THEORETICAL PHYSICS FOR UNDERGRADS

IMPORTANT TOPICS SIMPLY EXPLAINED

ALEXANDER FUFAEV

Hello there!

My name is Alexander Fufaev. You may know me from YouTube or from my website "fufaev.org". I'm a freelance theoretical physicist, and in this book I explain theoretical physics - the stuff you're going to study, are studying, or have studied in your physics degree, but haven't quite grasped yet. This book aims to provide an **intuitive understanding of the key topics in theoretical physics at undergraduate level** and serves as a perfect entry point for deepening your understanding. You should be familiar with the **basics of vector, differential and integral calculus**. If these concepts are new to you, it's a good idea to understand them first before delving into this book.

Weekly assignments at university are not easy for beginners and require a lot of time. The content is rushed through, so it's easy to find yourself not only failing to understand many topics, but also struggling to keep up with the weekly assignments. This often leads to not completing the required credits or failing the exams at the end of the semester.

Having completed my M.Sc. in Physics, I know exactly in retrospect what I would have needed in the theoretical physics modules to meet the academic requirements and pass the exams. A fundamental, intuitive understanding of the topics would have been essential because often I didn't know what the assignments are talking about. In some cases, I understood the assignment but didn't know how to approach it. There were also assignments I could solve, but I wasn't sure what or why I was calculating something.

If you work through this book from the first to the last chapter, you'll find it much easier to **master the assignments in theoretical physics** and **pass the exams more easily**.

May physics be with you!

A. FufaeV

What you'll learn

Mathematical Tools

1. Differential equations

More: en.fufaev.org/differential-equations

If you plan to deal with...

- the atomic world,
- the movement of the planets,
- chemical processes,
- electrical circuits,
- weather forecasts
- or with the spread of a virus

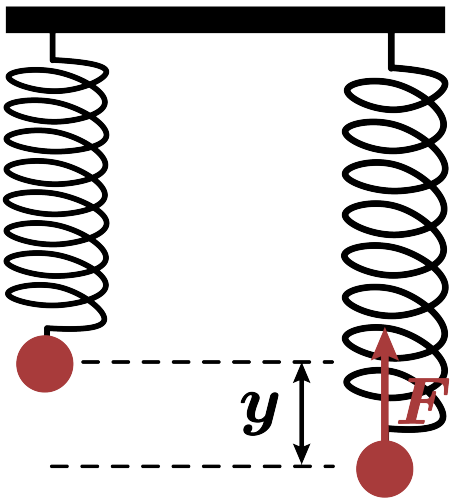
then you will eventually encounter **differential equations**. You will encounter differential equations in every part of theoretical physics, so it is important to understand how to work with differential equations.

1.1 What is a differential equation?

Let's take a look at Hooke's law as a simple example:

$$F = -D\,y \tag{1.1}$$

This law describes the **restoring force** F on a mass attached to a spring. The mass experiences this force when you deflect it from its rest position by the **distance** y. D is a constant coefficient that describes how hard it is to stretch or compress the spring.



The **mass** m is hidden in the force. We can write the force using Newton's second law as $m\,a$:

$$m\,a \;=\; -D\,y \tag{1.2}$$

Here a is the **acceleration** that the mass experiences when it is deflected by y from its rest position. As soon as you pull on the mass and release it, the spring will start to oscillate back and forth. Without friction, as in this case, it will never come to a standstill.

While the mass oscillates, the displacement y naturally changes. The displacement is therefore dependent on the time t. This means that the acceleration a also depends on the time. The mass naturally remains the same at all times, regardless of how much the spring is deflected. This also applies to a good approximation for the spring constant D:

$$m\,a(t) \;=\; -D\,y(t) \tag{1.3}$$

If we only bring m to the other side, we can use this equation to calculate the acceleration that the mass experiences for each displacement y:

$$a(t) = -\frac{D}{m} y(t) \tag{1.4}$$

But what if we are interested in the question: **At what displacement y will the spring be after 24 seconds?**

In order to answer such a future question, we need to know how exactly y depends on the time t. We only know THAT y depends on time, but not HOW.

Differential equations come into play for such future questions. We can easily show that the acceleration a is the second time derivative of the distance covered, so in our case it is the second derivative of y with respect to time t:

$$\frac{\mathrm{d}^2 y(t)}{\mathrm{d}t^2} = -\frac{D}{m} y(t) \tag{1.5}$$

We have now set up a differential equation for the displacement $y(t)$! You can recognize a differential equation by the fact that it **in addition to the unknown function $y(t)$, there are also derivatives of this function**. As in this case, the second derivative of y with respect to time t.

We can therefore conclude: **A differential equation is an equation that contains an unknown function and derivatives of this function**.

1.2 Different notations

You will certainly encounter many ways of writing a differential equation. We have written down our equation 1.5 for the spring oscillation of a mass in a **Leibniz notation**. Here it is again:

$$\frac{\mathrm{d}^2 y(t)}{\mathrm{d}t^2} = -\frac{D}{m} y(t) \tag{1.6}$$

You will often come across this **in physics**. We can also write it down a little more compactly without mentioning the time dependency:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = -\frac{D}{m} y \tag{1.7}$$

If the function $y(t)$ only depends on the time t, then we can write down the time derivative even more compactly using **Newton notation**. A time derivative of y corresponds to a point above the $\dot{y}$. With two derivatives, as in our case, there would therefore be two points above the function:

$$\ddot{y} = -\frac{D}{m} y \tag{1.8}$$

Obviously, this notation is rather unsuitable for considering the tenth derivative.

Another notation that you are more familiar with from mathematics is **Lagrange notation**. Here, dashes are used for the derivatives. So two dashes for the second derivative:

$$y'' = -\frac{D}{m} y \tag{1.9}$$

With Lagrange notation, it should be clear from the context with respect to which variable is being differentiated. If this is not clear, the variable on which y depends must be mentioned explicitly:

$$y''(t) = -\frac{D}{m} y(t) \tag{1.10}$$

Each notation has its advantages and disadvantages. However, it should be noted that these are only different **notations** that represent the **same physics**. Rearranging and renaming does not change the physics under the hood of a differential equation. We could also call the displacement x:

$$\frac{\mathrm{d}^2 x(t)}{\mathrm{d}t^2} = -\frac{D}{m} x(t) \tag{1.11}$$

1.3 What should I do with a differential equation?

In order to answer our original question: "At what displacement will the spring be after 24 seconds?", we need to **solve the differential equation**. Solving a differential equation means that you have to find out **how** the unknown function $y(t)$ depends exactly on the variable t: $y(t) = ...$

For simple differential equations, such as that of the oscillating mass, there are methods that you can use to get the unknown function $y(t)$. Keep in mind, however, that there is **no general recipe** for solving any differential equation. For some differential equations **there is not even an analytical solution**! »Non-analytical« means that you cannot write down a concrete equation for the function $y(t) = ...$

Differential equations that cannot be solved analytically can only be solved on the computer **numerically**. Then the computer does not spit out a concrete formula $y(t) = ...$, but data points $y(t_1), y(t_2), y(t_3), ...$, which you can plot in a y-t diagram and use to analyze the numerically solved function $y(t)$.

1.4 Recognize differential equations

As soon as you come across a differential equation, the first thing to do is to find out,

- what the **unknown function** is
- and **on which variables** it depends.

In the differential equation 1.5 of the oscillating mass, the function we are looking for is called $y(t)$ and it depends on the variable t:

$$\frac{\mathrm{d}^2 y(t)}{\mathrm{d}t^2} = -\frac{D}{m} y(t) \tag{1.12}$$

As an example, take a look at the three-dimensional wave equation that describes the electric field E of an electromagnetic wave propagating at the speed of light

c:

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} \tag{1.13}$$

What is the unknown function in this equation? It is the function E. Why is that? Because the differential equation contains the derivatives of E. On which variables does the function dependent? The dependency is not explicitly stated here, but you can immediately see from the curved del character ∂ that it must depend on several variables. You can immediately see from the derivatives in the differential equation that the unknown electric field depends on x, y, z and t. So on a total of four variables: $E(x, y, z, t)$.

Let's look at a slightly more complex example. This **system of differential equations** describes how a mass moves in a gravitational field:

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = G\frac{m}{x^2 + y^2 + z^2} \tag{1.14}$$

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = G\frac{m}{x^2 + y^2 + z^2} \tag{1.15}$$

$$\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = G\frac{m}{x^2 + y^2 + z^2} \tag{1.16}$$

You have a **coupled system of differential equations** in front of you. In this case, a single differential equation is not sufficient to describe the motion of a mass in the gravitational field. In fact, three functions are required here, namely the trajectories $x(t)$, $y(t)$ and $z(t)$, which determine a position $\boldsymbol{r}(t) = (x(t), y(t), z(t))$ of the mass in three-dimensional space. Each function describes the movement in one of the three spatial directions.

By »coupled« we mean that, for example, the first differential equation for the function $x(t)$, also contains the function $y(t)$. This means that we cannot simply solve the first differential equation independently of the second differential equation, because the second differential equation tells us how it behaves in the first differential equation. The functions we are looking for occur in all three differential equations, which means we have to solve all three

equations **simultaneously**. You may learn exactly how to do this in your math lectures.

1.5 Classify a differential equation

There are many different differential equations. But if you look closely, you will notice that some differential equations are similar to each other.

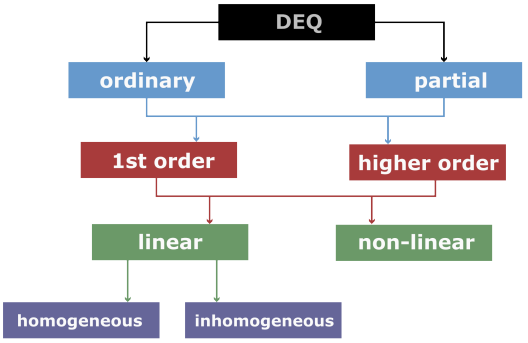
After you have found out what the unknown function is and on which variables it depends, you should answer a few more basic questions **to choose the appropriate solution method** for the differential equation. We need to **classify** the differential equation:

- **Is the differential equation ordinary or partial?** Partial differential equations describe multidimensional problems and are significantly more complex.

- **Of which order is the differential equation?** First-order differential equations are usually easy to solve and describe, for example, exponential behavior such as radioactive decay or the cooling of a liquid. Second-order differential equations, on the other hand, are somewhat more complex and also occur frequently in nature. Maxwell's equations in electrodynamics, the Schrödinger equation in quantum mechanics - these are all second-order differential equations.

- **Is the differential equation linear or non-linear?** The superposition principle applies to linear differential equations, which is very useful for describing electromagnetic phenomena, for example. Non-linear differential equations are much more complex and are used, for example, in non-linear electronics to describe superconducting currents. In addition, chaos can only occur with non-linear differential equations of third or higher order. As soon as you come across a non-linear differential equation, you can actually throw away pen and paper and treat the equation numerically directly on the computer. Most non-linear differential equations cannot even be solved analytically!

- **Is the linear differential equation homogeneous or inhomogeneous?** Homogeneous linear partial differential equations (PDEs) are simpler than inhomogeneous ones and describe, for example, an undisturbed oscillation, while inhomogeneous PDEs are also able to describe externally disturbed oscillations.



Once you have classified a differential equation, you can use a suitable solution method to solve the equation. Even if there is no specific solution method, the classification tells you how complex a differential equation is.

Let's classify the DGL for the oscillating spring, wave equation, mass in the gravitational field and for the decay law.

1.5.1 Ordinary or partial?

Let's take a look at the differential equation for the oscillating mass:

$$\frac{\mathrm{d}^2 y(t)}{\mathrm{d}t^2} + \frac{D}{m} y(t) = 0 \tag{1.17}$$

This is an **ordinary** differential equation. Ordinary means that the function $y(t)$ depends on only **one** variable. In this case on the time t.

What about the wave equation for the electric field?

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} \tag{1.18}$$

This is a **partial** differential equation. "Partial" means that the function you

are looking for depends on **at least two** variables and there **are derivatives with respect to these variables**. In this case, the function E depends on four variables: x, y, z and t. And derivatives with respect to these variables also occur in the differential equation.

1.5.2 Of what order?

The differential equation for the oscillating mass is a differential equation of **2nd order**. The order of a differential equation is the **highest occurring derivative** of the unknown function:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + \frac{D}{m} y = 0 \tag{1.19}$$

Since the second derivative is the highest and even the only derivative of in the differential equation, the oscillating mass at the spring is the 2nd order differential equation.

1.5.2.1 How to reduce the order

It is always possible to convert a higher order differential equation **into a system of 1st order differential equations**. Sometimes this procedure is helpful when solving a differential equation.

For example, we can convert the differential equation 1.19 for the oscillating mass on the spring (second-order differential equation) **into two coupled first-order differential equations**. All we have to do is introduce a new function, let's call it v and define it as the first time derivative of the unknown function y:

$$v = \frac{\mathrm{d}y}{\mathrm{d}t} \tag{1.20}$$

This is already **one of the two** coupled first-order differential equations. Now we just have to express the second derivative in the original differential equation with the derivative of v. Then we get the second 1st order differential equation.

The two coupled differential equations are as follows:

$$v - \frac{\mathrm{d}y}{\mathrm{d}t} = 0 \tag{1.21}$$

$$\frac{\mathrm{d}v}{\mathrm{d}t} + \frac{D}{m} y = 0 \tag{1.22}$$

The two differential equations are **coupled**. We must therefore solve them simultaneously. They are coupled because y occurs in the first equation and v in the second equation.

You can always proceed in this way if you want to **reduce the order of a differential equation**. The price you have to pay for this is additional coupled differential equations.

Let's continue. Of what order is the **differential equation for the decay law**?

$$-\lambda N = \frac{\mathrm{d}N}{\mathrm{d}t} \tag{1.23}$$

This is a differential equation **1st order**, because the highest occurring derivative of the function $N(t)$ is the first derivative.

1.5.3 Linear or non-linear?

The differential equation for the oscillating mass is **linear**:

$$\left(\frac{\mathrm{d}^2 y}{\mathrm{d}t^2}\right)^1 + \frac{D}{m} y^1 = 0 \tag{1.24}$$

»Linear« means that the **unknown function and its derivatives only contain powers of 1** and there are **no products of derivatives** with the function, such as y^2 or $y\,\ddot{y}$. There are also **no composition of functions**, such as $\sin(y(t))$ or $\sqrt{y(t)}$. For products, compositions and higher powers, we speak of **nonlinear** differential equations.

Note! The »to the power of two« in the second derivative in the Leibniz notation $\frac{\mathrm{d}^2 y}{\mathrm{d}t^2}$ is not a power of the derivative, but merely a notation for the

second derivative.

The decay law is also linear:

$$-\lambda N^1 = \left(\frac{\mathrm{d}N}{\mathrm{d}t}\right)^1 \tag{1.25}$$

What about the wave equation? It is also linear:

$$\left(\frac{\partial^2 E}{\partial x^2}\right)^1 + \left(\frac{\partial^2 E}{\partial y^2}\right)^1 + \left(\frac{\partial^2 E}{\partial z^2}\right)^1 = \frac{1}{c^2}\left(\frac{\partial^2 E}{\partial t^2}\right)^1 \tag{1.26}$$

The coupled differential equation system for the motion of a mass in the gravitational field, on the other hand, is non-linear because higher powers of the functions $x(t)$, $y(t)$ and $z(t)$ occur there, namely x^2, y^2 and z^2.

1.5.4 Homogeneous or inhomogeneous?

For homogeneous and inhomogeneous types of differential equations, the **coefficients** multiplied by the unknown function and its derivatives are important. For some solution methods, it is important to distinguish between...

- **constant** coefficients - these **do not** depend on the variables on which the unknown function also depends.
- **non-constant** coefficients - these **depend** on the variables on which the unknown function depends.

A coefficient does not necessarily have to be multiplied by the unknown function or its derivative. The coefficient can also stand alone! In this case, the coefficient is referred to as **perturbation function**.

Let's take another look at the oscillating mass:

$$1\,\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + \frac{D}{m}\,y = 0 \tag{1.27}$$

In this differential equation, there is an interesting coefficient that is multiplied

by the unknown function y, namely $\frac{D}{m}$. Strictly speaking, the second derivative is also preceded by a coefficient, namely 1 and the single coefficient, i.e. the perturbation function, is 0. If the **perturbation function is zero**, then we call the linear differential equation **homogeneous**.

The wave equation also has no perturbation function (no stand-alone coefficient). The differential equation is therefore homogeneous:

$$1\frac{\partial^2 E}{\partial x^2} + 1\frac{\partial^2 E}{\partial y^2} + 1\frac{\partial^2 E}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} \tag{1.28}$$

The differential equation for a **forced oscillation**, on the other hand, is **inhomogeneous**:

$$1\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + \mu\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{D}{m}y = F(t) \tag{1.29}$$

Here, the external force $F(t)$ corresponds to the perturbation function. As you can see, it stands alone without being multiplied by the function $y(t)$ or its derivatives. In addition, the perturbation function is time-dependent, so it is a non-constant coefficient.

1.6 Constraints

A differential equation alone is not sufficient to describe a physical system **uniquely**. The solution of a differential equation describes **many** possible systems that exhibit a certain behavior.

For example, the solution $N(t)$ of the decay law describes an **exponential behavior**. However, the knowledge of an exponential **behavior is not sufficient** to be able to say specifically how many nuclei $N(t = 10)$ have decayed after 10 seconds.

For this very reason, every differential equation usually has **constraints**. These are additional information that must be given for a differential equation in order to determine the **unique solution of the differential equation**. The number of necessary constraints depends on the order of the equation.

- Only **one constraint** is required for a 1st order differential equation: One function value of the unknown function: $y(t)$. For the decay law, for example, it should be known how many undecayed nuclei $N(t = 0)$ there were at the time $t = 0$. For example, 1000 nuclei. Then the constraint is: $N(0) = 1000$.

- For a 2nd order differential equation, **two constraints** are necessary: A function value of the unknown function $y(t)$ and $y'(t)$. For the oscillating mass, for example, it should be known what displacement $y(t = 0)$ the spring had at the time $t = 0$, e.g. $y(0) = 1$ and what the velocity $y'(t = 0)$ of the mass was at this time, e.g. $y'(0) = 0$.

- For a 3rd-order differential equation, **three constraints** would then be necessary so that the solution of the differential equation uniquely describes the system under consideration: A function value $y(t) = A$ of the unknown function, a function value $y'(t) = B$ e.g. its first derivative and a function value $y''(t) = C$ e.g. its second derivative.

- For a 4th order, **four constraints** would then be necessary and so on...

We can therefore state: **In order to uniquely determine the solution of a nth order differential equation, n constraints are necessary**.

Most of the time you will come across **initial conditions** and **boundary conditions**. These are also just names for constraints that tell you what kind of information you have about the system.

1.6.1 Initial conditions

Sometimes, for example, you know what the system **was at a certain point in time** $t = t_0$. This can be the initial time at which you deflected and released an oscillating mass. In this case, we speak of **initial conditions**. You determine what the displacement $y(t_0)$ was at **a** specific point in time. And since we need two constraints for the 2nd order equation, you also specify what the displacement (i.e. the velocity) $y'(t_0)$ was at time t_0.

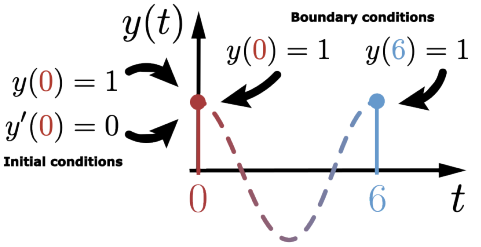
We call a **differential equation together with its initial conditions** as

initial value problem. If we solve the initial value problem, we can use the solution to predict the future behavior of a system.

1.6.2 Boundary conditions

Let's take another look at the second-order differential equation for the oscillating mass. And let's assume that we know the displacement $y(t_0) = y_0$ at time t_0. Sometimes we are unlucky and do not know **what velocity** the oscillating mass had at the initial time t_0. We therefore do not know the derivative $y'(t_0)$ at the time t_0 at which we know the displacement. However, we need **two** constraints, otherwise the solution is not unique and we cannot use the function $y(t)$ to calculate specific numbers for displacement.

However, we may know that after $t = 6$ seconds, for example, the oscillating mass was in the maximum deflected state. We therefore know the displacement $y(6) = y_6$.



If we know the constraints, such as $y(t_1) = y_1$ and $y(t_2) = y_2$, which **at two different times** t_1 and t_2 describe the system, then we speak of **boundary conditions**: $y(t_1) = y_1$ and $y(t_2) = y_2$.

We call a **differential equation together with two boundary conditions** as **boundary value problem**. If we solve the boundary value problem, we can use the solution to predict how the system will behave **within** these boundary values.

The 'function value at two different **times**' was of course just an example. Instead of time, it could be any variable that defines the system, usually at the boundaries: at different times, at different **locations**, at different **angles** and

so on.

You have now learned the most important basics of differential equations. This knowledge will help you as undergrad.

2. Tensors

More: en.fufaev.org/tensors

Before we fully understand tensors in their general definition, let's first get to know them from an engineering perspective. As long as you understand scalars, vectors, and matrices, you'll find it easy to understand tensors from this perspective because tensors are nothing but a **gegeneralization of scalars, vectors, and matrices**. Just as we use scalars, vectors, and matrices to describe physical laws, we can use tensors to describe physics. Tensors are an even more powerful tool with which we can describe physics that cannot be described solely with scalars, vectors, and matrices. In order to develop the modern theory of gravitation, Albert Einstein had to first understand the concept of tensors. Only then could he mathematically formulate the general theory of relativity.

2.1 Tensors of Zeroth and First Order

Let's start with the simplest tensor: The **zeroth order tensor**. This is a scalar σ, that is an ordinary number. This tensor has a single component and represents, for example, the electrical conductivity of a wire. This zero-order tensor σ indicates how well a wire conducts electricity in this case.

A slightly more complex tensor, let's call it $\boldsymbol{j}$, is a first-order tensor. This is a vector with three components j_1, j_2 and j_3 in three-dimensional space:

$$\boldsymbol{j} = \begin{bmatrix} j_1 \\ j_2 \\ j_3 \end{bmatrix} \tag{2.1}$$

In Eq. 2.1 we have represented the first order tensor as a column vector. Of course, we can also represent it as a row vector:

$$\boldsymbol{j} = [j_1,\ j_2,\ j_3] \tag{2.2}$$

At this stage, it doesn't matter how we write down the components. But remember that it will play a role later!

The notation of first-order tensors as row or column vectors only makes sense if we are working with concrete numbers, such as in computer physics, where we use tensors to obtain concrete numbers. In order to calculate theoretically, for example to derive equations or simply to formulate a physical theory, the tensors are formulated compactly in index notation. You are probably already familiar with this from vector calculus. Instead of writing out all three components of the first-order tensor, we write them with an index k, that is, j_k. What we call the index does not matter. j_k stands for the first component j_1, second j_2 or third j_3 component, depending on what we actually use for index k. In theoretical physics, we usually do not use anything specific because we want to write the physics as generally and compactly as possible.

From this index notation j_k it is not clear whether it represents a column or row vector. This is not good, because later it will be important to distinguish between column and row vectors. But we can easily introduce this distinction into our index notation by noting the **index below** j_k if we mean a **row vector**. And we note the **index top** j^k if we mean a **column vector**. The notation of indices above and below has a deeper meaning, which we will get to know later. At this stage, we only distinguish the representation of the first-order tensor.

2.2 Second-order tensors

The next more complex tensor is **second-order tensor**. Let us also denote this tensor by σ, because a second-order tensor can describe electrical conductivity. Electrical conductivity as a zero-level tensor describes isotropic materials. The conductivity as a second-order tensor, on the other hand, describes a non-isotropic material in which the conductivity varies depending on the direction in which the current flows.

You have certainly already become familiar with this tensor in mathematics in the **matrix representation**. In a three-dimensional space, the **second-order tensor is a 3x3 matrix**:

$$\sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \tag{2.3}$$

We also use index notation for the second-order tensor and note the components of the matrix with σ_{mk}, for example. The indices m and k can have the values 1, 2 or 3. The index m specifies the row and the index k specifies the column.

2.3 Tensors of higher orders

We can continue the game and consider a **third-order tensor**. This then has three indices σ_{mkn}. The **fourth-order tensor** has four indices: σ_{mkni}. And so on. The indices of a tensor of any level can also be superscripted. For example, the indices mk of the fourth-order tensor can be at the top and the indices ni at the bottom: $\sigma^{mk}{}_{ni}$. You will find out exactly what this means in the next chapters.

The **number of components** d^r of a tensor depends on the **space dimension** d and on the **level (rank)** r of the tensor. In a three-dimensional space ($d = 3$), a second-order tensor ($r = 2$) therefore has $3^2 = 9$ components.

2.4 Symmetric and antisymmetric tensors

In theoretical physics, especially in the theory of relativity and quantum mechanics, we will regularly encounter symmetric and antisymmetric tensors.

A **symmetric tensor** t_{ij} remains the same if we swap its indices: $t_{ij} = t_{ji}$. Specifically, swapping the indices of the second-order tensor as a matrix means that the matrix remains the same if we transpose it, that is, mirror the rows and columns on the diagonal:

$$\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} = \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \tag{2.4}$$

This symmetry property of tensors is very useful and simplifies calculations in computer physics enormously. Moreover, this property is crucial in quantum mechanics, because symmetric matrices have real eigenvalues. They therefore represent physical quantities (we call them observables) that we can measure in an experiment. So if you have a symmetric tensor in front of you, as a theoretical physicist you should immediately get a dopamine kick. The **Kronecker delta** δ_{mk}, for example, is a concrete example of a simple symmetric tensor.

We have considered a second-order tensor. What if the tensor is of a higher order? What about its symmetry property then? For example, if the tensor has four indices: t_{mkni}, and it remains the same if we swap the first two indices: $t_{mkni} = t_{kmni}$, then we are talking about a **symmetric tensor in the first two indices** or more precisely: **symmetric in the mk indices**.

However, we will also encounter tensors that are antisymmetric. A **antisymmetric tensor** t_{ij} changes sign when we swap its indices: $t_{ij} = -t_{ji}$. If the antisymmetric tensor is represented as a matrix, then it is equal to its negative transpose:

$$\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} = - \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \tag{2.5}$$

Unfortunately, most tensors are neither symmetric nor antisymmetric. But the great thing is: Mathematically, we can split every **tensor t into a symmetric s and an antisymmetric a part**: $t = s + a$.

Let's take a look at how we can practically decompose a general tensor t_{ij} of the second-order into the two parts.

1. The **symmetric part** s_{ij} of the tensor t_{ij} is $s_{ij} = \frac{1}{2}(t_{ij} + t_{ji})$. Here we have swapped the two indices and added the two tensors together. The factor $\frac{1}{2}$ is important because we have counted the tensor twice here.
2. The **antisymmetric part** a_{ij} of the tensor t_{ij} ist $a_{ij} = \frac{1}{2}(t_{ij} - t_{ji})$. Here we have swapped the two indices, so the swapped tensor gets a minus sign. Then we add the two tensors together. The factor $\frac{1}{2}$ is also important here.
3. We then add the symmetric and antisymmetric components together to obtain the total tensor:

$$t_{ij} = \frac{1}{2}(t_{ij} + t_{ji}) \; + \; \frac{1}{2}(t_{ij} - t_{ji}) \tag{2.6}$$

The first term in Eq. 2.6 is the symmetric part of the tensor t_{ij} and the second term is the antisymmetric part.

2.5 Combine tensors

We can do little with tensors alone. We need to be able to perform calculations with them. There are several arithmetic operations that we can use to combine two tensors a and b into a new tensor c.

2.5.1 Addition of tensors

We can **add** two tensors a_{ij} and b_{ij} of the same order:

$$c_{ij} = a_{ij} + b_{ij} \tag{2.7}$$

The result is a new tensor c_{ij} of the same order. If we represent the tensors a and b as matrices, then adding tensors is nothing other than adding matrices

component by component. The component a_{11} of the matrix a in the **first row** and **first column** is added with the component b_{11} of the matrix b, which is also in the same column and the same row. This is how matrix addition works. We proceed in the same way with all other components. The result is the matrix c:

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}}_{a} + \underbrace{\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}}_{b} = \underbrace{\begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \\ a_{31}+b_{31} & a_{32}+b_{32} & a_{33}+b_{33} \end{bmatrix}}_{c} \tag{2.8}$$

2.5.2 Subtraction of tensors

We can **subtract** two tensors a_{ij} and b_{ij} of the same order:

$$c_{ij} = a_{ij} - b_{ij} \tag{2.9}$$

The result is a new tensor c_{ij} of the same order. Subtraction works in the same way as addition. Simply replace the plus sign in Eq. 2.9 with a minus sign.

2.5.3 The outer product of tensors (tensor product)

The next operation is probably new to you, namely the **outer product** $\otimes$. Sometimes it is also called **tensor product**. Here, the same components are not offset against each other as with the addition and subtraction of tensors. For this operation, the indices of the tensor a_{ij} and b_{km} must therefore be designated differently. The tensor b_{km} has therefore been given the indices k and m.

$$c_{ijkm} = a_{ij} \otimes b_{km} \tag{2.10}$$

If we form the tensor product 2.10 of second-order tensors, then **the result c_{ijkm} is a fourth-order tensor**. If, on the other hand, we form the tensor product of tensors a_i and b_k of the first order, then the result is a tensor of the second

order:

$$c_{ik} = a_i \otimes b_k = a_i b_k \tag{2.11}$$

This is how the tensor product works with two tensors of any order. The only exceptions are zero-order tensors. In this case, the result remains a zero-order tensor. The tensor sign $\otimes$ is usually omitted in 2.10 and 2.11.

Let's make a concrete example of the tensor product that we can illustrate well, namely the tensor product of first-order tensors as in Eq. 2.11. They are represented by the vectors: $a = [a_1, a_2, a_3]$ and $b = [b_1, b_2, b_3]$. The result is a second-order tensor represented by a matrix:

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \otimes \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \underbrace{\begin{bmatrix} a_1 b_1 & a_1 b_2 & a_1 b_3 \\ a_2 b_1 & a_2 b_2 & a_2 b_3 \\ a_3 b_1 & a_3 b_2 & a_3 b_3 \end{bmatrix}}_{c} \tag{2.12}$$

The first index, the index i, numbers the rows of the matrix by definition and the second index, the index k, numbers the columns.

The tensor product does not necessarily have to be between two tensors of the same order. For example, we can also form the tensor product with the third-order tensor A_{ijm} and the second-order tensor B_{kn}. The result is a fifth-order tensor C_{ijmkn}:

$$C_{ijmkn} = A_{ijm} B_{kn} \tag{2.13}$$

As you have probably already noticed, for example, B_{kn} specifically represents the kn-th component of the tensor B. And A_{ijm} is the ijm-th component of the tensor A. If we form the tensor product as in 2.13, then this is the tensor product of the components. The result is the $ijmkn$-th component of the tensor C. When we write a tensor with indices, we always mean its components. Nevertheless, we casually say »tensor« for its component notation.

The tensor product naturally works in the same way with superscript indices, which we will get to know in the next chapter. If the indices ij are at the top of the tensor A, then they must also be at the top of the resulting tensor C:

$$C^{ij}{}_{mkn} = A^{ij}{}_{m} B_{kn} \tag{2.14}$$

2.5.4 Contraction of tensors

The next arithmetic operation we can perform is the **contraction of a tensor**. Let's take the fourth order tensor as an example: t_{ijmk}. The contraction of this tensor means the following:

1. We select two of its indices. For example, the index i and m: t_{ijmk}.
2. Then we set the two indices equal: $i = m$. For example, we can call them both i: t_{iiik}.
3. We then sum over the index i:

$$\boxed{+}_{i=1} t_{iiik}$$

In three-dimensional space, the index i ranges from 1 to 3, so the contraction of the tensor t_{ijmk} results in the following sum:

$$\boxed{+}_{i=1} t_{iiik} = t_{1j1k} + t_{2j2k} + t_{3j3k}$$

If we want to communicate these three steps to another physicist, then we say: Contraction of the indices i and m of the tensor t_{ijmk}. Or: **contraction of the first and third index of the tensor** t_{ijmk}.

The contraction is very useful because it reduces the **order of a tensor**. For example, the contraction of the fourth-order tensor t_{ijmk} has reduced its order by two. The result of the contraction is a second-order tensor: $t_{iik} = c_{jk}$.

In physics, we use the **Einstein sum convention**, which states that we can omit the sum sign in $\boxed{+}_{i=1} t_{iiik}$ to simplify the notation if two identical indices

appear in a tensor. With the tensor t_{ijik} in combination with the Einstein summation convention, summation is therefore performed using the index i:

$$t_{ijik} = t_{1j1k} + t_{2j2k} + t_{3j3k} \tag{2.15}$$

If we contract a second-order tensor t_{ii} then the contraction is also called **trace of the tensor**:

$$t_{ii} = t_{11} + t_{22} + t_{33} = \mathrm{Tr}\,(t) \tag{2.16}$$

The result is a zero-order tensor, that is, a scalar.

Of course, we can also contract the indices of two different tensors. For example, let's take a tensor M_{ij} and a tensor v_k. The tensor product $M_{ij}v_k$ without contraction results in a third-order tensor. Now we contract the indices j and k. Then, in the matrix and vector representation, this corresponds exactly to the **multiplication of a matrix M with a vector $\boldsymbol{v}$**. The result u_i is a first-order tensor, that is, a vector:

$$\underbrace{M_{ij}}_{M}\underbrace{v_j}_{\boldsymbol{v}} = \underbrace{u_i}_{\boldsymbol{u}} \tag{2.17}$$

2.6 Kronecker Delta

More: en.fufaev.org/kronecker-delta

The Kronecker delta δ_{ij} has become indispensable in theoretical physics. You will encounter this relatively simple, yet powerful tensor in practically all areas of theoretical physics. It is used, for example, to make long expressions more compact and to simplify complicated expressions. In combination with the Levi-Civita symbol, which you will learn in the next chapter, the two tensors are very useful!

Kronecker delta δ_{ij} is a small Greek delta that is either 1 or 0, depending on the value of the two indices i and j. Kronecker delta is equal to 1 if i and j are

equal. And the **Kronecker delta is equal to 0 if i and j are not equal**:

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{2.18}$$

Let's take a few examples:

- $\delta_{11} = 1$, as both indices are equal.
- $\delta_{23} = 0$, as both indices are different.
- $a\delta_{33} = a \cdot 1 = a$
- $\delta_{23}\delta_{22} = 0$

Also note that, unless otherwise stated, we use the Einstein summation convention we learned earlier. The same index is used for summation:

$$\delta_{ij}\delta_{jk} = \delta_{i1}\delta_{1k} + \delta_{i2}\delta_{2k} + \delta_{i3}\delta_{3k} \tag{2.19}$$

However, there are exceptions to the Einstein summation convention. For example, with the differential operator ∂_i. You are not allowed to move ∂_i in front of f_i if ∂_i acts as a derivative of f_i:

$$\partial_i f_i \neq f_i \partial_i$$

So be careful with operators in index notation!

Let's look at four useful rules with Kronecker delta that you can always use when summing over double indices.

2.6.1 Kronecker delta is symmetric

Indices, here i and j, **may be swapped**:

$$\delta_{ij} = \delta_{ji} \tag{2.20}$$

2.6.2 Contracting with Kronecker delta

If the product of two or more Kronecker deltas contains a summation index, here j, then the **product can be combined**, whereby the summation index j disappears:

$$\delta_{ij}\,\delta_{jk} \;=\; \delta_{ik} \tag{2.21}$$

An example with two summation indices:

$$\delta_{ij}\delta_{kj}\delta_{in} \;=\; \delta_{kn} \tag{2.22}$$

This should make it clear that the **order of contraction** of the Kronecker delta is irrelevant.

We can also apply this rule to the contraction of the Kronecker delta with another tensor, here a_i:

$$a_i\,\delta_{ij} \;=\; a_j \tag{2.23}$$

Other example: $\Gamma_{jmk}\,\delta_{nk} \;=\; \Gamma_{jmn}$.

2.6.3 Kronecker delta sum

If i takes the values from 1 to n, then the following rule applies:

$$\delta_{ii} \;=\; \underbrace{\delta_{11} \;+\; \delta_{22} \;+\; \ldots}_{n} \;=\; n \tag{2.24}$$

In the **four-dimensional spacetime**: $\delta_{ii} \;=\; 4$.

2.6.4 Scalar product in index notation

We can easily illustrate how useful the Kronecker delta is in theoretical physics using the scalar product. Let's look at any three-dimensional vector:

$$\boldsymbol{a} = [a_1, a_2, a_3] \tag{2.25}$$

$$= a_1\hat{\boldsymbol{e}}_1 + a_2\hat{\boldsymbol{e}}_2 + a_3\hat{\boldsymbol{e}}_3 \tag{2.26}$$

$$= a_i\hat{\boldsymbol{e}}_i \tag{2.27}$$

Here, $\hat{\boldsymbol{e}}_1$, $\hat{\boldsymbol{e}}_2$ and $\hat{\boldsymbol{e}}_3$ are three **basis vectors** that are normalized and orthogonal to each other. In this case, they span an orthogonal three-dimensional coordinate system. For the third equal sign, we have used the Einstein summation convention and represented the vector $\boldsymbol{a}$ in index notation: $a_i\hat{\boldsymbol{e}}_i$.

Let's now take another vector $\boldsymbol{b}$ and also represent it in index notation: $b_j\hat{\boldsymbol{e}}_j$. Note that we have to name the indices of the two vectors differently.

Now we form the scalar product of the two vectors:

$$\boldsymbol{a}\cdot\boldsymbol{b} = (a_i\hat{\boldsymbol{e}}_i)\cdot(b_j\hat{\boldsymbol{e}}_j) \tag{2.28}$$

We can sort the objects in index notation in Eq. 2.28 as we like. Let's take advantage of this and put brackets around the basis vectors to emphasize their importance when introducing the Kronecker delta:

$$\boldsymbol{a}\cdot\boldsymbol{b} = a_ib_j\,(\hat{\boldsymbol{e}}_i\cdot\hat{\boldsymbol{e}}_j) \tag{2.29}$$

Thus we have converted the scalar product $\boldsymbol{a}\cdot\boldsymbol{b}$ of the two vectors to the scalar product of the basis vectors $\hat{\boldsymbol{e}}_i\cdot\hat{\boldsymbol{e}}_j$. The basis vectors are orthonormal (i.e. pairwise orthogonal and normalized). Let's remember what the property of being orthonormal means for two vectors:

- Scalar product $\hat{\boldsymbol{e}}_i\cdot\hat{\boldsymbol{e}}_j$ equals 1 if i and j **equal**. In this case, it is the same vector.
- Scalar product $\hat{\boldsymbol{e}}_i\cdot\hat{\boldsymbol{e}}_j$ results in 0 if i and j **are not equal**. In this case,

there are two different basis vectors and they are orthogonal to each other.

The two behaviors can be written compactly in mathematical terms as follows:

$$\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j \;=\; \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{2.30}$$

Doesn't this property sound familiar to you? The scalar product 2.30 of two orthonormalized vectors behaves exactly like the definition 2.18 of Kronecker delta! Therefore, you may replace the scalar product between two basis vectors with the Kronecker delta:

$$\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j \;=\; \delta_{ij} \tag{2.31}$$

This allows us to calculate the **scalar-product** 2.29 **by using the Kronecker delta**:

$$\boldsymbol{a} \cdot \boldsymbol{b} \;=\; a_i \, b_j \, \delta_{ij} \tag{2.32}$$

If you remember the rules for the contraction, we can contract one of the summation indices i or j in 2.32. For example, let's contract (eliminate) the j. We get the **scalar product in index notation**:

$$\boldsymbol{a} \cdot \boldsymbol{b} \;=\; a_i \, b_i \tag{2.33}$$

And eq. 2.33 is exactly the definition of the scalar product, where the vector components are summed component by component.

Now you know how the scalar product is written in index notation and what role the Kronecker delta plays. It represents the scalar product of the basis vectors: $\hat{\boldsymbol{e}}_i \cdot \hat{\boldsymbol{e}}_j \;=\; \delta_{ij}$.

■ **Example 2.1 — Kronecker delta in quantum mechanics.** The spin-up state $|1\rangle$ and the spin-down state $|2\rangle$ are orthonormal to each other. The word "orthonormal" should trigger a thought in your mind: Kronecker delta can be

used here! Why again? Orthonormal vectors result in either 1 or 0, just like the Kronecker delta.

The scalar product $|i\rangle \cdot |j\rangle$ in quantum mechanics is represented in Bra-Ket notation $\langle i\,|j\rangle$ (we will learn this notation in the chapter 16):

$$|i\rangle \cdot |j\rangle \;=\; \langle i\,|j\rangle \;=\; \delta_{ij} \tag{2.34}$$

Here, i and j take the values 1 (spin up) or 2 (spin down). ∎

You can keep the following in mind: As soon as you discover an expression in the index notation of an equation that results in either 0 or 1 depending on the indices, replace it with Kronecker-Delta and use the Kronecker-Delta rules above to simplify the equations further or to represent them in index notation.

2.7 Levi Civita symbol

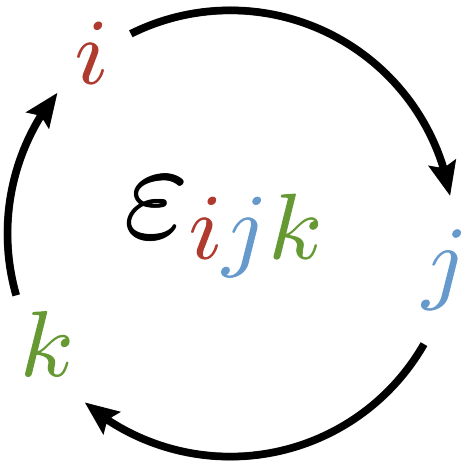
More: en.fufaev.org/levi-civita-symbol

In addition to the Kronecker delta δ_{ij}, the Levi-Civita symbol ε_{ijk} is a very common symbol in theoretical physics that is used in all areas of physics, from classical mechanics to quantum field theory.

With the Levi-Civita symbol, which is sometimes also called the epsilon tensor, you can easily transform and simplify complicated vector equations, such as multiple cross products, and represent equations more compactly.

Levi-Civita symbol ε_{ijk} is notated with a small Greek epsilon that has three indices i, j and k. The Levi-Civita symbol can take on three different values: +1, 0 or -1. When does it take on which value? That depends on how the indices ijk are arranged in relation to the original order. What do I mean exactly? Let's take a closer look. You can permute (swap) the indices ijk. We can permute the indices in two ways.

In the **straight (cyclic) permutation**, all indices ijk are rotated clockwise or anti-clockwise. With this permutation, all indices change their position.



For example:

- An even permutation of ijk in a clockwise direction results in kij. Can you see how the indices have been rotated here?
- An even permutation of kij in a clockwise direction results in jki.
- An even permutation of jki would again result in ijk. Remember that an anticlockwise rotation of the indices is also an even permutation.

In an **odd permutation**, two indices are swapped with each other. In this permutation, only two of the three indices ijk change position. For example:

- An odd permutation of ijk is jik. Here, i and j have been swapped, while k has remained in the same place.
- Another odd permutation of ijk is kji. Here, i and k have been swapped, while j has remained in the same place.
- And the last possible odd permutation of ijk is ikj. Here, i has been left as it is, while j and k have been swapped.

With this knowledge, you will be able to understand the definition of the Levi-Civita symbol. The permutations refer to a starting position of the indices. Here we assume $(ijk) = (123)$ as the starting position. Then **the Levi-Civita**

symbol behaves as follows:

$$\varepsilon_{ijk} = \begin{cases} 1 & (\text{ijk}) \text{ is even permutation of } (123) \\ -1 & (\text{ijk}) \text{ is odd permutation of } (123) \\ 0 & \text{at least two indices are equal} \end{cases} \tag{2.35}$$

Here are a few examples:

- $\varepsilon_{112} = 0$, since the first two indices are equal.
- $\varepsilon_{313} = 0$, since the first and third indices are the same.
- $\varepsilon_{222} = 0$, since all three indices are equal.
- $\varepsilon_{123} + \varepsilon_{213} = 1 + (-1) = 0$, since the indices of the first epsilon are in the start position and the indices of the second epsilon are an odd permutation of this.
- $\varepsilon_{123}\,\varepsilon_{231} = 1 \cdot 1 = 1$, since the indices of the first epsilon are in the starting position and the indices of the second epsilon have just been permuted counterclockwise.

2.7.1 Cross product in index notation

The enormous benefit of the Levi-Civita symbol can be seen by looking at the **double cross product** $\boldsymbol{a} \times \boldsymbol{b} \times \boldsymbol{c}$ or the **parallelepipedial product** $(\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c}$. But even for the simple cross product $\boldsymbol{a} \times \boldsymbol{b}$ of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, we need the Levi-Civita symbol to be able to represent the cross product compactly in index notation.

The cross product, written using the orthonormal basis vectors $\hat{\boldsymbol{e}}_1$, $\hat{\boldsymbol{e}}_2$ and $\hat{\boldsymbol{e}}_3$

looks like this:

$$\begin{aligned} \boldsymbol{a} \times \boldsymbol{b} &= \begin{bmatrix} a_2\, b_3 & - & a_3\, b_2 \\ a_3\, b_1 & - & a_1\, b_3 \\ a_1\, b_2 & - & a_2\, b_1 \end{bmatrix} \\ &= (a_2\, b_3 - a_3\, b_2)\, \hat{\boldsymbol{e}}_1 \\ &+ (a_3\, b_1 - a_1\, b_3)\, \hat{\boldsymbol{e}}_2 \\ &+ (a_1\, b_2 - a_2\, b_1)\, \hat{\boldsymbol{e}}_3 \end{aligned} \tag{2.36}$$

The i**-th component** $(\boldsymbol{a} \times \boldsymbol{b})_i$ **of the cross product** $\boldsymbol{a} \times \boldsymbol{b}$, represented in the orthonormal basis, we can write compactly in index notation as follows:

$$(\boldsymbol{a} \times \boldsymbol{b})_i = \varepsilon_{ijk}\, \hat{\boldsymbol{e}}_i\, a_j\, b_k \tag{2.37}$$

Take a look at the indices in eq. 2.37. All three indices i, j and k occur **double**. Here we have used the Einstein summation convention, therefore we **sum over duplicate indices**. If we insert concrete values for the indices in 2.37, we get exactly the first ($i = 1$), second ($i = 2$) or third ($i = 3$) component of the cross product. But eq. 2.37 is not only a compact notation of the cross product, it is also a clever notation for the cross product, with which we can easily derive relations for the parallelepipedial product and the double cross product.

For fun, write out the double cross product $(\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c}$ with vector notation and then write it out in index notation. And prove the following **BAC-CAB rule** with one and the other method:

$$(\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c} = \boldsymbol{b}\,(\boldsymbol{a} \cdot \boldsymbol{c}) + \boldsymbol{c}\,(\boldsymbol{a} \cdot \boldsymbol{b}) \tag{2.38}$$

You will be grateful to have learned about the Levi Civita symbol, as you will encounter it regularly during your undergraduate and master's studies.

3. Dirac delta

More: en.fufaev.org/dirac-delta

The Dirac delta $\delta(x)$ (sometimes also called the Dirac delta function, although it is not a function) is a useful mathematical object that is used in many areas of theoretical physics. Starting in electrodynamics in the **description of electric point charges** as a charge density concentrated in a single point, up to quantum field theory in the **description of quantum fields** as operators.

Let us consider a one-dimensional electric charge density $\rho(x)$ that depends on the position x. The charge density is therefore smeared on a line. $\rho(x)$ can also represent a mass density or any other **density function**. Here we look at the charge density as an example.

To calculate how large the **total charge** Q is on this line, we must integrate (sum up) the charge density $\rho(x)$ on this line. Let's assume that the charge density is smeared on the line from $x = a$ to $x = b$. These are our integration limits. The total charge is therefore calculated as follows:

$$\int_a^b \rho(x)\,\mathrm{d}x \;=\; Q \tag{3.1}$$

But what if Q is not a smeared charge, but a charge localized at a single point?

What if Q is a singularity? The entire charge density $\rho(x)$ is then concentrated in a single point and zero everywhere else. And this is where the problem arises: We cannot mathematically use the integral 3.1 for singularities. But we must somehow be able to mathematically describe a point charge.

The charge density $\rho(x)$ must fulfill two properties if it is to describe a single point charge:

- Charge density $\rho(x)$ must disappear at every location x, except at the location where the point charge is located. Let us assume that the charge is located at the coordinate origin $x = 0$, that is: $\rho(x) = 0$ for $x \neq 0$.
- The integral 3.1 over the charge density must give us the value Q if the point charge lies within the integration limits $x = a$ and $x = b$.

If we normalize the **charge to the value** $Q = 1$ and observe the **both properties of the charge density of a point charge**, then we note the density with a Greek delta $\delta(x)$ and call it the **Dirac delta**. The Dirac delta therefore describes a density and has the following properties, which we have chosen so that we can use it to describe a point charge ($Q = 1$):

- The Dirac delta is zero everywhere except at the origin:

$$\delta(x) = 0, \quad x \neq 0 \tag{3.2}$$

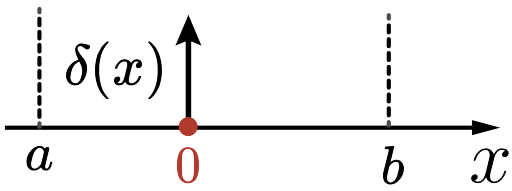
- If the integration of the Dirac delta includes the coordinate origin $x = 0$, then the integral has the value 1:

$$\int_a^b \delta(x)\,\mathrm{d}x \;=\; 1, \quad a < 0 < b \tag{3.3}$$

Even though the name »delta function« may suggest that it is a function, $\delta(x)$ is mathematically not a function, but another mathematical object that can be understood as a Dirac distribution or as a Dirac measure. Let us therefore continue to call $\delta(x)$ a Dirac delta so as not to upset the mathematicians.

The Dirac delta is graphically illustrated with an arrow that is located at the

position $x = 0$ of the unit point charge $Q = 1$. The height of the arrow is usually chosen so that it represents the value of the integral, in this case $Q = 1$.



3.1 Dirac delta in the coordinate origin

Let us now consider an integral of the delta function together with any function $f(x)$:

$$\int_a^b f(x)\,\delta(x)\,\mathrm{d}x \;= ? \tag{3.4}$$

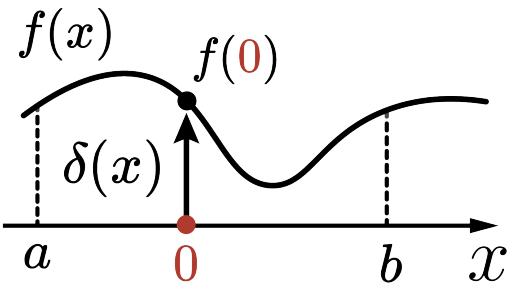
Such an integral is very easy to calculate, because due to the property 3.3 the Dirac delta is zero everywhere except at the point $x = 0$. This means that the product $f(x)\delta(x)$ is also zero everywhere except at the point $x = 0$. In the integral 3.4, only the function value $f(0)$ remains. Since $f(0)$ no longer depends on x, we can move this constant in front of the integral:

$$\begin{aligned}\int_a^b f(x)\delta(x)\,\mathrm{d}x \;&= \int_a^b f(0)\,\delta(x)\,\mathrm{d}x \\ &= f(0)\int_a^b \delta(x)\,\mathrm{d}x \\ &= ?\end{aligned} \tag{3.5}$$

The integral over the Dirac delta results in 1 if $x = 0$ lies between a and b (otherwise the integral is zero). This is exactly the property of the Dirac delta. So we know what the Dirac delta does in the integral 3.4 when multiplied by

a function $f(x)$. The **Dirac delta picks the value of the function at the origin** $x = 0$:

$$\int_a^b f(x)\,\delta(x)\,\mathrm{d}x \;=\; f(0) \tag{3.6}$$

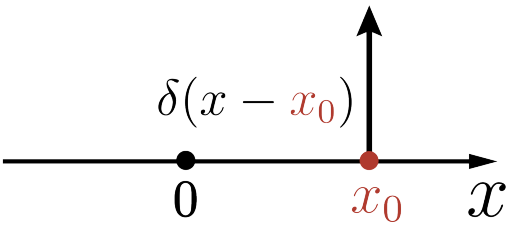


3.2 Shifted Dirac delta

Of course, we can also move the charge $Q = 1$ to another position on the x-axis, for example to the position $x = x_0$. To indicate the charge shifted outside the coordinate origin, we write the argument of the Dirac delta as $\delta(x - x_0)$. Why not $\delta(x + x_0)$? Because we have shifted the Dirac delta in the positive direction. Then the delta function must be zero everywhere except at the point x_0.

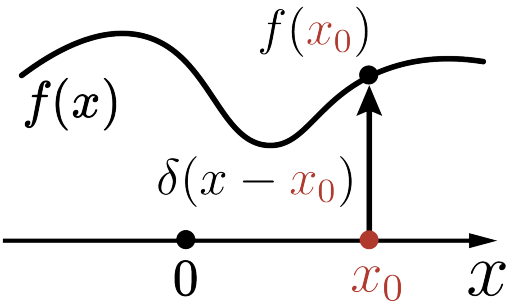
Even with a shifted charge, the integral over the delta function is equal to 1 if the charge at $x = x_0$ lies between the integration limits. We have only shifted the Dirac delta to x_0, so the value of the integral with $\delta(x - x_0)$ is the same as in the case of $\delta(x)$:

$$\int_a^b \delta(x - x_0)\,\mathrm{d}x \;=\; 1, \quad a < x_0 < b \tag{3.7}$$



What happens if the shifted Dirac delta is multiplied by another function $f(x)$ in the integral? $\delta(x-x_0)$ is zero everywhere except at the point x_0. This means: **Shifted Dirac delta $\delta(x - x_0)$ in the integral picks the function value $f(x_0)$ at the point where the Dirac delta is located**:

$$\int_a^b f(x)\,\delta(x - x_0)\,\mathrm{d}x \;=\; f(x_0), \quad a < x_0 < b \tag{3.8}$$



3.3 Properties of the Dirac delta

The Dirac delta has two important properties that we will need in theoretical physics when dealing with equations:

- The Dirac delta is **symmetric**:

$$\delta(-x) \;=\; \delta(x) \tag{3.9}$$

- The **factor** k in the argument of Dirac delta **can be pulled out**:

$$\delta(k\,x) \;=\; \frac{1}{k}\,\delta(x) \tag{3.10}$$

3.4 Analogy to the Kronecker delta

The defining properties 3.2 and 3.3 of the Dirac delta $\delta(x - x_0)$ are somewhat reminiscent of the definition of Kronecker delta δ_{ij}, if we call the letters the same: $x := i$ and $x_0 := j$. The Dirac delta then looks like this: $\delta(i - j)$.

Recall what the Kronecker delta (with Einstein summation convention) does in a sum $f_i\delta_{ij}$ with a vector component f_i. It selects the j-th vector component of the vector $\boldsymbol{f}$ (we have named the vector $\boldsymbol{f}$ to make the analogy clearer):

$$f_i\,\delta_{ij} \;=\; f_j \tag{3.11}$$

And now compare that with what the Dirac delta does in the integral:

$$\int_a^b f(i)\,\delta(i - j) \;=\; f(j) \tag{3.12}$$

While we can use the Kronecker delta δ_{ij} to pick a vector component f_j from a finite number of vector components f_i, we can use the Dirac delta $\delta(i - j)$ to pick a function value $f(j)$ from an infinite number of function values $f(i)$.

- The Kronecker delta δ_{ij} is used when we are dealing with vectors $\boldsymbol{f}$ and their finite number of vector components f_i.
- The delta function $\delta(i - j)$ is used when we are dealing with functions f and their infinite number of function values $f(i)$.

3.5 Three-dimensional Dirac delta

So far, we have only considered a one-dimensional Dirac delta $\delta(x)$ that can be moved back and forth along the x-axis. The charges or other density singularities such as black holes are usually located in a three-dimensional space with three

spatial axes: (x, y, z). Fortunately, the generalization of the Dirac delta to three-dimensional space is quite simple.

$$\delta(x)\,\delta(y)\,\delta(z) \;=\; 0, \quad (x, y, z) \neq (0, 0, 0) \tag{3.13}$$

If our unit charge $Q = 1$ is located in the coordinate origin $(x, y, z) = (0, 0, 0)$, then we can describe the corresponding density, that is, the three-dimensional Dirac delta, with the product of three one-dimensional Dirac deltas $\delta(x)$, $\delta(y)$ and $\delta(z)$:

$$\int_V \delta(x)\,\delta(y)\,\delta(z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \;=\; 1, \quad (0, 0, 0) \in V \tag{3.14}$$

To avoid having to write three Dirac deltas, we combine them into one Dirac delta with a superscript that specifies the spatial dimension. And in the argument of the Dirac delta, we write the position vector $\boldsymbol{r} = (x, y, z)$:

$$\delta^3(\boldsymbol{r}) \;:=\; \delta(x)\delta(y)\delta(z) \tag{3.15}$$

The Dirac delta shifted to the location $\boldsymbol{r}_0 = (x_0, y_0, z_0)$ then looks as follows:

$$\delta^3(\boldsymbol{r} - \boldsymbol{r}_0) \;:=\; \delta(x - x_0)\,\delta(y - y_0)\,\delta(z - z_0) \tag{3.16}$$

If the three-dimensional delta function appears in the integral in a product with a scalar three-dimensional function $f(\boldsymbol{r}) = f(x, y, z)$, then the three-dimensional Dirac delta $\delta^3(\boldsymbol{r} - \boldsymbol{r}_0)$ works in the same way as in the one-dimensional case. The **Dirac delta picks the value $f(\boldsymbol{r}_0) = f(x_0, y_0, z_0)$ of the function at the point $\boldsymbol{r}_0$**:

$$\int_V f(\boldsymbol{r})\,\delta^3(\boldsymbol{r} - \boldsymbol{r}_0)\,\mathrm{d}v \;=\; f(\boldsymbol{r}_0) \tag{3.17}$$

With the knowledge of the Dirac delta, we can theoretically describe density singularities (for example point charges and black holes).

4. Vector fields

A vector function $\boldsymbol{F}$ (or vector-valued function) is a vector that depends on the (Cartesian) coordinates (x, y, z) and has three components in three-dimensional space:

$$\boldsymbol{F}(x, y, z) = \begin{bmatrix} F_x(x, y, z) \\ F_y(x, y, z) \\ F_z(x, y, z) \end{bmatrix} \tag{4.1}$$

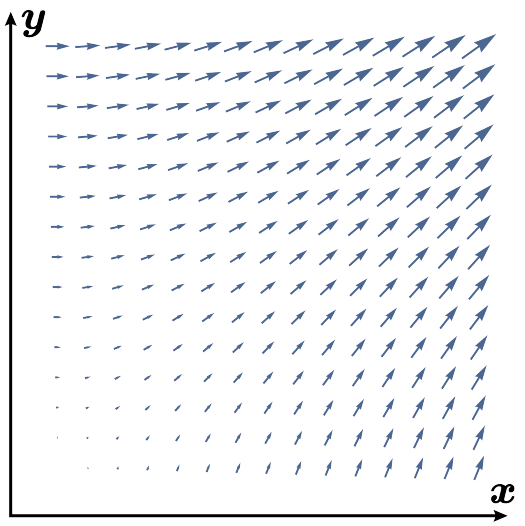
Here, $F_x(x, y, z)$, $F_y(x, y, z)$ and $F_z(x, y, z)$ are three scalar functions and they represent the three components of the vector function $\boldsymbol{F}$. Sometimes we also write briefly: $\boldsymbol{F}(\boldsymbol{r}) = \boldsymbol{F}(x, y, z)$, where $\boldsymbol{r} = (x, y, z)$ is the position vector.

In theoretical physics, we will mainly work with vector fields. A vector function can depend on any coordinates (x, y, z), such as angles. And, if we are talking about a vector field, then (x, y, z) represents the spatial coordinates. We represent vector functions and vector fields either with an arrow $\vec{F}$ above the symbol or more compactly, in bold $\boldsymbol{F}$. The vector field could, for example, represent the electric field $\boldsymbol{F} = \boldsymbol{E}$ or a magnetic field $\boldsymbol{F} = \boldsymbol{B}$.

■ **Example 4.1** A two-dimensional vector field could look like this:

$$\boldsymbol{F}(x, y, z) = \begin{bmatrix} 2x + 5y \\ 5x \end{bmatrix} \tag{4.2}$$

Here, $F_x(x, y) = 2x + 5y$ is the first component and $F_y(x) = 5x$ is the second component of the vector field. The second component depends only on the position coordinate x. If we represent 4.2 graphically, the vector field looks like this:



Each point (x, y) is assigned a vector $\boldsymbol{F}(x, y)$. For example, at the point $(x, y) = (1, 1)$ the vector looks like this: $\boldsymbol{F}(1,\ 1) = (7, 5)$. Simply insert $x = 1$ and for $y = 1$ into the vector field 4.2 to get this example vector. If you insert a large number of locations in this way, you will get the graphical representation of the vector field 4.2. ■

5. Nabla operator

More: en.fufaev.org/nabla-operator

We will encounter the **nabla operator** ∇ (inverted large delta) in every branch of theoretical physics when it comes to multidimensional derivatives, especially in electrodynamics when we get to know Maxwell's equations. The **three-dimensional Nabla operator** is notationally similar to a vector and looks like this in three-dimensional space when we express it **with Cartesian coordinates** (x, y, z):

$$\nabla = \begin{bmatrix} \partial_x \\ \partial_y \\ \partial_z \end{bmatrix} \tag{5.1}$$

The three components of the Nabla operator are **partial derivatives with respect to** x, y and z. We have notated the partial derivatives more compactly with ∂_x instead of $\frac{\partial}{\partial x}$. This notation is common in theoretical physics. The single derivatives ∂_x, ∂_y and ∂_z are called differential operators. You can apply a differential operator to a function f. The result is the derivative of the function. For example: $\partial_x f = \frac{\partial f}{\partial x}$.

We can apply the Nabla operator in three different ways to a scalar function f

or to a vector field $\boldsymbol{F}$:

- As **multiplication** with a scalar function: ∇f. The result ∇f is called **gradient** of the scalar function f.
- As **scalar product** with a vector field: $\nabla \cdot \boldsymbol{F}$. The result $\nabla \cdot \boldsymbol{F}$ is called **divergence** of the vector field $\boldsymbol{F}$.
- As **cross product** with a vector field: $\nabla \times \boldsymbol{F}$. The result $\nabla \times \boldsymbol{F}$ is called **curl** of the vector field $\boldsymbol{F}$.

5.1 Gradient

More: en.fufaev.org/gradient

Let's take a look at the first application of the nabla operator in the form of the **gradient ∇f of a scalar function** f. Here we apply the nabla operator ∇ to the function f. We will encounter the gradient in Maxwell's equations, for example:

$$\nabla f(x,y,z) \;=\; \begin{bmatrix} \partial_x f(x,y,z) \\ \partial_y f(x,y,z) \\ \partial_z f(x,y,z) \end{bmatrix} \tag{5.2}$$

The result 5.2 is called **gradient** and represents a **three-dimensional vector field** ∇f with three components:

- The first component contains the gradient $\partial_x f$ of the function $f(x,y,z)$ in the x direction.
- The second component contains the gradient $\partial_y f$ of the function $f(x,y,z)$ in the y direction.
- The third component contains the gradient $\partial_z f$ of the function $f(x,y,z)$ in the z direction.

Of course, we can also use a **two-dimensional Nabla operator** $\nabla_{2\mathrm{d}}$, which only has two components. A two-dimensional gradient of a function $f(x,y)$ then

looks like this:

$$\nabla_{2\mathrm{d}} f(x,y,z) \;=\; \begin{bmatrix} \partial_x f(x,y,z) \\ \partial_y f(x,y,z) \end{bmatrix} \tag{5.3}$$

And the **one-dimensional Nabla operator** $\nabla_{1\mathrm{d}}$ has only one component. Applied to a one-dimensional function $f(x)$, the gradient is simply the **partial derivative of the function**:

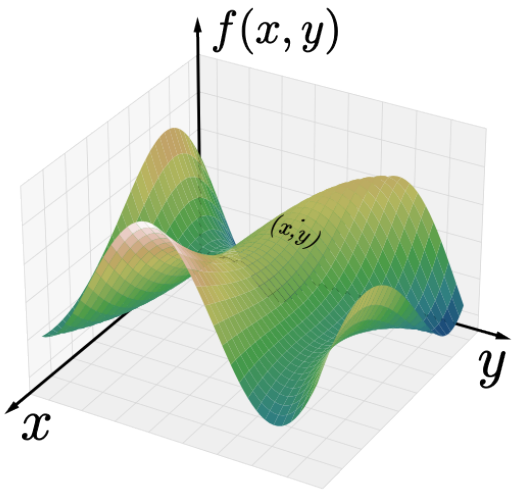
$$\nabla_{1\mathrm{d}} f(x,y,z) \;=\; \partial_x f(x,y,z) \tag{5.4}$$

■ **Example 5.1 — Calculating the gradient of a function.** Given is a scalar function $f(x,y,z) = x^2 + 5xy + z$. The gradient of this scalar function is:

$$\nabla f(x,y,z) \;=\; \begin{bmatrix} 2x + 5y \\ 5x \\ 1 \end{bmatrix} \tag{5.5}$$

■

The resulting vector $\nabla f(x,y,z)$ points at each location (x,y,z) **to the steepest slope of the function** $f(x,y,z)$. For example, look at the following plot of a two-dimensional scalar function $f(x,y)$:



At each location on the green function, imagine a vector that shows you the direction of the steepest ascent or descent at that location.

5.2 Divergence

More: en.fufaev.org/divergence

Let's look at the second application of the Nabla operator, namely the **divergence $\nabla \cdot \boldsymbol{F}$ of a vector field $\boldsymbol{F}$**. Here we apply the nabla operator ∇ to the vector-valued function $\boldsymbol{F}(x, y, z)$. Just like the gradient ∇f, we will also encounter the divergence in Maxwell's equations, for example.

For the divergence, we form the scalar product $\nabla \cdot \boldsymbol{F}$ between the nabla operator and the vector field $\boldsymbol{F}$:

$$\begin{aligned} \nabla \cdot \boldsymbol{F}(x,y,z) &= \begin{bmatrix} \partial_x \\ \partial_y \\ \partial_z \end{bmatrix} \cdot \begin{bmatrix} F_x(x,y,z) \\ F_y(x,y,z) \\ F_z(x,y,z) \end{bmatrix} \\ &= \partial_x F_x + \partial_y F_y + \partial_z F_z \end{aligned} \tag{5.6}$$

In the last step, we omitted the arguments for more compactness. The result $\nabla \cdot \boldsymbol{F}$ of the scalar product is a three-dimensional scalar function. By forming the gradient, a vector field was generated from a scalar function. And by formig the divergence, we make **a scalar function was generated from a vector field**. So the other way around!

▪ **Example 5.2 — Calculate the divergence of a vector field.** The following three-dimensional vector field is given:

$$\boldsymbol{F}(x,y,z) = \begin{bmatrix} 2x^3 \\ zy \\ 5xy \end{bmatrix} \tag{5.7}$$

The divergence of this vector field is the following scalar function:

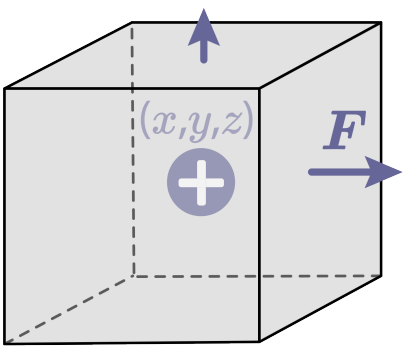
$$\nabla \cdot \boldsymbol{F}(x,y,z) = \underbrace{\partial_x(2x^3)}_{6x^2} + \underbrace{\partial_y(zy)}_{z} + \underbrace{\partial_z(5xy)}_{0} \tag{5.8}$$

■

So if a specific location (x, y, z) is inserted into the scalar function $\nabla \cdot \boldsymbol{F}(x, y, z)$, then this function spits out a number. This number is a measure of the divergence of the vector field at the considered location (x, y, z). The result can be a positive or negative number or even zero. Depending on whether the number is positive, negative or zero, it has a different physical meaning.

5.2.1 Positive divergence = source

We assume that we have inserted a specific location, for example something like $(x, y, z) = (1, 0, 3)$, into the result $\nabla \cdot \boldsymbol{F}(x, y, z)$ and obtained a positive number: $\nabla \cdot \boldsymbol{F}(x, y, z) > 0$. Then the location (x, y, z) is a **source of the vector field** $\boldsymbol{F}$.



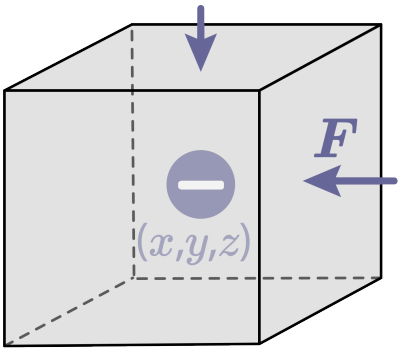
Why do we call the location a source? If we were to enclose this location point in an imaginary cube, then the vector field would mainly point out of the cube.

You can visualize the source as a hole from which the water comes and leaves the surface of the cube. Even though we can use the divergence to describe a water source, in this book we use the divergence to describe electric charges. In

this case, the vector field $\boldsymbol{F}$ corresponds to the electric field $\boldsymbol{F} = \boldsymbol{E}$. Then the source at the location (x, y, z) represents **a positive electric charge**.

5.2.2 Negative divergence = sink

If, on the other hand, we obtain a negative number after inserting the location (x, y, z) into $\nabla \cdot \boldsymbol{F}(x, y, z)$: $\nabla \cdot \boldsymbol{F}\,(x, y, z) < 0$, then we are talking about a **sink of the vector field** $\boldsymbol{F}(x, y, z)$.



If we enclose the location with an imaginary cube, then **the vector field flows into the surface**. We can imagine the sink as a hole into which the water flows. To do this, the water must flow into the cube. If we assume that the vector field is an electric field: $\boldsymbol{F} = \boldsymbol{E}$, then the sink at the location (x, y, z) corresponds to a **negative electric charge**.

■ **Example 5.3 — Sink of a vector field.** Let's take a look at the following vector field:

$$\boldsymbol{F}(x, y, z) \;=\; \begin{bmatrix} 2x \\ y \\ 4 \end{bmatrix} \tag{5.9}$$

Let's calculate the divergence of this vector field:

$$\nabla \cdot \boldsymbol{F}(x, y, z) \;=\; \underbrace{\partial_x(-2x)}_{-2} \;+\; \underbrace{\partial_y(y)}_{1} \;+\; \underbrace{\partial_z(4)}_{0} \;=\; -1 \tag{5.10}$$

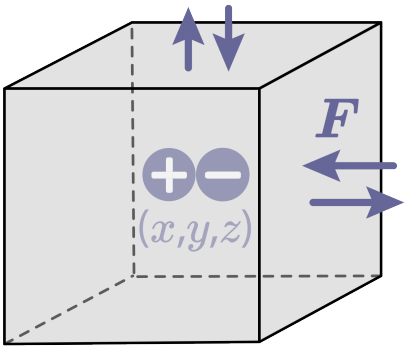
The vector field under consideration has a constant, negative divergence at every location (x, y, z). This means that no matter which location is used for (x, y, z), each location has a negative divergence with the value -1. Sinks of the vector field 5.9 are distributed everywhere. If the vector field were an electric field $\boldsymbol{F} = \boldsymbol{E}$, then this result would mean that a negative electric charge is smeared everywhere in space.

■

5.2.3 Divergence-free vector field

Now assume that we get zero after we have inserted a concrete location (x, y, z) into the divergence field: $\nabla \cdot \boldsymbol{F}(x, y, z) = 0$. Then the location (x, y, z) is **divergence-free**.

If we enclose this location with a cube surface, then the vector field neither flows out nor in. Or just as much of the vector field points into the surface as points out, so that the two opposite contributions cancel each other out and the divergence is net zero.



We can imagine this as if the cube enclosed a source (e.g. water source) and a sink (e.g. drain), so that the amount of water flowing in and water flowing out cancel each other out. If we interpret the vector field as the electric field $\boldsymbol{F} = \boldsymbol{E}$, then there could be an **electric dipole** at the considered divergence-free location. It consists of a positive (source) and negative (sink) charge.

■ **Example 5.4 — Divergence-free vector field.** Let's calculate the divergence

at the location $(x, y, z) = (1, 1, 1)$ of the following vector field:

$$\boldsymbol{F}(x, y, z) = \begin{bmatrix} -2x \\ 0.5y^2 \\ 0.5z^2 \end{bmatrix} \tag{5.11}$$

Let's calculate the divergence of this vector field:

$$\begin{aligned} \nabla \cdot \boldsymbol{F}(x, y, z) &= \underbrace{\partial_x(-2x)}_{-2} + \underbrace{\partial_y(0.5y^2)}_{y} + \underbrace{\partial_z(0.5z^2)}_{z} \\ &= -2 + y + z \end{aligned} \tag{5.12}$$

Insert the location $(1, 1, 1)$ into the calculated scalar function:

$$\nabla \cdot \boldsymbol{F}(1, 1, 1) = -2 + 1 + 1 = 0 \tag{5.13}$$

The divergence of the vector field at this location is zero. There is therefore neither a source nor a sink or an ideal electric dipole at the location $(1, 1, 1)$. ∎

5.3 Curl

As with divergence (scalar product $\nabla \cdot \boldsymbol{F}$), we also apply the nabla operator to a vector field $\boldsymbol{F}$ in the case of **curl (cross product $\nabla \times \boldsymbol{F}$)**:

$$\begin{aligned} \nabla \times \boldsymbol{F}(x, y, z) &= \begin{bmatrix} \partial_x \\ \partial_y \\ \partial_z \end{bmatrix} \times \begin{bmatrix} F_x(x, y, z) \\ F_y(x, y, z) \\ F_z(x, y, z) \end{bmatrix} \\ &= \begin{bmatrix} \partial_y F_z - \partial_z F_y \\ \partial_z F_x - \partial_x F_z \\ \partial_x F_y - \partial_y F_x \end{bmatrix} \end{aligned} \tag{5.14}$$

In the last step, we omitted the arguments for more compactness. The result of the cross product is again a vector field with three components. The curl $\nabla \times \boldsymbol{F}(x, y, z)$ of a vector field thus gives us another vector field.

We can visualize the curl $\nabla \times \boldsymbol{F}(x, y, z)$ of the vector field at the location (x, y, z) actually as the name tells us, as the **circulation of the vector field around the location** (x, y, z).

■ **Example 5.5 — Calculate the curl of a vector field.** Let us again consider the following vector field at the location $(1, 1, 1)$:

$$\boldsymbol{F}(x, y, z) = \begin{bmatrix} 2x^3 \\ zy \\ 5xy \end{bmatrix} \tag{5.15}$$

The vector field $\boldsymbol{F}(1, 1, 1)$ at this location is:

$$\boldsymbol{F}(1, 1, 1) = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} \tag{5.16}$$

The curl of the vector field is:

$$\nabla \times \boldsymbol{F} = \begin{bmatrix} \partial_y(5xy) - \partial_z(zy) \\ \partial_z(2x^3) - \partial_x(5xy) \\ \partial_x(zy) - \partial_y(2x^3) \end{bmatrix} = \begin{bmatrix} 5x - y \\ -5y \\ 0 \end{bmatrix} \tag{5.17}$$

Inserting the location $(1, 1, 1)$ gives the curl vector:

$$\nabla \times \boldsymbol{F}(1, 1, 1) = \begin{bmatrix} 4 \\ -5 \\ 0 \end{bmatrix} \tag{5.18}$$

■

Hopefully you can now imagine what the cross product $\nabla \times \boldsymbol{F}$ with the vector field $\boldsymbol{F}$ means. We will encounter curl in the chapter on Maxwell's equations.

6. Gauss Divergence Theorem

We will encounter the Gauss Divergence Theorem in Maxwell's equations. This theorem states that the sum of the sources and sinks in a volume is equal to the flow through the volume surface. Mathematically speaking:

$$\int_V (\nabla \cdot \boldsymbol{F}) \, \mathrm{d}v \;=\; \oint_A \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{a} \tag{6.1}$$

Please what? You're probably asking yourself. Don't worry. We break down the Divergence Theorem into its components so that you understand it one hundred percent.

6.1 Surface Integral in the Divergence Theorem

Let's first look at the right-hand side of the Divergence Theorem 6.1, namely the surface integral:

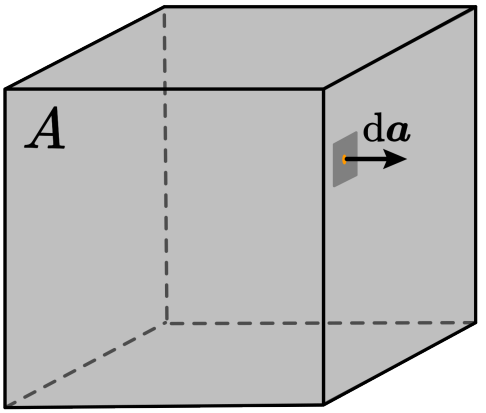
$$\oint_A \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{a}$$

The A stands for a **surface that encloses any volume**, for example the surface of a cube, a sphere or the surface of any three-dimensional shape you can think of. The small circle around the integral is intended to indicate that this surface

must fulfill a condition: It must be **closed**, that is, it must not contain any holes, so that the equality in 6.1 is mathematically fulfilled. The surface A is therefore a closed surface.

The $\boldsymbol{F}$ is any **vector field**: $\boldsymbol{F} = \boldsymbol{F}(x, y, z)$, that is, a vector with three components $F_x(x, y, z)$, $F_y(x, y, z)$ and $F_z(x, y, z)$, as shown in Eq. 4.1. For example, the vector field could represent an **elecetric field** $\boldsymbol{F} = \boldsymbol{E}$ or a **magnetic field** $\boldsymbol{F} = \boldsymbol{B}$.

The $\mathrm{d}\boldsymbol{a}$ is an **infinitesimal surface element**, that is, an infinitely small area of the surface A under consideration. As you may have noticed, the $\mathrm{d}\boldsymbol{a}$ element is shown in bold, so it is a vector with three components $\mathrm{d}a_x$, $\mathrm{d}a_y$ and $\mathrm{d}a_z$. The vector naturally also has a magnitude and a direction. The magnitude $|\mathrm{d}\boldsymbol{a}| = \mathrm{d}a$ indicates the area of this small piece of surface. The $\mathrm{d}\boldsymbol{a}$ vector is orthogonal to the surface area and points **out of the surface**.

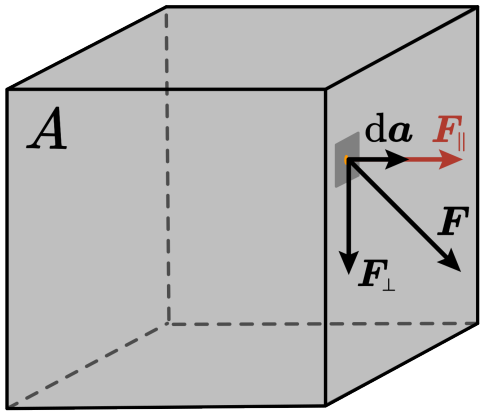


The point $\cdot$ in $\boldsymbol{F} \cdot \mathrm{d}\boldsymbol{a}$ is the **scalar product** (also called dot product). You should be familiar with this vector operation from basic mathematics. The scalar product is a way of multiplying two vectors together. In the Divergence Theorem, the scalar product is therefore formed between the vector field $\boldsymbol{F}$ and the $\mathrm{d}\boldsymbol{a}$ area. Written out, the scalar product looks like this:

$$\boldsymbol{F} \cdot \mathrm{d}\boldsymbol{a} = F_x \mathrm{d}a_x + F_y \mathrm{d}a_y + F_z \mathrm{d}a_z \tag{6.2}$$

The task of this scalar product is to pick out the part of the vector field $\boldsymbol{F}$ at point (x, y, z) that is perpendicular to the surface, i.e. that points parallel to the $\mathrm{d}\boldsymbol{a}$ surface element. How can we understand this? Mathematically, we can split the vector field $\boldsymbol{F} = \boldsymbol{F}_{||} + \boldsymbol{F}_{\perp}$ into two parts:

- In the component $\boldsymbol{F}_{||}$, which points **parallel to the $\mathrm{d}\boldsymbol{a}$ surface element**.
- In the component $\boldsymbol{F}_{\perp}$ that points **perpendicular to the $\mathrm{d}\boldsymbol{a}$ surface element**.



The scalar product $\boldsymbol{F} \cdot \mathrm{d}\boldsymbol{a}$ at the location (x, y, z) on the surface eliminates the perpendicular part of the vector field and leaves only the component of the vector field parallel to the element:

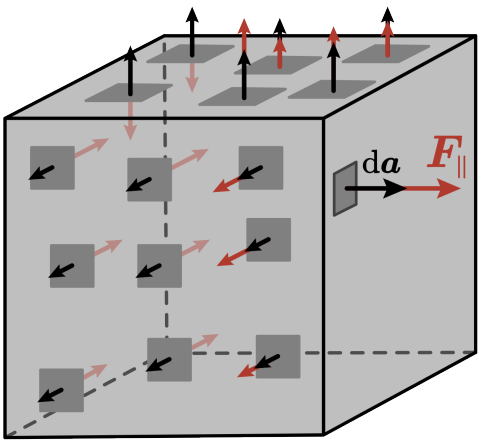
$$\begin{aligned} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{a} &= \left(\boldsymbol{F}_{||} + \boldsymbol{F}_{\perp}\right) \cdot \mathrm{d}\boldsymbol{a} \\ &= \boldsymbol{F}_{||} \cdot \mathrm{d}\boldsymbol{a} + \underbrace{\boldsymbol{F}_{\perp} \cdot \mathrm{d}\boldsymbol{a}}_{0} = \boldsymbol{F}_{||} \cdot \mathrm{d}\boldsymbol{a} \end{aligned} \tag{6.3}$$

Why again is the perpendicular component zero? Because the scalar product of two perpendicular vectors $\boldsymbol{F}$ and $\mathrm{d}\boldsymbol{a}$ is mathematically zero.

The scalar product $\boldsymbol{F} \cdot \mathrm{d}\boldsymbol{a}$ thus ensures that in the Divergence Theorem we only take the component $\boldsymbol{F}_{||}$ of the vector field that leaves or enters the surface

perpendicularly. Everything that »passes by the surface« (by this I mean the component $\boldsymbol{F}_\perp$ parallel to the surface) is omitted in the Divergence Theorem.

Then, on the right-hand side of the Divergence Theorem 6.1, the scalar products $\boldsymbol{F}(x,y,z) \cdot \mathrm{d}\boldsymbol{a}(x,y,z)$ are summed up for each point (x,y,z) on the surface A using the integral in 6.1.



Let us briefly denote the right-hand side of the Divergence Theorem by Φ:

$$\Phi = \oint_A \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{a} \tag{6.4}$$

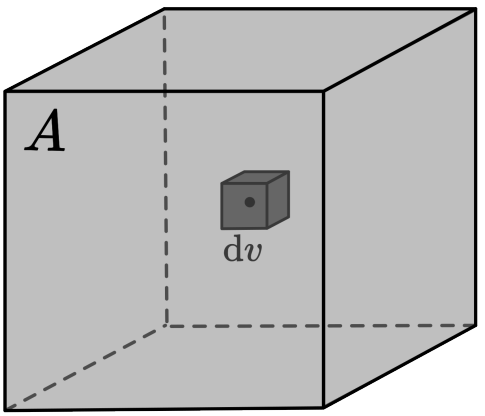
The surface integral therefore results in a number Φ, which is a measure of how much of the vector field $\boldsymbol{F}$ flows in or out of the surface A. The surface integral is the **flux Φ of the vector field $\boldsymbol{F}$ from the surface** A. In the chapter on Maxwell's equations, we will get to know the electric and magnetic flux.

6.2 Volume integral in the Divergence Theorem

Let us now look at the left-hand side of the Divergence Theorem 6.1, namely at the volume integral:

$$\int_V (\nabla \cdot \boldsymbol{F})\, \mathrm{d}v \tag{6.5}$$

The V stands for a **volume**, but not just any volume, it is the volume that is **enclosed by the surface** A. For example, if A is the surface of a cube, then V is the volume of this cube. The $\mathrm{d}v$ is an **infinitesimal volume element**, that is, an infinitely small volume piece of the volume V.



In the integrand $\nabla \cdot \boldsymbol{F}$ of the volume integral, ∇ stands for the **nabla operator**, which we got to know in the chapter 5. Although this operator is not a vector from a mathematical point of view, it looks like a vector. An operator such as the nabla operator is only useful if it is applied to a field. And this also happens in the integrand $\nabla \cdot \boldsymbol{F}$. The nabla operator ∇ **is applied to the vector field** $\boldsymbol{F}$ by forming the scalar product between the nabla operator and the vector field. Written out, this scalar product corresponds to the sum of the derivatives of the vector field with respect to the coordinates x, y and z:

$$\nabla \cdot \boldsymbol{F} \;=\; \partial_x F_x \;+\; \partial_y F_y \;+\; \partial_z F_z \tag{6.6}$$

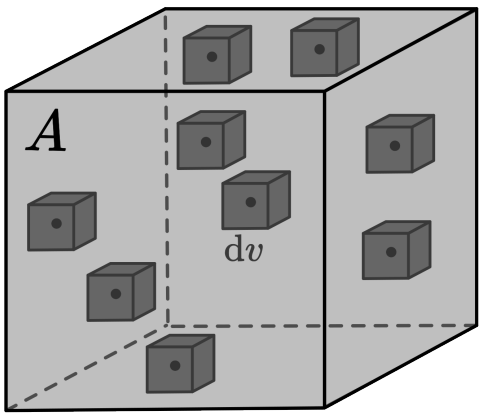
The integrand $\nabla \cdot \boldsymbol{F}$ is therefore the **divergence of the vector field** $\boldsymbol{F}$. We learned what divergence is in the chapter 5.2. The result $\nabla \cdot \boldsymbol{F}$ is no longer a vector, but a scalar that can be either positive, negative or zero:

- If the divergence **at location** (x, y, z) **is positive**: $\nabla \cdot \boldsymbol{F}(x, y, z) > 0$, then there is a **source of the vector field** at the location. In electrodynamics, the source corresponds to a positive charge.

- If the divergence **at location** (x, y, z) **is negative**: $\nabla \cdot \boldsymbol{F}(x, y, z) < 0$, then there is a **sink of the vector field** at the location. In electrodynamics, the source corresponds to a negative charge.

- If the divergence **at location** (x, y, z) **is zero**: $\nabla \cdot \boldsymbol{F}(x, y, z) = 0$, then this location **is neither a sink nor a source** of the vector field. The vector field does not enter or leave the surface **or** the vector field enters as much as it leaves, so that the two parts cancel each other out.

Then, in the volume integral 6.5, the divergences $\nabla \times \boldsymbol{F}$ (all sources and sinks) at each location (x, y, z) within the volume V are summed up with an integral.



The volume integral 6.5 in the Divergence Theorem is a number that measures **how many sinks and sources can be found within the volume** V.

Let us summarize the statement of the Divergence Theorem 6.1:

- The volume integral on the left-hand side describes the **sum of sources and sinks** of the vector field within a volume V:

$$\int_V (\nabla \cdot \boldsymbol{F}) \, \mathrm{d}v$$

- The area integral on the right-hand side describes the **flux** Φ **of the**

vector field through the surface A of this volume V:

$$\oint_A \boldsymbol{F} \cdot \mathrm{d}\,\boldsymbol{a}$$

According to the Divergence Theorem, both integrals are equal.

The Gauss Divergence Theorem therefore states: **The sum of the sources and sinks of a vector field $\boldsymbol{F}$ within a volume V corresponds to the flux Φ through the surface A of this volume**.

7. Stokes' Curl Theorem

More: en.fufaev.org/stokes-curl-theorem

Besides the Divergence Theorem, we will also need the **Stokes' Curl Theorem** (or shorter: Curl Theorem) in order to understand Maxwell's equations in depth. The Curl Theorem states that the **curl of a vector field within a surface is equal to the curl of the vector field along the edge of this surface**. Expressed mathematically, this theorem looks like this:

$$\int_A (\nabla \times \boldsymbol{F}) \cdot \mathrm{d}\boldsymbol{a} \;=\; \oint_L \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l} \tag{7.1}$$

If you have understood the Divergence Theorem, the Curl Theorem should no longer seem totally cryptic to you. You already know the **vector field** $\boldsymbol{F}(x, y, z)$. It depends on three spatial coordinates and has three components as a vector. The **scalar product** $\boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l}$, but also the **nabla operator** ∇ and the **infinitesimal surface** $\mathrm{d}\boldsymbol{a}$ should be familiar to you if you have read the chapter 6 on the Divergence Theorem.

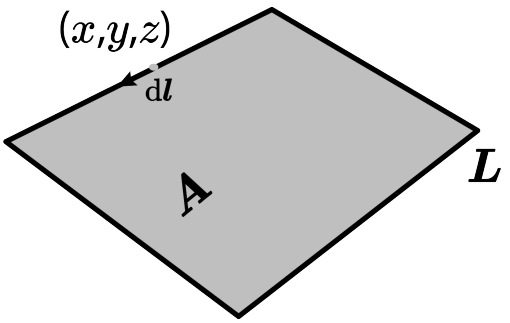
7.1 Line integral in the Curl Theorem

Let us first consider the line integral on the right-hand side of the Curl Theorem 7.1, namely:

$$\oint_L \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l}$$

The symbol L on the integral represents a **line in three-dimensional space**. The circle on the integral symbol indicates that this **line must be closed**, which means that its beginning and end are connected. We refer to such a closed line as a **loop** for short.

The $\mathrm{d}\boldsymbol{l}$ is an **infinitesimal line element** of the line, in other words an infinitely small piece of the line. You should also notice here that the $\mathrm{d}\boldsymbol{l}$ line element is shown in bold, which means it is a vector with three components: $\mathrm{d}l_x$, $\mathrm{d}l_y$ and $\mathrm{d}l_z$. The magnitude $\mathrm{d}l$ of the line element indicates the length of $\mathrm{d}\boldsymbol{l}$, while its direction points **along the line**.



Then, on the right-hand side, the scalar product $\boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l}$ is formed between a vector field and the line element. The scalar product looks like this when written out:

$$\boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l} \;=\; F_x\,\mathrm{d}l_x \;+\; F_y\,\mathrm{d}l_y \;+\; F_z\,\mathrm{d}l_z \tag{7.2}$$

You have already learned what the purpose of this scalar product is in the chapter 6 on the Divergence Theorem. Here is a quick recap: First, divide the

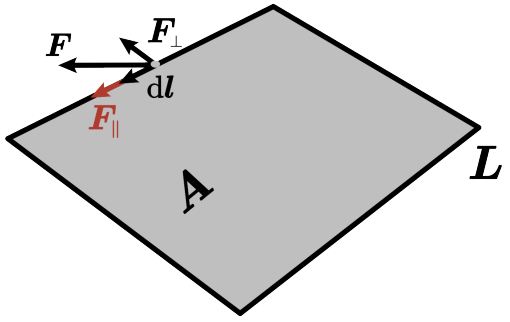
vector field $\boldsymbol{F}$ into two parts:

- Into the component $\boldsymbol{F}_{||}$, which points **parallel to the d$\boldsymbol{l}$ line element**.
- Into the component $\boldsymbol{F}_{\perp}$ that points **perpendicular to the d$\boldsymbol{l}$ line element**.

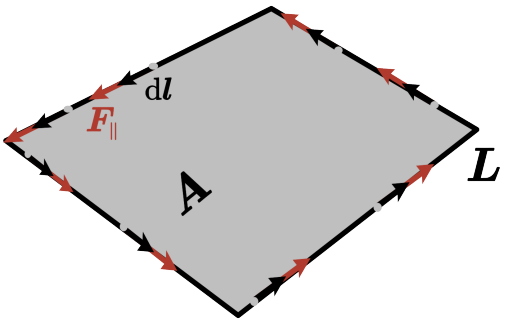
The scalar product $\boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l}$ **eliminates the perpendicular part** of the vector field $\boldsymbol{F}$ and leaves only the part $\boldsymbol{F}_{||}$ of the vector field that is parallel to the d$\boldsymbol{l}$ element. The scalar product of two perpendicular vectors $\boldsymbol{F}_{\perp}$ and d$\boldsymbol{l}$ is mathematically zero:

$$\begin{aligned} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l} &= \left(\boldsymbol{F}_{||} + \boldsymbol{F}_{\perp}\right) \cdot \mathrm{d}\boldsymbol{l} \\ &= \boldsymbol{F}_{||} \cdot \mathrm{d}\boldsymbol{l} + \underbrace{\boldsymbol{F}_{\perp} \cdot \mathrm{d}\boldsymbol{l}}_{0} = \boldsymbol{F}_{||} \cdot \mathrm{d}\boldsymbol{l} \end{aligned} \tag{7.3}$$

Since the line element d$\boldsymbol{l}$ is parallel to the line at every point of the line, in the scalar product $\boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l}$ only the parallel component $\boldsymbol{F}_{||}$ of the vector field remains, which of course also runs along the line L. All other components of the vector field are absent.



The scalar products for each point (x, y, z) on the line L are then added up on the right-hand side of the Curl Theorem using the line integral.



Let us briefly denote the right-hand side of the Curl Theorem by U:

$$U = \oint_L \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l} \tag{7.4}$$

The line integral therefore results in a number U, which is a measure of how much of the vector field runs along the line. Because the line L is closed, the summation returns to the same point (x, y, z) where the summation began. The closed line integral U therefore indicates **how much of the vector field $\boldsymbol{F}$ circulates along the closed line** L.

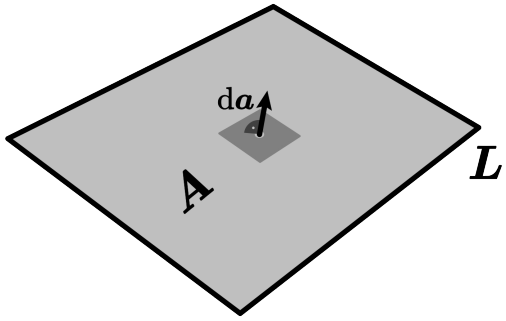
7.2 Surface integral in the Curl Theorem

Let us now consider the surface integral on the left-hand side of the Curl Theorem 7.1, namely:

$$\int_A (\nabla \times \boldsymbol{F}) \cdot \mathrm{d}\boldsymbol{a}$$

The surface A appears in the area integral. In contrast to the surface integral with a circle around the integral sign, as in the Divergence Theorem, here we consider an **open surface**. It therefore does not include a volume. This is merely a surface that is **enclosed by the loop** L.

The vector $\mathrm{d}\boldsymbol{a} = (\mathrm{d}a_x, \mathrm{d}a_y, \mathrm{d}a_z)$ represents an infinitely small element of the surface A and is perpendicular to every location point (x, y, z) on this surface.



The **cross product** $\nabla \times \boldsymbol{F}$ between the nabla operator and the vector field also appears in the surface integral. You should already know what the cross product means from the basics of mathematics. In addition to the scalar product, the cross product is the second way to multiply vectors with each other. The cross product $\nabla \times \boldsymbol{F}$ is the **curl of the vector field** $\boldsymbol{F}$.

In contrast to the scalar product, the result of the cross product $\nabla \cdot \boldsymbol{F}$ is again a **vector field that is perpendicular to** $\boldsymbol{F}$. Why perpendicular? Because that is the property of the cross product! If we write out the cross product in concrete terms, the result vector $\nabla \times \boldsymbol{F}$ looks like this:

$$\nabla \times \boldsymbol{F} = \begin{bmatrix} \partial_y F_z - \partial_z F_y \\ \partial_z F_x - \partial_x F_z \\ \partial_x F_y - \partial_y F_x \end{bmatrix} \tag{7.5}$$

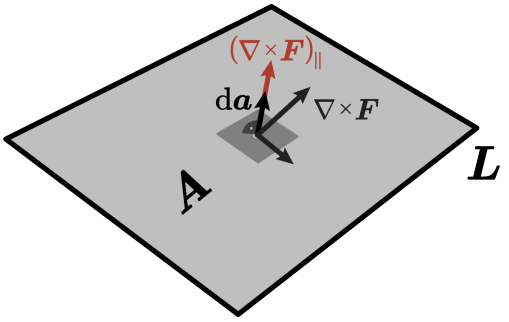
What does curl mean?

The vector $\nabla \times \boldsymbol{F}(x, y, z)$ indicates **how strongly the vector field** $\boldsymbol{F}$ **circulates** at the point (x, y, z), which is located within the area A.

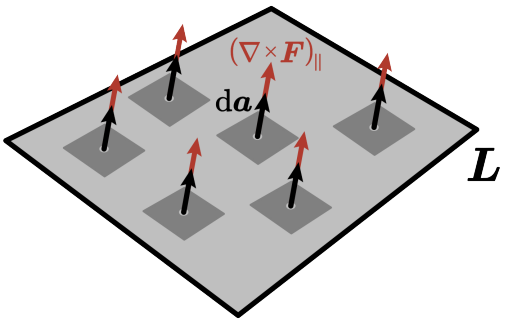
Then the scalar product $(\nabla \times \boldsymbol{F}) \cdot \mathrm{d}\boldsymbol{a}$ between the curl vector field $(\nabla \times \boldsymbol{F})$ and the infinitesimal surface element $\mathrm{d}\boldsymbol{a}$ is formed inside the surface integral of the Curl Theorem. As we already know, the scalar product is only used to pick out the component $(\nabla \times \boldsymbol{F})_{||}$ of the curl vector field that runs parallel to the surface element:

$$
\begin{aligned}
(\nabla \times \boldsymbol{F}) \cdot \mathrm{d}\boldsymbol{a} &= \big((\nabla \times \boldsymbol{F})_{||} + (\nabla \times \boldsymbol{F})_{\perp}\big) \cdot \mathrm{d}\boldsymbol{a} \qquad (7.6)\\
&= (\nabla \times \boldsymbol{F})_{||} \cdot \mathrm{d}\boldsymbol{a} + \underbrace{(\nabla \times \boldsymbol{F})_{\perp} \cdot \mathrm{d}\boldsymbol{a}}_{0}\\
&= (\nabla \times \boldsymbol{F})_{||} \cdot \mathrm{d}\boldsymbol{a}
\end{aligned}
$$

Since the surface element $\mathrm{d}\boldsymbol{a}(x, y, z)$ at a point (x, y, z) is perpendicular to the respective piece of surface, the scalar product **??** only picks out the component of the vector field $\mathrm{d}\boldsymbol{F}$ that is also perpendicular to the surface element or, in other words, the field component that is parallel to the $\mathrm{d}\boldsymbol{a}(x, y, z)$ vector. Therefore, only the component $(\nabla \times \boldsymbol{F})_{||} \cdot \mathrm{d}\boldsymbol{a}$ remains in the surface integral.



The scalar products $(\nabla \times \boldsymbol{F})_{||}(x, y, z) \cdot \mathrm{d}\boldsymbol{a}(x, y, z)$ are then summed up on the left-hand side of the Curl Theorem using the surface integral at each point (x, y, z).



Let us now summarize the statements of the surface integral (right-hand side) and line integral (left-hand side) of the Stokes' Curl Teorem:

- On the left-hand side, the curl $(\nabla \times \boldsymbol{F})$ of the vector field $\boldsymbol{F}$ is summed up at each individual location within the area A:

$$\int_A (\nabla \times \boldsymbol{F}) \cdot \mathrm{d}\boldsymbol{a}$$

- On the right-hand side, the vector field $\boldsymbol{F}$ is summed up along the boundary L of the surface A. The right-hand side therefore corresponds to a number that measures the curl of the vector field on the boundary:

$$\oint_L \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l}$$

Both integrals should be equal according to the Curl Theorem.

The Stokes' Curl Theorem thus clearly states: **The total curl of a vector field $\boldsymbol{F}$ within the surface A corresponds to the curl of the vector field along the edge L of this area**.

8. Fourier Series

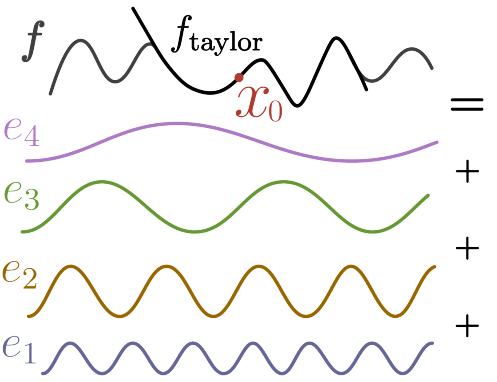
More: en.fufaev.org/fourier-series

You are certainly familiar with the Taylor expansion, with which we can approximate a function $f(x)$ at a point $x = x_0$ using a simpler Taylor series. Let us denote the approximation of the exact function as f. The more terms we take in the Taylor series, the better the approximation f in the vicinity of the selected point x_0.

As you can see in the image below, the Taylor series, represented by f_{taylor}, is a good approximation of the function f in the immediate vicinity of x_0. However, if we move further away from the point, we see that the Taylor series is not a good approximation there. The **Taylor expansion** is therefore a method with which we can approximate **a function only locally**.

However, if it is important for us to approximate a **function f on a whole interval**, then **we need a Fourier series of the function**. As we will see,

the Fourier series is a linear combination of simple periodic **basis functionsn** $(e_1, e_2, e_3, e_4, ...)$ such as cosine and sine or complex exponential functions, which in sum (series) can approximate the function f in a chosen interval. In the following, we assume a **interval of length** L.



8.1 The concept of Fourier series

We can represent a **vector** $\boldsymbol{v}$ that lives in an n-dimensional vector space as a **linear combination of basis vectors** $\{\boldsymbol{e}_k\}$ that span the vector space:

$$\boldsymbol{v} = v_1\boldsymbol{e}_1 + v_2\boldsymbol{e}_2 + v_3\boldsymbol{e}_3 + ... + v_n\boldsymbol{e}_n = \boxed{+}_{n}^{k} v_k\boldsymbol{e}_k \tag{8.1}$$

You should be familiar with the representation of the vector as a linear combination from linear algebra! Using a basis $\{\boldsymbol{e}_k\}$ we can represent every possible vector $\boldsymbol{v}$ in this vector space. Here v_k are the **components of the vector in the chosen basis**. The choice of **basis is not unique**, therefore the components v_k can be different. By choosing a different basis, the vector has different components! You should already know this.

We can also apply this concept of linear combination to **infinite-dimensional** vectors. A **function** f, for example from the illustration above, can be interpreted as an infinite-dimensional vector $\boldsymbol{f}$, which we can represent as a linear combination. The **components** v_k of a finite vector **become Fourier**

coefficients $\hat{f}_k$ if we represent a function and not a finite vector as a linear combination:

$$\begin{aligned} \boldsymbol{f} &= \hat{f}_1\boldsymbol{e}_1 + \hat{f}_2\boldsymbol{e}_2 + \hat{f}_3\boldsymbol{e}_3 + ... + v_n\boldsymbol{e}_n \\ &= \boxed{+}_{n}^{k} \hat{f}_k\boldsymbol{e}_k \end{aligned} \tag{8.2}$$

If we represent a function f as a linear combination 8.4 of basis functions $\boldsymbol{e}_k$, then we denote the sum 8.4 as **Fourier series of the function** f. For a linear combination **for a function**, the basis vectors $\boldsymbol{e}_k$ are more appropriately called **basis functions**. In optics, the basis functions are also called **Fourier modes**.

When considering the function $\boldsymbol{f}$ as a vector, the **function values** $f(x_0)$, $f(x_1)$, $f(x_2)$ and so on until $f(x_n) = f(x_0 + L)$, represent the components of f. We can imagine the **function** f **as a column vector**:

$$\boldsymbol{f} = \begin{bmatrix} f(x_0) \\ f(x_1) \\ f(x_2) \\ . \\ . \\ . \\ f(x_n) \end{bmatrix} \tag{8.3}$$

Of course, the representation is not exact. The argument x of the function $f(x)$ is real and there are therefore **theoretically infinitely many values**, even between x_0 and x_1.

We have omitted all these values in the representation of the function as a column vector 8.3. The column vector is therefore only an approximation vector for the function f. By the way: In this way, as in 8.3, we represent a quantum mechanical wave function as a state vector in computer physics.

8.2 Fourier coefficients

We can determine the Fourier coefficients in the same way as in linear algebra. How do we do this again in linear algebra? To get the k-th component of a finite-dimensional vector $\boldsymbol{v}$, we have to form the **scalar product between the k-th basis vector and the vector $\boldsymbol{v}$**:

$$\begin{aligned} v_k &= \boldsymbol{e}_k \cdot \boldsymbol{v} \\ &= e_{k0}v_0 + e_{k1}v_1 + \ldots + e_{kn}v_n \\ &= \boxed{\overset{n}{+}}_{j} e_{kj}v_j \end{aligned} \tag{8.4}$$

In the last step, we have written out the scalar product a little more compactly with a summation sign and selected the summation index as j. Here, e_{k0} to e_{kn} are the components of the basis vector $\boldsymbol{e}_k = [e_{k0}, e_{k1}, \ldots, e_{kn}]$.

If we are not working with finite-dimensional vectors but with functions, then we have to form the **scalar product between the k-th basis function and the function f** to obtain the k-th Fourier coefficient of f:

$$\hat{f}_k = \boldsymbol{e}_k \cdot \boldsymbol{f} = \langle \boldsymbol{e}_k | \boldsymbol{f} \rangle = \begin{bmatrix} \boldsymbol{e}(x_0) \\ \boldsymbol{e}(x_1) \\ \boldsymbol{e}(x_2) \\ . \\ . \\ . \\ \boldsymbol{e}(x_n) \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{f}(x_0) \\ \boldsymbol{f}(x_1) \\ \boldsymbol{f}(x_2) \\ . \\ . \\ . \\ \boldsymbol{f}(x_n) \end{bmatrix} \tag{8.5}$$

To indicate that we may be working with an infinite-dimensional vector space here, we can call the operation 8.5 not scalar product but inner product and use it as physicist in the **Bra-Ket Notation** $\langle \boldsymbol{e}_k \boldsymbol{f} \rangle$. We learn Bra-Ket Notation in the chapter 16.

Note that we represent the vectors in 8.5 up to the n component because, as

already mentioned, with a Fourier series we can only work with functions in a certain interval. Our chosen interval $(x_0, x_n) = (x_0, x_0 + L)$ has the length L.

Let's write out the inner product 8.5 as a sum:

$$\hat{f}_k = \langle \boldsymbol{e}_k | \boldsymbol{f} \rangle \approx \boxed{+}_{x=x_0}^{x_n} e_k(x) f(x) \tag{8.6}$$

You have certainly seen the approximation sign in Eq. 8.6. The sum is therefore only an approximation of the Fourier coefficient $\hat{f}_k$. Can we represent the Fourier coefficients $\hat{f}_k$ **exact**? It's quite simple! Since we are dealing with a **continuous summation** in the case of exact Fourier coefficients, we must replace the summation sign **with an integral**. So instead of summing discretely over x as in 8.6, we integrate over x:

$$\hat{f}_k = \langle \boldsymbol{e}_k | \boldsymbol{f} \rangle = \int_{x_0}^{x_n} e_k^*(x) f(x) dx \tag{8.7}$$

We have only made a small mathematical upgrade in the integral. The basis function $e_k^*(x)$ has been **complex conjugate**. We can also omit the asterisk if we are working with real basis functions, as $e^* = e$ applies to real basis functions. However, to allow complex basis functions, we must append an asterisk to the basis function. The **asterisk in the case of complex-valued functions is important** so that the integral 8.7 fulfills the **properties of an inner product**.

So, now we know how we can calculate the Fourier coefficients with the integral 8.7 and how the integral formula 8.7 comes about in the first place.

8.3 Fourier basis

Now let's get to know the basis functions. **Which basis functions e_k can we use in the Fourier series of f?**

$$f = \boxed{+}_{k}^{n} \hat{f}_k e_k \tag{8.8}$$

All functions that fulfill the **properties of a basis**! In order for a set of vectors or, as in our case, a set of functions $\{e_k\}$ to be called a basis, these functions must fulfill two conditions:

- If we take two basis functions e_k and e_m from the set $\{e_k\}$, then they must be **orthonormal** to each other, in other words **orthogonal and normalized**. This property can be expressed with the Kronecker delta: $\langle e_k | e_m \rangle = \delta_{km}$.
- The set $\{e_k\}$ of basis functions must be **complete**. In other words, they must span the space in which the functions f live. We must be able to **represent each function f exactly** with the set $\{e_k\}$.

Only when these two properties are fulfilled by the functions $\{e_k\}$ can we take these functions as basis functions and thus represent a function f as a Fourier series 8.8.

A typical basis $\{e_k\}$ used in physics are the **complex exponential functions**:

$$e_k = \frac{1}{\sqrt{L}} \mathrm{e}^{\mathrm{i}kx} \tag{8.9}$$

The factor $\frac{1}{\sqrt{L}}$ ensures that the basis functions are **normalized**, that is, they exactly fulfill the necessary **1.** property. In the context of physics, especially in optics, we refer to k as **wavenumber**. And remember that e in $\mathrm{e}^{\mathrm{i}kx}$ is the **Euler number** and not the label of the basis function e_k! I'm just saying...

Depending on what we use for the wavenumber k, we get a different basis function in 8.9. Of course, we can also choose a different basis for the Fourier series, such as cosine and sine functions. **We are free to choose a basis**. Here we have chosen complex exponential functions as a basis because they can be written in a nice compact way, especially for the explanation of the Fourier series.

The Fourier series 8.8 of the function f would look like this in the exponential

basis 8.9:

$$f = \boxed{\overset{n}{+}}_{k} \hat{f}_k e_k = \frac{1}{\sqrt{L}} \boxed{\overset{n}{+}}_{k} \hat{f}_k \, \mathrm{e}^{\mathrm{i}kx} \tag{8.10}$$

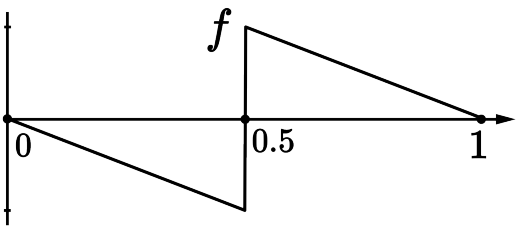
What can we do with the Fourier series 8.10 in the exponential basis? As I said, we can use it to approximate any function f in an interval. Let's take a look at a concrete example, then you'll know what I mean.

8.4 Example: Fourier series for the sawtooth function

As an example, let us consider the **sawtooth function** in the interval (0, 1):

$$f = \begin{cases} -x & (0, 0.5) \\ 1 - x & (0.5, 1) \end{cases} \tag{8.11}$$

This saw function looks like this:



Let's choose the exponential basis functions as our basis for the Fourier series of the sawtooth function:

$$f = \frac{1}{\sqrt{L}} \boxed{\overset{n}{+}}_{k} \hat{f}_k \, \mathrm{e}^{\mathrm{i}kx}$$

The total interval length is $L = 1$. This means that the normalization factor for

exponential basis functions is also 1:

$$f = \boxed{+}_{k}^{n} \hat{f}_k \, \mathrm{e}^{\mathrm{i}kx} \tag{8.12}$$

When determining Fourier series, we always have to do two things:

- **Choose a basis** and insert it into the Fourier series. We have already done this in eq. 8.12.
- **Calculate the Fourier coefficients** $\hat{f}_k$ with the integral 8.7 and insert it into the Fourier series 8.12. We determine the k-th Fourier coefficient 8.7 with the inner product between the k-th basis function and the sawtooth function f:

$$\hat{f}_k = \int_{x_0}^{x_n} e_k^*(x) \, f(x) \, dx \tag{8.13}$$

$$= \int_0^1 \mathrm{e}^{-\mathrm{i}kx} \, f(x) \, dx \tag{8.14}$$

Note that the exponential basis function must be complex conjugate in the integral. This is where the minus sign in the exponent of the exponential function comes from. And the integration limits $x_0 = 0$ and $x_n = 1$ are our free decision. We want to approximate the sawtooth function in this region.

Now it's up to you to solve the integral 8.14 to determine the Fourier coefficients concretely. I can't do it, so I'll leave it to you as an exercise.

In any case, here is the solution we need to illustrate it right away:

$$\hat{f}_k = \frac{1}{\mathrm{i}k} \mathrm{e}^{-\mathrm{i}k/2} \tag{8.15}$$

Since we have not entered a specific value for the wavenumber k in 8.15, we have determined all Fourier coefficients $\hat{f}_k$. For a different k value, we get a different Fourier coefficient in Eq. 8.15.

Let's just insert the Fourier coefficients 8.14 into the Fourier series 8.12 and

combine the two exponential functions:

$$f = \boxed{\overset{n}{+}}_{k} \hat{f}_k \, \mathrm{e}^{\mathrm{i}kx} \tag{8.16}$$

$$= \boxed{\overset{n}{+}}_{k} \frac{1}{\mathrm{i}k} \mathrm{e}^{-\mathrm{i}k/2} \mathrm{e}^{\mathrm{i}kx} \tag{8.17}$$

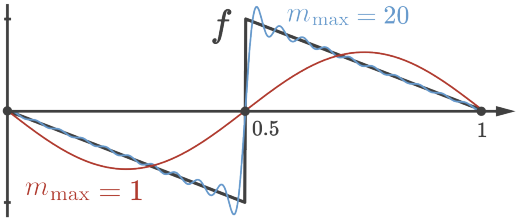
$$= \boxed{\overset{n}{+}}_{k} \frac{1}{\mathrm{i}k} \mathrm{e}^{\mathrm{i}k(x-0.5)} \tag{8.18}$$

$$= \boxed{\overset{\infty}{+}}_{m=-\infty} \frac{1}{\mathrm{i}2\pi m} \mathrm{e}^{\mathrm{i}2\pi m(x-0.5)} \tag{8.19}$$

In the last step, we selected periodic boundary conditions for $k = 2\pi m/L$, where $m = \ldots -2, -1, 0, 1, 2, \ldots$ takes whole numbers. We therefore sum over both positive and negative m.

We can decide up to which $m_{\max}$ we want to sum in the Fourier series 8.19 of the sawtooth function. The higher we choose $m_{\max}$, the better our approximation of the function f will be.

Look at the plots for the approximation $m_{\max} = 1$ and for a better approximation $m_{\max} = 20$:



With this Fourier series for the sawtooth function, we have basically gained two things:

- We can now sum the series up to a certain maximum value: $m = m_{\max}$ and thus obtain a continuously differentiable **good approximation for the sawtooth function**.
- Since we have determined the Fourier coefficients, we know which m

values are contained in the sawtooth function ($m = 0$, for example, is not included). We therefore know which building blocks (basis functions) the sawtooth function is composed of. This breaking down of the function into individual components is known as **Fourier analysis**.

II

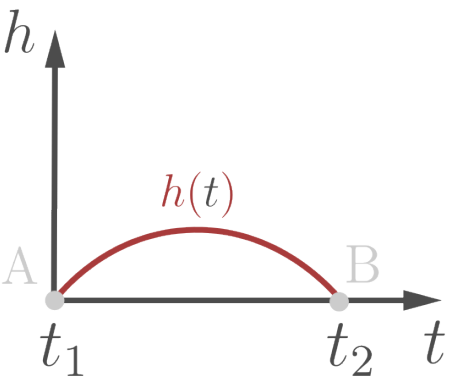
Nature is extreme

9. Action Functional

More: en.fufaev.org/euler-lagrange-equation

The Euler-Lagrange equation is a powerful tool with which we can set up **differential equations** (which you should be familiar with) **for a specific problem**. We will encounter the Euler-Lagrange equation not only in mechanics, but in all areas of theoretical physics. Let's first look at the motivation for this equation.

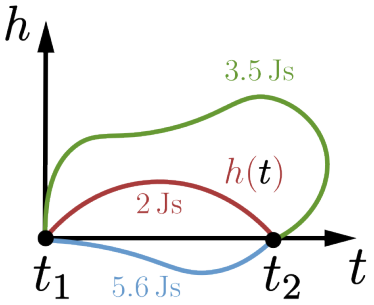
Let us consider a **particle in the gravitational field** that is thrown vertically upwards from the height $h(t_1) = 0$ at the time $t_1 = 0$. Marked as point A in the following image. It moves straight along a spatial direction and arrives on the ground at the same location $h(t_2) = h(t_1) = 0$ at time t_2 (marked as point B in the image):



The **connection between A and B**, that is, the trajectory $h(t)$ must be a **parabola** in this problem. But why is this trajectory a parabola and not some other trajectory? Why does nature or the particle between points A and B choose this path of all paths? And not for any other path?

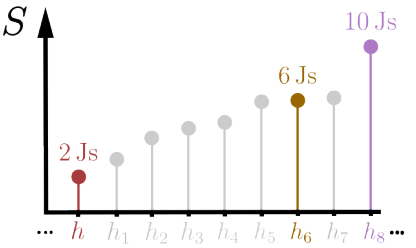
In order to answer this question, we need a physical quantity called **action**, which is abbreviated with an S and has the unit Js (Joule second).

We can assign an action $S[h]$ to each of the conceivable trajectories h. The action takes an entire function h in the argument and outputs a number $S[h]$, namely the value of the action for the corresponding function. For example, some trajectory h_1 could have the value $S[h_1] = 3.5\,\text{Js}$, another trajectory h_2 could have the value $S[h_2] = 5.6\,\text{Js}$ and the parabolic trajectory h could have the value $S[h] = 2\,\text{Js}$:



So now back to the question: **Why a parabola**? Experience shows that nature

follows the **principle of extreme action**. This means that if we calculate the corresponding action S for all possible trajectories $h(t)$, $h_1(t)$, $h_2(t)$ and so on between points A and B, then nature takes the value of the action that is **maximum**, **minimum** or a **saddle point**.



All other actions are out of the question for nature. Nature chooses one of these extremal actions. This is exactly what "extreme" means. Which of the extreme paths (minimum, maximum, saddle point) nature actually takes depends on the problem under consideration.

So we can answer the question: Why does the particle thrown upwards in the gravitational field take the path of the parabola in the space-time diagram? Because the parabolic trajectory h has the smallest action $S[h]$!

But **how do we actually calculate the value of the action**? For this we need the **lagrange function** $L(t, h, \dot{h})$. It depends on the time t, on the function value (position) $h(t)$ and on the time derivative (velocity) $\dot{h}(t)$ at time t. The Lagrange function has the unit of energy, that is, Joule (J).

If we integrate the Lagrange function L over the time t between t_1 and t_2, we get a quantity that has the unit Joule second. We interpret this as the action S:

$$S[h] = \int_{t_1}^{t_2} \mathrm{d}t\, L(t, h, \dot{h}) \tag{9.1}$$

The letter q is usually used instead of h and $\dot{q}$ instead of $\dot{h}$ and is called q **generalized coordinate** and the derivative $\dot{q}$ **generalized velocity**. The

generalized coordinate (the trajectory you are looking for) does not necessarily have to be the height h above the ground. For example, it can represent an angle $q = \varphi$ or any other quantity that may depend on the time t.

$$S[q] \;=\; \int_{t_1}^{t_2} \mathrm{d}t\, L(t, q, \dot{q}) \tag{9.2}$$

With this **formula for the action functional** we can calculate the value $S[q]$ of the action for every possible trajectory q that the particle can take. We only need to determine the Lagrange function L.

There are an infinite number of possible trajectories that a particle can take from A to B. Do I really have to calculate an infinite number of action functionals 9.2? No, there is a faster way to find the trajectory with the most extreme action and for this we need the Euler-Lagrange equation.

10. Euler-Lagrange Equation

Of course, it is totally cumbersome to calculate the action functional 9.2 for all possible trajectories and to take the trajectory that yields the smallest value of the integral. To save us this huge task, the **Euler-Lagrange equation** comes into play:

$$\frac{\partial L}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}} = 0 \tag{10.1}$$

This is one of the **most important equations in physics**. It is best to memorize it immediately. The derivation of the Euler-Lagrange equation is based on the definition of the action functional 9.2 and the principle of extremal action. In this chapter, we do not want to know how to derive the Euler-Lagrange equation, but rather how to use it to determine the desired extreme trajectory h.

The Euler-Lagrange equation 10.1 contains the partial derivative $\frac{\partial L}{\partial \dot{q}}$ of the Lagrange function with respect to the generalized velocity $\dot{q}$. This derivative is also referred to as **generalized momentum** and is abbreviated as p. You

may also encounter the Euler-Lagrange equation in the following form:

$$\frac{\partial L}{\partial q} - \frac{\mathrm{d}p}{\mathrm{d}t} = 0 \tag{10.2}$$

We call p »generalized momentum« because this **must not necessarily be mechanical momentum**. p can also represent a **torque**, for example. The generalized momentum p is then differentiated with respect to time t in the Euler-Lagrange equation.

If we rearrange the Euler-Lagrange equation 10.2 with respect to the time derivative of the momentum, we can read from it whether the momentum is conserved:

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{\partial L}{\partial q} \tag{10.3}$$

For the momentum to be conserved, the time derivative of the momentum must disappear. We therefore only have to calculate whether $\frac{\partial L}{\partial q}$ is zero, then the generalized momentum is obtained.

Using the form 10.3, we can also read off a **possible interpretation of the Euler-Lagrange equation**. It is a **condition for the conservation of generalized momentum**.

To be able to use the Euler-Lagrange equation at all, we need to know the Lagrange function L for a chosen system.

10.1 Lagrange function

The Lagrange function L is a scalar function that cannot be derived for any problem, but **can only be guessed or motivated**. If you think you have discovered a suitable Lagrange function for a problem, be it from quantum mechanics, classical mechanics or relativity, then you can easily use the Euler-Lagrange equation to check whether the Lagrange function you have

found correctly describes your problem or not. If you want to find the »Theore of Everything« formula that unites quantum mechanics with the general theory of relativity, then you should derive or dream up the corresponding Lagrange function.

In **classical mechanics**, the Lagrange function is the **difference between the kinetic energy $W_{\mathbf{kin}}$ and the potential energy $W_{\mathbf{pot}}$** of a particle:

$$L(t, q, \dot{q}) \;=\; W_{\text{kin}}(t, q, \dot{q}) \;-\; W_{\text{pot}}(t, q, \dot{q}) \tag{10.4}$$

So if we know the kinetic and potential energy of a particle, we can determine the Lagrange function 10.4 of mechanics and then use it in the Euler-Lagrange equation 10.1.

10.2 How To: Euler-Lagrange equation

Let's take a look at our example in the introduction, namely how we can calculate the parabola using the Lagrange function 10.4 and the Euler-Lagrange equation 10.1. To do this, we must always carry out the following five steps:

10.2.1 First step: Set generalized coordinates

First of all, we need to know what q and $\dot{q}$ actually represent. In our example, $q = h$ and $\dot{q} = v$, where v is the velocity of the thrown particle. Velocity is nothing other than the time derivative of the trajectory function, in other words $v = \dot{h}$.

10.2.2 Second step: Set up the Lagrange function

Next, we need to specify the Lagrange function 10.4 by giving the kinetic energy W_{kin} and the potential energy W_{pot} of the particle in the gravitational field as a function of q and $\dot{h}$. The kinetic energy W_{kin} of the thrown particle is given by

$$W_{\text{kin}} \;=\; \frac{1}{2}\, m\, \dot{h}^2 \tag{10.5}$$

Here m is the mass of the particle. The potential energy W_{pot} of the particle in

the gravitational field is given by

$$W_{\text{pot}} = m\,g\,h \tag{10.6}$$

The Lagrange function L for our problem is thus:

$$L = \frac{1}{2}\,m\,v^2 - m\,g\,h \tag{10.7}$$

10.2.3 Third step: Calculate derivatives

Now we can use the Lagrange function 10.7 to calculate the derivatives of the Lagrange function occurring in the Euler-Lagrange equation 10.1:

$$\frac{\partial L}{\partial h} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial v} = 0 \tag{10.8}$$

Differentiate the Lagrange function 10.7 with respect to h:

$$\frac{\partial L}{\partial h} = \frac{\partial L}{\partial h}\left(\frac{1}{2}\,m\,v^2 - m\,g\,h\right) \tag{10.9}$$

$$= \frac{\partial L}{\partial h}\left(\frac{1}{2}\,m\,v^2\right) - \frac{\partial L}{\partial h}\,(m\,g\,h) \tag{10.10}$$

$$= -mg \tag{10.11}$$

Differentiate the Lagrange function 10.7 with respect to v:

$$p = \frac{\partial L}{\partial v} = \frac{\partial L}{\partial v}\left(\frac{1}{2}\,m\,v^2 - m\,g\,h\right) \tag{10.12}$$

$$= \frac{\partial L}{\partial v}\left(\frac{1}{2}\,m\,v^2\right) - \frac{\partial L}{\partial v}\,(m\,g\,h) \tag{10.13}$$

$$= m\,v \tag{10.14}$$

Differentiate the calculated momentum $p = \frac{\partial L}{\partial v}$ with respect to time:

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(m\,v) \tag{10.15}$$

$$= m\,\dot{v} \tag{10.16}$$

$$= m\,\ddot{h} \tag{10.17}$$

Let's insert the calculated derivatives 10.11 and 10.14 into the Euler-Lagrange equation:

$$-m\,g \;-\; m\,\ddot{h} \;=\; 0 \tag{10.18}$$

Let's cancel the mass and bring $\ddot{h}$ to the right-hand side of the equation:

$$-g \;=\; \ddot{h} \tag{10.19}$$

What we have obtained in 10.19 is a differential equation for the desired trajectory $h(t)$. Hopefully you can see the usefulness of the Euler-Lagrange equation here: It is there **to set up differential equations for the extremal trajectory** $h(t)$**.**

Note that our example is a **one-dimensional problem** and therefore we only got **one differential equation**. For more complex multidimensional problems, we get several differential equations.

10.2.4 Fourth step: Solve the differential equations

Now we have to solve the differential equation 10.19 set up using the Euler-Lagrange equation. We can do this by integrating both sides twice. The solution is the unknown extremal trajectory:

$$h(t) \;=\; -\frac{1}{2}g\,t^2 \;+\; C_1\,t \;+\; C_2 \tag{10.20}$$

Here, C_1 and C_2 are the integration constants.

10.2.5 Fifth step: Set boundary conditions

The last step is to insert the constraints of the problem under consideration into the solution of the differential equation and determine the unknown constants C_1 and C_2.

In our problem, we have thrown the particle from the height $h_1 = 0$ at the time $t_1 = 0$. The first boundary condition is therefore: $h(0) = 0$. If we insert it into the solution 10.20, we get the second constant of integration: $C_2 = 0$. This simplifies the solution:

$$h(t) \;=\; -\frac{1}{2} g\, t^2 \;+\; C_1\, t \tag{10.21}$$

We know that the trajectory $h(t)$ ends at point B. Point B corresponds to the time t_2 at which the particle landed on the ground at $h(t_2) = 0$. This is the second constraint. If we insert this boundary condition into 10.21, we can determine the first constant of integration: $C_1 = \frac{1}{2} g\, t_2$. We are done!

The required extremal trajectory is therefore given by :

$$h(t) \;=\; -\frac{1}{2} g\, t^2 \;+\; C_1\, t \;+\; \frac{1}{2} g\, t_2\, t \tag{10.22}$$

This trajectory has the smallest value $S[h]$ of the action. If we plot the result 10.22 in the space-time diagram, we get a parabola.

Let us summarize: **The Euler-Lagrange equation helps us to set up differential equations for a desired trajectory between two fixed points**. The solution of these differential equations yields the exact shape of the trajectory that is allowed by nature.

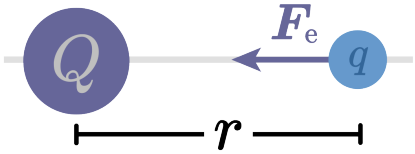
III Electromagnetism

11. The Electric Vector Field

Consider an electrically charged sphere with a large **source charge** Q and a sphere with a small **test charge** q. The test charge is at a certain point in time at a **distance** r from the source charge. The source charge exerts an **electric force** F_e on the test charge, which is given by the **Coulomb's law**:

$$F_\mathrm{e} = \frac{1}{4\pi\varepsilon_0}\frac{Q\,q}{r^2} \tag{11.1}$$

Here, $4\pi\varepsilon_0$ is a constant pre-factor with the vacuum permittivity ε_0, which ensures the correct unit of force on the right-hand side of Coulomb's law, namely the unit Newton (N).



What if we know the value of the big charge Q and want to know what force this

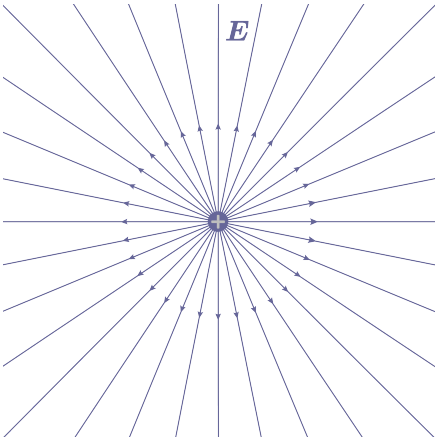
big charge exerts on another small charge q? But we don't know the exact value of this small charge. Or we deliberately leave this value open and only want to consider the electric force that would be exerted by the big charge if we place the test charge q near it. To do this, q must somehow be eliminated from the Coulomb's law. We achieve this by dividing the Coulomb's law on both sides by q so that the test charge on the right-hand side disappears:

$$\frac{F_\mathrm{e}}{q} = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2} \tag{11.2}$$

The quotient between force and charge on the left-hand side is defined as **electric field** E of the source charge Q:

$$E = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2} \tag{11.3}$$

So what is the electric field? **The electric field E indicates the electric force that WOULD act on a test charge q if it were placed at a distance r from the source charge Q.**



We have called Q the source charge to indicate that it is the **source of the electric field**. And so that Q is really the only source, we have chosen the test charge q **to be very small**.

The electric field in Eq. 11.3 is only the **magnitude**, that is, the value of the

electric field. For the cherry on the cake of electrodynamics, Maxwell's equations, we need the electric field as a vector quantity in order to also take into account the **direction of the electric field** at all locations in space. Therefore, the electric field E must be transformed into a vector $\boldsymbol{E}$. Vectors are shown in this book **in bold**.

The electric field $\boldsymbol{E}$ as a vector in three-dimensional space has three components E_1,E_2 and E_3.

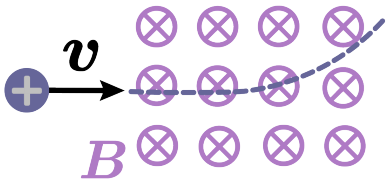
$$\boldsymbol{E} = \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \tag{11.4}$$

The first component $E_1(x, y, z)$ depends on the spatial coordinates (x, y, z) and it indicates the magnitude of the electric force that would act on a test charge along the x-axis if the test charge were placed at the location (x, y, z). The same applies to the other two field components $E_2(x, y, z)$ and $E_3(x, y, z)$, which each determine the electric force on a test charge along the y and z spatial directions.

We can summarize: **The electric vector field $\boldsymbol{E}$ assigns a vector $\boldsymbol{E}(x, y, z)$ to each point in space (x, y, z), which represents the electric field at that location. If a test charge is placed there, it is accelerated in the direction of this vector**.

12. The Magnetic Vector Field

Another important fundamental physical quantity that appears in the second and fourth Maxwell equation is the **magnetic field**. Experiments show that a particle with **electric charge** q moving in a straight line with **velocity** v in an external magnetic field, experiences a **magnetic force** F_{m}, which deflects the particle from its straight-line path.



The magnetic force on the particle increases proportionally to its charge $F_{\mathrm{m}} \sim q$ and proportional to its speed $F_{\mathrm{m}} \sim v$. This means that if the charge or speed is doubled, the magnetic force on the particle also doubles.

The force also increases in proportion to the applied magnetic field. To describe this proportionality of the force and the magnetic field, we introduce the quantity B. Overall, the magnetic force (also called **Lorentz force**) is given by:

$$F_\mathrm{m} = q\,v\,B \tag{12.1}$$

The unit of the quantity B must be such that the right-hand side of the equation results in the unit of the force, that is $\mathrm{N} = \mathrm{kg} \cdot \mathrm{m/s2}$. A simple transformation results in the unit of B: $\mathrm{kg/As^2}$. We refer to this unit as **Tesla** for short:

$$\mathrm{T} = \frac{\mathrm{kg}}{\mathrm{As}^2} \tag{12.2}$$

We refer to the proportionality constant B as **magnetic flux density** or short: **magnetic field** or even shorter **B-field**.

The equation 12.1 only represents the **magnitude** of the magnetic force F_m. To formulate the magnetic force **vectorially**, the force, the velocity and the magnetic field are written as vectors:

$$\boldsymbol{F}_\mathrm{m} = \begin{bmatrix} F_{\mathrm{m}1} \\ F_{\mathrm{m}2} \\ F_{\mathrm{m}3} \end{bmatrix}, \qquad \boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} \tag{12.3}$$

Now the three variables are not scalars, but three-dimensional vectors with the components in $x-$, $y-$ and z spatial direction. The question now is: **How must the velocity vector $\boldsymbol{v}$ be vectorially multiplied by the magnetic vector field $\boldsymbol{B}$?**

If the deflection of the charge in the magnetic field is investigated more closely in an experiment, it can be determined that the magnetic force deflects it always **orthogonally**, in other words perpendicularly to the direction of velocity and to the magnetic field lines. This orthogonality can be easily established with the **cross product $\boldsymbol{v} \times \boldsymbol{B}$**.

The cross product between the velocity vector and the magnetic field vector is

defined in such a way that the result of the cross product, which is a vector, is always **orthogonal** on the two vectors $\boldsymbol{v}$ and $\boldsymbol{B}$:

$$\boldsymbol{v} \times \boldsymbol{B} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \times \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} v_2 B_3 - v_3 B_2 \\ v_3 B_1 - v_1 B_3 \\ v_1 B_2 - v_2 B_1 \end{bmatrix} \tag{12.4}$$

So that the magnetic force $\boldsymbol{F}_\mathrm{m}$ is always orthogonal to $\boldsymbol{v}$ and $\boldsymbol{B}$, their cross product must be formed. The magnetic force is therefore given as a vector field by the following equation:

$$F_\mathrm{m} = q\,\boldsymbol{v} \times \boldsymbol{B} \tag{12.5}$$

So what is the magnetic vector field $\boldsymbol{B}$? **The magnetic vector field assigns a vector $\boldsymbol{B}(x, y, z)$ to each point in space (x, y, z), which determines the magnitude and direction of the magnetic force $F_\mathrm{m}(x, y, z)$ on a moving charge q.**

13. Maxwell's Equations

More: en.fufaev.org/maxwell-equations

The **four Maxwell equations** together with the **Lorentz force** contain the **entire knowledge of electrodynamics**. There are so many applications of this that I can't list them all, but some of them are, for example

- **Electronic devices** such as computers and mobile phones. They contain electrical capacitors, coils and entire circuits that make use of Maxwell's equations.
- **Power generation** - whether from nuclear, wind or hydroelectric power plants, the energy released must first be converted into electrical energy so that people can use it. This happens with electric generators. These in turn are based on Maxwell's equations.
- **Power supply**. AC voltages and transformers are needed to transport electricity to households with as little energy loss as possible.
- **And much more!** Electric welding for assembling car bodies, motors for electric cars, magnetic resonance imaging in medicine, kettles in the kitchen, the charger for your smartphone, radio, Wi-Fi and so on.

Isn't that incredible? Every device that utilizes electricity or magnetism is fundamentally based on Maxwell's equations. Here, take a look:

$$\begin{aligned}
\nabla \cdot \boldsymbol{E} &= \frac{\rho}{\varepsilon_0} \\
\nabla \cdot \boldsymbol{B} &= 0 \\
\nabla \times \boldsymbol{E} &= -\frac{\partial \boldsymbol{B}}{\partial t} \\
\nabla \times \boldsymbol{B} &= \mu_0 \boldsymbol{j} + \mu_0 \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}
\end{aligned}$$

They may still seem a little cryptic to you, but after this lesson you will be able to translate each of these four equations into a picture, which will be easier to internalize.

As you can see from Maxwell's equations, the **electric field** $\boldsymbol{E}$ and the magnetic field $\boldsymbol{B}$ appear there. You will hopefully have become familiar with these two quantities in the chapters 11 and 12. Of course, I also assume that you have read the chapter 5 on the nabla operator, the Stokes' Curl Theorem 7 and the Gauss Divergence Theorem 6. If this is the case, then you will have no problem understanding the following chapters.

13.1 Integral and Differential Representation

The four Maxwell equations can be represented in two different ways:

- We can represent Maxwell's equations in **integral form**. Here we express them with integrals. This allows us to understand Maxwell's equations **macroscopically**. This is what the integral form looks like. Let it affect

you briefly:

$$\begin{aligned}
\oint_A \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{a} &= \frac{Q}{\varepsilon_0} \\
\oint_A \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{a} &= 0 \\
\oint_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} &= -\int_A \frac{\partial \boldsymbol{B}}{\partial t} \cdot \mathrm{d}\boldsymbol{a} \\
\oint_L \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} &= \mu_0 I + \mu_0 \varepsilon_0 \int_A \frac{\partial \boldsymbol{E}}{\partial t} \cdot \mathrm{d}\boldsymbol{a}
\end{aligned}$$

What feelings do these equations generate in you? Maybe fear? Trust me - soon no more!

- We can represent Maxwell's equations in **differential form**. Here we express them with derivatives. This allows us to understand Maxwell's equations **microscopically**. This is what the differential form looks like:

$$\begin{aligned}
\nabla \cdot \boldsymbol{E} &= \frac{\rho}{\varepsilon_0} \\
\nabla \cdot \boldsymbol{B} &= 0 \\
\nabla \times \boldsymbol{E} &= -\frac{\partial \boldsymbol{B}}{\partial t} \\
\nabla \times \boldsymbol{B} &= \mu_0 \boldsymbol{j} + \mu_0 \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}
\end{aligned}$$

What is the exact difference between these two representations? Both representation **are not physically different**, but mathematically they are. And they are useful for different problems in different ways:

- While the differential form of a Maxwell equation **applies to a single point** in space, the integral form applies to a **entire spatial area**.
- The integral form is well suited for calculating **symmetric problems**, such as calculating the electric field of a charged sphere, a charged cylinder or a charged plane. The differential form is more suitable for calculating **complicated numerical problems** or **for various derivations**, such as the derivation of electromagnetic waves.

- In addition, the differential representation **is much more compact** than the integral form.

Both representations are useful and can be **converted into each other with the help of two mathematical theorems (Gauss and Stokes)**, which you learned about in the chapters 6 and 7. Once you have understood the two theorems, it will be easy for you to convert the integral form into the differential form and vice versa.

To understand Maxwell's equations, it is helpful to understand the electric and magnetic flux and the voltage.

13.2 Electric and Magnetic Flux

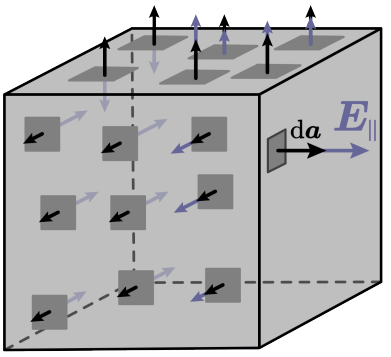
In the chapter 6 on the Gauss Divergence Theorem, you learned about the **flux Φ of a vector field $\boldsymbol{F}$ through the surface** A (we do not need the circle at the integral sign to define the flux):

$$\Phi = \int_A \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{a} \tag{13.1}$$

The surface integral over the vector field $\boldsymbol{F}$ results in a number Φ, which indicates how much of the vector field $\boldsymbol{F}$ passes through the surface A.

If the vector field $\boldsymbol{F}$ in the surface integral is an **electric field $\boldsymbol{F} = \boldsymbol{E}$**, then this surface integral is called **electric flux Φ_e through the surface** A:

$$\Phi_\mathrm{e} = \int_A \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{a} \tag{13.2}$$



If, on the other hand, the vector field $\boldsymbol{F}$ in the surface integral is a **magnetic field** $\boldsymbol{F} = \boldsymbol{B}$, then this surface integral is referred to as **magnetic flux** Φ_m **through the surface** A:

$$\Phi_\mathrm{m} = \int_A \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{a} \tag{13.3}$$

13.3 Electric and Magnetic Voltage

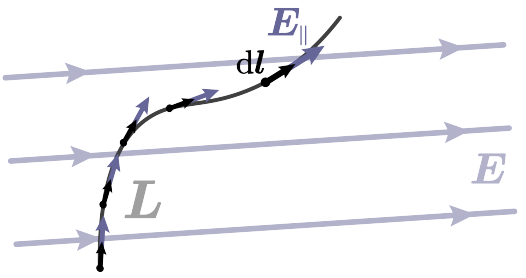
In the chapter 7 on the Stokes' Curl Theorem, we denoted the following line integral over the vector field $\boldsymbol{F}$ by U:

$$U = \int_L \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{l} \tag{13.4}$$

The number U indicates how much of the vector field circulates along the line L. It was no coincidence that we gave it the same letter as the voltage.

If the vector field $\boldsymbol{F}$ in the line integral is an **electric field** $\boldsymbol{F} = \boldsymbol{E}$, then this line integral is referred to as **electric voltage** U_e **along the line** L:

$$U_\mathrm{e} = \int_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} \tag{13.5}$$



The voltage 13.5 in the case of an electric field is **proportional to the kinetic energy**:

- A **positively charged particle gains energy** when it passes through the line L.
- A **negatively charged particle loses energy** when it passes through the line L.

The line integral 13.5 of the electric field, that is the voltage U_{e}, **measures the kinetic energy gain or energy loss** of a charged particle when it passes through the considered line L in the electric field. Note, however, that this kinetic energy does not come from nothing, but is **withdrawn from or added to the electric field**.

If the vector field $\boldsymbol{F}$ in the line integral is a **magnetic field** $\boldsymbol{F} = \boldsymbol{B}$, then this line integral is referred to as **magnetic voltage** U_{m} **along the line** L:

$$U_{\mathrm{m}} = \int_L \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} \tag{13.6}$$

In contrast to electric voltage, magnetic voltage has no interpretation as energy, because here the particle does not change its energy when it passes through the line L. Magnetic fields do not perform any work on moving charges. Nevertheless, the analogous definition to electric voltage makes mathematical sense.

We will need this knowledge of electric and magnetic flux and voltage in a moment if we want to understand Maxwell's equations in integral form.

13.4 First Maxwell Equation

13.4.1 Macroscopic form

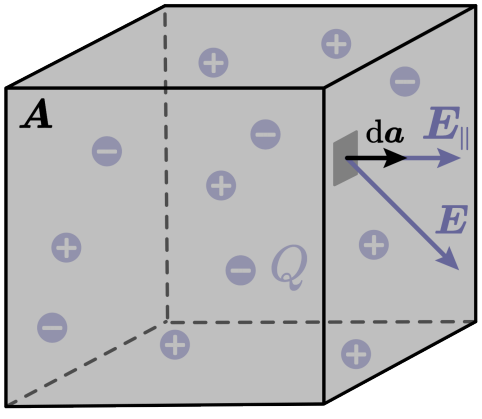
Let's take a look at the **first Maxwell equation in integral form**:

$$\oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{a} = \frac{Q}{\varepsilon_0} \tag{13.7}$$

You should be familiar with the left-hand side of Maxwell's equation 13.7. It is the **electric flux** Φ_e through an imaginary surface A that encloses something. The left-hand side of Maxwell's equation therefore tells you how much net of the **electric field** $\boldsymbol{E}$ exits and enters the surface A:

$$\Phi_\mathrm{e} = \frac{Q}{\varepsilon_0} \tag{13.8}$$

On the right-hand side of the first Maxwell equation is the **total electric charge** Q, which is **enclosed by the surface** A. The **vacuum permitivity** ε_0 is only there to have the correct unit "voltmeter" on both sides of the Maxwell equation. The interesting thing is: It doesn't matter **how** this enclosed charge is distributed.



So what does Maxwell's first equation mean in integral form? The electric flux through a surface is determined by the electric charge enclosed by

that surface.

13.4.2 Microscopic form

This is the **macroscopic interpretation** of Maxwell's first equation. For a **microscopic** interpretation, we need to convert the **intergral form into a differential form**. How do we do that? We must first convert both sides of Maxwell's equation 13.7 into a volume integral.

With the Gauss Divergence theorem 6.1, which links a volume integral with a surface integral, the surface integral on the left-hand side of the first Maxwell equation can be rewritten as a volume integral:

$$\int_V (\nabla \cdot \boldsymbol{E}) \, \mathrm{d}v \;=\; \frac{Q}{\varepsilon_0} \tag{13.9}$$

The enclosed charge Q can also be expressed with a volume integral. The charge corresponds to the **charge density** ρ over the considered volume V, because charge density is by definition **charge per volume**. This means that the volume integral of the charge density ρ over a volume V corresponds to the charge enclosed in this volume. This transforms the right-hand side of the first Maxwell equation 13.9 into a volume integral:

$$\int_V (\nabla \cdot \boldsymbol{E}) \, \mathrm{d}v \;=\; \frac{1}{\varepsilon_0} \int_V \rho \, \mathrm{d}v \tag{13.10}$$

On both sides in Eq. 13.10 we integrate **over the same volume** V. To ensure that this equation is always fulfilled for any chosen volume, the integrands on both sides must be the same (whereby the right integrand is multiplied by the constant $\frac{1}{\varepsilon_0}$). This results in the **differential form of the first Maxwell equation**:

$$\nabla \cdot \boldsymbol{E} \;=\; \frac{\rho}{\varepsilon_0} \tag{13.11}$$

On the left-hand side of the differential form is the **divergence** $\nabla \cdot \boldsymbol{E}$ **of the electric field**. The divergence at the location (x, y, z) can be positive, negative

or zero. We learned what this means in the 5.2 chapter:

- If the **divergence** $\nabla \cdot \boldsymbol{E}(x,y,z) > 0$ at the location (x,y,z) is **positive**, then the **charge density** $\rho(x,y,z)$ at the location (x,y,z) is also **positive**. There is therefore a **positive electric charge** at the location (x,y,z), which is the **source of the electric field**.
- If the **divergence** $\nabla \cdot \boldsymbol{E}(x,y,z) < 0$ at the location (x,y,z) is **negative**, then the **charge density** $\rho(x,y,z)$ at the location (x,y,z) is also **negative**. There is therefore a **negative electric charge** at the location (x,y,z), which is the **sink of the electric field**.
- If the **divergence** $\nabla \cdot \boldsymbol{E}(x,y,z) = 0$ at the location (x,y,z) is **zero**, then the **charge density** $\rho(x,y,z)$ at the location (x,y,z) is also **zero**. There is therefore **neither a negative nor a positive electric charge** at the location (x,y,z) or there is just as much positive as negative charge there, so that the total charge at this point (x,y,z) cancels out. In this case, there is a **ideal electric dipole** at this point.

So what does Maxwell's first equation mean in differential form? **The electric charges are the sources and sinks of the electric field. Charges generate an electric field.**

13.5 Second Maxwell Equation

13.5.1 Macroscopic form

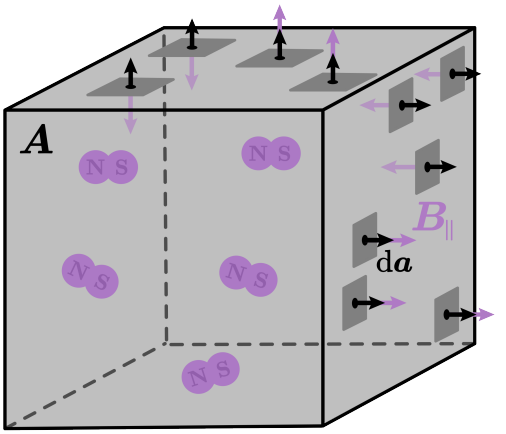
The **second Maxwell equation in integral form** looks like this:

$$\oint_A \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{a} \;=\; 0 \tag{13.12}$$

Nothing here should be unfamiliar to you. On the left-hand side is a surface integral. However, it is not integrated over an electric vector field, as in the first Maxwell equation, but over a **magnetic vector field $\boldsymbol{B}$**. This surface integral corresponds to the **magnetic flux Φ_m through the closed surface A**.

So what does Maxwell's second equation mean in integral form from a

macroscopic point of view? **The magnetic flux through a closed surface is always zero. There are as many magnetic field vectors pointing into the surface as out.**



13.5.2 Microscopic form

To obtain the **differential form** of the second Maxwell equation, we must convert the surface integral in the second Maxwell equation 13.12 **into a volume integral**. To do this, we simply replace the surface integral with the volume integral using the Gauss Divergence Theorem 6.1. The second Maxwell equation then looks like this:

$$\int_V (\nabla \cdot \boldsymbol{B})\, \mathrm{d}v \;=\; 0 \tag{13.13}$$

The integral 13.13 for **any volume** V is always zero only if the integrand $\nabla \cdot \boldsymbol{B}$ is zero. This is how the **second Maxwell equation in its differential form** emerges:

$$\nabla \cdot \boldsymbol{B} \;=\; 0 \tag{13.14}$$

The magnetic counterpart to the electric charge is the **magnetic charge**: a magnetic south pole and north pole. We call them **magnetic monopoles**. They have never been observed experimentally, which is why it is initially assumed

that no magnetic monopoles exist. Their non-existence is captured in the second Maxwell equation (the right-hand side is zero).

The differential representation of Maxwell's second equation allows us a **microscopic interpretation** of the non-existence of magnetic monopoles. The vanishing divergence $\nabla \cdot \boldsymbol{B}(x, y, z)$ of the magnetic field means: No matter which point in space (x, y, z) we look at - **there is no magnetic monopole in any point in space**. Or, when the divergence also disappears, is the case when there is **an ideal magnetic dipole** at the point (x, y, z). A magnetic dipole is a combination of the south and north poles. These are inseparably connected to each other.

Since there are no magnetic monopoles, **there are no sources and sinks** of the magnetic field. Consequently, there are no points in space where **magnetic field vectors originate** or **diverge to a point**. The **magnetic field lines must therefore always be closed**.

The second Maxwell equation is just like the other Maxwell equations **an experimental observation**. This means that if at some point a magnetic monopole is found, for example a single north pole without an associated south pole, then Maxwell's second equation must be modified. That would be nice for us, because then Maxwell's equations would be even more symmetrical!

13.6 Third Maxwell Equation

13.6.1 Macroscopic form

The **third Maxwell equation in integral form** looks like this:

$$\oint_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} \;=\; -\int_A \frac{\partial \boldsymbol{B}}{\partial t} \cdot \mathrm{d}\boldsymbol{a} \tag{13.15}$$

On the left-hand side is a line integral of the **electric field** $\boldsymbol{E}$ over a **closed line** L that borders the **surface** A. This line integral sums up all components $\boldsymbol{E}_{||}$ of the electric field that run along the line L. From the chapter 13.3 we know that this line integral corresponds to the **electric voltage** U_e along the loop L.

We can also write the second Maxwell equation as follows:

$$U_\mathrm{e} = -\int_A \frac{\partial \boldsymbol{B}}{\partial t} \cdot \mathrm{d}\boldsymbol{a} \tag{13.16}$$

This form of Maxwell's third equation is referred to as **law of induction**. On the right-hand side of the law of induction is a surface integral of the time derivative of the magnetic field. If the surface A through which the magnetic field penetrates does not change over time, we can put the **time derivative** $\frac{\partial}{\partial t}$ in front of the integral:

$$U_\mathrm{e} = -\frac{\partial}{\partial t} \int_A \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{a} \tag{13.17}$$

Now we can interpret the surface integral on the right-hand side as **magnetic flux Φ_m through the surface** A:

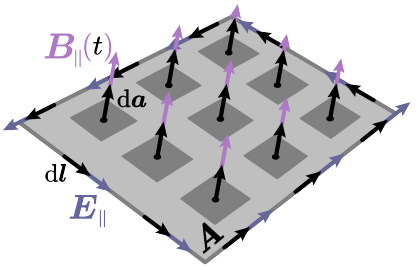
$$U_\mathrm{e} = -\frac{\partial \Phi_\mathrm{m}}{\partial t} \tag{13.18}$$

The time derivative in front of the magnetic flux is also still there. The magnetic flux is therefore differentiated with respect to time in the third Maxwell equation. The time derivative of the magnetic flux indicates **how quickly the magnetic flux changes** when time passes. The third Maxwell equation therefore tells us two equivalent things:

- The faster the magnetic flux Φ_m changes through the enclosed surface A, the greater the voltage U_e generated along the edge of the surface L.
- The faster the magnetic flux through the enclosed surface A changes, the stronger the **parallel field component** $\boldsymbol{E}_{||}$, which runs along the edge of the surface L. This electric field along the edge is also referred to as **electric vortex field**, because this field component swirls around the edge of the surface:

$$\oint_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} = -\frac{\partial \Phi_\mathrm{m}}{\partial t} \tag{13.19}$$

Of course, we can also interpret Maxwell's third equation 13.19, expressed with the vortex field, the other way round: **The stronger the electric vortex field $\boldsymbol{E}$** (or more precisely $\boldsymbol{E}_{||}$) around the surface boundary L **the faster the magnetic flux** Φ_m (or equivalently $\boldsymbol{B}_{||}$) changes through the surface A.



You are probably also wondering what the **minus sign before the time derivative** means? The minus sign takes into account the **circulation direction of the vortex field**:

- If the change in magnetic flux is **positive**, that is $\frac{\partial \Phi_\mathrm{m}}{\partial t} > 0$, then the voltage is **negative** due to the minus sign, that is $U_\mathrm{e} < 0$.
- If the change in magnetic flux is **negative**, that is $\frac{\partial \Phi_\mathrm{m}}{\partial t} < 0$, then the voltage is **positive** due to the minus sign, that is $U_\mathrm{e} > 0$.

The vortex component $\boldsymbol{E}_{||}$ of the electric field $\boldsymbol{E}$ thus swirls around in such a way that the **change in magnetic flux is impeded**. Nature tries to prevent the change in flux with a vortex field. You probably remember this minus sign in the third Maxwell equation from school as the **Lenz rule**. The minus sign takes into account the law of conservation of energy.

What would happen if we omitted the minus sign? The electric vortex field (with the **field energy** W_e) would generate a magnetic flux change $\frac{\partial \Phi_\mathrm{m}}{\partial t}$. This in turn would amplify the electric vortex field. This would increase the field energy W_e. The increased vortex field leads to an increased change in flux. This in turn leads to an even larger vortex field and thus to greater field energy.

This mutual amplification does not stop and the field energy W_e becomes infinitely large. We could tap into this with a capacitor, for example, and have

an inexhaustible source of energy. This not only violates the law of conservation of energy, but also the third law of thermodynamics, which states that it is impossible to build such a perpetual motion machine.

If the magnetic flux does not change over time $\frac{\partial \Phi_\mathrm{m}}{\partial t} = 0$, then of course there is also **no electric vortex field** and no electric voltage. The right-hand side of the third Maxwell equation 13.19 is therefore zero:

$$\oint_L \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{l} \;=\; 0 \tag{13.20}$$

Now it is stated in 13.20 that the line integral over the electric field, that is the electric voltage U_e, is always zero along a closed line L. **There is therefore no electric vortex field as long as there is no time-varying magnetic field!** This means: If an electron were to pass through the closed line L in the electric field $\boldsymbol{E}$, the electron would not change its energy.

13.6.2 Microscopic form

Let us now convert the third Maxwell equation 13.15 in integral representation into a **differential representation** in order to be able to interpret the Maxwell equation **microscopically**. To do this, we use the Stokes' Curl Theorem from the chapter 7, which links a **line integral with a surface integral**. So let's replace the line integral with a surface integral in 13.15:

$$\int_A (\nabla \times \boldsymbol{E}) \cdot \mathrm{d}\boldsymbol{a} \;=\; -\int_A \frac{\partial \boldsymbol{B}}{\partial t} \cdot \mathrm{d}\boldsymbol{a} \tag{13.21}$$

This brings the **curl** $\nabla \times \boldsymbol{E}$ into play. Since the equation 13.21 **applies to any** surfaces A, the integrands on both sides must be equal. This yields the **third Maxwell equation in differential representation**:

$$\nabla \times \boldsymbol{E} \;=\; -\frac{\partial \boldsymbol{B}}{\partial t} \tag{13.22}$$

The differential representation 13.22 states: If the **magnetic field** $\boldsymbol{B}(t, x, y, z)$ **changes in time at the point in space** (x, y, z)**, then an electric vortex**

field $\boldsymbol{E}(t, x, y, z)$ is generated around this point in space, which attempts to suppress this change in the magnetic field. Of course, the interpretation also works the other way round: an electric vortex field around a point in space generates a magnetic field that changes over time.

If the **magnetic field** does not change, that is, if it is **static**, the right-hand side in 13.22 is zero and the third Maxwell equation is simplified to an **electrostatic Maxwell equation**. »Electrostatic« here means that the electric field $\boldsymbol{E}$ is time-independent:

$$\nabla \times \boldsymbol{E} \;=\; 0 \tag{13.23}$$

As long as there is no changing magnetic field, the **electric field is always $\boldsymbol{E}$ vortex-free**. We know this from mathematics: If the rotation $\nabla \cdot \boldsymbol{E}$ of a vector field $\boldsymbol{F}$ vanishes, then the **vector field is conservative**, that is, it conserves energy. The **electrostatic electric field** $\boldsymbol{E}$ in Eq. 13.23 is therefore conservative. Electric charges are neither accelerated nor decelerated in this electric field - they do not change their energy.

13.7 Fourth Maxwell Equation

13.7.1 Macroscopic form

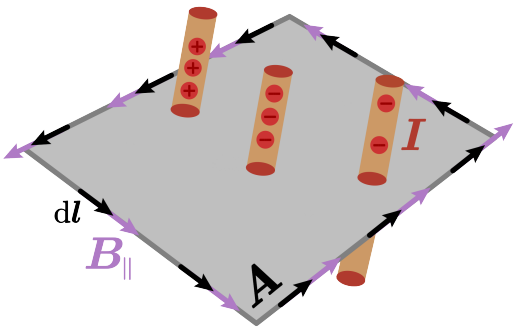
The **fourth Maxwell equation in integral form** looks like this:

$$\oint_L \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} \;=\; \mu_0 \boldsymbol{I} \;+\; \mu_0 \varepsilon_0 \int_A \frac{\partial \boldsymbol{E}}{\partial t} \cdot \mathrm{d}\boldsymbol{a} \tag{13.24}$$

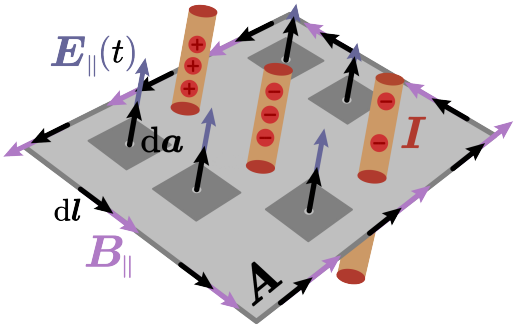
On the left-hand side of Maxwell's fourth equation is a line integral of the **magnetic field** $\boldsymbol{B}$ along the closed line (loop) L. We already know what this line integral means from the line integral over the electric field in Eq. 13.19. It indicates the **vortex component** $\boldsymbol{B}_{||}$ of the magnetic field $\boldsymbol{B}$, which swirls **around the loop** L.

The right-hand side of Maxwell's fourth equation 13.24 tells us how we can generate this magnetic vortex field:

- We can generate it with a **electric current** $\boldsymbol{I}$ through the surface A. This current **does not have to change in time** to generate a magnetic vortex field.



- We can generate it with a **time-changing electric field** $\boldsymbol{E}(t)$ through the surface A.
- We can generate it with both contributions, $\boldsymbol{I}$ and $\boldsymbol{E}(t)$.



The physical constants ε_0 and μ_0 in the fourth Maxwell equation are irrelevant for understanding the fourth Maxwell equation. These constants merely ensure that the right-hand side also has the unit "Tesla times meter", like the left-hand side of the equation.

On the right-hand side of 13.24 we can pull the time derivative in front of the integral if the surface A does not change:

$$\oint_L \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} \;=\; \mu_0 \boldsymbol{I} \;+\; \mu_0 \varepsilon_0 \frac{\partial}{\partial t} \int_A \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{a} \tag{13.25}$$

Then the surface integral, integrated over the electric field, corresponds exactly to the **electric flux** Φ_e through the surface A:

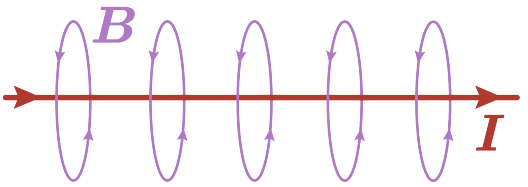
$$\oint_L \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} \;=\; \mu_0 \boldsymbol{I} \;+\; \mu_0 \varepsilon_0 \frac{\partial \Phi_e}{\partial t} \tag{13.26}$$

13.7.2 Ampere's Law

An important special case arises if the electric flux does not change over time, that is, $\frac{\partial \Phi_e}{\partial t} = 0$, then Maxwell's fourth equation is simplified to **Ampere's Law**:

$$\oint_L \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{l} \;=\; \mu_0 \boldsymbol{I} \tag{13.27}$$

According to the **Ampere's Law**, a current-carrying wire generates a magnetic vortex field around itself.



13.7.3 Microscopic form

Let us now derive the differential form of the fourth Maxwell equation. Using the Stokes' Curl Theorem 7.1, we convert the line integral inside the integral form 13.24 into a surface integral. In this way, the curlof the magnetic field $\nabla \times B$ comes into play:

$$\int_A (\nabla \times \boldsymbol{B}) \cdot \mathrm{d}\boldsymbol{a} \;=\; \mu_0 \boldsymbol{I} \;+\; \mu_0 \varepsilon_0 \int_A \frac{\partial \boldsymbol{E}}{\partial t} \cdot \mathrm{d}\boldsymbol{a} \tag{13.28}$$

Now the term with the electric current $\boldsymbol{I}$ must be converted into a surface integral. To do this, we simply have to express the current with the **electric**

current density $\boldsymbol{j}$. The current density is defined as the current per cross-sectional area. Consequently, the surface integral over the current density corresponds to the current through the cross-sectional area A. The fourth Maxwell equation thus becomes:

$$\int_A (\nabla \times \boldsymbol{B}) \cdot \mathrm{d}\boldsymbol{a} \;=\; \mu_0 \int_A \boldsymbol{j} \cdot \mathrm{d}\boldsymbol{a} \;+\; \mu_0 \varepsilon_0 \int_A \frac{\partial \boldsymbol{E}}{\partial t} \cdot \mathrm{d}\boldsymbol{a} \tag{13.29}$$

Note that the scalar product of the current density with the surface element $\mathrm{d}\boldsymbol{a}$ is taken in the integral. The scalar product therefore only picks the component $\boldsymbol{j}_{||}$ of the current density vector that runs **parallel to the surface element** $\mathrm{d}\boldsymbol{a}$. Only this current density component contributes to the current through the cross-sectional area A.

Now we have a surface integral in each term of the fourth Maxwell equation. We can combine the two surface integrals on the right-hand side into one surface integral because both summands are integrated over the same area:

$$\int_A (\nabla \times \boldsymbol{B}) \cdot \mathrm{d}\boldsymbol{a} \;=\; \int_A \left(\mu_0 \boldsymbol{j} \;+\; \mu_0 \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} \right) \cdot \mathrm{d}\boldsymbol{a} \tag{13.30}$$

For the Maxwell equation 13.30 to be fulfilled for **any-surface** A, the integrands on both sides must be equal. This results in the **fourth Maxwell equation in differential representation**:

$$\nabla \times \boldsymbol{B} \;=\; \mu_0 \boldsymbol{j} \;+\; \mu_0 \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} \tag{13.31}$$

What does Maxwell's fourth equation **microscopically** mean? **If the electric field $\boldsymbol{E}(t, x, y, z)$ changes over time at the point in space (x, y, z)** or if the **current density $\boldsymbol{j}(t, x, y, z)$ is not zero**, then a **magnetic vortex field $\boldsymbol{B}(t, x, y, z)$ is generated** around this point in space.

You should now have an intuitive understanding of Maxwell's four equations. Next, let's take a look at where exactly the electromagnetic waves are hidden in Maxwell's equations.

14. Electromagnetic Waves

More: en.fufaev.org/electromagnetic-waves

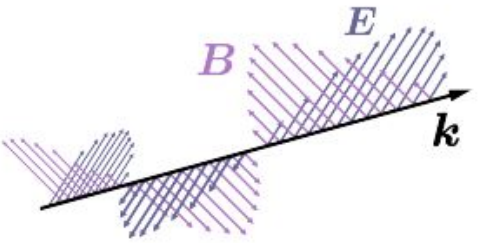
An electromagnetic wave (short: EM wave) consists of a **electric field component** $\boldsymbol{E}(t,x,y,z)$ and a **magnetic field component** $\boldsymbol{B}(t,x,y,z)$. The two field components assign an electric and magnetic field strength and its direction to each point (x,y,z) in three-dimensional space at each time t. The two field components are therefore three-dimensional vector fields:

$$\boldsymbol{E}(t,x,y,z) = \begin{bmatrix} E_1(t,x,y,z) \\ E_2(t,x,y,z) \\ E_3(t,x,y,z) \end{bmatrix} \qquad \boldsymbol{B}(t,x,y,z) = \begin{bmatrix} B_1(t,x,y,z) \\ B_2(t,x,y,z) \\ B_3(t,x,y,z) \end{bmatrix}$$

- The magnitude (field strength) $E(t,x,y,z)$ (not shown in bold) indicates the **electric amplitude** of an electromagnetic wave.
- The magnitude (field strength) $B(t,x,y,z)$ indicates the **magnetic amplitude** of an electromagnetic wave.

The amplitudes are generally not only dependent on location, but they also change with time t. This is the only way to obtain an electromagnetic **oscillation in space and time**. The following image shows the oscillation of

the $\boldsymbol{E}$ and $\boldsymbol{B}$ vectors in space. The **wave vector** $\boldsymbol{k}$ indicates the **propagation direction of the electromagnetic wave**:



How an electromagnetic wave changes exactly in space and time, that is, how it moves and propagates **in space**, is described by the **wave equations** for $\boldsymbol{E}$ and $\boldsymbol{B}$ vectors. Let's take a look at how we can extract these wave equations from Maxwell's equations:

$$\begin{aligned}
\nabla \cdot \boldsymbol{E} &= \frac{\rho}{\varepsilon_0} \\
\nabla \cdot \boldsymbol{B} &= 0 \\
\nabla \times \boldsymbol{E} &= -\frac{\partial \boldsymbol{B}}{\partial t} \\
\nabla \times \boldsymbol{B} &= \mu_0 \boldsymbol{j} + \mu_0 \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}
\end{aligned}$$

We assume that the electromagnetic wave propagates in an **empty space**, **without charges** ($\rho = 0$) **and currents** ($\boldsymbol{j} = 0$). We therefore set both the charge density ρ and the current density $\boldsymbol{j}$ in Maxwell's equations to zero. This simplifies them to **charge- and current-free Maxwell's equations**:

$$\nabla \cdot \boldsymbol{E} = 0 \tag{14.1}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{14.2}$$

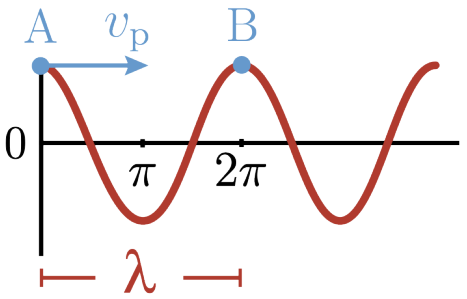
$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{14.3}$$

$$\nabla \times \boldsymbol{B} = \mu_0 \varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} \tag{14.4}$$

The general form of a **wave equation for a vector field $\boldsymbol{F}$** looks like this:

$$\nabla^2 \boldsymbol{F} = \frac{1}{v_\mathrm{p}} \frac{\partial^2 \boldsymbol{F}}{\partial t^2} \tag{14.5}$$

Here $\boldsymbol{F}$ is an arbitrary vector field that satisfies the **wave equation** and v_p is the **phase velocity** of the wave. It indicates how fast a point of the wave moves in space. Since we are not considering dispersion (meaning that the wave moves apart), the phase velocity describes the **propagation speed of the wave**.



A relation that is necessary for the derivation of the wave equation is the following relationship for the **curl of the curl of the vector field $\boldsymbol{F}$** (double cross product):

$$\nabla \times \nabla \times \boldsymbol{F} = \nabla(\nabla \cdot \boldsymbol{F}) - \nabla^2 \boldsymbol{F} \tag{14.6}$$

The four Maxwell equations are **coupled differential equations**. "Coupled" here means that the third and fourth Maxwell equations contain both the $\boldsymbol{E}$ field and the $\boldsymbol{B}$ field. To obtain the wave equations for the $\boldsymbol{E}$ and $\boldsymbol{B}$ field component of an electromagnetic wave, we need to decouple the two coupled **Maxwell equations**. Let's do that. It's quite simple.

14.1 Wave equation for the $\boldsymbol{E}$-field

To arrive at the wave equation for the electric field $\boldsymbol{E}$, we have to decouple the **third Maxwell equation**:

$$\nabla \times \boldsymbol{E} \;=\; -\frac{\partial \boldsymbol{B}}{\partial t} \tag{14.7}$$

Let's apply the **curl operator with cross product** $\nabla\times$ to both sides of the third Maxwell equation:

$$\nabla \times \nabla \times \boldsymbol{E} \;=\; \nabla \times \left(-\frac{\partial \boldsymbol{B}}{\partial t}\right) \tag{14.8}$$

The time derivative together with the minus sign may be placed in front of the Nabla operator, since the Nabla operator **only contains spatial derivatives** and thus **does not depend on time**:

$$\nabla \times \nabla \times \boldsymbol{E} \;=\; -\frac{\partial}{\partial t}\left(\nabla \times \boldsymbol{B}\right) \tag{14.9}$$

Now we can replace the curl $\nabla \times \boldsymbol{B}$ of the magnetic field using the fourth current-free Maxwell equation 14.4:

$$\nabla \times \nabla \times \boldsymbol{E} \;=\; -\frac{\partial}{\partial t}\left(\mu_0\,\varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}\right) \tag{14.10}$$

$$\;=\; -\mu_0\,\varepsilon_0\,\frac{\partial}{\partial t}\left(\frac{\partial \boldsymbol{E}}{\partial t}\right) \tag{14.11}$$

$$\;=\; -\mu_0\,\varepsilon_0\,\frac{\partial \boldsymbol{E}^2}{\partial t^2} \tag{14.12}$$

We are finished with the right-hand side. It has the same form as the general wave equation 14.5. Now we have to replace the double cross product on the left-hand side with the relation 14.6:

$$\nabla\left(\nabla \cdot \boldsymbol{E}\right) \;-\; \nabla^2 \boldsymbol{E} \;=\; -\mu_0\,\varepsilon_0\,\frac{\partial \boldsymbol{E}^2}{\partial t^2} \tag{14.13}$$

On the left-hand side is the divergence $\nabla \cdot \boldsymbol{E}$ of the electric field. According to

Maxwell's first equation 14.2, the divergence of the electric field in charge-free space is always zero. This simplifies 14.13 to **wave equation for the electric field component of an electromagnetic wave**:

$$\nabla^2 \boldsymbol{E} \;=\; \mu_0 \varepsilon_0 \frac{\partial \boldsymbol{E}^2}{\partial t^2} \tag{14.14}$$

The wave equation **thus links spatial derivatives** $\nabla^2 \boldsymbol{E}$ of the electric field with the time derivatives $\frac{\partial \boldsymbol{E}^2}{\partial t^2}$ and thus represents a system of **three partial differential equations**.

If we compare the wave equation 14.14 for the electric field with the general form 14.5 of a wave equation, we find out how the **propagation velocity** v_p **is related to the two field constants** μ_0 **and** ε_0:

$$\frac{1}{{v_\mathrm{p}}^2} \;=\; \mu_0\, \varepsilon_0 \quad \leftrightarrow \quad v_\mathrm{p} \;=\; \frac{1}{\sqrt{\mu_0\, \varepsilon_0}} \tag{14.15}$$

If we specifically calculate the propagation velocity of an electromagnetic wave, we get the **velocity of light** c:

$$v_\mathrm{p} = \frac{1}{\sqrt{\mu_0\, \varepsilon_0}} \;=\; 3 \times 10^8\, \frac{\mathrm{m}}{\mathrm{s}} \;=\; c \tag{14.16}$$

From Maxwell's equations and the derived wave equation for the $\boldsymbol{E}$ field, we can conclude that the **electric field component of the electromagnetic wave propagates at the speed of light**. We will see that it also applies to the $\boldsymbol{B}$ field component. We can therefore express the $\boldsymbol{E}$ **wave equation with the speed of light**:

$$\nabla^2 \boldsymbol{E} \;=\; \frac{1}{c^2} \frac{\partial \boldsymbol{E}^2}{\partial t^2} \tag{14.17}$$

14.2 Wave equation for the $\boldsymbol{B}$ field

To derive the wave equation for the **magnetic field** $\boldsymbol{B}$, we have to decouple the fourth Maxwell equation 14.4. Decoupling is done in the same way as we

did with the $\boldsymbol{E}$ field.

Apply the curl operator with cross product $\nabla\times$ on both sides of the fourth current-free Maxwell equation:

$$\nabla \times \nabla \times \boldsymbol{B} \;=\; \nabla \times \left(\mu_0\,\varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}\right) \tag{14.18}$$

Now let's move the time derivative and the two constants on the right-hand side in front of the Nabla operator:

$$\nabla \times \nabla \times \boldsymbol{B} \;=\; \mu_0\,\varepsilon_0 \frac{\partial}{\partial t}\left(\nabla \times \boldsymbol{E}\right) \tag{14.19}$$

Now we can replace the curl $\nabla \times \boldsymbol{E}$ of the electric field with the third Maxwell equation:

$$\nabla \times \nabla \times \boldsymbol{B} \;=\; \mu_0\varepsilon_0 \frac{\partial}{\partial t}\left(-\frac{\partial \boldsymbol{B}}{\partial t}\right) \tag{14.20}$$

We have now decoupled the third Maxwell equation. The time derivative on the right-hand side is combined and the double cross product on the left-hand side is replaced using the calculation rule 14.6:

$$\nabla\left(\nabla \cdot \boldsymbol{B}\right) \;-\; \nabla^2 \boldsymbol{B} \;=\; -\mu_0\,\varepsilon_0\,\frac{\partial \boldsymbol{B}^2}{\partial t^2} \tag{14.21}$$

The divergence $\nabla \cdot \boldsymbol{B} \;=\; 0$ disappears according to the second Maxwell equation (there are no magnetic monopoles). The term $\nabla\left(\nabla \cdot \boldsymbol{B}\right)$ therefore disappears and what remains is the **wave equation for the magnetic field of an electromagnetic wave**:

$$\nabla^2 \boldsymbol{B} \;=\; \mu_0\,\varepsilon_0\,\frac{\partial \boldsymbol{B}^2}{\partial t^2} \tag{14.22}$$

In order to find concrete electromagnetic waves (spherical waves emanating from a radio tower, for example), we have to solve these wave equations for

certain initial or boundary conditions. Mathematicians or Python can do this for us. It is important that you now know **how to get wave equations from Maxwell's equations** and that their **solution describes electromagnetic waves that propagate at the speed of light**.

14.3 A few hints

The derived wave equations 14.14 and 14.22 are **partial differential equations of the second order**. For example, if we look at the wave equation 14.14 for the $\boldsymbol{E}$-field, then, strictly speaking, these are **three** partial differential equations:

$$\nabla^2 \boldsymbol{E} \;=\; \frac{1}{c^2}\frac{\partial \boldsymbol{E}^2}{\partial t^2}$$

Why? Because the $\boldsymbol{E}$ field is a **vector with three components** E_1, E_2 and E_3. Let's write out this wave equation to understand what I mean:

$$\begin{bmatrix} \frac{\partial^2 E_1}{\partial x^2} + \frac{\partial^2 E_1}{\partial y^2} + \frac{\partial^2 E_1}{\partial z^2} \\ \frac{\partial^2 E_2}{\partial x^2} + \frac{\partial^2 E_2}{\partial y^2} + \frac{\partial^2 E_2}{\partial z^2} \\ \frac{\partial^2 E_3}{\partial x^2} + \frac{\partial^2 E_3}{\partial y^2} + \frac{\partial^2 E_3}{\partial z^2} \end{bmatrix} \;=\; \frac{1}{c^2}\begin{bmatrix} \frac{\partial^2 E_1}{\partial t^2} \\ \frac{\partial^2 E_2}{\partial t^2} \\ \frac{\partial^2 E_3}{\partial t^2} \end{bmatrix} \tag{14.23}$$

There are **three** differential equations for the $\boldsymbol{E}$ field that you have to solve. Fortunately, they are **not coupled** and can therefore be solved independently of each other. Physically, non-coupled differential equations mean:: The three **field components** E_1, E_2 and E_3 oscillate **independently of each other**. They **do not interfere with each other**!

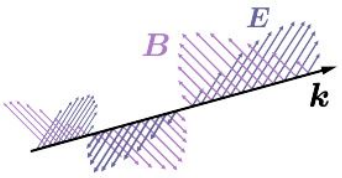
The **solution** $\boldsymbol{E}(t, x, y, z)$ of the wave equation 14.17 is an electric wave, but it does **not necessarily represent the $\boldsymbol{E}$-field of an electromagnetic wave**, only because $\boldsymbol{E}(t, x, y, z)$ solves the wave equation. The solution $\boldsymbol{E}(t, x, y, z)$ only describes the $\boldsymbol{E}$ field of an electromagnetic wave in a vacuum if the **solution also satisfies all four Maxwell equations**.

From the fourth, current-free Maxwell equation 14.4, for example, we can directly read off the orientation of the $\boldsymbol{E}$ and $\boldsymbol{B}$ field components. Here is the

Maxwell equation again:

$$\nabla \times \boldsymbol{B} \;=\; \mu_0\,\varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}$$

We know from mathematics that the result vector $\nabla \times \boldsymbol{B}$ of the cross product is always orthogonal to the vectors between which the cross product is formed. In this case, the $\boldsymbol{B}$ field vector is therefore orthogonal to the derivative of the $\boldsymbol{E}$ field vector. However, the time derivative does not change the direction of a vector. The $\boldsymbol{E}$ field vector and its derivative therefore point in the same direction. Thus **the solutions** $\boldsymbol{E}(t, x, y, z)$ and $\boldsymbol{B}(t, x, y, z)$ of the wave equation **are perpendicular to each other** at any time and at any place.



Now you should have an intuitive understanding of electromagnetic waves and the associated wave equations!

IV The Quantum World

15. Schrödinger equation

More: en.fufaev.org/schrodinger-equation

Most phenomena in our everyday world can be described using classical mechanics. The goal of classical mechanics is to find out how a **body moves over time**. Classical mechanics therefore determines the **trajectory** $\boldsymbol{r}(t)$, that is, the path of this body. With the trajectory, we can predict where this body was, is and will be at any time t. We thus describe the movement of the body.

Here are some examples of the motion of bodies whose trajectory $\boldsymbol{r}(t)$ can be predicted using classical mechanics:

- Movement of our earth around the sun
- Movement of a satellite around the Earth
- Motion of a rocket
- Motion of a swinging pendulum
- Motion of a thrown stone

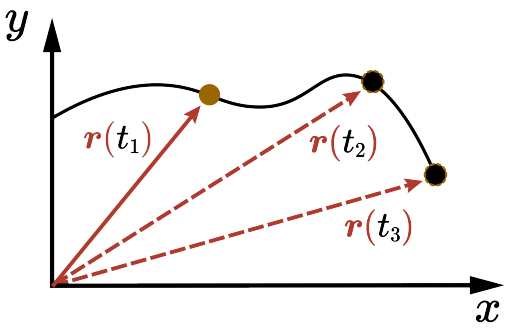
These are all **classical problems** that can be solved **with the help of**

Newton's second law of motion, that is, with the following differential equation:

$$m\,\boldsymbol{a} \;=\; \boldsymbol{F} \quad\leftrightarrow\quad m\frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2} \;=\; -\nabla W_{\mathrm{pot}} \tag{15.1}$$

Here, W_{pot} is the potential energy of a body of mass m. For example, this could be the potential energy in the Earth's gravitational field.

By **solving the Newton differential equation** 15.1 we can find the **unknown trajectory $\boldsymbol{r}(t)$ of a body**. The solution is a position vector $\boldsymbol{r}(t) = [x(t), y(t), z(t)]$, which specifies the three-dimensional position of the body at any time t.



Once we have determined the trajectory $\boldsymbol{r}(t)$ by solving the differential equation, we can extract **all other physical quantities**. Here are a few examples of these quantities:

- **Velocity** of the body: $\boldsymbol{v}(t) = \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}$
- **Momentum** of the body: $\boldsymbol{p}(t) = m\,\boldsymbol{v}(t)$
- **Kinetic energy** of the body: $W_{\mathrm{kin}}(t) = \frac{1}{2}m\,\boldsymbol{v}^2$

In order to be able to specify the solution $\boldsymbol{r}(t)$, the **initial conditions** that characterize the problem to be solved must also be **known**. In classical physics, these are the **initial position** $\boldsymbol{r}(t_0)$ and the **initial velocity** $\boldsymbol{v}(t_0)$ of the body.

In quantum mechanics, on the other hand, it would not even be possible to

specify a **initial position and initial velocity due to the Heisenberg uncertainty principle**. However, the procedure with Newton's second law and the determination of the trajectory $\boldsymbol{r}(t)$ is not possible for a quantum mechanical particle. This is because a quantum particle (such as an electron) **behaves like a wave** under many conditions. The position $\boldsymbol{r}(t)$ of an electron cannot be determined precisely due to this wave character, because **a wave is not concentrated at a single point**. And, if we try to squeeze the wave to a fixed point, then, according to the Heisenberg uncertainty principle, the momentum $\boldsymbol{p}(t)$ of an electron can no longer be determined precisely.

So we cannot determine the trajectory of a quantum particle as in classical mechanics and then deduce all other physical quantities from this, but **must find another way** to describe a quantum particle. And this other way is the development of quantum mechanics and the Schrödinger equation.

It was only through the novel approach to nature with the help of the Schrödinger equation that humans succeeded in making part of the microcosm controllable. This has enabled humans to build **lasers**, which are now indispensable in medicine and research. Or **scanning tunneling microscopes**, which significantly exceed the resolution of conventional light microscopes. It was only through the Schrödinger equation that the **periodic table of elements** and **nuclear fusion in our sun** were precisely understood. But this is only a fraction of the applications that the Schrödinger equation and quantum mechanics have brought us. So let's get to know this powerful equation a little better.

Take a look at the following **time-dependent Schrödinger equation in one spatial dimension** and let it sink in:

$$\mathrm{i}\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial x^2} + W_{\mathrm{pot}} \Psi \tag{15.2}$$

We can already state mathematically that the Schrödinger equation is a **partial differential equation of second order**:

- The Schördinger equation is a **differential equation**. The **unknown**

quantity is a function and derivatives of this function occur in the equation. The unknown function in the Schrödinger equation is the **wave function** $\Psi(x,t)$. It depends on the **spatial coordinate** x and the **time** t and describes a quantum mechanical particle with mass m and potential energy W_{pot}. Note that in the Schrödinger equation x specifies a spatial coordinate and not, as in classical mechanics the (one-dimensional) trajectory $x(t)$ of the position vector $\boldsymbol{r}(t) = [x(t), y(t), z(t)]$.

- The Schördinger equation is a **partial** differential equation. It therefore contains derivatives of Ψ with respect to different variables, namely the **derivative with respect to the spatial variable** x and the **derivative with respect to the time** t.

- The Schördinger equation is a partial differential equation **of second order**. By "second order" it is meant that the **maximum derivative** that occurs in the differential equation is **the second order derivative**. The wave function is differentiated twice with respect to the spatial coordinate in the Schrödinger equation.

As is the case with any differential equation, our goal is to solve the Schrödinger equation to find the desired wave function Ψ and then apply the initial conditions for a specific quantum mechanical problem (e.g. an electron in a potential well).

However, there is **no general recipe** for how to solve the Schrödinger differential equation for a given problem. Most quantum problems cannot even be solved analytically (exactly), but **require approximate methods or numerical solutions** using a computer.

15.1 Time-Independent Schrödinger Equation

Unfortunately, it is not possible to derive the Schrödinger equation from classical mechanics alone. We still need the **wave-particle duality**, which does not occur within classical mechanics. In the following, let us motivate the Schrödinger equation and thus understand the fundamental principles behind it.

We make our lives easier by looking at a **one-dimensional** movement. In one dimension, a quantum particle can only move along a straight line, namely along the local axis x.

15.1.1 Energy conservation

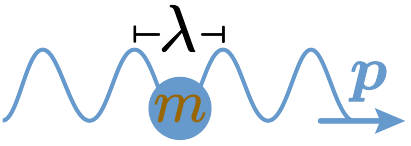
Let us now take a particle of **mass** m, which flies with a **velocity** v in x-space direction. The particle therefore has a **kinetic energy** W_{kin}. It can also be in a conservative (i.e. energy-conserving) field, for example in a gravitational field or in the electric field of a plate capacitor. The particle can therefore also have a **potential energy** W_{pot}. The **total energy** W of the particle is made up of the kinetic and potential energy and is constant in time (meaning that the total energy remains constant):

$$W = W_{\text{kin}} + W_{\text{pot}} \tag{15.3}$$

You should be familiar with the total energy and its conservation over time. You already know this from the basics of classical mechanics. The **law of conservation of energy** is a fundamental principle of physics, which is **also fulfilled in quantum mechanics** in a modified form in conservative fields.

15.1.2 Wave-Particle Duality

The peculiarity of quantum mechanics is added by the wave-particle duality. This allows us to view the **particle as a matter wave**.



The wave-particle duality links the **"particle-like" momentum** p with the

"wave-like" quantity, namely with the **de-Broglie wavelength** λ:

$$\lambda = \frac{h}{p} \leftrightarrow p = \frac{h}{\lambda} \tag{15.4}$$

The two quantities are linked by the **Planck's constant** $h = 6.6 \cdot 10^{-34}$ Js with each other. Because of the **tiny value** of h, it is understandable why we **do not observe wave-particle duality in our macroscopic everyday life**.

In theoretical physics, it is common to express the momentum 15.4 not with the de-Broglie wavelength λ, but with the **wavenumber** k. The momentum looks like this:

$$p = \frac{h\,k}{2\pi} = \hbar\,k \tag{15.5}$$

Here $\hbar = \frac{h}{2\pi}$ is defined as **reduced Planck's constant** and is only used for shorter notation. Whether we define the particle momentum as in Eq. 15.4 or 15.5 is purely a matter of taste. We simply stick to the usual representation 15.5 in theoretical physics.

The momentum 15.5 is also a measure of whether the particle behaves more particle-like or wave-like:

- **The smaller the wavenumber** k (that is, the greater the de Broglie wavelength λ), the more likely the particle behaves quantum mechanically - more like a matter wave. In this case, we speak of a **quantum mechanical particle**.
- **The larger the wavenumber** k (that is, the smaller the de Broglie wavelength λ), the more likely the particle behaves classically - like a real particle. In this case, we speak of a **classical particle**.

The particle has a small wave number (in other words a large de Broglie wavelength) if it has a very small momentum p. So a small mass m and small velocity v. A perfect candidate for such a quantum mechanical particle is a **free electron**. By "free" we mean that it is not in an external field. The

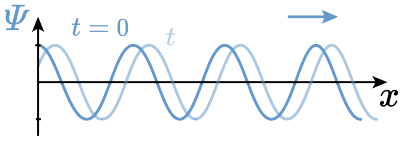
electron behaves like an extended matter wave, which we can describe mathematically with a **plane wave**. We denote the **plane matter wave** here with the capital Greek letter $\Psi(x,t)$. A plane matter wave generally depends on the spatial coordinate x and the time t.

15.1.3 Plane wave

We can describe a plane wave that has the **wavenumber** k, **(angular) frequency** ω and **amplitude** A by a cosine function (or sine function):

$$\Psi(x,t) = A\cos(k\,x - \omega\,t) \tag{15.6}$$

As time t progresses, the matter wave moves in the positive x direction, just like the electron we are looking at.



In order to perform calculations with such plane waves without any addition theorems, we convert the plane wave **into a complex exponential function**. This is an equivalent but extremely effective representation of the plane wave.

Als erstes: Addiere zur Cosinusfunktion die komplexe Sinusfunktion $\mathrm{i}\,A\,\sin(k\,x - \omega\,t)$:

$$\Psi(x,t) = A\cos(k\,x - \omega\,t) + \mathrm{i}\,A\,\sin(k\,x - \omega\,t) \tag{15.7}$$

$$= A\left[\cos(k\,x - \omega\,t) + \mathrm{i}\sin(k\,x - \omega\,t)\right] \tag{15.8}$$

We have thus converted a real function 15.6 into a complex function 15.8. Here, the **imaginary unit** i ensures that the plane matter wave becomes complex-valued and we can immediately represent it as a compact exponential function. The cosine term is the **real part** $\mathrm{Re}(\Psi)$ and the sine term is the **imaginary part** $\mathrm{Im}(\Psi)$ of the complex-valued function Ψ. The good thing is that we can

exploit the enormous advantages of the complex notation 15.8 and then agree that we are only interested in the real part (the cosine term) in the experiment. We can then simply ignore the imaginary part.

However, remember that a complex plane wave 15.8 is also a **possible solution** of the Schrödinger equation. Most solutions $\Psi(x,t)$ of the Schrödinger equation are **complex-valued wave functions**. Real-valued wave functions, as in Eq. 15.6, are then only a special case.

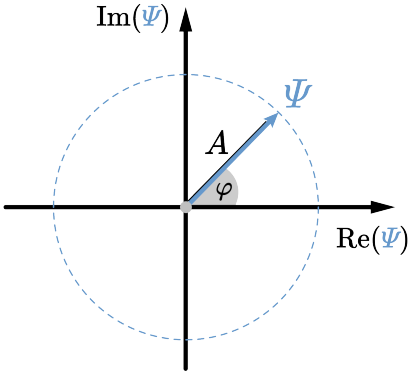
Next, we use the **Euler relation** $e^{i\varphi} = \cos(\varphi) + i\sin(\varphi)$ from mathematics, which links the **complex exponential function** with cosine and sine. In our case, $\varphi = k\,x + \omega\,t$. Let's use this to rewrite our complex plane wave:

$$\Psi(x,t) \;=\; A\,e^{i(k\,x-\omega\,t)} \tag{15.9}$$

Whenever you encounter such a complex exponential function 15.9, you know immediately that it always describes a plane wave - in this case a matter wave. Our original, real-valued plane wave 15.6 as a cosine function is **contained in the complex exponential function** 15.9, namely as the real part $\mathrm{Re}(\Psi)$ of the wave function.

15.1.3.1 Plane wave in a complex plane

Such a complex-valued wave function 15.9, at a fixed coordinate x, can be represented **in the complex number plane as an arrow** Ψ (a complex vector).



- The **amplitude** A corresponds to the **length of the arrow**.
- The **argument** $k\,x + \omega\,t$ corresponds to the **phase angle** φ, which is enclosed between the real axis and the Ψ pointer. As the angle changes with time t, the arrow rotates clockwise. This rotation represents the temporal propagation of the plane wave along the x-local axis.

The complex exponential function 15.9 is a **function** that describes a plane **wave**. This is why it is also called a **wave function** $\Psi(x,t)$, especially in the context of quantum mechanics. Sometimes we also say: The particle is **in the state** Ψ. By this we mean its equivalent representation as an infinite-dimensional vector (see chapter 16 on the Bra-Ket Notation).

There are, of course, a wide variety of wave functions that describe a wide variety of particles under a wide variety of conditions. The **plane** wave is only **a simple example** of a possible wave function.

Next, we multiply the total energy 15.3 by the wave function 15.9. In this way, we **combine the law of conservation of energy and the wave-particle duality** in one equation:

$$W\,\Psi \;=\; W_{\text{kin}}\,\Psi \;+\; W_{\text{pot}}\,\Psi \tag{15.10}$$

15.1.4 Wave equation

But this equation doesn't help us much yet. We still have to **convert it into a differential equation**. We regularly encounter a plane wave 15.9 in optics and electrodynamics when describing electromagnetic waves (see chapter 14). And from there we know that it is a possible solution to the following (one-dimensional) wave equation:

$$\frac{\partial^2 \Psi}{\partial x^2} \;=\; \frac{1}{c^2}\,\frac{\partial^2 \Psi}{\partial t^2} \tag{15.11}$$

In our case, $c = \omega/k$ is the **phase velocity of the matter wave**. On the left-hand side of the wave equation is the second derivative of the wave function with respect to the space x. It is therefore reasonable to differentiate the plane

wave function 15.9 twice with respect to x:

$$\frac{\partial^2 \Psi}{\partial x^2} = \frac{\partial^2}{\partial x^2}\left(A\,\mathrm{e}^{\mathrm{i}(k\,x-\omega\,t)}\right) = -k^2\,A\,\mathrm{e}^{\mathrm{i}(k\,x-\omega\,t)} \tag{15.12}$$

By taking the derivative twice, we get $\mathrm{i}^2 k^2$ as a factor and, because of $\mathrm{i}^2 = -1$, a minus sign. The wave function as an exponential function remains unchanged when differentiated - as you hopefully know. The second derivative therefore results in:

$$\frac{\partial^2 \Psi}{\partial x^2} = -k^2\,\Psi \tag{15.13}$$

Next, we carry out four seemingly arbitrary steps that will ultimately lead us to the Schrödinger equation. In these steps, we want to link the second derivative 15.13 of the wave function with the constant total energy W of the quantum particle:

- Let's use the de Broglie relation $p = \hbar\,k$ and replace k^2 in the second derivative 15.13:

$$\frac{\partial^2 \Psi}{\partial x^2} = -\frac{p^2}{\hbar^2}\,\Psi \tag{15.14}$$

- Next, we bring the kinetic energy expressed by the momentum $W_{\mathrm{kin}} = \frac{p^2}{2m}$ into play by substituting p^2 into Eq. 15.14 with $p^2 = 2mW_{\mathrm{kin}}$:

$$\frac{\partial^2 \Psi}{\partial x^2} = -\frac{2m}{\hbar^2}\,W_{\mathrm{kin}}\,\Psi \tag{15.15}$$

- If we now look at the total energy 15.10 multiplied by the wave function, we see that $W_{\mathrm{kin}}\,\Psi$ occurs there. We therefore rearrange 15.15 for $W_{\mathrm{kin}}\,\Psi$:

$$W_{\mathrm{kin}}\,\Psi = -\frac{\hbar^2}{2m}\,\frac{\partial^2 \Psi}{\partial x^2} \tag{15.16}$$

- If we only insert Eq. 15.16 into the total energy 15.10 multiplied by the

wave function, we get the **time-independent Schrödinger equation in one space dimension**:

$$W\,\Psi \;=\; -\frac{\hbar^2}{2m}\,\frac{\partial^2\,\Psi}{\partial x^2} \;+\; W_{\text{pot}}\,\Psi \tag{15.17}$$

We recognize the **one-dimensionality** of the Schrödinger equation 15.17 by the fact that only the **differentiation with respect to a single spatial coordinate** x occurs here. And we can recognize the **time-independence** of the Schrödinger equation by the fact that it contains **a constant total energy** W. However, the wave function $\Psi(x,t)$ in the time-independent Schrödinger equation may of course be time-dependent!

Let's summarize: To derive the time-independent Schrödinger equation refeq:timeindependent-schrodinger-equation, we needed two fundamental principles, **the law of conservation of energy** 15.3 and the **wave-particle-dualism** 15.4, which we introduced with the help of a plane matter wave 15.9. And since we started from the law of conservation of energy, the time-independent Schrödinger equation is also referred to as **law of conservation of energy in quantum mechanics**.

15.2 Interpretation of the wave function

Let's assume that we have solved the Schrödinger equation and thus found a concrete wave function. How exactly we did this doesn't matter at first. The wave function we have found can also be complex-valued. We should not neglect the imaginary part, as we agreed at the beginning with our plane wave. By omitting the imaginary part, the result of the Schrödinger equation would no longer agree with the results of experiments. For an experimenter, however, such **complex functions are unfavorable** because they are **not directly measurable**. In addition, **there is no direct interpretation of the wave function yet**. But how can we still use the calculated wave function in the experiment, even though the complex wave function cannot be measured directly?

This is where the **statistical interpretation of the wave function**, the so-

called **Copenhagen interpretation**, comes into play. Although it does not say what the wave function $\Psi(x,t)$ means, it **interprets its magnitude squared** $|\Psi(x,t)|^2$. By forming the absolute value square, we obtain a **real-valued** (measurable for the experimenter) function $|\Psi|^2$.

The statistical interpretation makes use of the mathematical fact that the **square of magnitude is always positive** $|\Psi|^2 > 0$ and interprets it as **probability density**. Because as you know: probabilities are always positive.

- In the one-dimensional case, the magnitude squared $|\Psi|^2$ is a **probability per length**.

- In the three-dimensional case, the magnitude squared $|\Psi|^2$ is a **probability per volume**.

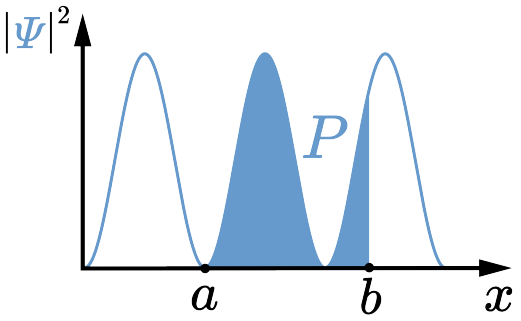
15.2.1 Probability

Let's stick to the simple one-dimensional case. If we integrate the probability density $|\Psi(x,t)|^2$ over the spatial coordinate x within the distance between the points $x = a$ and $x = b$, then we get a **probability** $P(t)$:

$$P(t) = \int_a^b |\Psi(x,t)|^2 \,\mathrm{d}x \tag{15.18}$$

The integral of the probability density $|\Psi(x,t)|^2$ indicates **with which probability** $P(t)$ **the particle is located in the region between** a **and** b **at time** t. The probability can generally change over time.

15.2.2 $|\Psi|^2$ graphically



If we plot the magnitude square $|\Psi(x,t)|^2$ as a function of the location x **in a diagram**, then we can extract the following information from it:

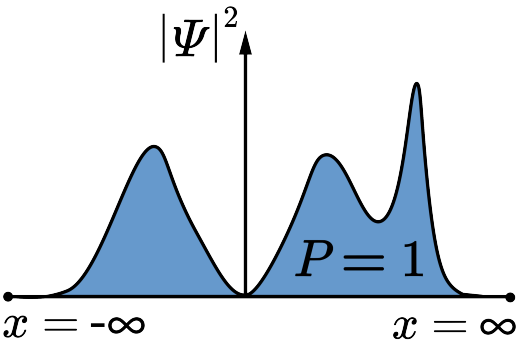
- The probability $P(t)$ at time t is the **area under the $|\Psi|^2$-curve**.
- It is most **likely** to find the particle at the **maxima** of the $|\Psi|^2$ curve.
- The most **unlikely** is to find the particle at the **minima** of the $|\Psi|^2$ curve.

Note, however, that it is not possible to specify the probability $P(t)$ of the particle **at a certain location** (for example at $x = a$), but **only for a spatial region** (here between $x = a$ and $x = b$). In the case of a single point in space, the integral 15.18 would be zero. That is the mathematical reason. The physical reason why we cannot specify a probability for a single point is that there are infinitely many points in space in the region between a and b. If each of these points in space were assigned a **finite** probability, then the sum (i.e. the integral 15.18) of all probabilities would be infinite, which would make no sense at all. Therefore, we always calculate the probability of being **in a spatial region**.

15.3 Normalization of the wave function

In order for the statistical interpretation to be compatible with the Schrödinger equation, the solution of the Schrödinger equation, that is, the wave function Ψ, must satisfy the **normalization condition**. This states that the particle must **exist** somewhere in space. In the one-dimensional case, it must therefore

be found one hundred percent somewhere on the line between $x = -\infty$ and $x = \infty$.



In other words: The normalization condition states that the integral 15.18 for the **probability, integrated over the entire space, must always result in 1**:

$$P = \int_{-\infty}^{\infty} |\Psi(x,t)|^2 \, \mathrm{d}x = 1 \tag{15.19}$$

The normalization condition 15.19 is a **necessary condition that every physically possible wave function must fulfil**. After solving the Schrödinger equation, the wave function $\Psi(x,t)$ **must be normalized with the help of the normalization condition**. "Normalization" means that we have to calculate the integral 15.19 and then choose the amplitude of the wave function so that the normalization condition is fulfilled.

It can be proven that the normalized wave function **remains normalized for all times** t. If this were not the case, then the Schrödinger equation and the statistical interpretation would be incompatible. There are of course solutions to the Schrödinger equation, such as $\Psi(x,t) = 0$, which are **not normalizable**. **Such solutions are unphysical** and we ignore them in quantum mechanics. Wave functions that can be normalized with Eq. 15.19 are called **square-integrable functions** in mathematics. You will definitely come across this term in your studies.

If you know with one hundred percent probability that the particle is located

between $x = a$ and $x = b$, then you may reduce the **integration limits in the normalization condition 15.19 to this spatial region** (this can sometimes be useful to solve the integral):

$$\int_a^b |\Psi(x,t)|^2 \, \mathrm{d}x \;=\; 1 \tag{15.20}$$

15.3.1 Example: Normalizing a wave function

An electron moves from the negative electrode to the positive electrode of a plate capacitor. The two electrodes are at a distance of d from each other. You have determined the following wave function by solving the Schrödinger equation:

$$\Psi(x,t) \;=\; A\,\mathrm{e}^{\mathrm{i}\,(kx-\omega t)} \tag{15.21}$$

Our goal is to determine the factor A so that the integral over the magnitude squared of this wave function is one.

You know with one hundred percent probability that the electron must be between the two electrodes. If we place the negative electrode at $x = 0$ and the positive electrode at $x = d$, then the electron is somewhere between these two points. The normalization condition becomes:

$$\int_0^d |\Psi(x,t)|^2 \, \mathrm{d}x \;=\; 1 \tag{15.22}$$

Next, we need to determine the magnitude squared $|\Psi(x,t)|^2$. The magnitude of the wave function is calculated in the same way as the magnitude of a vector. This is where the power of the complex exponential function becomes apparent for the first time. The following always applies: $|\mathrm{e}^{\mathrm{i}\varphi}| = 1$. The magnitude squared is therefore given by :

$$\begin{aligned} |\Psi(x,t)|^2 \;&=\; |A\,\mathrm{e}^{\mathrm{i}\,(kx-\omega t)}|^2 \\ &=\; A^2\,|\mathrm{e}^{\mathrm{i}\,(kx-\omega t)}|^2 \\ &=\; A^2 \end{aligned} \tag{15.23}$$

Let's insert the calculated magnitude squared into the normalization condition:

$$\int_0^d A^2 \,\mathrm{d}x \;=\; 1 \tag{15.24}$$

The amplitude A is independent of x, so it is a constant and we can place it in front of the integral. And the integral simply results in d:

$$\begin{aligned} A^2 \int_0^d 1 \,\mathrm{d}x \;&=\; 1 \\ A^2\, d \;&=\; 1 \\ A \;&=\; \frac{1}{\sqrt{d}} \end{aligned} \tag{15.25}$$

The **normalized wave function** for the electron is therefore:

$$\Psi(x,t) \;=\; \frac{1}{\sqrt{d}}\, \mathrm{e}^{\mathrm{i}\,(kx-\omega t)} \tag{15.26}$$

Once we have normalized the wave function of a quantum particle, we can not only find out the probability $P(t)$ of a particle, but also the **mean value $\langle x \rangle$ of the position** and the mean value of many other observables (physical quantities). For example, the mean value of the momentum $\langle p \rangle$, the velocity $\langle v \rangle$ or the kinetic energy $\langle W_{\mathrm{kin}} \rangle$ of a quantum particle.

15.4 Three-dimensional Schrödinger equation

In your physics course, you will not only encounter a one-dimensional Schrödinger equation, but also a two- or three-dimensional version. The **three-dimensional wave function** can depend not only on one spatial coordinate x, but **on three spatial coordinates**: $\Psi(x,y,z,t)$. We can combine the three spatial coordinates into one vector $\boldsymbol{r}$: $\Psi(\boldsymbol{r},t)$.

We can generalize the one-dimensional Schrödinger equation 15.17 to a three-dimensional Schrödinger equation. This is not difficult if you have read the chapter 5 about the Nabla operator. Here is the one-dimensional Schrödinger

equation again:

$$W\ \Psi\ =\ -\frac{\hbar^2}{2m}\frac{\partial^2\,\Psi}{\partial x^2}\ +\ W_{\text{pot}}\ \Psi \tag{15.27}$$

In it, we have to extend the second spatial derivative with respect to x so that the second spatial derivative with respect to y and z also appear in the three-dimensional Schrödinger equation. To do this, we simply **add** these derivatives to the spatial derivative with respect to x. Then we get the **three-dimensional, time-independent Schrödinger equation**:

$$W\ \Psi\ =\ -\frac{\hbar^2}{2m}\left(\frac{\partial^2\,\Psi}{\partial x^2}+\frac{\partial^2\,\Psi}{\partial y^2}+\frac{\partial^2\,\Psi}{\partial z^2}\right)\ +\ W_{\text{pot}}\ \Psi \tag{15.28}$$

We can write Eq. 15.28 a little **more compact with the Nabla operator**. To do this, factor out the wave function from the spatial derivatives:

$$W\ \Psi\ =\ -\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\ \Psi\ +\ W_{\text{pot}}\ \Psi \tag{15.29}$$

The sum of the spatial derivatives in the brackets form a **Laplace operator** $\nabla\cdot\nabla=\nabla^2$ (sometimes also noted as Δ). This operator is the scalar product of two nabla operators. This results in **three-dimensional Schrödinger equation expressed with Nabla operator**:

$$W\ \Psi\ =\ -\frac{\hbar^2}{2m}\nabla^2\ \Psi\ +\ W_{\text{pot}}\ \Psi \tag{15.30}$$

So far, we have only learned about the **time-independent** Schrödinger equation. You will have to use this regularly in the quantum mechanics lecture. For example, in problems such as particle in a potential well, quantum mechanical harmonic oscillator, tunnel effect and helium atom.

15.5 Time-dependent Schrödinger equation

A quantum particle described by the **time-independent** Schrödinger equation has a constant total energy W. We can therefore only use the time-independent Schrödinger equation to describe **quantum particles that do not change their total energy**.

But what if the total energy W of a quantum particle **is not constant in time**? This can happen, for example, if the particle **interacts with its environment** and its total energy increases or decreases as a result. For such a quantum system, we need the **time-dependent Schrödinger equation**. This is what it looks like in one spatial dimension:

$$\mathrm{i}\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi}{\partial x^2} + W_{\mathrm{pot}} \Psi \tag{15.31}$$

The only difference to the time-independent Schrödinger equation is that the **total energy W becomes an operator** $W = \mathrm{i}\hbar \frac{\partial}{\partial t}$. This operator is also called **time evolution operator**.

15.6 Stationary Wave Function

Solving the time-dependent Schrödinger equation 15.31 is not that easy. However, you can simplify the solution of this partial differential equation considerably if you convert it **into two ordinary differential equations**. One differential equation then only depends on the time t and the other only on the spatial coordiante x. We do this separation into two ordinary differential equations with the method »**Separation of Variables**«. This is a very important method in physics to simplify partial differential equations and make them easier to solve.

The only requirement for the Separation of Variables to work is that the **potential energy $W_{\mathrm{pot}}(x)$ does not depend on the time** t (but it may very well depend on the position x). The wave function itself can, of course, still depend on both position and time.

First, divide the time-dependent wave function $\Psi(x,t)$ (that is, the **total solution**) into two parts:

- Into a **partial solution** $\psi(x)$, which only depends on the location x.
- Into a **partial solution** $\phi(t)$, which only depends on the time t.

This **separation ansatz** (ansatz is a German word for approach) turns the total wave function into a product of the two partial solutions:

$$\Psi(x,t) \;=\; \psi(x)\,\phi(t) \tag{15.32}$$

The time-dependent Schrödinger equation thus becomes:

$$\mathrm{i}\hbar\,\frac{\partial}{\partial t}\,(\psi(x)\,\phi(t)) \;=\; -\frac{\hbar^2}{2m}\,\frac{\partial^2}{\partial x^2}\,(\psi(x)\,\phi(t)) \;+\; W_{\mathrm{pot}}\,\psi(x)\,\phi(t) \tag{15.33}$$

Not all wave functions can be separated into two partial solutions as in Eq. 15.32. However, since the Schrödinger equation is **linear**, we can form a linear combination of such solutions and thus obtain **all** wave functions (including those that cannot be separated). This is what makes variable separation so powerful.

As you can see from the time-dependent Schrödinger equation 15.31, the time derivative and the second spatial derivative occur there. Calculate the two derivatives of the separation ansatz 15.32:

- Differentiate the separated wave function 15.32 with respect to time t:

$$\frac{\partial\,\Psi}{\partial t} \;=\; \psi(x)\,\frac{\partial\phi(t)}{\partial t} \tag{15.34}$$

- Differentiate the separated wave function 15.32 once according to the position x:

$$\frac{\partial^2\,\Psi}{\partial x^2} \;=\; \phi(t)\,\frac{\partial^2\psi(x)}{\partial x^2} \tag{15.35}$$

We can insert the time derivative 15.34 and the spatial derivative 15.35 into the time-dependent Schrödinger equation 15.33:

$$\mathrm{i}\hbar\,\psi(x)\,\frac{\partial\phi(t)}{\partial t} \;=\; -\frac{\hbar^2}{2m}\,\phi(t)\,\frac{\partial^2\psi(x)}{\partial x^2} \;+\; W_{\mathrm{pot}}\,\psi(x)\,\phi(t) \tag{15.36}$$

In the following, we omit the position and time dependence in order to be able to write the Schrödinger equation in a more compact form. Now we have to reformulate the separated Schrödinger differential equation 15.36 so that its left-hand side depends only on the time t and its right-hand side only on the position x. We achieve this by dividing Eq. 15.36 by the product $\psi\,\phi$:

$$\mathrm{i}\hbar\,\frac{1}{\phi}\,\frac{\partial\phi}{\partial t} \;=\; -\frac{\hbar^2}{2m}\,\frac{1}{\psi}\,\frac{\partial^2\psi}{\partial x^2} \;+\; W_{\mathrm{pot}} \tag{15.37}$$

What do we get out of it? Quite a lot! If we change the time t (which **only occurs on the left** side), only the left side of the equation will change, while the right side remains unchanged. However, if the right-hand side does not change over time, it is constant. This constant is real, as a complex-valued constant would violate the normalization condition. The right-hand side corresponds to the time-constant total energy W:

$$\mathrm{i}\hbar\,\frac{1}{\phi}\,\frac{\partial\phi}{\partial t} \;=\; W \tag{15.38}$$

This is an ordinary differential equation for the partial solution $\psi(x)$. We can even write down the solution for this differential equation. It is easy to solve with pencil and paper. The **time-dependent partial solution is a plane wave**:

$$\psi(x) \;=\; \mathrm{e}^{\mathrm{i}\frac{W}{\hbar}t} \tag{15.39}$$

Now let's look at the right-hand side of Eq. 15.37. If you change the variable x on the right-hand side, the left-hand side of the equation remains constant because it is independent of x. Because of the equality, the left-hand side must

correspond to the same constant W:

$$W = -\frac{\hbar^2}{2m}\frac{1}{\psi}\frac{\partial^2\psi}{\partial x^2} + W_{\text{pot}} \tag{15.40}$$

If we multiply the differential equation 15.40 by ψ, we get the **stationary, time-independent Schrödinger equation for** ψ:

$$W\,\psi = -\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + W_{\text{pot}}\,\psi \tag{15.41}$$

By »stationary« we mean that the solution $\psi(x)$ does not depend on time. Therefore, we refer to the solution $\psi(x)$ of a stationary Schrödinger equation as **stationary wave function** $\psi(x)$ or as **stationary state**.

What have we achieved overall with the separation approach? Instead of having to solve a more complicated time-dependent Schrödinger equation for $\Psi(x,t) = \psi(x)\,\phi(t)$,

$$\mathrm{i}\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + W_{\text{pot}}\,\Psi \tag{15.42}$$

solve the stationary Schrödinger equation 15.41 for $\psi(x)$ instead and **multiply this position-dependent partial solution with the time-dependent partial solution** 15.39. As a result, we obtain the **total solution of the time-dependent Schrödinger equation**:

$$\Psi(x,t) = \psi(x)\,\mathrm{e}^{\mathrm{i}\frac{W}{\hbar}t} \tag{15.43}$$

The solution 15.43 is very special, because its magnitude squared $|\Psi(x,t)|^2$ is time-independent! All other observables that describe the particle are also time-independent. For example, a quantum particle described by the wave function 15.43 has a constant mean value of the energy $\langle W\rangle$, constant mean value of the momentum $\langle p\rangle$ and constant mean value of all other observables.

15.7 Hamilton operator

You will not only encounter the time-dependent and independent Schrödinger equation in this form:

$$W\ \Psi = -\frac{\hbar^2}{2m}\nabla^2\ \Psi + W_{\text{pot}}\ \Psi \tag{15.44}$$

$$\mathrm{i}\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\ \Psi + W_{\text{pot}}\ \Psi \tag{15.45}$$

If you factor out the wave function, you get:

$$W\ \Psi = \left(-\frac{\hbar^2}{2m}\nabla^2 + W_{\text{pot}}\right)\ \Psi \tag{15.46}$$

$$\mathrm{i}\hbar\frac{\partial\Psi}{\partial t} = \underbrace{\left(-\frac{\hbar^2}{2m}\nabla^2 + W_{\text{pot}}\right)}_{\hat{H}}\ \Psi \tag{15.47}$$

The operator in the brackets is called **Hamilton operator** $\hat{H}$ (sometimes also called **Hamiltonian**):

$$\hat{H} = \underbrace{-\frac{\hbar^2}{2m}\nabla^2}_{W_{\text{kin}}} + W_{\text{pot}} \tag{15.48}$$

The Hamilton operator describes the **total energy of a quantum particle**. You will also regularly encounter the representation of the Schrödinger equation with the Hamilton operator:

$$\hat{H}\ \Psi = W\ \Psi \tag{15.49}$$

$$\hat{H}\ \Psi = \mathrm{i}\hbar\frac{\partial\Psi}{\partial t} \tag{15.50}$$

With the Hamilton operator, we can interpret the time-independent Schrödinger equation as a **eigenvalue equation**. You should know what an eigenvalue equation is from linear algebra. So you apply the Hamilton operator (think of

it as **matrix**) to the **eigenfunction** Ψ (think of it as **eigenvector**). Then you get the eigenvector Ψ on the right-hand side of the Schrödinger equation, which is scaled with the corresponding energy eigenvalue W. The energy eigenvalues W depend on the Hamilton operator used $\hat{H}$ and are **discrete** for most of the Hamilton operators you will encounter in your studies. We say: The energy of the quantum particle is quantized.

Thus we have transferred the problem of **solving the Schrödinger differential equation** 15.47 to the problem of **solving the eigenvalue equation** 15.50.

15.8 What you've learned

Let's summarize what you should have learned from the chapter 15:

- You now know how to motivate the **time-independent Schrödinger equation**.
- You know **what the wave function is** and have become familiar with the plane wave as a simple **example of a wave function**.
- You know the **statistical interpretation** of the wave function.
- You can normalize a **wave function**.
- You know the **difference between the time-dependent and time-independent** Schrödinger equation.
- You know the **difference between the one-dimensional and three-dimensional** Schrödinger equation.
- You know what the **Hamilton operator** is.
- You know what a **stationary wave function** is.

Remember that the Schrödinger equation is a **non-relativistic equation**. It fails for quantum particles moving at almost the speed of light. Furthermore, it does not naturally take into account the spin of a particle. These two problems are only solved by its relativistic version, namely the **Dirac equation**. You will

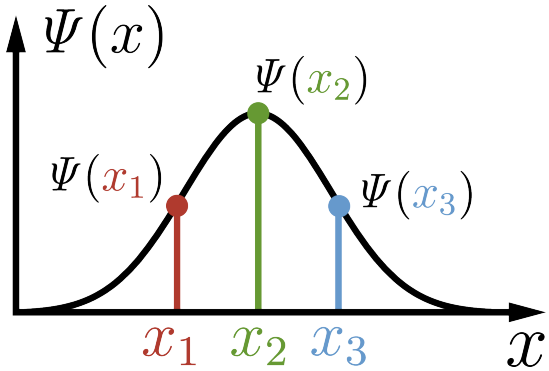
only learn about this in your Master's degree when you take course on quantum field theory.

In the following Chapter 16 you will learn the **representation of the wave function as a state vector** ("quantum mechanical state"). Advantage: You can work with the state vector in (almost) the same way as with the usual vectors that you know from linear algebra.

16. Bra-Ket Notation

More: en.fufaev.org/bra-ket-notation

Consider any **one-dimensional wave function** $\Psi(x)$ describing a quantum mechanical particle. We have omitted its time dependence $\Psi(t, x)$ because it is not relevant in this chapter. The value of the wave function, for instance, at the location x_1 is $\Psi(x_1)$, at the location x_2 the function value is $\Psi(x_2)$, at the location x_3 the function value is $\Psi(x_3)$, and so forth. In this manner you can assign to each point in space x the function value $\Psi(x)$ of the wave function. The sum of all these function values yields the shape of the wave function.



We can represent all these function values as a list of values. We can interpret

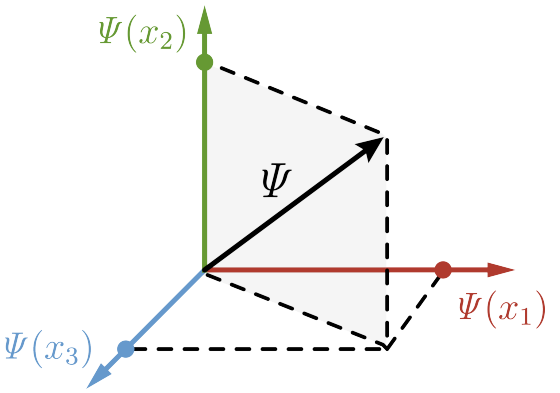
this list of values as a **column vector** Ψ. The column vector then has the following components:

$$\Psi = \begin{bmatrix} \Psi(x_1) \\ \Psi(x_2) \\ \Psi(x_3) \\ \dots \end{bmatrix} = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \dots \end{bmatrix} \tag{16.1}$$

At the second equality sign, we have represented the function values more compactly. Instead of writing the first component as $\Psi(x_1)$, we compactly write it as Ψ_1.

We can illustrate the column vector 16.1 as follows:

- The first component $\Psi(x_1)$ forms the first coordinate axis.
- The second component $\Psi(x_2)$ forms the second coordinate axis.
- The third component $\Psi(x_3)$ forms the third coordinate axis.
- and so forth.



We'll stick to only three components because I can't draw a four-dimensional coordinate system. So, each component is assigned a coordinate axis. In this way, the three components span a **three-dimensional space**. Once we add an additional function value $\Psi(x_4)$, the space becomes **four-dimensional**, and so on. We denote the vector Ψ representing a wave function $\Psi(x)$ as a **state vector**.

In theory, there are of course **infinitely many x-values**. Therefore, there are also **infinitely many associated function values** $\Psi(x)$ as components of the column vector. If there are infinitely many function values, then the space in which the state vector Ψ lives is **infinite-dimensional**. Remember that this space is not an infinitely-dimensional **position space** but an **abstract space**.

This abstract space, where various quantum mechanical state vectors Ψ live, is called a **Hilbert space**. In general, this is an infinite-dimensional vector space. However, it can also be **finite-dimensional**. For example, spin states $\Psi_\uparrow$ and $\Psi_\downarrow$, which describe the spin of a particle, reside in a **two-dimensional Hilbert space**. That is, state vectors like the spin-up state $\Psi_\uparrow$ have **only two components**:

$$\Psi_\uparrow = \begin{bmatrix} \Psi_{\uparrow 1} \\ \Psi_{\uparrow 2} \end{bmatrix} \tag{16.2}$$

However, even **approximating** an infinite-dimensional state with a column vector 16.1 is incredibly useful. In numerical computations, we have no other choice but to approximate the infinite-dimensional state with a finite number of function values. There's simply no other way since your computer would need infinite memory for that. The more components we take in numerical computation, the more accurate the state vector becomes, but the computations become slower and more memory-intensive.

So, we can represent a quantum mechanical particle in two ways:

- as a wave function $\Psi(x)$
- as a state vector Ψ

16.1 Bra- and Ket-State Vectors

16.1.1 Ket vector

To better distinguish the particle's description as a state vector from its description as a wave function, we enclose the state vector Ψ in arrow-like

brackets:

$$|\Psi\rangle = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \dots \end{bmatrix} \tag{16.3}$$

The **wave function** Ψ, represented as a **column vector** 16.3, is called a **ket vector** $|\Psi\rangle$, and the arrow-like bracket points to the right. It doesn't matter what you write **inside** the bracket. For example, you could have also noted the ket vector as $|\Psi(x)\rangle$. The only thing to consider is that the notation inside the bracket clarifies to other readers which quantum mechanical system this ket vector represents.

- So, when you see a **ket** $|\Psi\rangle$, you know that it refers to the representation of the quantum particle **as a state vector**.
- On the other hand, if you see $\Psi(x)$ without ket notation, you know that it refers to the representation of the quantum particle **as a wave function**.

16.1.2 Bra Vector

The vector $|\Psi\rangle^\dagger$, which is the **adjoint** of the ket vector, is called a **bra vector**. The symbol † is pronounced as »Dagger«. For a clever, compact notation, we write the bra vector with a **reversed arrow** $\langle\Psi|$ instead of using the dagger $|\Psi\rangle^\dagger$.

To obtain the bra vector $\langle\Psi|$ adjoint to the ket vector $|\Psi\rangle$, we need to perform two operations:

- **Transpose** the ket vector $|\Psi\rangle$. This turns it into a row vector:

$$|\Psi\rangle^{\mathrm{T}} = [\Psi_1,\ \Psi_2,\ \Psi_3,\ ...] \tag{16.4}$$

- **Complex-conjugate** the transposed ket vector $|\Psi\rangle^{\mathrm{T}}$. This operation

adds asterisks to the components to obtain the bra vector:

$$\langle \Psi | \;=\; [\Psi_1^*,\; \Psi_2^*,\; \Psi_3^*,\; ...] \tag{16.5}$$

In summary: The wave function $\Psi(x)$ corresponds in vector representation to the ket vector $|\Psi\rangle$, and the row vector adjoint to the ket vector, denoted as $\langle \Psi|$, is the bra vector.

Since we've interpreted the wave function $\Psi(x)$ as a ket vector $|\Psi\rangle$, we can practically work with the ket vector in much the same way as with ordinary vectors you're familiar with from linear algebra. For example, we can form a **scalar product or tensor product between the bra or ket vectors**. What may be new to you, however, is that unlike vectors from linear algebra, the **components of the ket vector can be complex**, and the **number of components can be infinite**.

16.2 Scalar and Inner Product

We can form the **scalar product** $\langle \Phi | \cdot | \Psi \rangle$ between a bra vector $\langle \Phi |$ and a ket vector $|\Psi\rangle$. Here, we don't need to write the scalar product dot and can omit a vertical line. We write $\langle \Phi | \Psi \rangle$ instead of $\langle \Phi | \cdot | \Psi \rangle$ for brevity.

When the state vectors between which you form the scalar product live in an **infinite-dimensional Hilbert space**, we call this operation not a scalar product but an **inner product**. However, the notation $\langle \Phi | \Psi \rangle$ for the inner product remains the same as in the case of the scalar product.

In a finite n**-dimensional Hilbert space**, the written out scalar product $\langle \Phi | \Psi \rangle$ between any bra vector $\langle \Phi |$ and ket vector $|\Psi\rangle$ looks like this:

$$\langle \Phi | \Psi \rangle \;=\; [{\Phi_1}^*, {\Phi_2}^*, {\Phi_3}^*, ..., {\Phi_n}^*] \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \\ \dots \\ \Psi_n \end{bmatrix} \tag{16.6}$$

We can multiply the row and column vectors in 16.6 just as we do with the usual matrix multiplication.:

$$\langle \Phi | \Psi \rangle = \Phi_1^* \Psi_1 + \Phi_2^* \Psi_2 + \Phi_3^* \Psi_3 + ... + \Phi_n^* \Psi_n \tag{16.7}$$

$$= \boxed{+}_{i=1}^{n} \Phi_i^* \Psi_i \tag{16.8}$$

In the last step, we abbreviated the scalar product **using a sum sign**. Here, n represents the **dimension of the Hilbert space**, that is, the **number of components** of a state vector living in this Hilbert space. The dimension $n = \infty$ of the Hilbert space can also be infinite.

16.3 Continuous Quantum States

So far, we have discretized a quantum state $|\Psi\rangle$ by omitting many function values of the wave function $\Psi(x)$. Just between the positions x_1 and x_2 alone, there are **infinitely many more values**. Why? Because the position coordinate is a real number. This means there are infinitely many components between $\Psi_1 = \Psi(x_1)$ and $\Psi_2 = \Psi(x_2)$ that we have omitted in the column vector representation:

$$|\Psi\rangle = \begin{bmatrix} \Psi_1 \\ ... \\ \Psi_2 \\ ... \\ \Psi_3 \\ ... \end{bmatrix} \tag{16.9}$$

This means that representing a wave function $\Psi(x)$ with a **real-valued argument** x as a column vector is only an approximation and serves merely for illustration purposes.

Similarly, the inner product 16.8 with the sum sign is not exact for states with real-valued arguments. How can we make the inner product exact for these

states? We need to **switch to an integral**. Therefore, we replace the sum sign with an integral sign. We now consider the function values Φ_i and Ψ_i not at discrete points x_i but at all points x:

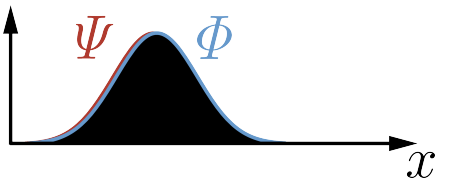
$$\langle \Phi \,|\, \Psi \rangle \;=\; \int \Phi(x)^* \, \Psi(x) \, \mathrm{d}x \tag{16.10}$$

So, to calculate the **exact inner product of two wave functions** $\Phi((x)$ and $\Psi(x)$, we need to evaluate the integral 16.10.

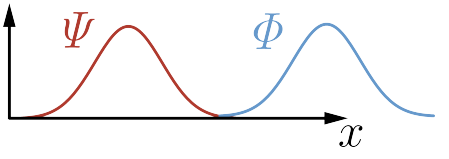
16.3.1 Overlap of Quantum States

What does this inner product of two quantum states actually tell us intuitively? Similar to the scalar product, the inner product is a number that measures **how much two quantum states overlap**. Let's consider two normalized quantum states Φ and Ψ for simplicity:

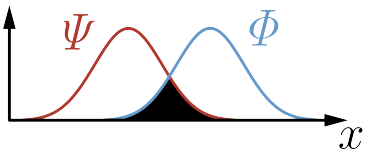
- If the inner product is $\langle \Phi \,|\, \Psi \rangle = 1$, then the corresponding wave functions $\Phi(x)$ and $\Psi(x)$ **completely overlap**.



- If the inner product is $\langle \Phi \,|\, \Psi \rangle = 0$, then the wave functions $\Phi(x)$ and $\Psi(x)$ **do not overlap** at all.



- All values of the inner product $\langle \Phi \,|\, \Psi \rangle$ between 1 and 0 indicate **partial overlap** of the two wave functions.



16.4 Orthonormal Quantum States

Let's consider two **normalized and orthogonal** (orthonormal) states $|\Psi_i\rangle$ and $|\Psi_j\rangle$, denoted by variable indices i and j instead of fixed values. Then, their scalar product $\langle\Psi_i|\Psi_j\rangle$ yields either 0 or 1. Therefore, they are suitable as **basis states**. You should know this property from linear algebra when calculating the scalar product of two basis vectors:

- The scalar product of two **different** orthonormal states, $i \neq j$, yields: $\langle\Psi_i|\Psi_j\rangle = 0$.
- The scalar product of two **identical** orthonormal states, $i = j$, yields: $\langle\Psi_i|\Psi_j\rangle = 1$.

These two cases can be combined **in a single equation using the Kronecker delta** δ_{ij}:

$$\langle\Psi_i|\Psi_j\rangle = \delta_{ij} \tag{16.11}$$

16.5 Tensorproduct in Bra-Ket Notation

Another important operation between a bra and ket vector is the **tensor product**: $|\Phi\rangle \otimes \langle\Psi|$. We can omit the **tensor symbol** $\otimes$, because it is immediately clear from the bra-ket notation that it is not a scalar or inner product: $|\Phi\rangle\langle\Psi|$, since **bra and ket vectors are interchanged** here.

The **result of the tensor product is a matrix**:

- If the states $|\Phi\rangle$ and $|\Psi\rangle$ each have **two components**, then $|\Phi\rangle\langle\Psi|$ is a **2x2 matrix**.
- If the states $|\Phi\rangle$ and $|\Psi\rangle$ each have **three components**, then $|\Phi\rangle\langle\Psi|$ is a **3x3 matrix**.

- If the states $|\Phi\rangle$ and $|\Psi\rangle$ each have n **components**, then $|\Phi\rangle\langle\Psi|$ is an $n \times n$ **matrix**.

As you know from matrix multiplication, in the tensor product, we multiply a ket vector $|\Phi\rangle$, which is a column vector, with a bra vector $\langle\Psi|$, which is a row vector. If the states have three components, then we get a 3x3 matrix:

$$|\Phi\rangle\langle\Psi| = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \end{bmatrix} [\Psi_1^*, \Psi_2^*, \Psi_3^*] = \begin{bmatrix} \Phi_1\,\Psi_1^* & \Phi_1\,\Psi_2^* & \Phi_1\,\Psi_3^* \\ \Phi_2\,\Psi_1^* & \Phi_2\,\Psi_2^* & \Phi_2\,\Psi_3^* \\ \Phi_3\,\Psi_1^* & \Phi_3\,\Psi_2^* & \Phi_3\,\Psi_3^* \end{bmatrix} \tag{16.12}$$

You will encounter such matrices in form of density matrices very often in quantum mechanics, for example, when learning about **quantum entanglement**.

16.6 Projection Matrices

Let's take a **normalized state** $|\Psi\rangle$, meaning the magnitude of this vector is 1, and form the tensor product of this state with itself, then we obtain a **projection matrix** $|\Psi\rangle\langle\Psi|$ (or **projection operator**, if specific components are not considered):

$$|\Psi\rangle\langle\Psi| = \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \end{bmatrix} [\Psi_1^*, \Psi_2^*, \Psi_3^*] = \begin{bmatrix} \Psi_1\,\Psi_1^* & \Psi_1\,\Psi_2^* & \Psi_1\,\Psi_3^* \\ \Psi_2\,\Psi_1^* & \Psi_2\,\Psi_2^* & \Psi_2\,\Psi_3^* \\ \Psi_3\,\Psi_1^* & \Psi_3\,\Psi_2^* & \Psi_3\,\Psi_3^* \end{bmatrix} \tag{16.13}$$

If we apply a projection matrix to any ket vector $|\Phi\rangle$ (which may not be normalized), then we **multiply a matrix** $|\Psi\rangle\langle\Psi|$ **by a column vector** $|\Phi\rangle$:

$$|\Psi\rangle\langle\Psi|\,|\Phi\rangle \;=\; |\Psi\rangle\langle\Psi|\Phi\rangle = \begin{bmatrix} \Psi_1\,\Psi_1^* & \Psi_1\,\Psi_2^* & \Psi_1\,\Psi_3^* \\ \Psi_2\,\Psi_1^* & \Psi_2\,\Psi_2^* & \Psi_2\,\Psi_3^* \\ \Psi_3\,\Psi_1^* & \Psi_3\,\Psi_2^* & \Psi_3\,\Psi_3^* \end{bmatrix} \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \Phi_3 \end{bmatrix} \tag{16.14}$$

The special feature of a projection matrix is that it **projects the state** $|\Phi\rangle$ **onto the state** $|\Psi\rangle$. In simple terms, it yields the part of the quantum state

$|\Phi\rangle$ that overlaps with the quantum state $|\Psi\rangle$. The result of the projection is thus a **ket vector** $|\Psi\rangle\langle\Psi|\Phi\rangle$, **which describes the overlap** of the quantum states $|\Phi\rangle$ and $|\Psi\rangle$.

16.6.1 Basis Transformation with Projection Matrices

Projection matrices are an important tool in theoretical physics for investigating the overlap of quantum states. However, perhaps the most important use of projection matrices is for effortless **basis transformation**. If we have any quantum state $|\Phi\rangle$ and want to **view it from a different perspective**, or mathematically speaking, represent it in a different basis, we first choose the **desired new basis**: $\{|\Psi_i\rangle\}$. As you hopefully know from linear algebra, this is a set of orthonormal vectors $|\Psi_1\rangle$, $|\Psi_2\rangle$, $|\Psi_3\rangle$, and so on. Their **number equals the dimension** of the Hilbert space in which these quantum states live. In quantum mechanics, we refer to the basis vectors as **basis states**. For describing particle spin, for example, we only need two basis states.

For the sake of illustration, let's assume that our desired basis consists of only three basis states: $\{|\Psi_1\rangle, |\Psi_2\rangle, |\Psi_3\rangle\}$. With each of these basis states, we can construct projection matrices: $|\Psi_1\rangle\langle\Psi_1|$, $|\Psi_2\rangle\langle\Psi_2|$, and $|\Psi_3\rangle\langle\Psi_3|$.

To represent a quantum state $|\Phi\rangle$ in this basis, we first form the sum of the projection matrices:

$$\boxed{\sum_{i=1}^{3}} |\Psi_i\rangle\langle\Psi_i| = |\Psi_1\rangle\langle\Psi_1| + |\Psi_2\rangle\langle\Psi_2| + |\Psi_3\rangle\langle\Psi_3| = \mathbb{I} \tag{16.15}$$

As we know from mathematics, the **sum of the projection matrices** forming a basis is an **identity matrix** $\mathbb{I}$. The fact that the sum yields an identity matrix is crucial during basis transformation because we do not want to alter the quantum state $|\Phi\rangle$. Multiplying an identity matrix by a column vector $|\Phi\rangle$ does not change this vector:

$$|\Phi\rangle = \mathbb{I}|\Phi\rangle \tag{16.16}$$

Now, let's substitute the sum of the basis projection matrices 16.15 for the identity matrix:

$$|\Phi\rangle = \boxplus_{i=1}^{3} |\Psi_i\rangle\langle\Psi_i|\Phi\rangle \tag{16.17}$$

$$= (|\Psi_1\rangle\langle\Psi_1| + |\Psi_2\rangle\langle\Psi_2| + |\Psi_3\rangle\langle\Psi_3|)\,|\Phi\rangle \tag{16.18}$$

$$= |\Psi_1\rangle\langle\Psi_1|\Phi\rangle + |\Psi_2\rangle\langle\Psi_2|\Phi\rangle + |\Psi_3\rangle\langle\Psi_3|\Phi\rangle \tag{16.19}$$

$$= |\Phi\rangle \tag{16.20}$$

The resulting state $|\Phi\rangle$, although denoted the same as the original state $|\Phi\rangle$, is now represented in the **new basis** $\{|\Psi_i\rangle\}$. If we want to emphasize the new basis, we can also assign it an index: $|\Phi\rangle_\Psi$. I hope you now understand the usefulness of the concept of projection matrices!

In general, we can represent a quantum state $|\Phi\rangle$ with n components using a **basis** $\{|\Psi_i\rangle\}$ **consisting of** n **basis states** as follows:

$$|\Phi\rangle_\Psi = \boxplus_{i=1}^{n} |\Psi_i\rangle\langle\Psi_i|\Phi\rangle \tag{16.21}$$

The basis change with a finite number of basis states is, of course, **only exact** when the states $|\Phi\rangle$ **live in a finite-dimensional** Hilbert space. For states with infinitely many components, Eq. 16.21 is **only an approximation of the old state in the new basis**. The approximation becomes more accurate as we choose n larger. Thus, in computational physics, we can save memory by not choosing n too large, but large enough to approximate the quantum state good enough in the new basis.

Guess how the basis change for states with infinitely many components can be made **exact**? With an integral! To do this, replace the discrete summation with a sum sign with a **continuous summation with an integral**:

$$|\Phi\rangle_\Psi = \int \mathrm{d}x\, |\Psi\rangle\langle\Psi|\Phi\rangle \tag{16.22}$$

16.7 Schrödinger Equation in Bra-Ket Notation

In Chapter 15.7, you learned about the Schrödinger equation as an eigenvalue equation. Here it is again:

$$\begin{aligned} \hat{H}\ \Psi &= W\ \Psi \\ \hat{H}\ \Psi &= \mathrm{i}\hbar \frac{\partial \Psi}{\partial t} \end{aligned}$$

You will encounter this eigenvalue equation regularly in Bra-Ket notation. To do so, replace the wave function Ψ with the ket vector:

$$\hat{H}\,|\Psi\rangle = W\,|\Psi\rangle \tag{16.23}$$

$$\hat{H}\,|\Psi\rangle = \mathrm{i}\hbar \frac{\partial}{\partial t}\,|\Psi\rangle \tag{16.24}$$

16.8 Mean Values in Bra-Ket Notation

We can utilize the familiar Bra-Ket notation to represent the **mean value $\langle \hat{H} \rangle$ of an operator $\hat{H}$ in the quantum state $|\Psi\rangle$**. Often physicists call the mean value as **expectation value** - however, this terminology is misleading and should not be used for the mean value $\langle \hat{H} \rangle$. The notation $\langle \hat{H} \rangle$ for the mean value is a short representation of $\langle \Psi \,|\, \hat{H} \,|\, \Psi \rangle$.

So, we obtain the mean value of an observable by sandwiching the operator $\hat{H}$ between a Bra vector $\langle \Psi \,|$ and a Ket vector:

$$\langle \hat{H} \rangle = \langle \Psi \,|\, \hat{H} \,|\, \Psi \rangle = \langle \Psi \,|\, \hat{H}\ \Psi \rangle \tag{16.25}$$

For the last equal sign, we have taken advantage of the fact that $\hat{H}$ applied to the ket vector $|\Psi\rangle$ results in a **new ket vector** $|\hat{H}\ \Psi\rangle$. As you know, the notation in $|\ \rangle$ is irrelevant as long as it is clear what this ket vector represents.

Now you should have a solid basic knowledge of bra-ket notation. You should have learned the following from this chapter:

- You know what bra and ket vectors are.
- You know how to form the scalar product and inner product.
- You know how to construct projection matrices in Bra-Ket notation.
- You know how to carry out a change of basis with projection matrices.

17. Represent Operators as Matrices

If we choose a basis, we can represent an operator $\hat{H}$ **as a matrix** H**.** For example, we can take the eigenbasis $\{|\varphi_i\rangle\}$ of $\hat{H}$. Eigenbasis is the set of eigenstates (eigenvectors) $|\varphi_1\rangle, |\varphi_2\rangle, |\varphi_3\rangle, ...$ of $\hat{H}$.

- If the operator $\hat{H} : \mathbb{H} \to \mathbb{H}$ is a mapping between the **two-dimensional** Hilbert spaces $\mathbb{H}$, then the eigenbasis of $\hat{H}$ has two eigenstates $\{|\varphi_1\rangle, |\varphi_2\rangle\}$ and can be represented by a **2x2-matrix**:

$$H = \begin{bmatrix} \langle\varphi_1 | \hat{H} | \varphi_1\rangle & \langle\varphi_1 | \hat{H} | \varphi_2\rangle \\ \langle\varphi_2 | \hat{H} | \varphi_1\rangle & \langle\varphi_2 | \hat{H} | \varphi_2\rangle \end{bmatrix}$$

- If the operator $\hat{H} : \mathbb{H} \to \mathbb{H}$ is a mapping between the **three-dimensional** Hilbert spaces $\mathbb{H}$, then the eigenbasis of $\hat{H}$ has three eigenstates $\{|\varphi_1\rangle, |\varphi_2\rangle, |\varphi_3\rangle\}$ and can be represented by a **3x3-matrix**:

$$H = \begin{bmatrix} \langle\varphi_1 | \hat{H} | \varphi_1\rangle & \langle\varphi_1 | \hat{H} | \varphi_2\rangle & \langle\varphi_1 | \hat{H} | \varphi_3\rangle \\ \langle\varphi_2 | \hat{H} | \varphi_1\rangle & \langle\varphi_2 | \hat{H} | \varphi_2\rangle & \langle\varphi_2 | \hat{H} | \varphi_3\rangle \\ \langle\varphi_3 | \hat{H} | \varphi_1\rangle & \langle\varphi_3 | \hat{H} | \varphi_2\rangle & \langle\varphi_3 | \hat{H} | \varphi_3\rangle \end{bmatrix}$$

18. Hermitian Operators

More: en.fufaev.org/hermitian-operators

Let's take a look at the mean value $\langle \hat{H} \rangle$ of the operator $\hat{H}$ in the state $|\Psi\rangle$. If $\hat{H}$ is the Hamilton operator, then $\langle \hat{H} \rangle$ describes the **mean value of the total energy** of a quantum particle:

$$\langle \hat{H} \rangle = \langle \Psi \,|\, \hat{H}\, \Psi \rangle \tag{18.1}$$

$$= \int \Psi(x,t)^* \left(\hat{H}\, \Psi(x,t) \right) \mathrm{d}x \tag{18.2}$$

The mean value of a physical quantity, as we know it from everyday life, is a **real number**. In quantum mechanics, however, complex mean values can also occur for some operators. Imagine that we want to calculate the local mean value $\langle \hat{x} \rangle$ or the energy mean value $\langle \hat{H} \rangle$ and obtain a complex value as a result. This is problematic, because what does a complex position or a complex energy mean?

In quantum mechanics, we are therefore only interested in mean values $\langle \hat{H} \rangle$ that are real. How can we demand this mathematically? Quite simple! A real number, for example the number 5, remains unchanged if we complex conjugate

it: $5 = 5^*$. A complex number, for example $4 + 2\mathrm{i}$, does not remain the same if we complex conjugate it: $(4 + 2\mathrm{i})^* = 4 - 2\mathrm{i}$.

So if the **mean value is real**, then it is **equal to its complex conjugate value**:

$$\langle \Psi \,|\, \hat{H}\, \Psi \rangle \;=\; \langle \Psi \,|\, \hat{H}\, \Psi \rangle^* \tag{18.3}$$

We call the operator $\hat{H}$ whose mean value $\langle \hat{H} \rangle$ is real, as a **Hermitian operator**. A Hermitian operator therefore represents a **measurable quantity**, such as momentum, position and energy. A quantity that we can measure in an experiment is called an **observable** in quantum mechanics.

What does this mean for the mean value integral 18.2 if $\hat{H}$ is a Hermitian operator? Let's take a look at this by calculating the complex conjugate mean value $\langle \Psi \,|\, \hat{H}\, \Psi \rangle^*$ by using the mean value integral 18.2:

$$\begin{aligned}
\langle \Psi \,|\, \hat{H}\, \Psi \rangle^* \;&=\; \left(\int \Psi^* (\hat{H}\, \Psi)\, \mathrm{d}x \right)^* \\
&=\; \left(\int (\hat{H}\, \Psi)\, \Psi^*\, \mathrm{d}x \right)^* \\
&=\; \int (\hat{H}\, \Psi)^*\, (\Psi^*)^*\, \mathrm{d}x \\
&=\; \int (\hat{H}\, \Psi)^*\, \Psi\, \mathrm{d}x \\
&=\; \langle \hat{H}\, \Psi \,|\, \Psi \rangle
\end{aligned} \tag{18.4}$$

We have discovered the following important property of Hermitian operators:

$$\langle \Psi \,|\, \hat{H}\, \Psi \rangle \;=\; \langle \hat{H}\, \Psi \,|\, \Psi \rangle \tag{18.5}$$

In order for our mean-has-to-be-real requirement 18.2 to be fulfilled, the operator $\hat{H}$ **must be interchangeable in the scalar product**. It must therefore not matter whether we first apply $\hat{H}$ to the Ket or Bra vector in the mean

value calculation. So if you know that an operator is Hermitian, then move the operator wherever you want in the Bra-Ket notation.

18.1 Useful properties of Hermitian operators

In addition to the useful property 18.5, a Hermitian operator has a bunch of other useful properties.

Usually, if we want to move an operator $\hat{H}$ that acts on a Ket vector to the Bra vector, we have take the **adjoint** $\hat{H}^\dagger$ of the operator:

$$\langle \Psi \,|\, \hat{H}^\dagger \, \Psi \rangle \;=\; \langle \hat{H} \, \Psi \,|\, \Psi \rangle \tag{18.6}$$

As you have learned, in the case of a Hermitian operator we do not have to do this. A **Hermitian operator is a self-adjoint operator**:

$$\hat{H} \;=\; \hat{H}^\dagger \tag{18.7}$$

To make it clear that $\hat{H}$ is a Hermitian operator, the mean value is also noted as follows:

$$\begin{aligned} \langle \Psi \,|\, \hat{H} \, \Psi \rangle \;&=\; \langle \hat{H} \, \Psi \,|\, \Psi \rangle \\ &=\; \langle \Psi \,|\, \hat{H} \,|\, \Psi \rangle \\ &=\; \langle \Psi \,|\, \hat{H}^\dagger \,|\, \Psi \rangle \end{aligned} \tag{18.8}$$

A Hermitian operator has another important property that you must remember for your quantum mechanics courses: The **set of eigenstates** $\{|\varphi_i\rangle\}$ (eigenvectors) of a Hermitian operator **can be used as a basis**. This property is so important that it has a name, namely the **spectral theorem**.

Take this to heart: You have a Hermitian operator in front of you. You can take its eigenstates $\{|\varphi_i\rangle\}$ as a basis and thus represent any other state in this basis. That's incredible and super useful!

18.2 Examples of Hermitian Matrices

Most of the operators that you will encounter in your undergrad courses are Hermitian operators. These include, for example, the momentum operator $\hat{p}$, the position operator $\hat{x}$, the Hamiltonian operator $\hat{H}$, the kinetic energy operator $\hat{W_{\text{kin}}}$ and so on.

As you learned in Chapter 17, we can represent an operator as a matrix if we choose a concrete basis in which to represent the operator. Let's look at a few concrete examples of Hermitian matrices.

The σ_y spin matrix is Hermitian. If you transpose and complex-conjugate it, you get the same matrix:

$$\sigma_y = \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix} = \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}^* = (\sigma_y)^* \tag{18.9}$$

The σ_x spin matrix is also a Hermitian matrix:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^* = (\sigma_x)^* \tag{18.10}$$

And here is an example of a non-Hermitian matrix. If you transpose and complex conjugate it, you get a completely different matrix that is not equal to the original matrix:

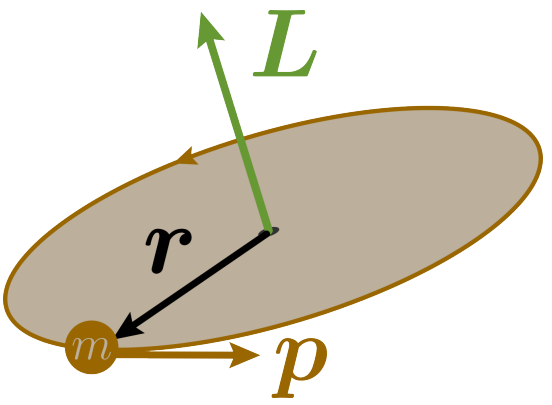
$$\begin{bmatrix} 1 & 2 \\ -3\mathrm{i} & 0 \end{bmatrix} \neq \begin{bmatrix} 1 & 2 \\ -3\mathrm{i} & 0 \end{bmatrix}^* = \begin{bmatrix} 1 & 3\mathrm{i} \\ 2 & 0 \end{bmatrix} \tag{18.11}$$

19. Angular Momentum

More: en.fufaev.org/quantum-angular-momentum

The **angular momentum** (more precisely: **orbital angular momentum**) $\boldsymbol{L}$ of a classical particle is given by the cross product between the **distance** $\boldsymbol{r}$ of the particle from the axis of rotation and the linear **momentum** $\boldsymbol{p} = m\boldsymbol{v}$ of the particle:

$$\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p} = \begin{bmatrix} y\,p_{\mathrm{z}} - z\,p_{\mathrm{y}} \\ z\,p_{\mathrm{x}} - x\,p_{\mathrm{z}} \\ x\,p_{\mathrm{y}} - y\,p_{\mathrm{x}} \end{bmatrix} \tag{19.1}$$



The angular momentum $\boldsymbol{L}$ is therefore perpendicular to the vectors $\boldsymbol{r}$ and $\boldsymbol{p}$ due to the cross product.

If we write out the cross product, we get the individual components L_1, L_2 and L_3 of the angular momentum vector, which each indicate the magnitude of angular momentum in the x, y and z directions:

- Angular momentum in the x direction is:

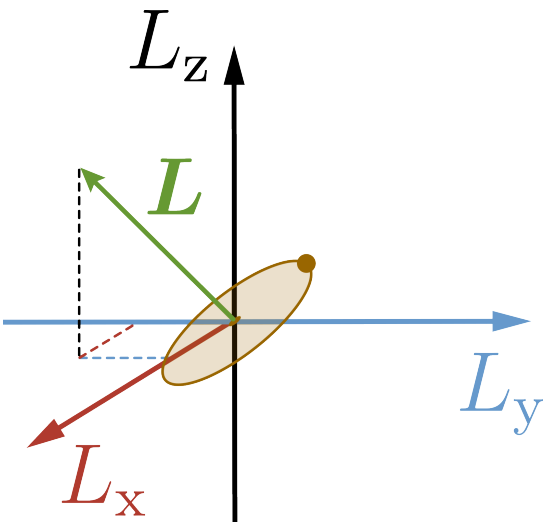
$$L_{\mathrm{x}} = y\,p_{\mathrm{z}} - z\,p_{\mathrm{y}} \tag{19.2}$$

- Angular momentum in the y direction is:

$$L_{\mathrm{y}} = z\,p_{\mathrm{x}} - x\,p_{\mathrm{z}} \tag{19.3}$$

- Angular momentum in the z direction is:

$$L_{\mathrm{z}} = x\,p_{\mathrm{y}} - y\,p_{\mathrm{x}} \tag{19.4}$$



How do we turn these classical angular momentum components into **quantum mechanical** angular momentum components? By putting little hats on them: $\hat{L}_x$, $\hat{L}_y$ and $\hat{L}_z$, in other words turning the angular momentum components into operators. The spatial components $x = \hat{x}$, $y = \hat{y}$ and $z = \hat{z}$ remain the same. And the momentum components are replaced by the following **axiomatic mappings**:

- Momentum component p_x becomes the operator: $\hat{p}_x = -i\hbar\,\partial_x$
- Momentum component p_y becomes the operator: $\hat{p}_y = -i\hbar\,\partial_y$
- Momentum component p_z becomes the operator: $\hat{p}_z = -i\hbar\,\partial_z$

Here, ∂_x,∂_y and ∂_z are **derivative operators**. When applied to a function, they result in the derivative of this function with respect to x, y or z. Of course, a single derivative makes no sense. For this reason, a momentum operator only becomes useful when it is applied to a wave function. The result is a new wave function modified by the operator. With the axiomatic mappings, we have quantized the **classical orbital angular momentum**:

- $\hat{L}_x = -i\hbar\,y\,\partial_z + i\hbar\,z\,\partial_y = i\hbar\,(z\,\partial_y - y\,\partial_z)$
- $\hat{L}_y = -i\hbar\,z\,\partial_x + i\hbar\,x\,\partial_z = i\hbar\,(x\,\partial_z - z\,\partial_x)$
- $\hat{L}_z = -i\hbar\,x\,\partial_y + i\hbar\,y\,\partial_x = i\hbar\,(y\,\partial_x - x\,\partial_y)$

19.1 Can We Measure Angular Momentum?

We have constructed the angular momentum operators $\hat{L}_\mathrm{x}$, $\hat{L}_\mathrm{y}$ and $\hat{L}_\mathrm{z}$. The question now is: Do they represent observables? Or to put it another way: **Are they Hermitian operators?** This is important because the angular momentum components can only be measured in the experiment if they are Hermitian operators.

A Hermitian operator $\hat{L}_\mathrm{x}$ is equal to its complex conjugate $\hat{L}_\mathrm{x}{}^\dagger$. Let's check that:

$$\hat{L}_\mathrm{x}{}^\dagger \;=\; (\hat{y}\,\hat{p}_\mathrm{z} \;-\; \hat{z}\,\hat{p}_\mathrm{y})^\dagger \tag{19.5}$$

First, we apply the property of **anti-linearity**:

$$\hat{L}_\mathrm{x}{}^\dagger \;=\; (\hat{y}\,\hat{p}_\mathrm{z})^\dagger \;-\; (\hat{z}\,\hat{p}_\mathrm{y})^\dagger \tag{19.6}$$

In the next step **we will use the anti-distributivity**. This swaps the two operators in the parenthesis and the parenthesis disappears:

$$\hat{L}_\mathrm{x}{}^\dagger \;=\; \hat{p}_\mathrm{z}{}^\dagger\,\hat{y}^\dagger \;-\; \hat{p}_\mathrm{y}{}^\dagger\,\hat{z}^\dagger \tag{19.7}$$

We know that the position operators $\hat{y}$, $\hat{z}$ and momentum operators $\hat{p}_\mathrm{z}$, $\hat{p}_\mathrm{y}$ are Hermitian. Hermitian operators are **equal to their adjoint**. We can therefore omit †:

$$\hat{L}_\mathrm{x}{}^\dagger \;=\; \hat{p}_\mathrm{z}\,\hat{y} \;-\; \hat{p}_\mathrm{y}\,\hat{z} \tag{19.8}$$

The momentum and position operators may be interchanged here, because operator $\hat{p}_\mathrm{z} = -\mathrm{i}\hbar\,\partial_z$ differentiates with respect to the z-coordinate and $\hat{y} = y$ does not depend on z and therefore acts like a constant that can be moved forward. The argumentation for the term $\hat{p}_\mathrm{y}\,\hat{z}$ is the same:

$$\hat{L}_\mathrm{x}{}^\dagger \;=\; \hat{y}\,\hat{p}_\mathrm{z} \;-\; \hat{z}\,\hat{p}_\mathrm{y} \;=\; \hat{L}_\mathrm{x} \tag{19.9}$$

The expression obtained corresponds exactly to the $\hat{L}_x$ operator. Similarly, we can show that the $\hat{L}_y$ and the $\hat{L}_z$ operator are also Hermitian. Conclusion: **We can measure the angular momentum components of quantum particles in an experiment**. Perfect!

19.2 Can We Determine ALL Angular Momentum Components?

In **classical physics**, in our macroscopic world, the values of all three angular momentum components L_x, L_y and L_z exist - for example of a spinning particle. All three components can therefore be determined **exactly and simultaneously**.

In the **quantum world**, on the other hand, we have the **Heisenberg uncertainty principle**, which makes it impossible to determine certain observables exactly at the same time because one of the observables does not inherently have an exact value if the other observable is measured exactly. Momentum $\hat{p}_y$ and position $\hat{x}$ are an example of two observables that cannot be determined exactly at the same time.

Formulated mathematically, the Heisenberg uncertainty principle states: If we first apply the position operator $\hat{x}$ to the wave function $\Psi(x, y, z)$ and then the momentum operator: $\hat{p}_x \, \hat{x} \, \Psi(x, y, z)$, then we get something different than if we first apply the momentum operator and then the position operator: $\hat{p}_x \, \hat{x} \, \Psi(x, y, z)$. **It matters whether we first measure the position or the momentum of a quantum particle**. As soon as we reverse the order of the measurement, we get something completely different for the momentum and position. We say that the momentum and position are subject to the Heisenberg uncertainty principle. The **difference between the two measurements is provided by the commutator** $[\hat{x}, \hat{p}_x]$. To do this, we calculate the difference between the two measurements and factor out the wave

function. The difference of the operators is the commutator of $\hat{x}$ and $\hat{p}_\mathrm{x}$:

$$\begin{aligned} \hat{p}_\mathrm{x}\,\hat{x}\,\Psi \;-\; \hat{p}_\mathrm{x}\,\hat{x}\,\Psi \;&=\; (\hat{p}_\mathrm{x}\,\hat{x} \;-\; \hat{p}_\mathrm{x}\,\hat{x})\,\Psi \\ &=\; [\hat{x}, \hat{p}_\mathrm{x}]\;\Psi \\ &=\; \mathrm{i}\,\hbar\,\Psi \end{aligned} \tag{19.10}$$

In the last step, we used the commutator of the position and momentum operator $[\hat{x}, \hat{p}_\mathrm{x}] = \mathrm{i}\,\hbar$.

- If the commutator is zero, then in principle it is possible to determine both observables exactly at the same time.
- If the commutator is not zero, as in the case of $[\hat{x}, \hat{p}_\mathrm{x}] = \mathrm{i}\,\hbar$, then it is impossible to determine both observables exactly at the same time. Only one of the observables, either $\hat{p}_\mathrm{x}$ or $\hat{x}$, can be determined exactly.

With this knowledge, we can now ask: **Can we know all the angular momentum components of a quantum particle exactly at the same time?**

Short answer: **No!** To do this, we must use the commutators $[\hat{L}_\mathrm{x}, \hat{L}_\mathrm{z}]$, $[\hat{L}_\mathrm{y}, \hat{L}_\mathrm{z}]$ and $[\hat{L}_\mathrm{x}, \hat{L}_\mathrm{y}]$ of the angular momentum components. We will find that none of the commutators is zero. It is therefore impossible to know two angular momentum components at the same time.

To demonstrate this, let's look at the commutator $[\hat{L}_\mathrm{x}, \hat{L}_\mathrm{z}]$ and show that $\hat{L}_\mathrm{x}$ and $\hat{L}_\mathrm{z}$ are subject to the uncertainty principle. First, we use the definition of the commutator:

$$\begin{aligned} [\hat{L}_\mathrm{x}, \hat{L}_\mathrm{z}] \;&=\; \hat{L}_\mathrm{x}\,\hat{L}_\mathrm{z} \;-\; \hat{L}_\mathrm{z}\,\hat{L}_\mathrm{x} \\ &=\; (\hat{y}\,\hat{p}_\mathrm{z} \;-\; \hat{z}\,\hat{p}_\mathrm{y})\,(\hat{x}\,\hat{p}_\mathrm{y} \;-\; \hat{y}\,\hat{p}_\mathrm{x}) \\ &-\; (\hat{x}\,\hat{p}_\mathrm{y} \;-\; \hat{y}\,\hat{p}_\mathrm{x})\,(\hat{y}\,\hat{p}_\mathrm{z} \;-\; \hat{z}\,\hat{p}_\mathrm{y}) \\ &=\; \hat{y}\,\hat{p}_\mathrm{z}\,\hat{x}\,\hat{p}_\mathrm{y} \;-\; \hat{y}\,\hat{p}_\mathrm{z}\,\hat{y}\,\hat{p}_\mathrm{x} \;-\; \hat{z}\,\hat{p}_\mathrm{y}\,\hat{x}\,\hat{p}_\mathrm{y} \\ &+\; \hat{z}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{p}_\mathrm{x} \;-\; \hat{x}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{p}_\mathrm{z} \;+\; \hat{x}\,\hat{p}_\mathrm{y}\,\hat{z}\,\hat{p}_\mathrm{y} \\ &+\; \hat{y}\,\hat{p}_\mathrm{x}\,\hat{y}\,\hat{p}_\mathrm{z} \;-\; \hat{y}\,\hat{p}_\mathrm{x}\,\hat{z}\,\hat{p}_\mathrm{y} \end{aligned} \tag{19.11}$$

For the second equal sign, we have expressed the angular momentum operators with position and momentum operators. For the third equal sign, we have multiplied out the brackets.

Then we swap the operators so that some terms are canceled out. In the first term, we can place $\hat{x}$ at the beginning without problems: $\hat{x}\,\hat{y}\,\hat{p}_\mathrm{z}\,\hat{p}_\mathrm{y}$, because $\hat{x}$ commutates with both $\hat{y}$ and $\hat{p}_\mathrm{z}$ (their commutator is zero, so we can move them back and forth). We can place the operator $\hat{p}_\mathrm{y}$ in front of $\hat{p}_\mathrm{z}$ without problems: $\hat{x}\,\hat{y}\,\hat{p}_\mathrm{y}\,\hat{p}_\mathrm{z}$, but not in front of $\hat{y}$, because the commutator of $[\hat{y}, \hat{p}_\mathrm{y}] = \mathrm{i}\hbar$ is not zero. Therefore, we must replace $\hat{y}, \hat{p}_\mathrm{y}$ with $\mathrm{i}\hbar + \hat{p}_\mathrm{y}\,\hat{y}$: $\hat{x}\,(\mathrm{i}\hbar + \hat{p}_\mathrm{y}\,\hat{y})\,\hat{p}_\mathrm{z} = \mathrm{i}\hbar\,\hat{x}\,\hat{p}_\mathrm{z} + \hat{x}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{p}_\mathrm{z}$. This cancels out the term $\hat{x}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{p}_\mathrm{z}$:

$$\begin{aligned}[\hat{L}_\mathrm{x}, \hat{L}_\mathrm{z}] \;=\; & \mathrm{i}\hbar\,\hat{x}\,\hat{p}_\mathrm{z} \;-\; \hat{y}\,\hat{p}_\mathrm{z}\,\hat{y}\,\hat{p}_\mathrm{x} \;-\; \hat{z}\,\hat{p}_\mathrm{y}\,\hat{x}\,\hat{p}_\mathrm{y} \\ & +\; \hat{z}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{p}_\mathrm{x} \;-\; \hat{x}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{p}_\mathrm{z} \;+\; \hat{x}\,\hat{p}_\mathrm{y}\,\hat{z}\,\hat{p}_\mathrm{y} \\ & +\; \hat{y}\,\hat{p}_\mathrm{x}\,\hat{y}\,\hat{p}_\mathrm{z} \;-\; \hat{y}\,\hat{p}_\mathrm{x}\,\hat{z}\,\hat{p}_\mathrm{y}\end{aligned} \tag{19.12}$$

In the expression $\hat{y}\,\hat{p}_\mathrm{z}\,\hat{y}\,\hat{p}_\mathrm{x}$ we can swap all operators without any problems and cancel it out with the other term $\hat{y}\,\hat{p}_\mathrm{x}\,\hat{y}\,\hat{p}_\mathrm{z}$:

$$\begin{aligned}[\hat{L}_\mathrm{x}, \hat{L}_\mathrm{z}] \;=\; & \mathrm{i}\hbar\,\hat{x}\,\hat{p}_\mathrm{z} \;-\; \hat{z}\,\hat{p}_\mathrm{y}\,\hat{x}\,\hat{p}_\mathrm{y} \;+\; \hat{z}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{p}_\mathrm{x} \\ & +\; \hat{x}\,\hat{p}_\mathrm{y}\,\hat{z}\,\hat{p}_\mathrm{y} \;-\; \hat{y}\,\hat{p}_\mathrm{x}\,\hat{z}\,\hat{p}_\mathrm{y}\end{aligned} \tag{19.13}$$

And also in the expression $\hat{z}\,\hat{p}_\mathrm{y}\,\hat{x}\,\hat{p}_\mathrm{y}$ operators can be swapped so that the operator that is related to the expression $\hat{x}\,\hat{p}_\mathrm{y}\,\hat{z}\,\hat{p}_\mathrm{y}$ cancels out:

$$[\hat{L}_\mathrm{x}, \hat{L}_\mathrm{z}] \;=\; \mathrm{i}\hbar\,\hat{x}\,\hat{p}_\mathrm{z} \;-\; \hat{z}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{p}_\mathrm{x} \;-\; \hat{y}\,\hat{p}_\mathrm{x}\,\hat{z}\,\hat{p}_\mathrm{y} \tag{19.14}$$

Now we come back to a term $\hat{y}\,\hat{p}_\mathrm{x}\,\hat{z}\,\hat{p}_\mathrm{y}$, where the swapping is not simply possible. First, we can swap $\hat{y}$ and $\hat{p}_\mathrm{x}$: $\hat{p}_\mathrm{x}\,\hat{y}\,\hat{z}\,\hat{p}_\mathrm{y}$ and then $\hat{z}$ with $\hat{p}_\mathrm{y}$, so that we have the following expression: $\hat{p}_\mathrm{x}\,\hat{y}\,\hat{p}_\mathrm{y}\,\hat{z}$. To now swap $\hat{y}$ with $\hat{p}_\mathrm{y}$, we have to use their product because of the non-vanishing commutator $[\hat{y}, \hat{p}_\mathrm{y}] = \mathrm{i}\hbar$ with $\mathrm{i}\hbar + \hat{p}_\mathrm{y}\,\hat{y}$:

$\hat{p}_\mathrm{x}\,(\mathrm{i}\hbar + \hat{p}_\mathrm{y}\,\hat{y})\;\hat{z} = \mathrm{i}\hbar\,\hat{p}_\mathrm{x}\,\hat{z} - \hat{p}_\mathrm{x}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{z}$. This turns the commutator into:

$$[\hat{L}_\mathrm{x}, \hat{L}_\mathrm{z}] \;=\; \mathrm{i}\hbar\,\hat{x}\,\hat{p}_\mathrm{z} \;-\; \hat{z}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{p}_\mathrm{x} \;-\; \mathrm{i}\hbar\,\hat{p}_\mathrm{x}\,\hat{z} - \hat{p}_\mathrm{x}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{z} \tag{19.15}$$

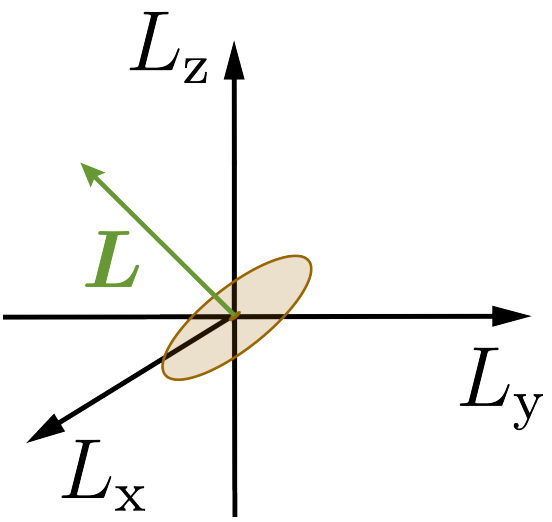
In this expression $\hat{z}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{p}_\mathrm{x}$ we swap $\hat{z}$ with $\hat{p}_\mathrm{x}$: $\hat{p}_\mathrm{x}\,\hat{p}_\mathrm{y}\,\hat{y}\,\hat{z}$ and can thus cancel it out:

$$\begin{aligned}
[\hat{L}_\mathrm{x}, \hat{L}_\mathrm{z}] \;&=\; \mathrm{i}\hbar\,\hat{x}\,\hat{p}_\mathrm{z} \;-\; \mathrm{i}\hbar\,\hat{p}_\mathrm{x}\,\hat{z} \\
&=\; \mathrm{i}\hbar\;(\hat{x}\,\hat{p}_\mathrm{z} \;-\; \hat{p}_\mathrm{x}\,\hat{z}) \\
&=\; \mathrm{i}\hbar\,\hat{L}_\mathrm{y}
\end{aligned} \tag{19.16}$$

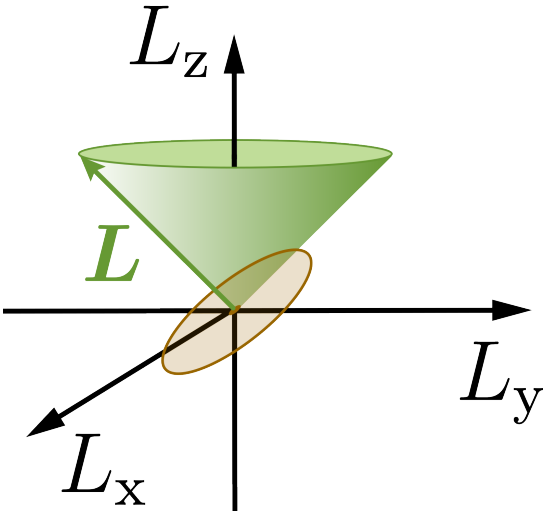
As you can see, the commutator $[\hat{L}_\mathrm{x}, \hat{L}_\mathrm{z}]$ is not zero, so it is impossible to know $\hat{L}_\mathrm{x}$ and $\hat{L}_\mathrm{z}$ simultaneously with arbitrary precision. The other two commutators can be derived in the same way:

$$\begin{aligned}
[\hat{L}_\mathrm{y}, \hat{L}_\mathrm{z}] \;&=\; \mathrm{i}\hbar\,\hat{L}_\mathrm{x} \\
[\hat{L}_\mathrm{x}, \hat{L}_\mathrm{y}] \;&=\; \mathrm{i}\hbar\,\hat{L}_\mathrm{z}
\end{aligned} \tag{19.17}$$

We can easily illustrate this fundamental uncertainty of the angular momentum components. Let's consider a classical particle that moves on a circular path. It therefore has an angular momentum $\boldsymbol{L}$. All three angular momentum components are exactly fixed, so $\boldsymbol{L}$ has a fixed direction.



What if it were a quantum particle? Let's assume that we have measured the angular momentum component $\hat{L}_z$ of the quantum particle. We have thus exactly determined its angular momentum component L_z. Due to the non-vanishing commutators, the other two angular momentum components L_x and L_y have no concrete value. The direction of the total angular momentum vector $\boldsymbol{L}$ is no longer clearly given, but lies **anywhere on a cone mantle**.



Based on the cone in the illustration, we can already guess that although the **direction of $\boldsymbol{L}$ is not unique**, the **length of the $\boldsymbol{L}$ vector is unique**. We

can determine the length of the $\boldsymbol{L}$ vector using the sum of the squares of the angular momentum operators:

$$\hat{\boldsymbol{L}}^2 = \hat{L}_\mathrm{x}{}^2 + \hat{L}_\mathrm{y}{}^2 + \hat{L}_\mathrm{z}{}^2 \tag{19.18}$$

This sum is briefly noted as the $\hat{\boldsymbol{L}}^2$ operator. This operator is Hermitian, so it represents an observable, namely the **length of the angular momentum vector squared**. And the great thing is: This operator **commutes with each angular momentum component** $\hat{L}_\mathrm{x}$, $\hat{L}_\mathrm{y}$ and $\hat{L}_\mathrm{z}$. This is very good, because it allows us to determine not only one of the angular momentum components of a quantum particle exactly, but also the **magnitude of the total angular momentum**:

$$\hat{L} = \sqrt{\hat{L}_\mathrm{x}{}^2 + \hat{L}_\mathrm{y}{}^2 + \hat{L}_\mathrm{z}{}^2} \tag{19.19}$$

This would be very bad for physics if the magnitude of the total angular momentum did not exist exactly at all times. Without a fixed, exact total angular momentum, the law of conservation of angular momentum in quantum mechanics would not work at all.

19.3 Quantum Numbers l and m

Let's use the Bra-Ket notation and treat operators as matrices and wave functions as Ket vectors (states). A commutator not only tells us whether two observables can be measured exactly at the same time, but also whether the associated operators **share eigenstates** $\{|\Psi_i\rangle\}$.

If the commutator $[\hat{\boldsymbol{L}}^2, \hat{L}_\mathrm{z}]$ vanishes (and it does), then we know that there is a state $|\Psi\rangle$, which is simultaneously both an eigenstate of $\hat{\boldsymbol{L}}^2$ and an eigenstate of $\hat{L}_\mathrm{z}$.

- If $\hat{\boldsymbol{L}}^2$ operator is applied to the state $|\Psi\rangle$, which is an eigenstate of this operator, then the result is an eigenstate scaled by the eigenvalue: In the case of $\hat{\boldsymbol{L}}^2$, the eigenvalue returns the magnitude of the total angular

momentum squared:

$$\hat{\boldsymbol{L}}^2 \,|\,\Psi\rangle \;=\; L^2 \,|\,\Psi\rangle \tag{19.20}$$

- And if $\hat{L}_z$ is applied to the state $|\,\Psi\rangle$, which is also an eigenstate of $\hat{\boldsymbol{L}}^2$, then we again get the scaled eigenstate with a different eigenvalue. In the case of $\hat{L}_z$, this eigenvalue represents the magnitude of the angular momentum component in the z-direction:

$$\hat{L}_z \,|\,\Psi\rangle \;=\; L_z \,|\,\Psi\rangle \tag{19.21}$$

Using the **ladder operators**, we can derive the eigenvalues L^2 and L_z a little more precisely. Here I give the famous result that probably every chemist knows. The **eigenvalues L^2 are a multiple of $\hbar^2$**:

$$\begin{aligned} \hat{\boldsymbol{L}}^2 \,|\,\Psi\rangle \;&=\; L^2 \,|\,\Psi\rangle \\ &=\; l\,(l+1)\,\hbar^2 \,|\,\Psi\rangle \end{aligned} \tag{19.22}$$

The eigenvalue $l\,(l+1)\,\hbar^2$ is determined by a **integer or half-integer number** l, which we call **angular momentum quantum number**. This quantum number can have the following values:

$$l \;=\; 0,\; \frac{1}{2},\; 1,\; \frac{3}{2},\; 2,\; \ldots \tag{19.23}$$

Consequently, the total angular momentum squared L^2 cannot take on any continuous value, but only the following discrete values, which are determined by the angular momentum quantum number:

$$L^2 \;=\; 0,\; 0.75\hbar^2,\; 2\hbar^2,\; 3.75\hbar^2,\; 6\hbar^2,\; \ldots \tag{19.24}$$

Let's take the square root of the magnitude square. The **magnitude L of the**

total angular momentum is quantized:

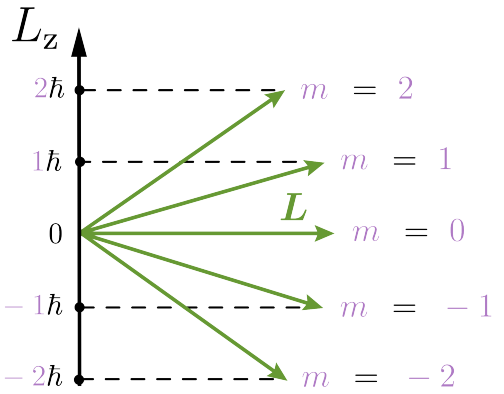
$$L = 0,\ \sqrt{0.75}\hbar,\ \sqrt{2}\hbar,\ \sqrt{3.75}\hbar,\ \sqrt{6}\hbar^2,\ ... \tag{19.25}$$

The **eigenvalues L_z of the $\hat{L}_z$ angular momentum operator are a multiple of $\hbar$**:

$$\begin{aligned} \hat{L}_z |\Psi\rangle &= L_z |\Psi\rangle \\ &= m\hbar |\Psi\rangle \end{aligned} \tag{19.26}$$

The quantum number m is called the **magnetic quantum number** and it can only take on values between $m = -l$ and $m = l$ in $+1$ steps. The $\hat{L}_z$ angular momentum component can therefore not take on continuous values as in classical physics - the $\hat{L}_z$ angular momentum component is quantized.

Example: If a quantum particle has a total angular momentum $L = 2(2+1)\hbar^2 = 6\hbar^2$, represented by the angular momentum quantum number $l = 2$, then its magnetic quantum number can take on $m = -2, -1, 0, 1, 2$ values and no others. The L_z angular momentum component of this quantum particle can only have 5 possible values: $L_z = -2\hbar, -1\hbar, 0, 1\hbar, 2\hbar$.



Let's summarize what you should take away from the 19 chapter:

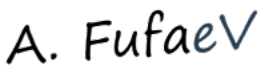
- You now know how to quantize **classical angular momentum**.

- You know how to show that the **angular momentum components are Hermitian**.
- You have learned that **all angular momentum components are subject to the uncertainty principle** and why only one of the components can be determined exactly.
- You have learned how to work **with angular momentum commutators**.
- You know what the $\hat{\boldsymbol{L}}^2$ operator is good for.
- You know the possible eigenvalues of $\hat{L}_z$ and $\hat{\boldsymbol{L}}^2$.

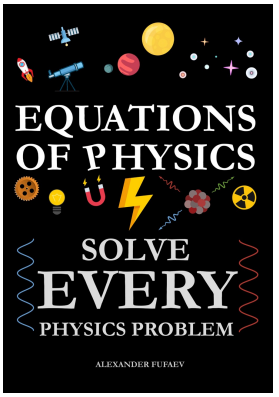
The End

If you enjoyed the book, it would immensely help me if you could leave a **brief review on Amazon** with a rating. Even more important is that you send me **any mistakes, suggestions for improvement, or any unclear sections** as soon as possible to the email alexander@fufaev.org so that I can address them immediately.

May physics be with you!



Equations of Physics: Solve Every Physics Problem



- Contains **over 500 illustrated formulas** on only 140 pages
- Includes **reference tables** with examples and constants
- **Easy to use**, because without vectors and integrals
- Perfect for courses with **basic physics**

Get this book: en.fufaev.org/physics-equations

Independently Published
ISBN for paperback: 9798320488516
ISBN for premium hardcover: 9798320491899

Imprint
Alexander Fufaev
Peiner Straße 86
31137 Hildesheim, Germany
alexander@fufaev.org

fufaev.org

Published, March 2024