct result. When he was in school and the teacher asked the class to add the numbers from 1 to 100, he derived a formula for the sum of an arithmetic series. In statistics, this is known as the Gaussian distribution. In integral calculus, Gauss theorem for converting surface and volume integrals is named in his honor. Gauss perfected the representation of complex numbers in a plane, which we now call the Gaussian plane. The unit of magnetic field induction, the gauss, is also named after Gauss. He worked in number theory, integral and differential calculus, geometry, mathematical analysis, electrostatics, astronomy, optics, geodesy, and many other fields of the natural sciences. He wrote his seminal work on number theory (*Disquisitiones Arithmeticae*) at the age of 21.

Green, George (1793–1841), a British mathematician and physicist who studied solids. In mathematics, he focused on the use of multidimensional integrals to describe electromagnetic fields. He derived Gauss theorem in two dimensions independently of Mikhail Vasilyevich Ostrogradsky. The Green function – the solution to a partial differential equation with a unit impulse on the right-hand side – is named after Green. He lived in Sneinton, which is now part of Nottingham. He worked there as a miller, but found a passion for mathematics, which he pursued more and more as a self-taught scholar. Eventually, he decided to study mathematics, and in 1832, at nearly 40 years of age, he was admitted to Cambridge.

Hamilton, William Rowan (1805–1865), see section Theoretical Mechanics



Hilbert, David (1862–1943), a German mathematician who studied geometry (Hilbert spaces, used in quantum theory, are named after him), number theory, mathematical logic, and differential equations. His scope of work was truly enormous. He was a co-founder of the calculus of variations. In physics, he attempted to solve the three-body problem. In 1900, he formulated 23 fundamental mathematical problems as a challenge to be solved in the 20th century. Some of them remain unsolved to this day. Hilbert studied at the University of Königsberg.

Kronecker, Leopold (1823–1891), a German mathematician, worked on number theory, groups, quadratic forms, differential equations, probability theory, determinants, and elliptic functions. In particular, he tried to find connections between the various mathematical disciplines. He became famous for the statement, "*God created the integers; everything else is the work of man*." He studied astronomy, chemistry, and meteorology. In 1860, he became a member of the Berlin Academy and began lecturing at the university. However, but his lectures were difficult to follow. The Kronecker delta symbol δ_{kn} is named after Kronecker; it is equal to 1 if $k = n$, and 0 if $k \neq n$.

Ladyzhenskaya, Olga Alexandrovna (1922–2004), a Russian mathematician who specialized in partial differential equations, particularly in the field of fluid flow. She investigated numerical schemes for computer simulations of fluid using the Navier–Stokes equations. She studied the regularity of parabolic and elliptic partial differential equations. She investigated solutions to partial differential equations in terms of distributions and was one of the first to formulate the problem of so-called *weak solu-*

tions to these equations. For her work, she was awarded the Lomonosov Gold Medal in 2002. She is the author of more than 200 scientific papers, including 6 monographs.

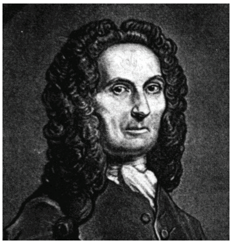
Laplace, Pierre Simon de (1749–1827), a French mathematician and physicist who synthesized the mathematical findings of all his predecessors in astronomy in the five-volume work *Mécanique Céleste* (Celestial Mechanics), published between 1799 and 1825. Laplace systematized and further developed the theory of probability in his work *Essai Philosophique sur les Probabilités* (1814). He was the first to calculate the Gauss integral as the square root of π. He studied the integral transform named after him. Oliver Heaviside perfected it. Laplace contributed to Lavoisier calorimetric theory. He designed a calorimeter for measuring specific heats of many substances. He discovered and introduced gravitational potential and showed that in a vacuum it can be calculated using Laplace equation. After being appointed Minister of the Interior by Napoleon, he was soon dismissed with the words: "*You offer infinitesimally little hope of solving great problems*."

Laurent, Pierre Alphonse (1813–1854), a French mathematician who generalized Taylor series to negative powers, which enabled to describe functions of a complex variable in the annulus between poles. He studied the polarization of light waves. His father was French, and his mother was English. Laurent graduated from the École Polytechnique in Paris and became a military engineer. He worked at the University of Metz, then he participated in the Algerian War. He returned in 1840 and spent six years supervising expansion of the port of Le Havre. In 1846, Cauchy nominated Laurent for membership in the Academy of Sciences, but the nomination failed. Laurent's works were published after his death.

Legendre Adrien-Marie (1752–1833), a versatile French mathematician, worked on number theory, solution of differential equations, elliptic integrals, and in other topics. His work was continued by Niels Henrik Abel and Friedrich Gauss. He generalized the concept of the factorial to the gamma function. He discovered the Legendre dual transformation, which converts Lagrange equations into Hamilton. Legendre polynomials, which are important in solving the eigenvalues of the Laplace operator on the sphere, are named after him. He published an extensive three-volume treatise, *Exercices de Calcul Intégral*.

Levi-Civita, Tullio (1873–1941), Italian mathematician and physicist. He refined Riemann relation for the number of prime numbers in a given interval. He studied the application of differential and tensor calculus in general relativity. The Levi-Civita tensor (totally antisymmetric third-order tensor) is named after him. Together with Weyl, he attempted to develop a unified theory of the electromagnetic field and gravity. He also studied celestial mechanics, particularly the three-body problem, quantum theory (the Dirac equation), hydrodynamics, etc.

Lie, Marius Sophus (1842–1899), a Norwegian mathematician who studied group theory and its significance for geometry. He collaborated with Oskar Klein and Peter Sylow. He prepared for publication previously unknown works by Niels Abel (also a Norwegian mathematician) on groups. He also studied the properties of partial differential equations from the perspective of symmetries. He introduced Lie groups and Lie algebras. He studied in Berlin. He received his PhD from the University of Christiania (now Oslo) in 1871. In 1878, he became an honorary member of the London Mathematical Society, in 1892 a member of the French Academy of Sciences, and in 1895 a foreign member of the British Royal Society in London. That same year, he became a member of the U.S. National Academy of Sciences.



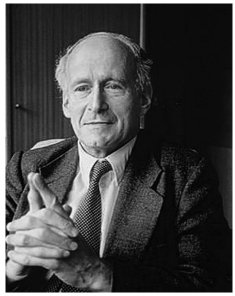
Moivre, Abraham de (1667–1754), a French mathematician who laid the foundations of probability theory. His work, published in 1711, *was titled Philosophical Transactions*. It was later expanded into *Doctrine of Chances* (published in 1718). It contained important findings, Stirling formula, and the Gauss integral. He also published Moivre theorem, by which the powers and roots of complex numbers can be calculated. He was a friend of Isaac Newton, Edmond Halley, and James Stirling.

Morgan, Augustus de (1806–1871), a British mathematician who focused primarily on set theory and logic. His well-known De Morgan laws originate from this field. He also formulated a correct procedure for proofs using mathematical induction. He made significant contributions to the theory of quaternions (a generalization of complex numbers to 4D space). He was born in India to a family of British colonizers. He studied at Cambridge. He spent most of his career at the University of London, where he became a professor of mathematics.

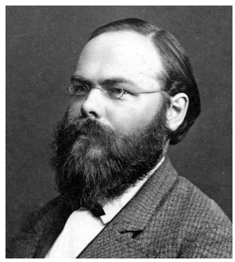
Pfaff, Johann Friedrich (1765–1825), a German mathematician who was a predecessor of Friedrich Gauss. Pfaff graduated from the University of Göttingen and later studied astronomy in Berlin under Johann Bode. He calculated data on the positions of astronomical objects. In 1788, he became a professor of mathematics at the University of Helmstedt. He worked there until the university was closed in 1810. He then found a position at the University of Halle. He focused on mathematical series and integral calculus. His most significant achievements came from his study of first-order partial differential equations. He systematically analyzed differential forms that today bear his name.

Riemann, Bernhard George Fridrich (1826–1866), German mathematician (died in Italy). He produced significant works in differential geometry, complex analysis, and mathematical physics. His work paved the way for the general relativity. He also studied Abelian functions and elliptic integrals, and was an outstanding teacher. His influence on mathematics was enormous. Riemann initially studied theology. In 1846, his father secured sufficient funds for him to begin his studies at the University of Göttin-

gen, where he attended lectures by Gauss. In 1847, he left to study in Berlin for two years. In 1849, he returned to Göttingen. There, in 1857, he became an extraordinary professor. Several dozen mathematical concepts that he discovered are named after him: foremost among these is non-Euclidean Riemannian geometry, which he began to develop in 1854 and which paved the way for general relativity, as well as the Riemann curvature tensor describing the properties of this geometry. Also named after Riemann are: the Cauchy-Riemann conditions for the existence of a derivative of a complex function, the generalized Riemann integral, the Riemann-Stilties integral, the Riemann zeta function, the Riemann theta function, the Riemann sphere, the Riemann surface, and also a crater on the Moon and the asteroid 4167.



Schwartz, Laurent (1915–2002), a French mathematician and pioneer of the theory of distributions on the Western side of the Iron Curtain. Distributions are functions that make sense only in scalar peoduct with a so-called test function. The better the properties of the test function, the more "wild" the properties of the original function can be. The space of test functions is today called Schwartz (Sobolev) space. Laurent Schwartz graduated from the prestigious *École Normale Supérieure*. He taught at the *École Polytechnique in Paris* for several years. He often spoke out against Stalin's totalitarian regime in former Soviet Union.

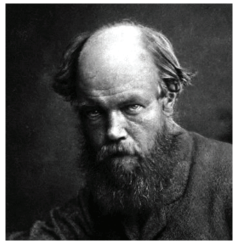


Schwarz Hermann Amandus (1843–1921), a German mathematician specializing in complex analysis, the calculus of variations, and differential geometry. Schwarz originally studied chemistry in Berlin, but eventually, under the influence of the prominent mathematician Karl Weierstrass, he decided to abandon his chemistry studies and began studying mathematics. He worked at the University of Halle, then at the Swiss Federal Institute of Technology in Zurich (ETH). In mathematics, the Schwarz mapping and the Schwarz lemma are named after him.

Sobolev, Sergej (1908–1989), a Soviet mathematician specializing in mathematical analysis and differential equations. Among his most significant contributions was the formulation of the theory of distributions, which he developed on the eastern side of the Iron Curtain independently of the French mathematician Laurent Schwartz. He was the first to formulate what a weak solution of a partial differential equation is (a solution in the sense of a scalar product). This concept was later elaborated in detail by Schwartz. Sobolev graduated from Leningrad University and subsequently worked at a number of Soviet universities. He was awarded the Gold Lomonosov Medal. Naturally, he was also awarded the title of Hero of Socialist Labor.

Stokes, George Gabriel, sir (1819–1903), an Irish mathematician and physicist. He studied continuum mechanics, waves in elastic bodies, acoustics, optics, wave diffraction, polarization, and more. In his honor, the fundamental equations of continuum mechanics are known as the Navier–Stokes equations. In addition to hydrodynamics, he

studied measurements of changes in the gravitational field on the Earth's surface. His interests extended beyond physics; he also studied chemistry and botany. He graduated from Pembroke College in Cambridge, England, where he was appointed to a professorship in 1849. The following are named after him: the Navier-Stokes equations (fluid motion equations), Stokes' theorem (conversion between surface and line integrals), the unit of viscosity, the stokes, the Stokes vector (describing polarization), the Stokes shift (the difference between the absorption and emission lines), the Stokes line (a separatrice in the complex plane), and Stokes' law (the viscous force acting on a body moving in a fluid).



Tait, Peter Guthrie (1831–1901), a Scottish mathematician and physicist, a pioneer of thermodynamics. Together with Kelvin, he authored the *Treatise on Natural Philosophy*. In mathematics, he worked on topology, graph theory, knot theory, and quaternions. He disputed with Maxwell regarding the naming of the gradient operator. Tait's proposal to call the operator "nabl" after the Assyrian harp prevailed. Tait graduated from the University of Edinburgh and then continued his studies at Cambridge.

Taylor, Brook, sir (1685–1731), an English mathematician, formulated the method of polynomial expansion of functions, known as the Taylor series. It was generalized by Pierre Laurent. Taylor also studied finite differences, which today form the basis of many numerical methods. Using differences, he studied the motion of a vibrating string. He was also interested in physics (magnetism and capillary phenomena), philosophy, and religion. Taylor studied mathematics at Cambridge. In 1712, he was elected a Fellow of the Royal Society.

List of Terms

References

Lecture recordings

Server www.aldebaran.cz, Section Study; www.vsprednasky.cz (Czech Language only

Subsequent textbooks

[1] P. Kulhánek: *Vybrané kapitoly z teoretické fyziky II (Statistická fyzika, Elektromagnetické pole, Relativita)*; AGA 2020, ISBN: 978-80-906638-3-1 online verze: http://www.aldebaran.cz/studium/tf2.pdf

[2] P. Kulhánek: *Vybrané kapitoly z teoretické fyziky III* (*Teorie plazmatu, Numerické simulace)*; AGA 2020, ISBN: 978-80-906638-4-8, online verze: http://www.aldebaran.cz/studium/tf3.pdf

Theoretical Mechanics

[3] M. Brdička, A. Hladík: *Teoretická mechanika*; Academia, Praha 1987

[4] J. Kvasnica a kol.: *Mechanika*; Academia, Praha 1988

[5] L. D. Landau, J. M. Lifšic: *Úvod do teoretickej fyziky 1* (mechanika, elektrodynamika); ALFA, Bratislava 1980

[6] D. Halliday, R. Resnick, J. Walker: *Fyzika; Část 1: Mechanika*; VUTIUM/Prometheus, Brno 2000

[7] W. Greiner: *Classical Mechanics – Systems of Particles and Hamiltonian Dynamics*; Second edition, Springer, 2010

[8] E. Tonti: *Variational Formulations of Nonlinear Differential Equations*; Bull. Sciences Acad. R. de Belgique (1969) pp 137–165, 262–278

[9] N. J. Cornish: *The Lagrange Points*; Montana State University/NASA, 1999; online: http://www.physics.montana.edu/faculty/cornish/lagrange.pdf

[10] T. Münch: *The Three-Body Problem and the Lagrangian Points*; Umeå Universitet Institutionen för Fysik, 2008; online: http://www.tp.umu.se/space/Proj_08/T_Munch.pdf.

[11] M. Robnik: *Theory of Adiabatic Invariants*; Socrates Lecture Course at the Physics Department, University of Marburg, Germany, 2004; online: http://www.camtp.uni-mb.si/socrates/marburg2004/robnik.pdf

[12] S. D. Mathur: *The Hamilton-Jacobi Equation*; Ohio State University, 2007. online: http://www.physics.ohio-state.edu/~mathur/821hj.pdf

[13] P. Kulhánek, J. Maloch: *Inverse Variational Problem for the Rail Plasma Accelerator*; Czech. J. Phys. B37 (1987) 561

[14] E. W. Weisstein: *Lorenz Attractor*; MathWorld – A Wolfram Web Resource; online: http://mathworld.wolfram.com/LorenzAttractor.html

[15] J. Horák, L. Krlín, A. Raidl: *Deterministický chaos a podivná kinetika*; Academia, Praha 2007

Quantum Physics

[16] J. Formánek: *Úvod do kvantové teorie I.,II.*; Academia, Praha 2004

[17] J. Formánek: *Úvod do relativistické kvantové mechaniky a kvantové teorie pole*; Karolinum, Praha 2000

[18] R. P. Feynman, R. B. Leighton, M. Sands: *Feynmanovy přednášky z fyziky*, díl 3/3, Fragment, 2006

[19] M. Dušek: *Koncepční otázky kvantové teorie*; Vydavatelství Univerzity Palackého, Olomouc 2002, 238 stran

[20] T. Hey, P. Walters: *Nový kvantový vesmír*; Argo/Dokořán 2003

[21] J. Novotný a kol.: *Základy teorie relativity*; Masarykova Univerzita 2006, online: http://is.muni.cz/elportal/estud/prif/ps06/f5010/zaklady_TR.pdf

[22] P. Cejnar, M. Dušek: *Kvantové hlavolamy I až V*; Vesmír 77/3–7 (1998), online: http://muj.optol.cz/dusek/clanky/popular.htm

[23] B. Simons: *Advanced Quantum Mechanics*; University of Cambridge Course, Cavendish Lab. 2009; online: http://www.tcm.phy.cam.ac.uk/~bds10/aqp.html

[24] Ch. Kittel: *Introduction to Solid State Physics*; Willey, New York 1996

[25] J. Singleton: *Band theory and electronic properties of solids*; Oxford University Press, 2001

[26] E. W. Weisstein: *Airy Functions*; MathWorld – A Wolfram Web Resource, online: http://mathworld.wolfram.com/AiryFunctions.html

[27] M. V. Fedoryuk: *Airy equation; Encyclopedia of Mathematics*; online: http://www.encyclopediaofmath.org/index.php/Airy_equation

[28] T. Jenke et al.: *Realization of a gravity-resonance-spectroscopy technique*; Nature Physics 7, 468–472 (2011)

[29] H. Batelaan, A. Tonomura: *The Aharonov–Bohm effects: Variations on a subtle theme*; Physics Today **62**/9 38 (2009)

[30] J. S. Bell: *Speakable and unspeakable in quantum mechanics*, Cambridge University Press, 1987

[31] J. A. Wheeler, W. H. Zurek: *Quantum Theory and Measurement*; Princeton University Press, 1983

[32] A. Einstein, B. Podolsky, N. Rosen: *Can Quantum-Mechanical Description of Physical Reality be Considered Complete?*; Physical Review **47** (1935) 777–780

[33] Vojtěch Hála: *Kvantová kryptografie*; Aldebaran bulletin 14/2005; online: https://aldebaran.cz/bulletin/2005_14_kry.php

[34] Petr Kulhánek: *Kvantové počítače*; Aldebaran bulletin 21/2003; online: https://www.aldebaran.cz/bulletin/2003_21_qua.php

[35] Petr Kulhánek: Teleportace; Aldebaran bulletin 33/2017 online: https://www.aldebaran.cz/bulletin/2017_33_tel.php

[36] Serge Haroche: *Controlling Photons in a Box and Exploring the Quantum to Classical Boundary*; Nobel Lecture (2012)

[37] Miroslav Havránek: *Strašidelné interakce ve kvantovém světě*; Aldebaran bulletin 19/2020; online: https://www.aldebaran.cz/bulletin/2020_19_kva.php

[38] Jennifer Chu: *Light from ancient quasars helps confirm quantum entanglement*; MIT News 19 Aug 2018; online: http://news.mit.edu/2018/light-ancient-quasars-helps-confirm-quantum-entanglement-0820/

[39] D. Rauch et al.: *Cosmic Bell Test Using Random Measurement Settings from High-Redshift Quasars*; arXiv:1808.05966, 2018; online: https://arxiv.org/abs/1808.05966

Mathematics

[40] Jozef Kvasnica: *Matematický aparát fyziky*; Academia 1997

[41] Jiří Kopáček: *Matematická analýza nejen pro fyziky (I až IV)*; Matfyzpress 2010

[42] Pavel Čihák: *Matematická analýza nejen pro fyziky V*; Matfyzpress 2016

[43] Jiří Kopáček: *Příklady z matematiky nejen pro fyziky (I až V)*; Matfyzpress 2017

[44] Karel Rektorys a spol. *Přehled užité matematiky I, II*; Prometheus 2009

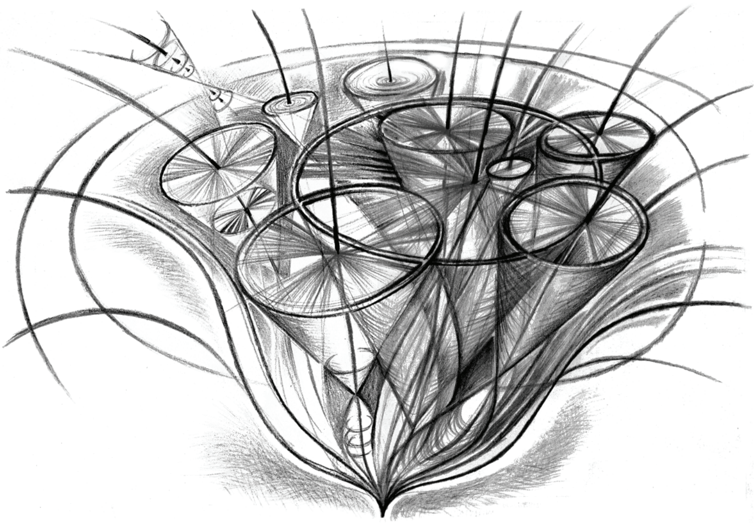
[45] Isaac Newton: *Philosophiae Naturalis Principia Mathematica*; 1687

[46] Ján Čižmár: *Dejiny matematiky od najstarších čias po súčasnosť*; Perfekt 2017

[47] Michael Woolfson, Malcolm Woolfson: *Mathematics for Physics*; Oxford University Press 2006

[48] Alexander Atlant, Jan von Delft: Mathematics for Physicists; Cambridge University Press 2019

[49] Michael Stone: *Mathematics for Physics – A Guided Tour for Graduate Students*; Cambridge University Press 2009



Petr Kulhánek
Selected Topics in Theoretical Physics I
(1. Theoretical Mechanics, 2. Quantum Theory, 3. Matematika pro fyziku)

Author: prof. RNDr. Petr Kulhánek, CSc.
Graphics and cover: Ing. arch. Ivan Havlíček
Format: 165×235 mm, 416 stran, 216 obrázků
Publisher: AGA (Aldebaran Group for Astrophysics)
Lecture recordings: aldebaran.cz (Daniel Handl, FEL ČVUT)
Edition: First, English version, 2026, Prague
Last updated: May 12, 2026

ISBN 978-80-906638-2-4