Introduction to Mathematics

Lecture in WS 2023/24 at the KFU Graz

Axel Maas

Contents

Chapter 1

Introduction

1.1 Prelude

This lecture is intended primarily to make you acquainted with the basic mathematical tools needed during the first semesters of physics. All of the fields covered will be discussed in much more details and depth during this time in specialized lectures. The purpose is here to provide every student with a set of basic tools to start directly with applications in physics. Later on, the specialized lectures will prepare you with the necessary tools to tackle the more complex parts, especially of theoretical physics.

After all, all of physics can be formulated with mathematics. Thus, this is the primary language of our current understanding of the laws of nature. This is not self-evident, and merely an empirical fact. Its effectiveness is as unreasonable, as it is powerful. Having a clear command of it is therefore necessary to speak about and understand physics. Thus, mathematics is mainly a tool to the physicist, but a tool of which fluency is indispensable.

To some extent the following will be known from school. However, to provide the same set of tools to everyone, this lecture will cover all topics. Thus, necessarily, there will be repetitions. But depending on your school background there will also be new stuff. Also, you will notice that in many respects a lot is dissected, and made by hand. Especially many, apparently, menial tasks, like differentiating, will be required of you to do by hand, even if you know already technical tools to do them for you. The reason is that you will need to extend many concepts far beyond what you learned so far and what technical tools can cover. This will require a familiarity with the basic versions, which can only be acquired by experience. Just like craftsmanship, mathematics is learned by doing. Thus, just like a mechanic needs to learn the in and outs of an engine by working on it, so you will need to learn the in and outs of mathematics. Even though later on you will again use technical tools, you will then be able to go on, where none has been before.

Still, the aim in this lecture is to explain the different subjects at a rather basic level, emphasizing definitions and concepts over proofs. The latter can be given and understood in the context of the more general math lectures later. But it should not be missed that this is essentially a tour-de-force. Though much may be known to you, it is compressed into a one-hour-lecture equivalent. At the same time, given that about 19 more hours will be provided this semester, this gives an idea of the scale.

One of the more important points is that, despite the topics seeming to be rather loosely connected, nothing is truly independent, and many relations between different subjects will only manifest themselves much later in the course of studying physics. Patience may be required to understand the full breadth and relevance of certain topics. But all of the topics appear in the daily life of both experimentalists and theoreticians.

Finally, a lot of mathematics and (theoretical) physics is to know exactly what kind of technique can be used when. Hence, a lot of the following will also introduce the basic repertoire of tricks used to manipulate mathematical expressions. Usually, most people experience them first as 'how can anyone get this idea?'. The answer is that the first ones found them by trial and error, and thus as a test of creativity and tenacity. The generations afterwards learned them again from experience. So do not despair if you do not see how anyone could have guessed - none did. Just put them in your portfolio for later use.

In the following, in favor of being able to cover all required more complex concepts, some more basic ingredients are assumed to be known. This especially covers the basic operations of addition, multiplication as well as their inverse subtraction and division; taking the absolute value; the knowledge of natural, rational and real/irrational numbers; the way how to manipulate fractions and calculations of percentages; rules of proportions; and geometry of lines.

One basic, and sometimes frustrating, truth in both mathematics and physics is that we are (not) yet able to write down a completely closed system. I. e. we always have to make at some point some external input, called axioms (in mathematics) or postulates (in physics). The rule of addition is an example of a (possible) such external input in mathematics. We can also proof that we can create in any sufficiently complicated setup situations which cannot be decided to be true or false, and therefore, we are not able to calculate everything, though we can calculate everything relevant to physics. To the latter again an 'in principle' has to be added, as we encounter many situations where the calculations are so complicated that we have not (yet) been able to do them. But one should always be so ambitious to strive for the everything.

But we can get far, and this is the first step.

1.2 Numbers

In most of the following, the basic entities will be numbers, though rarely an explicit numerical value will be needed. One can distinguish between the positive integer numbers $\mathbb{N}$, which includes zero[1], the integers of both signs, $\mathbb{Z}$, the rational numbers $\mathbb{Q}$, which can be written as the quotient of two integer numbers, and the real numbers $\mathbb{R}$, which are numbers which can be approximated arbitrarily good from above an below by rational numbers, but are not rational numbers. Adding subscripts $+$ or $-$ includes only numbers of either sign, and $/0$ indicates the explicit exclusion of zero, while a subscript 0 explicitly includes the zero. So, $\mathbb{N} = \mathbb{Z}_+$ and $\mathbb{N}/0$ are the non-zero, positive integers.

In the following familiarity with the basic operations on these numbers, addition, subtraction, multiplication, and division, is assumed.

1.3 Sets

Before performing explicit calculations, a few basic elements of set theory will be necessary.

A set S is defined as a collection $\{a_1, ..., a_n\}$ of elements a_i counted by an index i . The elements a_i are taken to be unique, i. e. every element a_i in S is taken to be in the set once and only once. An element is said to be in a set, $a_3 \in S$, if it is part of the set S. The converse is stated by $b_3 \notin S$. It is possible to have an empty set, i. e. a set without elements. This is usually called the null set, symbolized by $\emptyset = \{\}$. Note that the set elements can be anything. Especially, a set can be an element of another set.

Examples of sets are the positive integers, $\mathbb{N}$ or the real numbers $\mathbb{R}$. A set can be defined conditionally, e. g.

$$S_{\text{even}} = \{a|a \in \mathbb{N} \text{ and } a/2 \in \mathbb{N}\}$$

is the set of all positive, even integers. And can also be abbreviate[2] as $\vee$ and or as $\wedge$. Elements can be excluded from a set, e. g. $\mathbb{Z}/\{0\}$ (more brief $\mathbb{Z}/0$) is the set of all integers without zero. The symbol $\forall$ is used to indicate a feature applying to all elements of a set. E. g. $a_i > 0 \ \forall \ a_i \in S_{\text{even}}$ states that all elements of S_{even} are positive.

The size of the set is just the number of items. E. g. the set $S = \{1, 2, 3\}$ has size 3. If the size is a finite number, the set is called finite. If the set is not finite, but the

[1]This is not universally defined. It needs to be checked, which conventions are used in a given text. Sometimes, zero is excluded or considered to have both signs.

[2]Note that the simple statements of being true or false can be embedded in a more consistent framework, logic. While this can be necessary in physics, the common understanding of it is sufficient for now.

elements can be counted by an integer, e. g. S_{even} or $\mathbb{N}$ itself, the set is called countable or denumerable. If this is not the case, e. g. $\mathbb{R}$, it is called innumerable.

If there are two or, more, sets S_i, there are many meaningful operations on this set of sets. Especially, it can be asked, which element appear in all sets, i. e. what is the intersection of the sets. For two sets S_1 and S_2 this is expressed as

$$S_i = S_1 \bigcap S_2.$$

E. g. $\{1,2\}\bigcap\{2,3\} = \{2\}$ and $\{1,2\}\bigcap\{3,4\} = \emptyset$. Alternatively, a set can be defined which contains all elements of both sets, the union

$$S_u = S_1 \bigcup S_2.$$

Elements appearing in both sets still appear only once in S_u. E. g. $\{1,2\}\bigcup\{2,3\} = \{1,2,3\}$ and $\{1,2\}\bigcup\emptyset = \{1,2\}$.

Both definitions can be continued by chaining multiple sets in arbitrary ways. To identify precedence, parentheses can be used. E. g.

$$\begin{aligned} (\{1,2\}\bigcup\{2,3\})\bigcap\{1\} &= \{1,2,3\}\bigcap\{1\} = &\{1\} \\ \{1,2\}\bigcup(\{2,3\}\bigcap\{1\}) &= \{1,2\}\bigcup\emptyset = &\{1,2\}. \end{aligned}$$

Paying attention to parentheses is therefore quite important.

1.4 Sums, sequences, and limits

A fundamentally important concept created from sets of numbers is a sequence of numbers a_i, where the index i is an integer and counts the elements of the sequence (and also of the set). Usually, sequences start with index 0, or sometimes also 1, but this is merely a convention. There are finite sequences, i. e. sequences that run up to maximum value of the index. More important are infinite sequences, where the index runs from 0 to infinity (∞), or sometimes from $-\infty$ to ∞. The sequence is then called denumerable (infinite), as the set itself. E. g. the finite sequence a_0, a_1, and a_2 yields the set $\{a_0, a_1, a_2\}$.

For an infinite sequence the elements may, or may not, approach better and better a certain number with increasing index. If this is the case, the sequence is said to have a limit a, written as

$$\lim_{i\to\infty} a_i = a,$$

where a may be any number, including infinity. The question under which condition a sequence has a limit is highly non-trivial in general, and will be detailed in the lecture on analysis. Note that a may or may not be an element of the set defining the sequence.

Given any sequence, a possibility is to add its elements, creating a sum, written as

$$\sum_{i=i_1}^{i_N} a_i,$$

where i_1 is the first index to be included, and i_N the last element, and thus $M = i_N - i_1$ is the number of elements to be added. M can thus not be larger than the number of elements in the sequence. If clear from the context, parts or all of the labels of the sum are omitted.

While any sum of a finite number M of finite a_i is finite, this is not necessarily so if an infinite number of elements from an infinite sequence are summed. Again, the general conditions under which a sum with infinitely many terms is finite will be treated in more detail in the courses on analysis. It is useful to define the value of an infinite sum a as a limit of adding more and more elements as

$$\lim_{i_N \to \infty} \sum_{i_1}^{i_N} a_i = a.$$

A sum for which a is finite is called convergent. A sum with a infinite is called divergent. Both statements are for now defined only if all terms a_i are finite individually. Note that a can, but does not need to, be an element of the set.

Limits do not necessarily have to do something with sequences or series. It is well possible to ask what happens if some quantity x approaches a certain value a, written as

$$\lim_{x \to a} x.$$

This is called a limit in a general sense. Though here only x is written, the object of which to form a limit can be much more complicated. Especially, it can be some arbitrary function of x

$$\lim_{x \to a} f(x).$$

Properties of functions is therefore the concept to be turned to next.

Chapter 2

Functions of a single variable

2.1 Generalities

Functions are a central element of mathematical descriptions of physics. To be more precise, a function f is a description, called a map, which takes a quantity x, called the argument or variable, taken from a domain of definition D, and yields a function value $f(x)$ in target range I, where I is called the image. Both the domain of definition and the image are for now taken to be all or part of the real numbers. This is also written as

$$f(x) : D \to I.$$

The domain of definition and the image may be different, if they are not the entire set of real numbers. In fact, the trivial function is $f(x) = 1$, which maps all real numbers to just a single number. The generalization $f(x) = c$ with c some number is called the constant function. The quadratic function is $f(x) = x^2$. A function does not need to yield a unique value for every argument. Indeed, the quadratic function is of this type, as any function value is obtained for two different values of the variables (except $x = 0$). In all cases $D = \mathbb{R}$, while $I = \{1\}$, $I = \{c\}$, and $I = \mathbb{R}_+$, respectively. The function $1/x$ has $D = I = \mathbb{R}/\{0\}$.

Two (or more) functions can be added, subtracted or multiplied, defining a composite function. Composite functions have the smallest of the domain of definition of the involved functions, but the image may be the union of both images. A division by a function is also possible, as long as $f(x) \neq 0$. If the function by which it is divided has such a zero, the corresponding values of x have to be removed from the domain of definition[1].

A final possibility is to chain functions together, i. e. first evaluate a function $f(x) = y$, and then evaluate on the result another function $g(y) = z$, which yields the final result

[1] The set of values for which $f(x)$ vanishes is also called the kernel of f.

z. As shorthand, this is written as $g(f(x))$. Only the intersection of the image of f and domain of definition of g is the domain of definition of the chain, and the image is the image of g restricted to this domain. If the intersection of the image of f and the domain of definition of g is empty, the image of the chain is also empty. These compositions can of course be extended to an arbitrary number of compositions in a straightforward way.

As an example, consider the function $f(x) = x^2$, with $D = \mathbb{R}$ and $I = \mathbb{R}_+$ and $g(y) = 1/(4-y)$, with $D = \mathbb{R}/\{4\}$ and $I = \mathbb{R}$. The composite function $g(f(x)) = 1/(4-x^2)$ has $D = \mathbb{R}/\{\pm 2\}$ and $I = \mathbb{R}/[0, 1/4)$. Here the notation $[a, b]$ denotes all real numbers between a and b including a and b, and $[a, b)$ denotes all real numbers between a and b including a but not b. In the present case, the image is thus all real numbers, excluding the range from 0 and 1/4, excluding 1/4.

Note that, in the sense of a limit, 0 can be included when $x \to \infty$. This may depend on the context. In the present case it is not included.

2.2 Ordinary functions

Ordinary functions are the most basic types of functions. They involve only the basic mathematical operations, and are therefore the simplest functions. One important step is to recognize that particular numbers play little to no special role. Thus, most of the following will be 'letter calculations', i. e. a placeholder letter is used instead of a concrete number.

2.2.1 Polynomials

The basic possibility to create a function of a single variable is by the use of addition and multiplication. For this purpose, the basic entity is the monomial

$$ax^n,$$

where a is any real number, and the exponent n is a positive integer or zero, called the order of the monomial, with $x^0 = 1$ understood. The notation states that x should be multiplied with itself n times. Multiplying two monomials yields

$$x^n x^m = x^{n+m}$$

by reverting to the definition that this would be x multiplied n times multiplied by x multiplied m times, and thus $n + m$ times in total. Thus exponents of monomials can be added. In a similar fashion

$$(x^n)^m = x^{nm}$$

is the statement to multiply m times products of n times x, giving in total a product of nm times x. Hence, exponents can also be multiplied.

Such monomials can be added to create polynomials, e. g.,

$$P(x) = ax^0 + bx^1 + cx^2 = a + bx + cx^2.$$

Subtraction of monomials is automatically included by permitting the prefactors to have either sign. Since for powers of the same order the pre-factors can be combined, the most general polynomial is

$$P_N(x) = \sum_{i=0}^{N} a_i x^i,$$

where the a_i form a set. Note that while some a_i may have the same value, the a_i themselves are distinguished by being the coefficients of different powers in the original polynomial. Note that one or more of the a_i can also be zero, without special notice. This is very often convenient. The number N specifies the highest order appearing in the polynomial, and thus is called order of the polynomial. A polynomial with highest power x is called linear.

In practice, N does not need to be finite, and many cases will be encountered where it is not. This requires the sum still to be finite to make sense, and hence the monomials must vanish sufficiently quickly. This has been already discussed in section 1.4, but is here generalized to

$$\lim_{N\to\infty} P_N(x) = \lim_{N\to\infty} \sum_{i}^{N} a_i x^i = P(x)$$

for the case of N infinite. In this case, the sum is a sum of functions, rather than numbers. Note that this implies that $P(x)$ is infinite for all values of x where the sum does not converge.

If the context is clear, often short-hand notations are used, especially

$$P(x) = \sum_i a_i x^i = \sum a_i x^i,$$

as the only purpose i can serve here is the one of a running index. Of course, this requires to have the limits of the sum either obvious from the context or to be irrelevant.

Take as an example $N = 3$, $a_0 = a_2 = 0$, $a_1 = 1$, and $a_3 = 23$. The result is then

$$\sum a_i x^i = 0 + 1x + 0x^2 + 23x^3 = x + 23x^3.$$

In this case neither a constant term appeared, nor all possible exponents.

Polynomials can be added/subtracted

$$A_N(x) + B_M(x) = \sum a_i x^i + \sum b_j x^j = \sum^{\max(N,M)} (a_i + b_i) x^i.$$

If N and M are not equal, additional coefficients needed are set to zero in the polynomial of lower order.

Polynomials can also be multiplied

$$A_N(x)B_M(x) = \left(\sum_i^N a_i x^i\right)\left(\sum_j^M b_j x^j\right) = \sum_k^{N+M} c_k x^k = C_{N+M}(x), \tag{2.1}$$

where the values of the c_k are determined from the a_i and b_j. The result is also a polynomial. It is important to keep track of the fact that the indices differ. For example

$$(a_0 + a_1 x)(b_0 + b_1 x^2) = a_0 b_0 + a_1 b_0 x + a_0 b_1 x^2 + a_1 b_1 x^3 = \sum_i^3 c_i x^i$$

with $c_0 = a_0 b_0$, $c_1 = a_1 b_0$, $c_2 = a_0 b_1$, and $c_3 = a_1 b_1$. Hence, the polynomial now runs up to $N + M = 3$ with $N = 1$ and $M = 2$.

2.2.2 Rational functions

So far, the construction only included addition, subtraction, and multiplication. This leaves division. By dividing two polynomials the result is a rational function

$$R(x) = \frac{A_N(x)}{B_M(x)} = \frac{\sum_i a_i x^i}{\sum_j b_j x^j},$$

where one should keep attention not to mix the two independent summations. If the polynomial B_M has zeros, they must be excluded from the domain of definition. Otherwise, the domain of definition is the intersection of the domains of the individual domains, as for any composite function.

Also rational functions can be added, subtracted, and multiplied. Division by a polynomial is defined by

$$\frac{R(x)}{C_K(x)} = \frac{A_N(x)}{B_M(x)C_K(x)} = \frac{\sum_i a_i x^i}{\left(\sum_j b_j x^j\right)\left(\sum_k c_k x^k\right)}$$

where, as in (2.1), the multiplication has to be done in the usual form.

A very special case of a rational function is

$$A(x) = \frac{1}{x^n}$$

Which is thus just the division by a monomial. This is rewritten as x^{-n}. Dividing two monomials yields

$$A(x) = \frac{x^m}{x^n} = x^{m-n},$$

where the difference $m-n$ is the degree of the rational function, and which can be positive or negative.

2.2.3 Inverse functions

A question, which often arises, is if it is possible to find an x such that the equation

$$f(x) = y \tag{2.2}$$

is satisfied for a given y and given function $f(x)$. This requires, of course, y to be within the image of f, since otherwise there is no solution for x within the domain of definition of f. But this is not a sufficient condition, merely a necessary one. This distinction is a very important one in general. A necessary condition needs to be met for something to be true. But only if a sufficient criterion is met, it is guaranteed to occur.

In fact, if $A(x) = x^2$, then for x a real number there are two solutions for $y = 4$, $x = 2$ and $x = -2$. Thus, there is no unique solution to the equation (2.2). A unique solution only exists if for every element in the domain of definition exactly one element in the image exists, a relation called one-to-one. Then, such a solution exists, and it is called the inverse of the function f. This fact is written as

$$x = f^{-1}(y).$$

This shorthand notation should not be confused with dividing by f, which would be written as

$$\frac{1}{f(x)} = (f(x))^{-1}$$

and most importantly is a function of x and not of y. It is merely symbolic, and only in very few cases this may actually be literally. In fact, f^{-1} should be considered a different function than f, even though it is of course defined in terms of f.

The conditions under which such an inverse exist can be systematically discussed, but this is farther within the realm of analysis. However, subtraction can be regarded as the inverse function to addition, while division can be regarded as the inverse function to multiplication. Especially,

$$\begin{aligned} A(x) = x + b = y \quad &\Longrightarrow \quad x = y - b = A^{-1}(y) \\ B(x) = ax = y \quad &\Longrightarrow \quad x = \frac{y}{a} = B^{-1}(y). \end{aligned}$$

A solution for a polynomial of up to order four, which is invertible, can be explicitly constructed. In the case of a quadratic polynomial, this is the so-called *pq*-formula[2]

$$A(x) = ax^2 + bx + c = y \Longrightarrow x = A^{-1}(y) = \frac{-b \pm \sqrt{b^2 - 4a(c-y)}}{2a}$$

[2] A more formal discussion of roots will be given below in section 2.3.1. Here it is used that this is only a repetition, anticipating latter more formal developments.

The polynomial is only invertible if $b^2-4a(c-y)$ is zero, as otherwise there are two possible solutions. This should also emphasize that only the notion of being invertible is tied to the uniqueness of a solution. If more than one solution exists, it is said that the equation (2.2) is multivalued, and this is a quite common case.

Except for special values of the coefficient, it can be proven that it is not possible to invert a polynomial equation of order 5 or above such that the result becomes some closed formula, i. e. can be written in terms of elementary operations. In general, only a numerical solution is possible.

2.3 Special functions

Addition and multiplication, as well as their inverse, are just special cases of a more general class of mathematical operations. The most basic ones are the trigonometric ones, the exponential, and the power laws. These are the simplest example of so-called special functions, that is some kind of somehow defined mathematical operation which maps one number into another, but which cannot be expressed in a (finite number of) addition/subtraction and/or multiplication/division. These functions are sometimes called transcendental in opposition to (finite) polynomials.

2.3.1 Power laws and roots

The first example is the generalization of the monomials. So far exponents had to be integer numbers. It is a valid question whether this can be generalized. This can be best discussed with an example using inverse functions. Set x to be the solution of $x^2 = y$. In a sense, half a power of y solves this equation, and thus one defines the symbols

$$x = y^{\frac{1}{2}} = \sqrt{y} = \sqrt[2]{y}$$

to yield the solution to $y = x^2$. Of course, since the solution is multivalued, in principle the correct statement would either to be using a restricted domain of definition or to make both solutions explicit

$$x = \pm y^{\frac{1}{2}} = \pm\sqrt{y} = \pm\sqrt[2]{y}.$$

This defines what a half-integer power should mean, which is also called a (square)root. Especially, when restricting to positive x,

$$x = +\sqrt{x^2} = x = +(x^2)^{\frac{1}{2}} = +x^{2\frac{1}{2}} = x,$$

which implies that the multiplication of exponents proceeds as for monomials even for half-integer exponents.

This can be used to define what a rational exponent means. It instructs that $x^{p/q}$ is taking the qth root, i. e. the value which exponentiated by q will return back x, and then take this quantity to the pth power. Again, a convention must be chosen for possible signs. E. g. $27^{2/3}$ is calculated as $27^{1/3}$, which is 3, since 3^3 is again 27. It may not be -3, as -3^3 is -27. This leaves squaring 3, to arrive finally at $27^{2/3} = 9$.

To finally arrive at a definition for real numbers, it suffices to use that any real number can be arbitrarily approximated from above and below by a rational number, with the same limit. Thus, so can then x^a, where a is a real number, be determined by the results for taking the limit of x^{a_+} and x^{a_-}, where a_+ and a_- are rational numbers limiting a from above and below. This can be written as

$$\begin{aligned}
a_+ &= a + \epsilon \\
a_- &= a - \delta \\
a &= \lim_{\epsilon \to 0} a_+ = \lim_{\delta \to 0} a_- \\
x^a &= \lim_{\epsilon \to 0} x^{a_+} = \lim_{\delta \to 0} x^{a_-}
\end{aligned}$$

with ϵ and δ being chosen such that a_+ and a_- remain rational numbers. This defines finally a power-law x^a for arbitrary real numbers a.

These definitions ensure that calculating with real exponents remains the same as for integer exponents

$$\begin{aligned}
x^a x^b &= x^{a+b} \\
(x^a)^b &= x^{ab},
\end{aligned}$$

where division and taking a root translates into subtraction and division of exponents.

A logical possibility is to also consider $f(x) = a^x$. However, since for any fixed x this can be considered just as a function $f(a)$, this is not something new.

2.3.2 Logarithms

A question directly related to power-laws is, whether there is an inverse function for taking a power, in the sense that $f(x^a) = a$ for positive x. The answer to this is yes, though it is, like taking the root, implicitly defined. Such a function is called the logarithm, especially the logarithm to a special base. It is defined as

$$\log_x x^a = a,$$

that is, it depends in general on the x in question. Conversely,

$$x^{\log_x x^a} = x^a.$$

This implies $\log_x x = 1$ if $a = 1$ is selected and $\log_x 1 = 0$, since $\log_x x^0 = \log_x 1$. Note that since 0^a is zero for $a \geq 0$ or undefined for $a < 0$, the logarithm of zero is not defined. It also implies

$$\log_x x^a = a \log_x x.$$

Since it depends on x, this is called the logarithm to base x. Of course, using x as basis is rather inconvenient.

It is, however, possible to use a reference basis by

$$\log_x y = \frac{\log_z y}{\log_z x}, \tag{2.3}$$

which follows from

$$\frac{\log_z y}{\log_z x} = \frac{\log_z x^{\log_x y}}{\log_z x} = \log_x y \frac{\log_z x}{\log_z x} = \log_x y.$$

It is therefore possible to select a reference base. This basis is usually the Euler constant $e = 2.71828....$, for reasons which will become clear in section 4.7. A logarithm to base e is called a natural logarithm, and abbreviated by ln. In the following only this natural logarithm will be used. In case of need, it is always possible to revert to an arbitrary basis by usage of (2.3).

Note that $\ln x^a = a \ln x$. Also, $\ln e = 1$ and since $\ln 1^a = a \ln x$ for any a, this can only be true if $\ln 1 = 0$.

The asymptotics of the logarithm can be inferred as follows. If x becomes large, $\ln x$ becomes larger and larger, while it has to become negative infinite when x approaches zero. This can be seen from

$$x = e^{\ln x}.$$

Since $e > 1$, an increase on the left-hand requires an increase of $\ln x$. At the same time, there is no solution for $0 = e^a$ for any a, but if x is small, $\ln x$ must be negative, since $1/e^a$ becomes small for large a. Finally, $\ln 1 = 0$, to achieve $1 = e^{\ln 1}$.

The composition laws for exponents then immediately yields a further relation for logarithms

$$\ln xy = \ln e^{\ln x} e^{\ln y} = \ln e^{\ln x + \ln y} = (\ln x + \ln y) \ln e = \ln x + \ln y$$

and which can be generalized in a straightforward way to quotients.

A particular special importance has the function

$$\exp(x) = e^x,$$

which is called the exponential function, and is the inverse function to the natural logarithm,

$$x = \ln^{-1} \ln x = e^{\ln x} = x.$$

2.3.3 Trigonometric functions

Another important class of special functions are the trigonometric functions[3]. They are defined using elementary geometry. Start with a circle, drawn in a coordinate system with its origin coinciding with the point $x = y = 0$, and radius 1. Add a line which connects the center of the circle with any point on its rim. This line will enclose an angle α, measured in radians, i. e. from 0 to 2π instead of from 0 to 360°, as will be derived in more detail in section 8.1.2. The function cosine, $\cos\alpha$, is then defined as the one yielding the x coordinate of the point on the rim, while the function sine, $\sin\alpha$, is defined as the y coordinate. Thus, $\cos 0 = \cos 2\pi = 1$ and $\sin 0 = \sin 2\pi = 0$. Furthermore, $\cos\pi = -1$ and $\sin\pi = 0$ as well as $\cos\pi/2 = \cos 3\pi/2 = 0$ and $\sin\pi/2 = 1$ and $\sin 3\pi/2 = -1$. For angles larger than 2π a full rotation has been performed, and the values restart anew. The same is true by moving below 0. It is said that cosine and sine are periodic functions with a period of 2π, satisfying

$$\begin{aligned}\sin(x + 2\pi) &= \sin(x)\\ \cos(x + 2\pi) &= \cos(x)\end{aligned}$$

and anti-periodic over the half period of π

$$\begin{aligned}\sin(x + \pi) &= -\sin(x)\\ \cos(x + \pi) &= -\cos(x),\end{aligned}$$

and so on. Furthermore, $\sin(x + \pi/2) = \cos(x)$ and $\cos(x - \pi/2) = \sin(x)$.

Since the x and y coordinates are the edges of a right-angled triangle with hypotenuse of length 1, it directly follows from elementary geometry that

$$\sin^2\alpha + \cos^2\alpha = 1$$

Furthermore, it is possible to invert the functions sin and cos for a domain of definition $[-1, 1]$, which have an image $[0, 2\pi]$. These inverse functions are called arcsine and arccosine, denoted as arcsin and arccos or sometimes $\sin^{-1}$ and $\cos^{-1}$.

It is furthermore convenient to define also

$$\tan\alpha = \frac{\sin\alpha}{\cos\alpha},$$

the tangent of an angle[4]. This functions maps its argument to $[-\infty, \infty]$.

[3] Actually, they are related to power-laws, as will be discussed in section 6.

[4] Sometimes, there is also a function cotangent, $\cot x$ defined, which is just the inverse $1/\tan x$. Since its properties can be derived from those of $\tan x$, it will not be discussed here separately.

From the geometrical definition various trigonometric identities, or also called addition theorems, can be derived, e. g.

$$\cos^2 \alpha - \sin^2 \alpha = \cos(2\alpha). \tag{2.4}$$

These will not be derived or discussed here, as it will possible to obtain them much more conveniently later in section 6.2.

2.4 Function of multiple variables

It is possible to make a function dependent on two or more variables, e. g. $f(x, y)$ or $f(x_1, x_2, x_3)$. In this case, for every variable slot, the function has a separate domain of definition, and a final value is only obtained when values for each variable have been provided.

Consider

$$f(x, y) = x + \frac{1}{y}.$$

The domain of definition is $\mathbb{R}$ for x and $\mathbb{R}^+$ for y. Its final value will require to give a number for both x or y,

$$f(1, 2) = 1 + \frac{1}{2} = \frac{3}{2}.$$

However, if fewer are provided, the function can only be partially evaluated,

$$f(1, y) = 1 + \frac{1}{y}.$$

This result can be considered to be a new function with less variables. It is also valid to insert a new variable in both slots, e. g.

$$f(z, z) = z + \frac{1}{z},$$

or

$$f(z, 2z) = z + \frac{1}{2z}$$

creating yet again different functions of a single variable. If there are more than two variables, the same can be done for any subset of variables.

The image of such functions are usually harder to determine. In the present example, it is relatively straightforward and is $\mathbb{R}$. Functions of many variables are more usual than those of a single one in physics. If they are many variables, there carry usually an index, and writing the functions like

$$f(x_1, ..., x_n)$$
$$f(\{x_i\})$$

for a function depending on n variables is common. It is even possible to have functions depending on an infinite number of variables.

Chapter 3

Equations

The basic objects to be dealt with in (theoretical) physics will not only be functions, but equations like

$$h(x) = g(x),$$

with some arbitrary functions h and g. In this case, values of x are searched for, which this equality holds. Thus, manipulating equations is an important ingredient in doing (theoretical) physics.

3.1 Solving an equation

An example of an equation has been encountered before, e. g. the question what are the solutions x for the equation $y = x^2$ for a given functions $h(x) = x^2$ and $g(x) = y$?

Generically, any equation of a single variable x can be written as

$$0 = f(x) = h(x) - g(x), \tag{3.1}$$

sometimes called the normal form. E. g. the equation $x^2 = y$ can also be formulated like this with $f(x) = x^2 - y$. Equations can more generally be manipulated by performing the same operation on both sides. This can be regarded as adding a zero on one side,

$$\begin{aligned} f(x) &= a \\ f(x) + 0 &= a \\ f(x) + a - a &= a \\ f(x) - a &= 0, \end{aligned}$$

where in the last step it was recognized that the a on both sides can be dropped, as it is the same on either side. Thus, it is always possible to bring an equation to the form (3.1).

However, if the operation yields ambiguous results, it is necessary to track all possibilities. This is especially important when taking, e. g., a square-root,

$$x^2 - y + a = 0 \tag{3.2}$$

$$x^2 = y - a$$

$$x = \pm \sqrt{y-a} \tag{3.3}$$

since there are now two possible solutions to the equation.

Solving an equation is the reformulation of (3.1) such that it has the form

$$x = l, \tag{3.4}$$

where l is an expression involving only numbers and (known) constants. E. g. (3.3) is a case where the original equation has been solved for x as a function of the two constants y and a. It would then be possible to regard $l = \pm\sqrt{y-a}$ as a function of the two constants $l(y,a)$. Especially, this is the solution for any arbitrary values of y and a. It is not necessary to solve the equation again for each value of a and y. This is one of the advantages of working with placeholders instead of actual numbers.

However, this already shows an important feature. The equation (3.3) has no solution if y is smaller than a, as there is no(t yet a) root of a negative number. Thus, even if an equation is given in the normal form (3.1), as (3.3) for (3.2), this by no means guarantees that a solution exists. While this can be read off almost immediately from equation (3.2), this is in general a hard problem. Giving general answers under which conditions solutions to equations exist is an important part of mathematics. Formulating the equations, and determining actual solutions, if any exist, is an important part of physics.

It is also possible that there are more than one solution, and in fact any number of solution is possible. An extreme example is the equation

$$f(x) = 1 - \cos(x) = 0 \tag{3.5}$$

which has the solution

$$x = \cos^{-1}(1)$$

and thus yields the infinite set of solutions $2\pi \times n$ with n any integer number. Identifying the number of solutions (if any) is often as important (and sometimes even more important) than the actual solution.

While the equations (3.2) and (3.5) can be solved explicitly for x, this is in general not possible. Consider the equation in normal form

$$f(x) = x - \tan x = 0. \tag{3.6}$$

In this case, the function $f(x)$ involves the variable x in two different ways. Once as a monomial, and once again as an argument of the special function tan. Though it is not immediately obvious, one can proof that this equation cannot be solved such that the form (3.4) is obtained, which is called a closed form of the solution. The reason is that there is no possibility to isolate the x from the tan without at the same time making the monomial x a more involved function. This situation is actually quite common in physics. To find the solution to such equations, called implicit equations, requires other methods. In the present case, a solution could be found by drawing x and $\tan x$ in a coordinate system, and then locate the points where both cross. These will be the solutions of the equation, as there both functions have the same values. In more involved cases, so-called numerical methods are necessary, the domain of numerical mathematics to be explored in other courses. These form one of the most important tools in physics.

3.2 Inequalities

Equations like (3.1) are a special kind of relations. In general, a relation compares two expressions. A different kind of relations are inequalities, i. e.

$$f(x) \geq 0, \tag{3.7}$$

i. e. the requirement that the left-hand side is greater or equal than 0. Another possibility is excluding equality,

$$f(x) > 0,$$

the more stronger requirement that the left-hand side is greater than zero.

Since negative numbers are smaller than positive ones, any inequality of the type $a \leq b$ can always be turned into the type (3.7) by multiplying both sides by -1. Also, any contribution on the right-hand side can always be subtracted on both sides, to end up with the form (3.7),

$$\begin{aligned} f(x) &< a \\ f(x) - a &< 0 \\ a - f(x) = g(x) &> 0, \end{aligned}$$

and thus the normal form of an inequality.

Inequalities can be resolved, as before, thus getting conditions on the variable x e. g.

$$\begin{aligned} 2x + a &\geq 0 \\ x &\geq -\frac{a}{2}. \end{aligned}$$

The same as said on equations applies here as well. An inequality can have no solution, one solution, many solutions, or even an infinite number of solutions. But since inequalities are less demanding than equalities, the conditions on x to solve an inequality are often weaker than for an equation. Again, not all inequalities can be solved explicitly for x.

A thing which becomes more involved than for equations are operations with ambiguous results, like taking a root,

$$
\begin{aligned}
x^2 &\geq 4 \\
x &\geq 2 \text{ or } x \leq -2,
\end{aligned}
$$

and thus this may affect the type of relation. Great care needs to be taken here.

Chapter 4

Differentiation

One of the most fundamental questions in physics is the one of determining a rate of change, i. e. determining how much a certain quantity changes under the change of a parameter, in a very general sense. One of the most familiar examples is speed, which is the rate of change of position with time. But there are numerous (and often much more abstract) examples in physics.

In general, the question can be reformulated as: "For any given function $f(x)$, how much does this function changes when x is changed by some amount?". The calculation of this is called differentiation.

It should be noted that here only the differentiation of functions with a single variable will be addressed. There are additional complications for functions with multiple variables, which will be addressed in the lecture on analysis.

4.1 Definition and limiting process

The basic mathematical quantity of interest is how much a quantity f changes when its parameter x changes by some given amount. Denoting the changes as Δf and Δx, the searched-for quantity f' is given by their quotient

$$f' = \frac{\Delta f}{\Delta x}. \tag{4.1}$$

E. g. the function $f(x) = x^3$ changes for $\Delta x = 3 - 2 = 1$ by $\Delta f = 27 - 8 = 19$, and thus $f' = 19$.

A more interesting situation arises when the question is posed how large the rate of change of f is at a given point x, i. e. the local rate of change. Especially, this local rate of change is then again a function of x, i. e. the function $f'(x)$ is searched for, giving this quantity.

This quantity cannot be uniquely obtained by just taking ratios similar to (4.1), since the rate of change depends on how large the interval is (e. g. for $\Delta x = 4 - 1 = 3$ is $\Delta f = 64 - 1 = 63$ and thus $f' = 21 \neq 19$), and there is no unique way to specify where, within the interval, the point x should be located.

To avoid these problems, the solution is to shrink the interval further and further, such that it becomes arbitrarily small, called infinitesimally small, around the desired point. Take the size of the interval to be $\Delta x = h$ independent of x. Then this statement can be formulated as

$$f'(x) = \lim_{h\to 0} \frac{f\left(x + \frac{h}{2}\right) - f\left(x - \frac{h}{2}\right)}{h} = \frac{df(x)}{dx} \tag{4.2}$$

where the point to evaluate the function has been selected to be in the middle of the interval. That is actually not necessary, and any location within the interval will do. Intuitively, this is clear, as if the interval becomes arbitrarily small, any point within it should be equally well suited. A proof of this will be given in the analysis lecture. Note however, that, though this seems to be clear, there are functions for which this is not valid. Fortunately, such functions are encountered not too often in physics.

The version on the right hand side is a short-hand notation for the prescription of the left-hand side, and called the derivative. It should be noted that the value of x has not been specified in the process. Thus, the derivative itself is again a function of x. If the derivative should be evaluated at a certain value of x, it is often written as

$$\left.\frac{df(x)}{dx}\right|_{x=a}.$$

This requires to calculate the derivative and then to evaluate it at $x = a$. It must be kept in mind that this is no ordinary quotient, as the limiting process is involved. Thus, this is considered to be split into two, the operator d/dx, which is applied to/acted on the function $f(x)$, to yield the derivative $df(x)/dx$. Very often, the derivative is written just as df/dx, without the explicit marking of the dependency on x. There are also many other short-hand notations in use, like the already used f', $\dot{f}$, $d_x f$, and others. Since not all of them make the variable explicitly, they require context to correctly interpret.

To see how this works, try first $f(x) = x^n$ for $n = 0 - 3$,

$$\frac{d1}{dx} = \lim_{h\to 0} \frac{1-1}{h} = 0 \tag{4.3}$$

$$\frac{dx}{dx} = \lim_{h\to 0} \frac{x+\frac{h}{2}-x+\frac{h}{2}}{h} = \lim_{h\to 0} \frac{h}{h} = 1 \tag{4.4}$$

$$\frac{dx^2}{dx} = \lim_{h\to 0} \frac{\left(x+\frac{h}{2}\right)^2-\left(x-\frac{h}{2}\right)^2}{h} = \lim_{h\to 0} \frac{2hx}{h} = 2x \tag{4.5}$$

$$\frac{dx^3}{dx} = \lim_{h\to 0} \frac{\left(x+\frac{h}{2}\right)^3-\left(x-\frac{h}{2}\right)^3}{h} = \lim_{h\to 0} \frac{\frac{h^3}{4}+3hx^2}{h} = 3x^2 \tag{4.6}$$

There are a number of interesting observations.

The first appears obvious: A constant function does not change.

The second shows that a linear function has constant rate of change. More interestingly, it appears that the dx in the denominator has been canceled by the dx in the numerator. Though this is indeed correct in this case, much caution should be applied. This is already visible in the next line. Naive cancellation would yield x, but the correct result is $2x$. The final case shows another interesting feature. There two terms appear, but after cancellation, the first behaves as h^2, and therefore vanishes when the limit is taken, and only the second term survives this limit. Such situations, where multiple terms of different order appear in a limiting process, are quite common. Often the notation $\mathcal{O}(h)$ is used to indicate that something behaves 'like h'. In this case the first term behaved as $\mathcal{O}(h^2)$.

4.2 Differentiation of simple functions

The differentiation of sums is rather straightforward. Since in the definition (4.2) just the difference of two functions is required, the differentiation can be executed on each term separately, and the derivative is the sum of the derivatives term by term[1],

$$\frac{d}{dx}\sum_i^N f_i(x) = \sum_i^N \frac{df_i(x)}{dx} = \sum_i^N f_i'(x).$$

and likewise for subtractions.

Also, if a function is multiplied by some constant, the constant appears linearly in all terms, and can therefore be taken out of the differentiation,

$$\frac{daf(x)}{dx} = a\frac{df(x)}{dx} = af'(x),$$

and likewise for divisions by constants.

These leaves monomials as the elementary functions for the moment. Here, the result can be obtained by a process which is called 'proof by induction'. It is based on a guess/hypothesis/conjecture/whatever of the correct result.

Here, the interesting question is the derivative of x^n. Based on (4.3-4.6), a suitable assumption seems to be nx^{n-1}. For $n = 1$ (and $n = 0$), the answer is known. Assume now that the answer for x^{n-1} would be known, and check, whether from this the answer for x^n

[1]Note that infinite sums can be tricky, and care should be applied: For infinite sums, it is not always possible to exchange summation and differentiation as is done here.

can be inferred:

$$\begin{aligned}
\frac{dx^n}{dx} &= \lim_{h\to 0}\frac{\left(x+\frac{h}{2}\right)^n-\left(x-\frac{h}{2}\right)^n}{h}=\lim_{h\to 0}\frac{\left(x+\frac{h}{2}\right)\left(x+\frac{h}{2}\right)^{n-1}-\left(x-\frac{h}{2}\right)\left(x-\frac{h}{2}\right)^{n-1}}{h}\\
&= \lim_{h\to 0}\frac{x\left(\left(x+\frac{h}{2}\right)^{n-1}-\left(x-\frac{h}{2}\right)^{n-1}\right)+\frac{h}{2}\left(\left(x+\frac{h}{2}\right)^{n-1}+\left(x-\frac{h}{2}\right)^{n-1}\right)}{h}\\
&= x\lim_{h\to 0}\frac{\left(x+\frac{h}{2}\right)^{n-1}-\left(x-\frac{h}{2}\right)^{n-1}}{h}+\frac{1}{2}\lim_{h\to 0}\left(\left(x+\frac{h}{2}\right)^{n-1}+\left(x-\frac{h}{2}\right)^{n-1}\right)\\
&= x(n-1)x^{n-2}+x^{n-1}=nx^{n-1} \qquad (4.7)
\end{aligned}$$

where in step 4 it was used that x does not depend on h and can therefore be pulled out of the limit, and in step 5 that, by assumption, the derivative of x^{n-1} is known. The result implies that it is iteratively possible to reach the beginning, the so-called induction seed that $dx/dx = 1$, by applying the process repeatedly. Thus, the induction assumption is correct. This completes the proof.

4.3 Product rule

The previous results suggests that x^n could be viewed as xx^{n-1}, and the derivative would then be

$$\frac{dxx^{n-1}}{dx}=x\frac{dx^{n-1}}{dx}+\frac{dx}{dx}x^{n-1}=x(n-1)x^{n-2}+1x^{n-1}=nx^{n-1}$$

and thus that the derivative of a product is the sum of all possibilities to derive only one of the terms. This is indeed true, and called the product rule or Leibnitz rule,

$$\frac{d}{dx}(f(x)g(x))=\frac{df(x)}{dx}g(x)+f(x)\frac{dg(x)}{dx}.$$

For arbitrary polynomials, the product rule can be derived in the same way as before, as it is possible to break it down to a sum of monomials.

The general proof proceeds by the important concept of inserting a convenient zero,

$$\begin{aligned}
\frac{dfg}{dx} &= \lim_{h\to 0}\frac{f\left(x+\frac{h}{2}\right)g\left(x+\frac{h}{2}\right)-f\left(x-\frac{h}{2}\right)g\left(x-\frac{h}{2}\right)}{h}\\
&= \lim_{h\to 0}\frac{f\left(x+\frac{h}{2}\right)g\left(x+\frac{h}{2}\right)+f\left(x+\frac{h}{2}\right)g\left(x-\frac{h}{2}\right)-f\left(x+\frac{h}{2}\right)g\left(x-\frac{h}{2}\right)-f\left(x-\frac{h}{2}\right)g\left(x-\frac{h}{2}\right)}{h}\\
&= \lim_{h\to 0}\frac{\left(f\left(x+\frac{h}{2}\right)-f\left(x-\frac{h}{2}\right)\right)g\left(x-\frac{h}{2}\right)+f\left(x+\frac{h}{2}\right)\left(g\left(x+\frac{h}{2}\right)-g\left(x-\frac{h}{2}\right)\right)}{h}\\
&= \frac{df}{dx}g+f\frac{dg}{dx}.
\end{aligned}$$

The important step was adding a zero in step 2, which cannot change anything. In the final step, it was essential that the limit for both terms can be taken independently, and that $\lim_{h\to 0} g(x-h/2) = g(x)$, as g in this case does not appear as a difference.

The product rule can be used to generalize the result (4.7) to other exponents. First split some known exponent

$$nx^{n-1} = \frac{dx^n}{dx} = \frac{d}{dx}(x^a x^b) = x^a \frac{dx^b}{dx} + x^b \frac{dx^a}{dx}$$

This equality must hold for every $n \neq 0$ and x, and any splitting of n in $a+b=n$ with $a \neq 0$ and $b \neq 0$. Thus

$$\begin{aligned} x^a \frac{dx^b}{dx} &\sim x^{n-1} \\ x^b \frac{dx^a}{dx} &\sim x^{n-1}. \end{aligned}$$

This can only be true if $(x^c)' \sim x^{c-1}$. Likewise the prefactor must sum to n, which is then only possible for

$$\frac{dx^c}{dx} = cx^{c-1}.$$

Since no assumptions need to be made on the actual value of the exponents, this then necessarily works for any real numbers. E. g.,

$$\begin{aligned} \frac{d\frac{1}{x}}{dx} &= \tfrac{dx^{-1}}{dx} = -1x^{-1-1} = & -\frac{1}{x^2} \\ \frac{d\sqrt{x}}{dx} &= \tfrac{dx^{\frac{1}{2}}}{dx} = \tfrac{1}{2}x^{\frac{1}{2}-1} = & \frac{1}{2}\frac{1}{\sqrt{x}}. \end{aligned}$$

4.4 Chain rule

An often appearing situation is that it is necessary to differentiate a function of a function, $f(g(x))$, as introduced in section 2.1. In this case, differentiating with respect to x is not the same as differentiating with respect to the argument of $f(x)$. Take as an example $f(x) = x^2$ and $g(x) = 1 + x^2$. Then

$$\begin{aligned} \left.\frac{df}{dx}\right|_{x=g(x)} &= 2x|_{x=g(x)} = & 2(1+x^2) \\ \frac{d}{dx}f(g(x)) &= \tfrac{d}{dx}(1+x^2)^2 = \tfrac{d}{dx}(1+2x^2+x^4) = & 4(x+x^3), \end{aligned} \tag{4.8}$$

which is different. Both prescriptions are well defined, but the first is just taking an ordinary differential, and then apply the resulting function to some other function. It therefore yields nothing new. The second one, where the differentiation is applied to the argument of the argument of a function is different, and actually the much more interesting case in practice.

To obtain a general rule, the following formal manipulations can be done,

$$\begin{aligned}
&\lim_{h\to 0}\frac{f\left(g\left(x+\frac{h}{2}\right)\right)-f\left(g\left(x-\frac{h}{2}\right)\right)}{h}\\
=\ &\lim_{h\to 0}\frac{f\left(g\left(x+\frac{h}{2}\right)\right)-f\left(g\left(x-\frac{h}{2}\right)\right)}{g\left(x+\frac{h}{2}\right)-g\left(x-\frac{h}{2}\right)}\frac{g\left(x+\frac{h}{2}\right)-g\left(x-\frac{h}{2}\right)}{h}.
\end{aligned} \tag{4.9}$$

The first factor behaves as the ratio in the original definition of differentiation, (4.2), if the replacement $g(x\pm h/2)=x\pm h'(h)$ is made. Herein h' is some (usually unknown) function of h. However, the important statement is that $h'(0)=0$, and thus the limiting process can be performed likewise, though the approach to the limit may be slightly different[2]. The second term in (4.9) is just the ordinary expression for the differentiation of $g(x)$. Thus, taking the limit the chain rule

$$\frac{df(g(x))}{dx}=\frac{dg(x)}{dx}\left.\frac{df(x)}{dx}\right|_{x=g(x)}=g'(x)f'(g(x))$$

is obtained, where the last equality is the usual abbreviation.

For the example given above, $g'(x)=2x$ and $f'(x)=2x$, this yields $2x\times 2(1+x^2)=4(x+x^3)$, the same as (4.8), and thus as desired.

An interesting way to write the chain rule is

$$\frac{df(g(x))}{dx}=\frac{df(g(x))}{dg(x)}\frac{dg(x)}{dx},$$

which is just the statement that f has to be derived with respect to its argument, which in the present case just happens to be another function. This formally looks like an expansion of the fraction. Though it is usually possible to work in this context indeed as with the expansion of fraction, there are subtle cases where it is not true. Thus, in general, caution is advised. For the functions introduced here so far, this kind of expansion indeed works.

[2] There is a complication if both numerator and denominator go to zero in the limit. This requires some more careful work, done in the analysis courses, but yields the same result.

4.5 Quotient rule

An important combination of the product rule and the chain rule is the quotient rule,

$$\begin{aligned}\frac{d}{dx}\frac{f(x)}{g(x)} &= \frac{d}{dx}\left(f(x)\times\frac{1}{g(x)}\right) = \frac{f'(x)}{g(x)} + f(x)\frac{d}{dx}\frac{1}{g(x)}\\ &= \frac{f'(x)}{g(x)} - f(x)\frac{g'(x)}{g(x)^2} = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}\end{aligned}$$

where in the second step the product rule was used and in the third step the chain rule with $(1/x)' = -1/x^2$.

4.6 Rule of L'Hospitâl

Differentiation can also be helpful in a quite different context. Consider the case of $f(x)/g(x)$ at a point x where both $f(x) = g(x) = 0$. If only either of them would be zero, the situation is well-defined: If only f vanishes, the whole expression vanishes, if only g vanishes, the expression becomes infinite (or, more precisely, ill-defined). But what if both vanishes?

The situation could be written as

$$\lim_{h\to 0}\frac{f(x+h)}{g(x+h)} = \lim_{h\to 0}\frac{f(x+h)-f(x)}{g(x+h)-g(x)} = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}\frac{h}{g(x+h)-g(x)} = \frac{f'(x)}{g'(x)}. \tag{4.10}$$

Here, it was used that choosing to evaluate the functions in (4.2) at $x \pm h/2$ was arbitrary, and the same results would be obtained if $x+h$ and x would have been chosen instead. In the second step then it was used that at x both functions vanish, so adding them there is again just adding a zero. Finally, the rest is just an application of (4.2). Thus, in the situation at hand, the result of the fraction is given by the fraction of the derivatives, which is known as L'Hospitâl's rule.

If the result is not yet unique because still both derivatives vanish, this rule can be repeated as long as necessary. Take as an example $f(x) = x^2 - 1$ and $g(x) = x - 1$, which both vanish at $x = 1$. According to (4.10), the value of the expression is then $2x$, and therefore 2. If the situation in question is $f(x) = (x-1)^3$ and $g(x) = (x-1)^2$, the first application, using the chain rule, yields $3(x-1)^2$ and $2(x-1)$, which does not yet yields a result, but a second application yields $6(x-1)$ and 2, and thus the result is zero.

4.7 Differentiation of special functions

Being able to differentiate normal functions is quite important, but differentials of special functions beyond power-laws play a substantial role as well.

The differentiation of trigonometric functions can best be obtained from trigonometric identities. Start with

$$1 = \sin^2 x + \cos^2 x.$$

Differentiating on both sides yields

$$0 = \sin x \frac{d\sin x}{dx} + \cos x \frac{d\cos x}{dx},$$

which is equivalent to

$$\sin x \frac{d\sin x}{dx} = -\cos x \frac{d\cos x}{dx}.$$

This has to be true for any x. But this can only be if

$$\begin{aligned} \frac{d\sin x}{dx} &= g(x)\cos x \\ \frac{d\cos x}{dx} &= -g(x)\sin x \end{aligned}$$

where $g(x)$ is some common product function. However, differentiating (2.4) yields

$$4g(x)\sin x\cos x = 2g(2x)\sin(2x),$$

but there is a trigonometric identity

$$2\sin x\cos x = \sin 2x,$$

and thus the only possible solution is $g(x) = 1$.

The next special functions are logarithms. Since because of (2.3) any arbitrary base can be transformed into a quotient of logarithms to a different base, it is sufficient to consider the natural logarithm. Determining its derivative is actually quite non-trivial, and will be relegated to a different lecture. The result is

$$\frac{d\ln x}{dx} = \frac{1}{x},$$

and thus surprisingly a normal rational function.

With this at hand, the differentiation of $\exp(x)$ is straightforward,

$$1 = \frac{dx}{dx} = \frac{d\ln\exp(x)}{dx} = \frac{1}{\exp(x)}\frac{d\exp(x)}{dx} \tag{4.11}$$

and thus differentiating $\exp(x)$ yields again $\exp(x)$.

The logarithm can also be used to obtain a result for a^x,

$$\frac{1}{a^x}\frac{da^x}{dx} = \frac{d\ln a^x}{dx} = \frac{d(x\ln a)}{dx} = \ln a,$$

and thus

$$\frac{da^x}{dx} = a^x \ln a,$$

of which the derivative of $\exp x$ is thus just a special case.

4.8 Multiple differentiation

The derivative of a function is just another function. Thus, it is perfectly valid to differentiate it again. This yields the rate of change of the rate of change, written as

$$f'' = \frac{d}{dx}\frac{df(x)}{dx} = \frac{d^2 f(x)}{dx^2}.$$

It is very important in the last way to write a second derivative that this is not a differentiation with respect to x^2, which in the sense of a chain rule can also be done.

This can be repeated n times, which is written as

$$\frac{d^n f(x)}{dx^n}.$$

An example is

$$\frac{d^2 x^3}{dx^2} = \frac{d3x^2}{dx} = 6x.$$

4.9 Minima, maxima, and saddle points

Functions can, but do not need to, develop different particularly important structures. If they exist, they often signal particularly interesting physical phenomena. Being able to identify them is thus necessary.

One is the asymptotic behavior, i. e. what happens in the limits

$$\lim_{x\to\pm\infty} f(x).$$

If the function itself tends to $\pm\infty$ in either (or both) of the limits, it is called divergent in the corresponding limit(s). If there exists one or more special values x_i for which

$$\lim_{x\to x_i} |f(x)| = \infty$$

the function is said to have singularities at these points.

Two further structures are extrema: The maximum and minimum value of a function. Since a function can have multiple such extrema, it is necessary to distinguish between the concept of local (relative) and global (absolute) extrema. Local extrema are the most extreme function values in some (small) interval around them, while global extrema are the most minimal or maximal values of the function in its whole domain of definition. There can be multiple global extrema, if there are multiple points where the function takes its most extreme values. If there are multiple absolute minima or maxima, these are called degenerate. Note that a local minimum can have a larger function value than a local maximum. In these definitions the occurrences of divergencies or singularities are excluded. The latter are not part of the domain of definition of the function, and thus are excluded. Asymptotic divergency, on the other hand, implies that for any value of the argument, there is a value of the argument of larger absolute size for which the function is of larger absolute size. Thus, no matter how far the function is pursued, there is always a more extreme value, and thus the asymptotic behavior cannot yield extrema.

For example, the function x^4-5x^3+4x+2 has two minima, one around[3] $x \approx -0.49$ and one at $x \approx 3.7$. The one at negative x is much more shallow (≈ 0.68) than at positive x (≈ -49). It is therefore a local minimum, while the other is the global minimum. There is also a maximum at $x \approx 0.56$ (with value ≈ 3.5). This maximum is global, even though the function increases for $x \to \infty$ arbitrarily. It also illustrates that global extremes may not be the largest value of a function, if the function grows beyond all bounds for $x \to \pm\infty$. Take as another example $\cos(x)$. It is finite in the limits $x \to \pm\infty$. It has an infinite number of degenerate global minima and maxima at $x = 2n\pi$ and $x = (2n+1)\pi$.

An important observation is that the rate of change of the function at these points vanishes, i. e.

$$\frac{df}{dx} = 0. \tag{4.12}$$

This can be seen geometrically: The functions increases/decreases towards an extremum, and afterwards it needs to decrease/increase again, as otherwise it would keep on growing/diminishing, and therefore the point could not be an extremum. Therefore, the equation (4.12) can be used to determine the extrema. It is therefore a necessary criterion for an extremum. It is, however, not a sufficient criterion, as will be discussed below.

Before doing so, note that the equation (4.12) cannot distinguish between local extrema and global extrema. In both cases, the rate of change vanishes. The only way to determine this further classification requires in addition also the function values at all extrema of the

[3] Note that always the same number of significant, i. e. non-zero, digits is given in such approximations, no matter where the decimal point is. The $\approx$ signifies that the value is not exact.

same type, and compare them. That is a very important conceptual insight: The existence of extrema is a local information. To identify genuine extrema it is only necessary to know the value of the derivative at a point and in a (arbitrarily) small neighborhood[4] of this point. To know whether it is a global extrema requires to know all the extrema of a function and the value of the function at all its extrema. This is a global information. As will be seen, in physics it is often easy to get local information, but global information is (almost) impossible to get. Thus, answering the question whether an extremum is local or global belongs to the hardest questions in physics (and also mathematics).

As noted, there are special cases, where (4.12) is not a sufficient criterion to determine, whether there is an extremum or not. However, it remains a necessary condition, as at any extremum the rate of change still has to vanish. This will also illustrate the very important distinction of necessary and sufficient once more.

Consider the function x^3. Its derivative, $3x^2$ vanishes at $x = 0$. According to (4.12) it would therefore have an extremum. Plotting the function immediately shows that this is not true. What happens is that the function has a so-called saddle-point, or point of inflection, at $x = 0$, i. e. a point where the rate of change vanishes, but no extremum develops. This can only be true if the rate of change has afterwards again the same sign as before. Thus, the important information to distinguish saddle points is whether the rate of change of the rate of change is non-zero or not. If it is non-zero, but the rate of change is zero, the rate of change goes through zero, and has afterwards a different sign. Incidentally, this also permits to distinguish minima and maxima: A positive rate of change of the rate of change at an extremum is a minimum, if negative it is a maximum. If the rate of change of the rate of change, i. e. the second derivative of the function, vanishes

$$\frac{d^2 f(x)}{dx^2} = 0$$

it may be a saddle point. However, this is again not sufficient. E. g., for $-x^4$ both first derivatives vanish at $x = 0$, despite the fact that it has an extremum there. In this case, more information is needed.

This is obtained by performing further derivatives. The necessary condition to have an extremum is that the rate of change has a different sign on both sides of the extremum. This is the case when the first-non-vanishing derivative n is with n even, while there is a saddle-point if the first non-vanishing derivative has n odd. In the latter case the sign also indicates whether the rate of change at the given point is positive or negative. E. g. for x^3, the first non-vanishing derivative is the third and positive, and thus the rate of change

[4] Otherwise a constant function, with a derivative vanishing everywhere, could be said to have extrema everywhere.

before and after the saddle point is positive. The proof of this statement will be given in the analysis lecture.

However, this also illustrates that it is insufficient to know the value of the function at a given point to decide whether there is an extremum or a saddle-point. The definition of the derivative (4.2) requires knowledge of the function in an infinitesimal region around the point in question, and therefore probes a neighborhood of a point.

Chapter 5

Integration

Another question, which can be posed about a function, is what kind of area it encloses with the axis in a certain interval. In the simplest case, this question can be answered geometrically. E. g. the function x encloses with the x-axis on the interval $[0, 1]$ the area $1/2$, as it is an orthogonal triangle. The question becomes more involved when thinking about a function $x^8 \sin x$, and a geometrical solution appears to be at least cumbersome. The answer to this question, and its generalization, is integration.

Before embarking on a formal definition, there is an interesting question to be solved. What is the area enclosed by x with the x-axis in the interval $[-1, 0]$? It appears reasonable to just say again $1/2$. However, this solution turns out to be inconvenient when applying the concepts of integration to more general problems, as is required in physics. A better solution is to introduce the concept of a signed area, i. e. counting the area above the x-axis as positive and below the x-axis as negative. The result would then be $-1/2$, and the result for the interval $[-1, 1]$ would be zero. That x encloses zero area with the x-axis in this interval appears at first sight counter-intuitive, but, as stated, will be mathematical convenient. If indeed the area, rather than the signed area, is required, this could be obtained from the function $|x|$, having a total area of 1. Similarly, for any function $f(x)$ the area rather than the signed area can be obtained by calculating the area of $|f(x)|$ instead.

5.1 Riemann sum

Similar to differentiation, the key to calculate an integral is again performing a limiting procedure. Given a function $f(x)$, the (signed) area A in the interval $[a, b]$ is certainly approximated by

$$A=\sum_{i=0}^{N-1}\Delta x f(x_i),$$

where the interval has been dismantled into N equal subintervals, each of length $\Delta x = (b-a)/N$. The x_i are arbitrary points inside the interval $[a+i\Delta x, a+(i+1)\Delta x]$. In the end, how the points are selected will (usually) not matter, as will be shown in the analysis lecture. A convenient choice is at the center of each of the subintervals

$$x_i=a+\left(i+\frac{1}{2}\right)\Delta x.$$

Each term is therefore an approximation of the signed area in each interval. This is called a Riemann sum.

This approximation becomes better and better, just from geometry, when the size of the interval shrinks, i. e. N is made larger. Taking the limit

$$A=\lim_{N\to\infty}\sum_{i=0}^{N-1}\Delta x f(x_i)=\int\limits_a^b dx f(x) \tag{5.1}$$

will then yield the total signed area of the function. This is also called the integral of the function $f(x)$ over the interval $[a,b]$. The second expression is then a convention to express that the limit has been taken. Just like d/dx represents the differentiation, the expression $\int_a^b dx$ represents obtaining the integral, called performing an integration. Like differentiation, this is also called an operator: The differentiation operator and the integration operator. They act on the function $f(x)$. These are just two examples of operators; in physics (and mathematics); there will be many more.

Note that as a formal convention

$$\int\limits_a^b dx f(x)=-\int\limits_b^a dx f(x)$$

for any function $f(x)$ and

$$\int\limits_a^a dx f(x)=0,$$

as the area of a line is geometrically zero, no matter the sign.

5.2 Integral of a simple function

While the formal definition is nice, it is necessary to also make the results explicit. To show that this indeed calculates the area, start with $f(x) = x$ on the interval $[a, b]$. Then

$$\begin{aligned}
\int_a^b dxx &= \lim_{N\to\infty}\sum_{i=0}^{N-1}\frac{b-a}{N}\left(a+\left(i+\frac{1}{2}\right)\frac{b-a}{N}\right)\\
&= \lim_{N\to\infty}\frac{b-a}{N}\sum_{i=0}^{N-1}\left(a+\frac{1}{2}\frac{b-a}{N}+i\frac{b-a}{N}\right)\\
&= \lim_{N\to\infty}\frac{b-a}{N}\left(Na+\frac{1}{2}(b-a)+\frac{N-1}{2}(b-a)\right)\\
&= (b-a)a+\frac{1}{2}(b-a)^2 = ab-a^2+\frac{1}{2}(b^2-2ab+a^2) = \frac{b^2-a^2}{2}.
\end{aligned}$$

Here, in the third step use has been made of the fact that the finite sums can be calculated as

$$\begin{aligned}
\sum_{i=0}^{N-1} 1 &= N\\
\sum_{i=0}^{N-1} i &= \frac{N(N-1)}{2}.
\end{aligned}$$

In the last step, only those terms will survive, and stay finite, which are independent of N, yielding the result. This is precisely the result which is expected from geometry, as it is the area of the corresponding triangle, if $a \geq 0$. It is also visible that the signed area is zero, if $a = b$.

5.3 Integration and differentiation

Before continuing on, it is useful to consider the following question: Given a function $f(x)$, what is the integral on the interval $[a, b]$ of its derivative? Using the two definitions (4.2) and (5.1), this yields the following:

$$\int_a^b dx\frac{df}{dx} = \lim_{N\to\infty}\sum_{i=0}^{N-1}\Delta x\frac{df}{dx}(x_i) = \lim_{N\to\infty}\sum_{i=0}^{N-1}\Delta x\lim_{h\to 0}\frac{f\left(x_i+\frac{h}{2}\right)-f\left(x_i-\frac{h}{2}\right)}{h}$$

This expression can be simplified by noting that the way the intervals are created is free. It is therefore perfectly permissible to rewrite it as

$$\lim_{N\to\infty}\lim_{h\to 0}\sum_{i=0}^{N-1}\Delta x\frac{f\left(x_i+\frac{h}{2}\right)-f\left(x_i-\frac{h}{2}\right)}{h}.$$

Now, the only requirement is that the intervals Δx should shrink. Instead of using a division $\Delta x = (b-a)/N$, it is therefore possible to split the interval $[a,b]$ into intervals of size h with $N(h) = (b-a)/h$. Then

$$\begin{aligned}
&\lim_{h\to 0}\sum_{i=0}^{N(h)-1} h\frac{f\left(x_i+\frac{h}{2}\right)-f\left(x_i-\frac{h}{2}\right)}{h}\\
=&\lim_{h\to 0}\sum_{i=0}^{N(h)-1}\left(f\left(a+\left(i+\frac{1}{2}\right)h+\frac{h}{2}\right)-f\left(a+\left(i+\frac{1}{2}\right)h-\frac{h}{2}\right)\right)\\
=&\lim_{h\to 0}\sum_{i=0}^{N(h)-1}\left(f(a+(i+1)h)-f(a+ih)\right)=f(b)-f(a)=f(x)|_a^b
\end{aligned}$$

where in the second-to-last step it was used that now each function appears precisely twice, except for the evaluation of the functions at the end-point, which thus remain. The last equality is just a convention to write the result.

Thus, in a sense, the integral is the inverse to a differentiation[1]. This provides now a powerful way how to determine the integrals of functions f: Just find a function whose derivative is the function in question, the so-called primitive F, and evaluate it at the edges of the interval:

$$\int_a^b dx f(x) = F(b)-F(a) = F(x)|_a^b \tag{5.2}$$

$$\frac{dF(x)}{dx} = f(x). \tag{5.3}$$

This is the celebrated central theorem of integration and differentiation.

Before using this trick to determine the integrals of various functions, it is useful to introduce another concept.

5.4 Indefinite integrals

As is visible from (5.2), the actual interval on which the integral is performed plays not an important role at all. Just knowing the primitive is sufficient to solve the original question of determining the integral. Once the primitive is known, the calculation of the integral is merely an exercise in evaluating functions. Thus, the result for an arbitrary interval,

[1]As always, some subtleties are involved when playing around with the limits, but for most functions this procedure is well-defined.

the primitive itself, is really the interesting question. It has therefore its own name, the so-called indefinite integral, written as

$$\int dx f(x) = F(x) + C,$$

where no interval is indicated. However, the primitive is not uniquely defined. As the only requirement is that its derivative equals $f(x)$, (5.3), any function which has this derivative will do. Thus, it is always possible to add to a primitive a function with vanishing derivative. Since the only function with vanishing derivative is a constant, this implies that the function to be added to the primitive can be at most a constant, called the integration constant C. Note that this does not alter the definite integral on an interval $[a, b]$:

$$\int\limits_a^b dx f(x) = (F(x) + C)|_a^b = (F(b) + C - F(a) - C) = F(b) - F(a).$$

Knowing the indefinite integrals grants therefore all required knowledge.

Indefinite integrals can be considered also as a particular kind of definite integral. Assume that there is some x_0 for which $F(x_0) = -C$. Then

$$\int\limits_{x_0}^x dy f(y) = F(x) - F(x_0) = F(x) + C. \tag{5.4}$$

Of course, this is formally only correct if C can be rewritten in this form.

It is worthwhile to remark that actually finding the primitive of a given function is by no means simple in general. In fact, it can be proven that there are functions for which it is impossible to write down the primitive in closed form, i. e. as some, arbitrarily complicated, combination of the functions introduced in chapter 2.

5.5 Integrals of functions

Using the result of section 5.3 permits to calculate the primitives, and thus integrals, of all the basic functions, using the results from chapter 4. Just inverting the differentials yields

$$\begin{aligned}
\int dx \sum_i a_i x^i &= C + \sum_i \frac{a_i}{i+1} x^{i+1} \\
\int dx x^a &= C + \frac{1}{a+1} x^{a+1} \text{ if } a \neq -1 \\
\int dx \frac{1}{x} &= C + \ln x \\
\int dx \sin(x) &= C - \cos(x) \\
\int dx \cos(x) &= C + \sin(x) \\
\int dx \ln(x) &= C - x + x \ln x \\
\int dx e^{ax} &= C + \frac{e^{ax}}{a} \\
\int dx a^x &= C + \frac{a^x}{\ln(a)}
\end{aligned}$$

Obtaining definite integrals, provided the primitive exists on the domain of integration, is then a straightforward procedure by just evaluating the primitives at the corresponding edges of the intervals.

5.6 Multiple integrals

Since the indefinite integral of a function is again a function, it is possible to repeat the integration multiple times, e. g.

$$\int^x dy \int^y dz z = \int^x dy \left(C + \frac{y^2}{2} \right) = D + Cx + \frac{x^3}{6}$$

Of course, the integration variable needs to change for each of them. This is indicated by adding the new variable as an upper limit, in spirit of (5.4), but dropping the explicit form of the integration constant from the lower limit. The important thing to notice is that every integration produces a new integration constant, here C and D, which have to be integrated in every follow-up integration as well. Other than that, it is indeed always searching again the primitive of the function obtained as the primitive of the prior integral.

5.7 Partial integration

The product rule can be inverted using an integral as well. However, it is more interesting to use it in the following way

$$\int_a^b dx f(x)\frac{dg(x)}{dx} = -\int_a^b dx\frac{df(x)}{dx}g(x) + \left(f(x)g(x)\right)\Big|_a^b \tag{5.5}$$

which is obtained by integrating the product rule and putting one of the terms on the right-hand side. This implies that the differentiation can, up to a minus sign, be shifted from one function to the other, if an appropriate boundary term is added. This boundary term is the last term on the right-hand side. This is called partial integration. Note that this is also possible if the integral is indefinite, but then an additional constant has to take care of the the unspecified boundary term.

This result has two major applications. One is the integration of complex functions. E. g., integrating $x\sin(x)$ can be simplified by this:

$$\int_a^b dx x\sin(x) = -\int_a^b dx 1\times(-\cos(x)) + x(-\cos(x))\big|_a^b = (\sin(x) - x\cos(x))\big|_a^b$$

With this approach, it is often, though not always, possible to reduce complicated integrals to known integrals, at the expense of picking up boundary terms. Since the previous result is independent of the actual domain of integration, this also yields the indefinite integral

$$\int dx x\sin(x) = C + \sin(x) - x\cos(x).$$

Of course, this result could also be obtained by differentiating the primitive, but it is not so easy to guess it in general.

The other major application, very often encountered in physics applications, is when the boundary term vanishes. Then, the differentiation operator can be swapped around in the integral as desired, provided the minus sign is kept track of.

5.8 Reparametrization

Inverting the chain rule is another important result for integrals, the so-called reparametrization. It is in general very useful again to turn an integral into a simpler integral.

Start out first with an indefinite integral of a function $f(x)$. This function has a certain dependency on its variable, which may be involved. However, it may be that the primitive of f would be known, if the variable x could be replaced by a new variable y, which has a certain dependency on x. But this can then be used to place an inverse chain rule into the integral as

$$\int dx f(x) = \int dy \frac{dx}{dy} f(x) = \int dy \frac{dx}{dy} f(x(y))$$

where the appearing derivative $\frac{dx}{dy}$ is called the Jacobian[2]. It should be noted that this requires to calculate the inverse function $x(y) = y^{-1}(x) = x$ to calculate. Furthermore, the resulting expression will only be helpful if the resulting function to be integrated, the integrand is actually easier to integrate than the original. Surprisingly, this is often the case, but usually this is far from obvious.

It appears as if this has just extended the dx by $(dx/dy)dy$. This is actually true, and going back to the definition of both the Riemann sum and the derivative, this can be proven, but this will be skipped here. However, this has a further consequence for the limits of the integral, if this is done. They need to be modified too, as they are also transformed. This implies that reparametrization for a definite integral is given by

$$\int\limits_a^b dx f(x) = \int\limits_{y(a)}^{y(b)} dx \frac{dx}{dy} f(x(y)).$$

Fortunately, in most cases reparametrization is used in physics for indefinite integrals, and thus do not need to calculate the exchange boundaries. The whole procedure is somewhat involved, and thus requires some examples to understand.

Consider first the case $f(x) = x^2$ to be integrated from 1 to 2. The direct result is

$$\int\limits_1^2 dx x^2 = \left.\frac{x^3}{3}\right|_1^4 = \frac{7}{3}.$$

A possibility would be to select $x = +\sqrt{y}$, with the inverse $y = x^2$. The necessary derivative is then

$$\frac{dx}{dy} = \frac{1}{2\sqrt{y}}$$

yielding

$$\frac{1}{2}\int\limits_1^4 dy \frac{1}{\sqrt{y}} y = \frac{1}{2}\int\limits_1^4 dy \sqrt{y} = \left.\frac{y^{\frac{3}{2}}}{3}\right|_1^4 = \frac{7}{3}$$

While this does not directly show any advantage, the general program was straightforward.

[2]Actually Jacobian determinant, as will be discussed in the lecture on multiple integration.

To see an advantage, consider the following example, where $x = \sin(y)$,

$$\int dx \sin^{-1} x = \int dy \frac{d\sin(y)}{dy} \sin^{-1} \sin y = \int dy y \cos y = \int dy \frac{d}{dy}(y \sin y + \cos y)$$
$$= \quad y \sin y + \cos y = y \sin y + \sqrt{1 - \sin^2 y} = x \sin^{-1}(x) + \sqrt{1 - x^2}$$

Thus, it was possible to reduce the integration of the complicated function $\sin^{-1}$ back to the simpler ordinary trigonometric functions. There has also been used that the expression $y \cos x$ is the result of a product rule, i. e. integration by parts has been used. If limits would have been present, it would also only be necessary to know the function $\sin^{-1}$, but neither its integral, nor its derivative. Such applications are the dominant ones for the reparameterization: Make an integral simpler to perform.

Chapter 6

Complex functions

6.1 The imaginary unit

One of the problems encountered in the solution of equations is that within the real numbers the equation

$$x^2 = -1$$

has no solution. It is now by far a non-trivial statement that it is possible to solve this problem. The solution is to define a new quantity, called[1] i, the imaginary unit, as the solution to this equation. I. e., by definition, the symbol i has the meaning

$$i = +\sqrt{-1}$$

and it is therefore a new kind of number, as no real number has this property.

While it is certainly nice to define the solution to an equation, rather than to obtain it, it is then also necessary to show that this makes sense, i. e. that this is a number which can be used for anything else. This is obtained by first defining that i can be multiplied by a real number and added to a real number defining the so-called complex numbers

$$z = a + ib$$

for arbitrary real numbers a and b.

The next step is to define the addition/subtraction of two complex numbers

$$z = z_1 \pm z_2 = (a_1 \pm a_2) + i(b_1 \pm b_2).$$

I. e. if two complex numbers are added/subtracted, the parts proportional to i, called the imaginary parts and denoted by $\Im z_i$ are added/subtracted, and so are the remainder, the

[1] In engineering, it is sometimes called j.

so-called real parts $\Re z_i$, to define the new real and imaginary parts

$$\begin{aligned}\Re z &= \Re z_1 \pm \Re z_2\\ \Im z &= \Im z_1 \pm \Im z_2.\end{aligned}$$

This definition satisfies the ordinary rules of addition and subtraction. For vanishing imaginary parts this reduces to the ordinary addition/subtraction.

Multiplication is a bit more complicated. The basic tenant must again be that for zero imaginary part the original multiplication reappears. Furthermore, $i^2 = -1$ must be preserved for consistency. The solution is to use the binomial formula to obtain

$$\begin{aligned}z_1 z_2 &= (a_1 + ib_1)(a_2 + ib_2) = a_1 a_2 + ia_1 b_2 + ia_2 b_1 - b_1 b_2 = (a_1 a_2 - b_1 b_2) + i(a_1 b_2 + a_2 b_1)\\ \Re(z_1 z_2) &= \Re z_1 \Re z_2 - \Im z_1 \Im z_2\\ \Im(z_1 z_2) &= \Re z_1 \Im z_2 + \Im z_1 \Re z_2.\end{aligned}$$

This formula has all the desired properties. Interestingly, the square of a complex number is then

$$z^2 = (a_1 + ib_1)^2 = (a_1^2 - b_1^2) + 2ia_1 b_1.$$

Especially, the square of a number can be negative. This is necessary, as otherwise $i^2 = -1$, cannot be maintained, since i is just a particular complex number. Note that still only zero squares to zero.

Division by a complex number is another thing to define, as it should still be possible to invert multiplication. To start, note that if the division should be the inversion of the multiplication, dividing by the same number must yield 1, and thus

$$1 = \frac{i}{i} = i \times \frac{1}{i} \to -i = \frac{1}{i}.$$

Thus, the inverse of i is $-i$. Similarly, solving $zw = 1$ for the real and imaginary part of w yields

$$\begin{aligned}\Re w &= \Re\tfrac{1}{z} = \frac{\Re z}{(\Re z)^2 + (\Im z)^2}\\ \Im w &= \Im\tfrac{1}{z} = -\frac{\Im z}{(\Re z)^2 + (\Im z)^2}\end{aligned}$$

and thus division mixes both real and imaginary parts. Note that when $\Im z = 0$, this reduces to the ordinary division, and the inverse of a real number has no imaginary part. Also, all other properties of divisions are maintained.

This generalizes the basic mathematical operations to complex numbers. Before going to functions of complex numbers, it is worthwhile to investigate some of their geometric properties.

6.2 The complex plane and the Euler formula

Real numbers can be represented as a line. This is no longer possible for complex numbers, as for any point of a line identified, e. g. by the real part (or any function of the real and imaginary part), there is an infinite range of values the imaginary part can take. Hence there is also no ordering in the sense of bigger or lesser for complex numbers, only the question of equality, which requires both the real and imaginary parts to agree.

Thus, it is necessary to take this into account, by plotting any complex number inside a plane. The x coordinate can then be taken to be the real part, and the y coordinate is the imaginary part. Thus, any complex number is uniquely identified with a particular point in this so-called complex plane. E. g., the imaginary unit has the coordinates $x = 0$, $y = 1$. Any complex number obtained by a basic mathematical operation is then also uniquely mapped to a point in the plane. Especially, addition is adding the x coordinates and the y coordinates separately. Multiplication and division have no such simple geometrical interpretation, but the map exists nonetheless.

It leads to an interesting insight to observe that every complex number z can be seen as a point in a rectangular triangle. The one edge has then the length of the x coordinate or real part, and the other edge the length of the imaginary part. The hypotenuse of the triangle is

$$\rho = \sqrt{(\Re z)^2 + (\Im z)^2},$$

which is called the absolute value $|z|$ of the complex number z. The angle, measured with respect to the x-axis is given by

$$\tan\theta = \frac{\sin\theta}{\cos\theta} = \frac{\Im z}{\Re z} = \tan\arg z$$

The last abbreviation denotes the argument $\theta = \arg z$ of a complex number. This implies that a real number has argument zero. These two geometrical quantities permits to rewrite any complex number as

$$z = \rho\cos\theta + i\rho\sin\theta = \rho(\cos\theta + i\sin\theta),$$

with the only exception of $z = 0$, as there the angle is not well-defined. It is often conventionally used that zero is real, and thus $\theta = 0$, but this is strictly speaking not correct, and the domain of definition of arg is in principle only the complex numbers without zero.

There is an interesting result, which can be obtained from these geometrical ideas. Take a complex number on the unit circle, i. e. a number for which $\rho = 1$. Then derive the number with respect to the angle θ,

$$\frac{d(\cos\theta + i\sin\theta)}{d\theta} = -\sin\theta + i\cos\theta = i(\cos\theta + i\sin\theta).$$

Thus, up to a factor i, just the number is reproduced. But this behavior is the one of the exponential function. Thus

$$\cos\theta + i\sin\theta = e^{i\theta},$$

the so-called Euler's formula[2]. This implies that for any complex number

$$z = \rho e^{i\theta}$$

and establishes one of the most useful and important relations in complex function theory. It should be noted that due to Euler's formula $\exp(n2\pi i) = 1$ for n any integer, including zero. Thus, the argument is periodic.

To provide an example, this also permits a relatively simple way to derive trigonometric identities, using e. g.

$$e^{i(\theta+\omega)} = e^{i\theta}e^{i\omega} = (\cos\theta\cos\omega - \sin\theta\sin\omega) + i(\cos\theta\sin\omega + \sin\theta\cos\omega)$$

and comparing real and imaginary parts on both sides yields

$$\begin{aligned}\cos(\theta+\omega) &= \cos\theta\cos\omega - \sin\theta\sin\omega\\ \sin(\theta+\omega) &= \cos\theta\sin\omega + \sin\theta\cos\omega,\end{aligned}$$

which are not entirely trivial to derive geometrically.

6.3 Complex conjugation

The two independent elements of a complex number permit to define a new elementary operation on a single complex number, the complex conjugation

$$z^* = \Re z - i\Im z.$$

This, at first sight rather innocuous, definition has an intimate relation to the absolute value ρ of the complex number, since

$$\rho^2 = zz^* = z^*z = (\Re z + i\Im z)(\Re z - \Im z) = (\Re z)^2 + (\Im z)^2 = |z|^2$$

[2]There are more formal proofs to be encountered in other lectures.

where the last equality is the complex extension of the absolute value function of a real number where there it just delivered the number without its sign. Hence, the complex conjugate can be used to extract the absolute value of a complex number. In complex analysis many powerful results will be derived using the interplay of complex numbers and their conjugates.

Note also

$$z^* = \rho e^{-i\theta},$$

i. e. in Euler's formula complex conjugation just reverses the sign of the argument.

6.4 Simple functions of complex variables

So far, the operations defined on complex numbers are ordinary addition, multiplication and their inverse. This immediately also defines polynomials.

Powers of complex functions are most straightforwardly introduced using Euler's formula,

$$z^\alpha = \left(\rho e^{i\theta}\right)^\alpha = \rho^\alpha e^{i\alpha\theta}.$$

The exponentiation of the complex exponential assumes tacitly that this also works for complex numbers which it indeed does, as will be proven in another lecture. This reduces the powers of complex numbers to those of ordinary numbers for the length of the number, and to a multiplication for the argument. Note in particular that for an integer n

$$\sqrt[n]{e^{i\theta}} = \left(e^{i\theta}\right)^{\frac{1}{n}} = (1)^{\frac{1}{n}} e^{\frac{i\theta}{n}} = e^{i\frac{2\pi m}{n}+\frac{i\theta}{n}} \text{ for } m = 0...n-1$$

and thus the nth root of a complex number has n possible results, due to the periodicity of the argument. The standard case of a real number having two possible roots with either a plus sign or a minus sign is then a special case. Especially, for $\theta = 0$, these so-called roots of unity lie on n evenly spaced point on the unit-circle, beginning with 1. Especially, odd roots have only one real root, while even roots have two real roots, at $+1$ and -1.

Exponentials of a complex number can then also be directly defined as to be reduced to their real and imaginary part

$$e^z = e^{\Re z + i\Im z} = e^{\Re z} e^{i\Im z},$$

and thus for the exponential of a complex number its absolute value is the exponential of the real part, $\rho = \exp \Re z$, while its argument is the imaginary part $\theta = \Im z$.

Taking the sine or cosine of a complex number can be reduced to exponentials using Euler's formula, e. g.

$$\begin{aligned}
\cos z &= \frac{1}{2}\left((\cos z + i\sin z) + (\cos z - i\sin z)\right) = \frac{e^{iz}+e^{-iz}}{2} = \frac{e^{i\Re z-\Im z}+e^{-i\Re z+\Im z}}{2}\\
&= \frac{1}{2}\left(e^{-\Im z}\left(\cos\Re z + i\sin\Re z\right) + e^{\Im z}\left(\cos\Re z - i\sin\Re z\right)\right).
\end{aligned}$$

A similar formula can be obtained for the sine. This reduces trigonometric functions of complex numbers to functions of real numbers, thereby defining how to evaluate them.

A particular interesting case is the one of a pure imaginary number iy with y real. Then

$$\begin{aligned}
\sin iy &= i\frac{e^y - e^{-y}}{2} = i\sinh y\\
\cos iy &= \frac{e^y + e^{-y}}{2} = \cosh y\\
\tan iy &= \frac{\sin iy}{\cos iy} = i\frac{\sinh y}{\cosh y} = i\tanh y.
\end{aligned}$$

The so-defined hyperbolic functions sinh, cosh, and tanh play an important role in many aspects of physics. There is nothing special about them, given their definition in terms of e-functions. The later property is also very useful in determining their inverse,

$$\begin{aligned}
\sinh^{-1} x &= \ln\left(x + \sqrt{x^2+1}\right)\\
\cosh^{-1} x &= \ln\left(x + \sqrt{x^2-1}\right)\\
\tanh^{-1} x &= \frac{1}{2}\ln\frac{1+x}{1-x}
\end{aligned}$$

and their other properties, like derivatives and integrals.

A much more tougher problem is the definition of the logarithm of a complex number. It appears easy enough to define it just as the inverse of the exponential

$$\ln e^z = z$$

and thus

$$\ln z = \ln \rho e^{i\theta} = \ln\rho + i\theta.$$

But this shows already the problem. The argument θ can be changed by 2π without changing the original number, but the logarithm, where the argument is added, would change. Thus, the logarithm of a complex number is not well defined. There are deep reasons for that to be discussed in the lecture on function theory. The operative resolution of this is to define the argument of the logarithm of a complex number to be always between

π and $-\pi$, which leaves open of how to define the value on the negative real axis. As a consequence, the negative real axis, including zero, is not considered to be part of the domain of definition of the logarithm. Again, this issue will be taken up in the lecture on function theory.

Chapter 7

Probability

Probability plays a relevant role in classical physics, especially thermodynamics. It becomes totally indispensable in quantum physics. Though it then becomes quite different from what one usually understands as probability theory.

7.1 Combinatorics

The most basic questions in probability theory requires first to answer some very particular questions in combinatorics, i. e. the question of how to calculate the number of possibilities. So the following is a prelude to the problem of probability theory proper.

The first question is, if there are n distinct numbers, how many different ways are there to arrange them. For one number it is trivial, there is only one possibility. For two numbers, there are two possibilities: $\{1, 2\}$ can be arranged as 1,2 and 2,1, so there are two possibilities. For three, say $\{1, 2, 3\}$, there are 1,2,3, 1,3,2, 2,1,3, 2,3,1, and 3,2,1, and thus 6. The question can be answered by taking first one number out of the n. Then there remain $n - 1$ numbers to chose, and thus there are $n(n - 1)$ possibilities for arranging. Going on to the last, there are then

$$n(n-1)...(n-n+2)(n-n+1) = n!,$$

where the so-defined operation is called faculty. This can be proven, e. g., by induction. One furthermore defines $0! = 1$ for convenience, though from a combinatorics point of view this is an ill-defined question.

A related, but different, question is, how many different subsequences, without ordering, of fixed length k can be obtained from a set of n numbers. E. g. from the set $\{1, 2, 3\}$

there are three 1-element sets, 1, 2, and 3. The answer is given by the binomial coefficient,

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

which can be understood as following. In total, there are $n!$ possibilities to arrange the numbers. But since the sequence does not matter, the $k!$ possibilities to arrange that numbers are all equal, and therefore, this has to be divided by it. Also, it does not matter how the remaining numbers are arranged, and therefore the result has to be divided by the $(n-k)!$ possibilities to arrange them.

These two basic combinatoric formulas are essentially the basis for almost all combinatoric problems in physics.

7.2 Single experiments

An important application of combinatorics is to determine probabilities. Especially in quantum mechanics, but also in some cases of classical mechanics, it is extremely important to determine the probability of some event to occur.

In the simplest case, there is the situation that something occurs in $100p\%$ of the cases, where p is a number between, and including, zero and one. Then its probability to occur is just p. The total probability of something to occur is 1, and thus the probability for this not to occur is hence $1-p$.

In general, if an experiment can have n different outcomes with probabilities p_i each, then

$$\sum p_i = 1,$$

that is the probabilities add to one, and one of the probabilities is hence always determined by the rest.

7.3 Sequences of experiments

More interesting is to repeat experiments. If the experiments are independent, then the probability to have n times the same outcome is p^n, i. e. the probability shrinks with every experiment, provided the probability for an individual experiment p satisfies $p < 1$. If the question is how probable it is to have in n experiments an outcome with individual probability p k times, then this is at the same time a combinatorial question, as it corresponds to arrange the outcomes in all possible way. Thus, the result is

$$P_{nk}^{p} = \binom{n}{k} p^k (1-p)^{n-k}.$$

This can be generalized, if there is more than two possible outcomes, giving more complicated formulas.

A more important question is, however, what value one expects when doing an experiment, the so-called expectation value. If an experiment yields a number and has produced values n_i when performing it N times, the expectation value is given by the average

$$\langle n \rangle = \frac{1}{N} \sum_{i=1}^{N} n_i,$$

and thus if the experiment would be done another time, it would be expected that the value would be $\langle n \rangle$. This expectation value is not necessarily a true outcome of an experiment. For throwing a dice, it takes the value 3.5, which certainly is never obtained in any single throw. In this case this just means that the throw of 3 and 4 occur with equal probability (and actually that all numbers 3 and lower and 4 and higher are thrown equally often).

Of course, the numerical value of the expectation value depends on N. Throwing the dice a single time, and thus $N = 1$, will produce an expectation value with whatever value it has. Thus, the expectation value is an indication, but for any finite N it will depend on the number N. It has thus an error.

Most experiments belong to a class which is called well-behaved. In this case, the error σ of the expectation value can be estimated to be the so-called standard deviation

$$\sigma^2 = \frac{1}{N(N-1)} \sum_i (n_i - \langle n \rangle)^2,$$

and the true expectation value, i. e. the one obtained in the limit $N \to \infty$, will be found with 67% probability within the range $\langle n \rangle_{\text{finite } N} \pm \sigma$. Increasing the number of standard deviation increases the probability that the real value is inside the band. E. g. with 95% probability the actual value is within the range $\langle n \rangle_{\text{finite } N} \pm 2\sigma$. However, this probability reaches never unity as long as only a finite range, and thus a finite number of σ are considered. A proof of this is left to a different lecture.

This is only a first glimpse at these things, which will be discussed in much more detail during the experimental physics lectures, and are the gateway to the much richer and more complex topic of data analysis, an indispensable part of modern physics. E. g., the number of σ a measured value deviates from the expectation is used to quantify whether the expectation are based on an incomplete theory. E. g., in particle physics 3σ

are counted as evidence for this to happen and 5σ is considered to be a discovery of such a missing piece. Of course, this is still a statistical statement. Only in the limit of repeating the experiment infinitely often certainty could be achieved.

Chapter 8

Geometry

While ordinary geometry is a topic of some, but limited importance, in physics, the concepts of geometry are far more important. In generalizing many of the basic concepts of geometry deep insights in the laws of nature can be obtained. However, to truly grasp the nature of these generalization, and their implications, requires to be very familiar with the concepts in conventional geometry. The following will repeat some of the more pertinent features, which will then be generalized in the lecture on linear algebra and in many lectures on theoretical physics and advanced mathematics later.

8.1 Simple geometry

The basic objects of conventional geometry are lines and shapes on a plane and in three dimensions. The basic questions are usually related to the area or volume of shapes, as well as questions about how lines intersect.

8.1.1 Length and area

Probably the most basic object is a polygon, i. e. a closed, two-dimensional line, which is created from piecewise straight lines. The circumference l of such a polygon is the sum of the lengths l_i of its edges

$$l = \sum_i l_i.$$

To determine its area it is best to decompose it into triangles, which is always possible. This requires to calculate the area of triangles.

Any triangle can always be decomposed into two rectangular triangles, which have each half the area of the corresponding enclosing rectangle, which in turn just has as area the product of its edges, say a and b and thus $A = ab$. Hence, the area of a rectangular triangle with short edges a and b is

$$A_{\mathrm{rt}} = \frac{ab}{2}.$$

It is possible to reformulate it using the long edge, the hypotenuse, which is given by the formula of Pythagoras, $c = \sqrt{a^2 + b^2}$. Or, using trigonometric functions

$$\begin{aligned} c &= \sqrt{a^2+b^2} = \frac{a}{\sin\alpha} = \frac{b}{\cos\alpha} \\ \tan\alpha &= \frac{a}{b}, \end{aligned}$$

where α is the angle between c and b. The angle β between c and a is

$$\tan\beta = \frac{b}{a}.$$

This can be used to obtain formulas expressing the area using c.

To split any triangle with sides a, b, and c into two rectangular triangles, choose a side, e. g. b, and connect the opposite edge with it. This creates a dividing line of length h. To calculate its length requires to determine the angles. The angles α_{ij} enclosed by sides i and j can be calculated as

$$\begin{aligned} \cos\alpha_{ab} &= \frac{a^2+b^2-c^2}{2ab} && (8.1) \\ \cos\alpha_{bc} &= \frac{b^2+c^2-a^2}{2bc} && (8.2) \\ \cos\alpha_{ac} &= \frac{a^2+c^2-b^2}{2ac}, && (8.3) \end{aligned}$$

which can be obtained by using the elementary formulas for the constructed rectangular triangles and eliminating h. This also shows two important properties of triangles

$$\begin{aligned} \alpha_{ab} + \alpha_{bc} + \alpha_{ac} &= 180^\circ \\ \frac{a}{\sin\alpha_{bc}} = \frac{b}{\sin\alpha_{ac}} &= \frac{c}{\sin\alpha_{ab}}, \end{aligned}$$

which can be obtained by trigonometric identities. Finally, the desired length h is given by, e. g.,

$$h = a\sin\alpha_{ab}.$$

Inserting everything, this yields

$$\begin{aligned} A_t &= \frac{h\sqrt{a^2-h^2}}{2} + \frac{h\sqrt{c^2-h^2}}{2} = \frac{a\sin\alpha_{ab}}{2}\left(a\sqrt{1-\sin^2\alpha_{ab}} + c\sqrt{1-\sin^2\alpha_{bc}}\right) \\ &= \frac{a\sin\alpha_{ab}}{2}\left(a\cos\alpha_{ab} + c\cos\alpha_{bc}\right) = \frac{a\sin\alpha_{ab}}{2}\left(\frac{a^2+b^2-c^2}{2b} + \frac{b^2+c^2-a^2}{2b}\right) \\ &= \frac{a\sin\alpha_{ab}}{4b}\left(a^2+b^2-c^2+b^2+c^2-a^2\right) = \frac{ab\sin\alpha_{ab}}{2}. && (8.4) \end{aligned}$$

Just by exchange, this also implies

$$A = \frac{ab \sin \alpha_{ab}}{2} = \frac{bc \sin \alpha_{bc}}{2} = \frac{ac \sin \alpha_{ac}}{2}.$$

The triangulation of a polygon is then highly dependent on the details of the polygon, and will therefore not be detailed here. But there are algorithmic constructions for it. Manually this is of course also possible, but usually rather tedious.

8.1.2 Circles and π

There is one particular polygon, however, which should be considered. For this purpose, take a polygon, which is constructed from n elements each of length l, which are all connected in the same way. The simplest is an equilateral triangle with $n = 3$. In this case, the angle is 60°, and thus

$$\begin{aligned} l_3 &= \sum_i l = 3l \\ A_3 &= \frac{\sqrt{3}l^2}{4}, \end{aligned}$$

and thus a special value.

If continuing on, then the created polygon can be triangulated into n triangles with their tips meeting at the center. The angle there is

$$\alpha = \frac{360°}{n}.$$

If the distance from the center to the middle of the edge is r, then the outer lengths l_i are given by

$$l = 2r \tan \frac{360°}{2n}.$$

This yields as total circumference and length, respectively

$$\begin{aligned} l_n &= 2rn \tan \frac{360°}{2n} \\ A_n &= r^2 n \tan \frac{360°}{2n}. \end{aligned}$$

Of course, in the limit of $n \to \infty$, this polygon becomes the well-known circle. This requires to determine the number

$$\pi = \lim_{n \to \infty} n \tan \frac{360°}{2n}.$$

The limit cannot be analytically taken. It is therefore called π, and it is a so-called transcendental number, just as e, i. e. a non-periodic number with an infinite number of digits, with the first six being $\pi \approx 3.14159$.

For many practical purposes it is now convenient to measure angles rather than in degrees in units of π, defining

$$\pi \equiv 180^\circ.$$

and henceforth using these units, called radians, to measure angles. It is right now not at all obvious that this is a particularly useful convention, but this will become clearer over time. It is the first example of changing a system of units such that expressions becoming simpler, a practice to be encountered regularly in physics.

8.1.3 Volumes

Volumes are a three-dimensional extension of lengths and areas. For a brick of lengths a, b, and c it is defined as

$$V = abc,$$

and thus for an equilateral brick, a cube, as $V = a^3$.

This already shows one particular property of volumes: If the body whose volume should be calculated is just a three-dimensional extension of an area, it is sufficient to multiply the area by the height,

$$V = Ah.$$

This was visible for the cube and the brick. For a cylinder, this implies $V = \pi r^2 h$.

The situation is more complicated, if the volume has a less regular shape. As long as the volume has straight edges, it is possible again to decompose it into pyramids. The basic object is then a pyramid with a triangular base. In a similar, though more cumbersome way, as before for the triangle, a volume of an object can be determined by decomposing it into several pyramids. Here, therefore only the result for the pyramid will be quoted. It is

$$V_p = \frac{Ah}{3},$$

where A is the area of the base shape and h is its height. Note that this formula not only applies if the base shape is a triangle, but actually applies for any shape which is a concentric polygon, including a cone. In the later case, the volume is then just $V = \pi r^2 h/3$.

A little more involved is the situation for two other bodies appearing regularly in physics: The parallelepiped, which essentially is a skew brick, and the sphere.

The parallelepiped has thus three edge lengths, a, b, and c. Selecting the corner where all angles are smaller than $\pi/2$, then again the angles between two of the edges shall be α_{ij}. The volume of the parallelepiped is then given by

$$V = abc\sqrt{1 + 2\cos\alpha_{ab}\cos\alpha_{ac}\cos\alpha_{bc} - \cos^2\alpha_{ab} - \cos^2\alpha_{ac} - \cos^2\alpha_{bc}}, \tag{8.5}$$

which is symmetric under relabeling of the edges, as it ought to be.

The volume of the sphere is given by

$$V = \frac{4\pi}{3}r^3,$$

or, more generally, of an ellipsoid with three different axes a, b, and c

$$V = \frac{4\pi}{3}abc.$$

8.2 Vectors

So far, everything said about geometry was based on shapes. There is actually a much better suited language for this, the one of vectors. A vector is foremost an ordered set of n numbers, or in general elements, so-called coordinates. The number n is called the dimension of a vector. Especially, a single number can also be regarded as a vector of dimension 1.

8.2.1 Vectors in a space

More interesting is the case with $n > 1$. Then the usual way of writing a vector $\vec{a}$ is, e. g. for $n = 3$,

$$\vec{a} = \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

which thus has three coordinates x, y, and z. The coordinates are also written as $(\vec{a})_i$, or just a_i, with i running from 1 to 3 (or sometimes 0 to 2), and also called components or elements of the vector. Thus, e. g. $(\vec{a})_2 = a_2 = y$. The usage of the arrow $\rightarrow$ above a for the vector is also something optional if the context uniquely identifies a quantity as a vector, but for this lecture it will be kept.

The name coordinates for the elements indicates their origin. Take an n-dimensional space, say $n = 2$. Then any point of this plane is uniquely identified by its coordinates x and y, which can be read from the axes. This point can therefore be uniquely identified by a two-dimensional vector with the same numbers as coordinates. The usage of this idea of vectors assembled from coordinates is usually also called analytic geometry.

However, vectors are more than just this. It is the convention that the vector not only signifies this point, but also identifies a line connecting the origin to this point. Thus, a vector is also a direction. This is more than just a line or an edge.

8.2.2 Vector addition

Now consider a triangle in a plane, with one of the corners (often also called vertex) located at the origin. Then two of the edges can be described by two vectors, which have as coordinates the other two corners, call them $\vec{e}$ and $\vec{f}$. Is there a possibility to also describe the third edge with a vector?

To find a way, consider the following case. Take some point in the plane denoted by the vector $\vec{a}$. Now, select a second point in the plane, denoted by the coordinates x and y. It is certainly possible to draw a line between the point identified by $\vec{a}$ and the coordinates x and y, and give it a direction. The question is, whether there is some way to go first to the position indicated by $\vec{a}$, and then onward to the point signified by x and y.

There are certainly some numbers b_1 and b_2 such that

$$\begin{aligned} a_1 + b_1 &= x \\ a_2 + b_2 &= y, \end{aligned}$$

and which therefore are uniquely determined by $\vec{a}$ and x and y. Now, combine x and y into a vector $\vec{c}$ and b_1 and b_2 into a vector $\vec{b}$. Then define vector subtraction as

$$\vec{b} = \vec{c} - \vec{a} = \begin{pmatrix} x - a_1 \\ y - a_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

Define furthermore for any number d

$$d\vec{a} = \begin{pmatrix} da_1 \\ da_2 \end{pmatrix},$$

to deal with any appearing minus signs, so-called scalar multiplication. Then it would also be possible to state

$$\vec{c} = \vec{a} + \vec{b}.$$

Thus the point designated by x and y could be reached from the position described by $\vec{a}$ by addition with $\vec{b}$. This defines vector addition.

It is possible to worry now that the vector $\vec{b}$ is not a real vector, as it really not starts at the origin. It is defined such as to be continued from the point $\vec{a}$. This is actually not something which alters what a vector is. Rather, it is part of the definition of the vector addition. Geometrically vector addition is the statement that a vector, which originates from the origin is taken and moved (without changing the orientation) to the end of the first vector. The final vector is then the point which is described by the end-point of the second vector, but again taken to start from the origin. Thus, there are not different types of vectors.

Algebraically, the sum of two vectors is just the sum of its coordinates. The multiplication by a number d is just an elongation (or shortening if $|d| < 1$ and including a reversal if $d < 0$) of the vector.

The original problem of the triangle is then just a special case of the previous construction. The remaining edge is obtained by subtracting both vectors describing the first two edges.

Since vector addition is essentially defined by coordinate addition, which in turn is just ordinary addition of numbers, it retains all properties of ordinary addition. Also, since vector addition has been defined as a coordinate-wise operation, the number of dimensions n did not matter. Hence, it works the same way for arbitrary n, especially in $n = 3$.

8.3 Dot product

Given the example of the triangle, it is an interesting question whether it is possible to read off also the angles between the edges of the triangle. Geometrically, this is certainly possible, but is there a possibility to obtain it algebraically using vectors?

To find an answer, it will be necessary to define first the length of a vector. Considering the vector as a line, its length can be calculated geometrically, since it really is only a rectangular triangle when viewed with respect to the coordinate axes. Hence its length is just the hypotenuse of this triangle, and the formula of Pythagoras yields for a two-dimensional vector $\vec{a}$

$$|\vec{a}| = \sqrt{a_1^2 + a_2^2},$$

where the notation $|\vec{a}|$ is the statement of taking the length. Using the same notation as the absolute value originates from regarding a number as a one-dimensional vector. Then the length of this vector is just its absolute value. Generalizing this to more dimensions, the result is

$$|\vec{a}| = \sqrt{\sum_i a_i^2},$$

a straightforward geometrical extension.

The rest is then straightforward geometry, as the remaining calculation of an angle can be taken from (8.1-8.3), yielding for the angle at the origin

$$\cos\alpha = \frac{a_1}{|\vec{a}|},$$

and similarly for the other angels.

Take now two vectors, $\vec{a}$ and $\vec{b}$. Together with $\vec{c} = \vec{a} - \vec{b}$ they also form a triangle. The angle between $\vec{a}$ and $\vec{b}$ can also be calculated geometrically, using again (8.1-8.3), and is

$$\cos\alpha = \frac{a_1 b_1 + a_2 b_2}{|\vec{a}||\vec{b}|}.$$

There is now an interesting relation to the length of $\vec{c}$,

$$|\vec{c}| = |\vec{a} - \vec{b}| = \sqrt{|\vec{a}|^2 + |\vec{b}|^2 - 2|\vec{a}||\vec{b}|\cos\alpha}.$$

Thinking about

$$|a - b| = \sqrt{(a-b)^2} = \sqrt{a^2 + b^2 - 2ab}$$

this seems to suggest to define the quantity $|\vec{a}||\vec{b}|\cos\alpha$ as the product of two vectors,

$$\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}|\cos\alpha = a_1 b_1 + a_2 b_2, \tag{8.6}$$

to complete the analogy

$$|\vec{c}| = \sqrt{(\vec{a} - \vec{b})^2} = \sqrt{\vec{a}^2 + \vec{b}^2 - 2\vec{a}\cdot\vec{b}},$$

where it has been used that the definition (8.6) for a vector upon itself yields

$$\vec{a}^2 = \vec{a}\cdot\vec{a} = a_1 a_1 + a_2 a_2 = |\vec{a}|^2, \tag{8.7}$$

and thus the length.

This is indeed done, and the expression (8.6) is called the scalar product or dot product or inner product, depending on context. Though this looks like a multiplication, it is quite different from it. It does not map two vectors to a vector, like a multiplication of two numbers yields a number. Rather, it yields a number. Therefore, it does also not make sense to ask what happens when performing a dot product of three vectors. Since a dot product of two vectors does not yield a vector, and there is no meaning for a dot product of a vector and a number. Thus, while the square of a vector is well-defined, see (8.7), any other power is not. Especially, there is no inverse operation like a division. The scalar product looses information. It maps two vectors, described by at least four numbers, into a single number. There is no way to reconstruct from this single number the four original ones.

Geometrically, the dot-product determines the projection of one vector upon the other. Factoring out the length of one vector, the remainder is geometrically just a triangle with the second vector being the hypotenuse. Taking its length times the enclosed angle gives the length of the base line of the triangle. Thus, geometrically the scalar product is a projection. Of course, factoring out the other length gives the projection of the other vector.

What is possible is to generalize the dot product to more dimensions by

$$\vec{a} \cdot \vec{b} = \sum_i a_i b_i = |\vec{a}||\vec{b}| \cos\alpha,$$

which therefore also generalizes the length. Geometrically, since any two vectors always lie inside a plane, which is called coplanar, the obtained angle in the second equality is again the angle between both vectors in this plane.

Note that because $\cos\pi/2 = \cos 3\pi/2 = 0$, the dot product vanishes if both vectors are orthogonal to each other, no matter if to the left or the right.

8.4 Cross product

The interesting question is then, whether there can be constructed also some operation which maps two vectors into a vector. The answer to this question is actually much more subtle than it seems at first sight.

Since in one dimension the dot product actually reduces to the ordinary multiplication mapping two numbers to a number, there is actually no need for another product. Thus, at least two dimensions are required to even make the question meaningful.

But, geometrically, two dimensions are special. There, two vectors can only either be parallel or already addition can be used to reach every other vector from them. Thus, in two dimension any such multiplication operation would be just addition in disguise.

Hence, move on to three dimensions. Here it is for the first time really possible to have three vectors which do not have any trivial relation. Geometrically, this occurs by having a vector which is perpendicular to the plane where the other two vectors are lying in. Thus, the third vector should be perpendicular to both. Given two vectors $\vec{a}$ and $\vec{b}$, define the cross product, or vector product, or sometimes also called outer product, as

$$\vec{a} \times \vec{b} = \begin{pmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{pmatrix}, \tag{8.8}$$

which is indeed perpendicular to both, as can be tested using the dot product

$$\vec{a}\cdot(\vec{a}\times\vec{b}) = a_1a_2b_3 - a_1a_3b_2 + a_2a_3b_1 - a_2a_1b_3 + a_3a_1b_2 - a_3a_2b_1 = 0,$$

and in the same way

$$\vec{b}\cdot(\vec{a}\times\vec{b}) = a_2b_1b_3 - a_3b_1b_2 + a_3b_1b_2 - a_1b_2b_3 + a_1b_2b_3 - a_2b_1b_3 = 0.$$

Thus the cross product has the desired properties. Since it is orthogonal to the other two, this implies that in three dimensions it is perpendicular to the plane in which the other two vectors lie.

If they are parallel, and thus the plane does not exist, it is helpful to note that

$$(\vec{a}\times\vec{b})^2 = \vec{a}^2\vec{b}^2 - (\vec{a}\vec{b})^2 = \vec{a}^2\vec{b}^2(1-\cos^2\alpha) = \vec{a}^2\vec{b}^2\sin^2\alpha,$$

which follows by direct calculation. Thus, the cross product is proportional to the sine of the angle between the two vectors, and therefore vanishes if both are parallel. Thus, there is then also no ambiguity in its direction. By comparison to equation (8.4), this also implies that the absolute value of the cross-product of two vectors gives twice the area of the triangle formed by it.

The cross product has a number of rather surprising features. First, from its definition, it can be derived that

$$\vec{a}\times\vec{b} = -\vec{b}\times\vec{a}.$$

Thus the cross product is not commutative, but anti-commutative, very differently from the conventional product of two numbers.

Next, given three vectors it is possible to form a number from them by first performing a cross-product between two, and then form a scalar product with the third. Again, from the definition it follows that for this number

$$\vec{a}\cdot(\vec{b}\times\vec{c}) = \vec{b}\cdot(\vec{c}\times\vec{a}) = \vec{c}\cdot(\vec{a}\times\vec{b}), \tag{8.9}$$

where again, due to the anti-commutativity of the cross product, the ordering matters. Especially

$$\vec{a}\cdot(\vec{b}\times\vec{c}) = -\vec{a}\cdot(\vec{c}\times\vec{b}) = (\vec{b}\times\vec{c})\cdot\vec{a},$$

where the last step was possible because the dot product is commutative.

This combination has an interesting relation to the volume of a parallelepiped (8.5). Since three non-coplanar vectors can always be considered as the edges of a parallelepiped, it can be shown, using elementary geometry,

$$V = |\vec{a}\cdot(\vec{b}\times\vec{c})|, \tag{8.10}$$

where the absolute value is necessary due to the anti-commutativity of the cross product: Only for a certain (cyclic) ordering of the three vectors the result is positive, otherwise negative. The geometrical interpretation also elucidates why this is the volume. The cross product gives a vector perpendicular to the first two, with a length of the area of the parallelogram, thus twice the area of the triangle, formed by them, as noted above. The scalar product then determines the height of the parallelepiped, since the dot product determines the projection of the vector $\vec{a}$ on the vector perpendicular to the the base area. Then, this is just the area times height, and thus the volume.

Finally, since the cross product yields another vector, it is possible to perform another cross-product. However, again the order matters, it is non-associative, i. e. in general

$$\vec{a} \times (\vec{b} \times \vec{c}) \neq (\vec{a} \times \vec{b}) \times \vec{c}.$$

An explicit counter-example is if $\vec{b}$ and $\vec{c}$ are parallel, and orthogonal to $\vec{a}$. Then the left-hand side is zero, but the right-hand side is not: $\vec{a} \times \vec{b}$ is orthogonal to both, $\vec{a}$ and $\vec{b}$, and therefore to $\vec{c}$, and thus the combination is not zero.

It is an interesting feature that this product cannot be extended in any straightforward way into more than three dimensions, for which there are deeper geometric reasons. This will be addressed in the lecture on linear algebra.

Chapter 9

Special topics

Finally, there are a number of special topics, which will be essentially giving a few definitions and some practical insights. All of them will be explained in much more detail in the corresponding lectures.

9.1 Differential equations

So far, all equations have involved variables and functions of them,

$$x = f(x).$$

However, in physics it very often happens that it involves actual derivatives of x, if x is itself a function,

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

Such an equation is called a differential equation.

An example is

$$\frac{dx}{dt} = c, \tag{9.1}$$

where c is some constant. Inserting $x = ct + a$, where a is another constant yields

$$\frac{d(ct + a)}{dt} = c + 0 = c,$$

and therefore $ct + a$ is a solution to this differential equation. This differential equation is called first-order differential equation, as there only the first derivative of x appears.

An example of a second order differential equation is

$$\frac{dx^2}{dt^2} = a^2 x(t) \tag{9.2}$$

This equation has two possible solutions, $x(t) = d_\pm \exp(\pm at)$ where the numbers $d_\pm$ are constants. This is again shown by explicit calculations

$$\frac{d^2}{dt^2} d_\pm e^{\pm at} = \frac{d}{dt} d_\pm a e^{\pm at} = d_\pm a^2 e^{\pm at} = a^2 x(t).$$

The fact that there are now two solutions with two different constants has something to do with the fact that it is a second-order differential equation. In general, there are n solutions with n constants for a differential equation of nth-order. This will be shown in the lecture on differential equations.

Another infamous differential equation is

$$\frac{dx^2}{dt^2} = -a^2 x(t),$$

which looks, up to the minus sign, very similar to the previous one. However, it has the quite different solutions

$$\begin{aligned} x_1(t) &= d_s \sin(at) \\ x_2(t) &= d_c \cos(at). \end{aligned}$$

This is again shown by explicit insertion. This equation is known as the harmonic equation or oscillator equation

It is not possible to find a general recipe how to solve differential equations. It is possible for certain classes of them, and for these recipes will be derived in the lecture on differential equations. Other than that, its more (educated) guesswork. It is also entirely possible that there is no solution to a differential equation in closed form, but it can be shown that there is always a solution. This is again subject of the lecture on differential equations.

As a final remark, the constants appearing can be selected if boundary conditions are provided, i. e. conditions which the solutions must fulfill. In physics they are usually provided by knowledge of the described system at some time, and therefore known as initial conditions. For every constant there must be a boundary condition to make it well-defined. However, it is entirely possible that even if there are as many initial conditions as there are constants, it is possible that there is exactly one, some up to the order, or no solution for the constants, depending on whether the ensuing equations have a solution or not.

Take again the differential equation (9.1). A possible initial condition would be that $x(0) = s$, where s is some number. Then $a = s$ would be a suitable choice to satisfy this initial condition. An example which is impossible to satisfy is the solution to (9.2).

Require, e. g. that $x(0)=0$, This is only solved using

$$c(e^{at}-e^{-at}).$$

That the sum is also a solution can again be seen by direct insertion. If it is furthermore required that $x(1)=0$ as well, the only solution would be $c=0$, and thus, there is no real solution.

9.2 Matrices

9.2.1 Definition

Another topic to be introduced is matrices. Matrices will become a very important concept discussed in great detail in linear algebra, and will also play a central role in physics later. Here, a more pragmatic definition will be given.

A matrix is foremost a rectangular scheme of numbers, where here only a square one will be considered. If there are n^2 numbers, they can thus be rewritten as an $n\times n$ scheme

$$M=\begin{pmatrix} m_{11} & \dots & m_{1n} \\ \vdots & \ddots & \vdots \\ m_{n1} & \dots & m_{nn} \end{pmatrix},$$

where the n^2 numbers m_{ij} are called the elements of the matrix, and M itself the matrix.

The most important thing which can be done with a matrix is to combine it with a vector to obtain the operation of matrix-vector multiplication. It is given by

$$M\vec{v}=\begin{pmatrix} \sum_i m_{1i}v_i \\ \vdots \\ \sum_i m_{ni}v_i \end{pmatrix}.$$

This is a definition. The new vector has elements given by performing a scalar product of the original vector $\vec{v}$ and the ith row of the matrix M, interpreted as a vector.

9.2.2 Systems of linear equations

The real use of it becomes clearer when one starts to consider the following problem. So far, only the situation has been considered that there is a single equation for a single unknown. However, in general, there will be many equations and many unknowns. A particular example are such sets of equations where every unknown appears at most linearly, so-called systems of linear equations. In the case of two such equations with two variables, this looks like

$$\begin{aligned} m_{11}x_1 + m_{12}x_2 &= b_1 \\ m_{21}x_1 + m_{22}x_2 &= b_2 \end{aligned}$$

where the x_i are the unknowns, and the remainder are constants. Such a system of equations can then be written as

$$M\vec{x} = \vec{b}, \tag{9.3}$$

which is called matrix-vector form. So far, nothing has been gained but a more compact notation.

Such a system can be solved in a very similar fashion as for ordinary systems, by solving them one-by-one, treating the other parts always as constant. E. g.

$$\begin{aligned} x_1 + x_2 &= b_1 \\ x_1 - x_2 &= b_2 \end{aligned}$$

yields from the second equation

$$x_1 = b_2 + x_2. \tag{9.4}$$

Inserting this into the first equation yields

$$b_2 + x_2 + x_2 = b_1 \to x_2 = \frac{b_1 - b_2}{2}.$$

Finally inserting this into (9.4) yields

$$x_1 = \frac{b_1 + b_2}{2},$$

completing the solution.

Note that such systems of equations can also have infinitely many solutions. A trivial example is one where all coefficients in the second equation are zero, and only the equation

$$m_{11}x_1 + m_{12}x_2 = b_1$$

remains. Then the equation is solved for any x_1 if

$$x_2 = \frac{b_1 - m_{11}x_1}{m_{12}},$$

and since x_2 is a real number, there are infinitely many possibilities.

However, such systems of equations, in contrast to the situation of a single linear equation, are not necessarily solvable. Consider the case

$$\begin{aligned} x_1 + x_2 &= b_1 \\ x_1 + x_2 &= b_2 \end{aligned}$$

with $b_1 \neq b_2$. Solving the second equation for x_1, acting as if x_2 is just a constant, and thus like the case of a single-variable equation, yields

$$x_1 = b_2 - x_2.$$

Reentering this into the first equation yields

$$x_1 + x_2 = b_2 \neq b_1,$$

and therefore this system cannot be solved. However, it was necessary to solve the system of equations to figure this out.

9.2.3 Determinants and Cramer's rule

The concept of matrices now provides a possibility to check this without finding explicitly a solution. To this end, define the operation of determinant for a 1×1 matrix just the matrix element and for a 2×2 matrix

$$\det M = m_{11}m_{22} - m_{12}m_{21} \tag{9.5}$$

In the linear algebra lecture it will be shown that the system of equations has a solution only if $\det M \neq 0$. In the above case, $\det M = 0$, and therefore demonstrates this. Thus, calculating the determinant yields whether it is useful to search for a solution

The generalization of the determinant to $n > 2$ is not so straightforward, and there are multiple possibilities. One is given by

$$\det M = \sum_i (-1)^{i+1} M_{1i} \det M^{1i}, \tag{9.6}$$

and others will be discussed in the lecture on linear algebra. In (9.6) M^i denotes the matrix in which the first row and ith column is removed, and thus of size $n-1 \times n-1$, if the original matrix was of size $n \times n$. Thus, the calculation of a determinant is a so-called recursive process. To see how this works, consider

$$\det \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = (-1)^2 m_{11} \det(m_{22}) + (-1)^3 m_{12} \det(m_2 1) = m_{11}m_{22} - m_{12}m_{21},$$

and thus in agreement with (9.5). Likewise, this can be done for larger matrices, where for each determinant, for which the result is not yet known, the formula is applied again.

Besides checking whether a set of equation does have a solution, there is a further useful application of determinants, Cramer's rule, which allows to solve systems of equations: Given a system of equations in Matrix-vector form (9.3), the solution is given by

$$x_i = \frac{\det M^{i \to \vec{b}}}{\det M}.$$

In this $M^{i \to \vec{b}}$ is the matrix where the ith column has been replaced with the right-hand vector $\vec{b}$. This also shows again that $\det M$ needs to be non-zero for a solution. While this appears to easily resolve the issue of solving linear equations, the calculation of determinants very quickly escalates for larger matrices. Thus, for those more efficient methods will be needed, some of which will be introduced in the lecture of linear algebra.

9.3 Bodies, groups, and rings

In the beginning, sets and operations on sets were introduced. This is a quite abstract notion, but actually it is in mathematics possible to derive very many consequences just from these generic properties. They therefore apply to whatever the realization of the set and the operation is. It also gives criteria under which conditions and results from one setup can be transferred to another setup.

Over time, a number of especially useful combinations of sets and operations have been identified, and therefore have received definite names. If in a given situation it is possible to assure that the objects in questions belong to any such category, immediately all derived properties of these categories are at one's disposal. This is often very useful and the lecture on linear algebra will give powerful examples.

Here, therefore, a number of such categories will be defined for later use.

The basic starting point is always the combination of some sets $\mathcal{S}_i$ and one or two operations $\circ$ and $\bullet$ establishing certain relations.

A group is based on a single set with one operation $\circ$ such that for any element a, b, and c

- $a \circ b \in \mathcal{S}$, this is called closure
- $(a \circ b) \circ c = a \circ (b \circ c)$, this is called associativity
- There exists $e \in \mathcal{S}$ such that for any a $e \circ a = a \circ e = a$, which is called the existence of the identity element e

- For any a there exists an element, called a^{-1}, such that $a \circ a^{-1} = a^{-1} \circ a = e$, which is called the inverse

- If $a \circ b = b \circ a$, the group is called Abelian, otherwise it is called Non-Abelian

The group generalizes the usually multiplication of real numbers.

A monoid is a single set with one operation $\circ$ for which for any three elements a, b, and c $(a \circ b) \circ c = a \circ (b \circ c)$ holds, and where also an identity element exists such that $a \circ e = e \circ a = a$. A monoid therefore satisfies only some of the properties of a group. Thus, it is often also called a semigroup. The main difference is that there does not need to be an inverse element.

A ring is based on a single set with two operations $\circ : \mathcal{S} \to \mathcal{S}$ and $\bullet : \mathcal{S} \to \mathcal{S}$, which satisfies the following properties for any elements a, b, and c.

- It is an Abelian group under $\circ$

- It is a monoid/semigroup under $\bullet$

- $a \bullet (b \circ c) = (a \bullet b) \circ (a \bullet c)$ and $(b \circ c) \bullet a = (b \bullet a) \circ (c \bullet a)$, which is called distributivity

To distinguish the neutral elements under $\circ$ and $\bullet$, the one under $\circ$ is usually called the zero element and the one under $\bullet$ is called the unit element. A ring is the generalization of the conventional real numbers with multiplication and addition.

A body is a ring which also forms an Abelian group under $\bullet$. This is even closer to the real numbers with addition and multiplication. A body is also called a field.

Though the differences between these categories seem rather abstract at first, these structures play fundamental roles in physics. Especially Abelian and non-Abelian (semi-)groups and fields are essential structures in the formulation of modern physics.

Chapter 10

Exercises with solutions

This chapters contains further exercises with, partially worked out, solutions for the topics of the lecture.

10.1 Sets

Exercises

Determine the size of each of the following sets, as well as all possible intersections and unions, in any order:

- $S_1 = \emptyset$
- $S_2 = \{1\}$
- $S_3 = \{-1, a, 3, 1, 0.5\}$

Solutions

The sizes are 0, 1, and 6. The unions and intersections are

1. $S_1 \cup S_2 = \{1\}$ and $S_1 \cap S_2 = \emptyset$
2. $S_1 \cup S_3 = S_3$ and $S_1 \cap S_3 = \emptyset$
3. $S_2 \cup S_3 = S_2$ and $S_2 \cap S_3 = S_3$
4. $(S_1 \cup S_2) \cup S_3 = S_1 \cup (S_2 \cup S_3) = S_3$
5. $(S_1 \cap S_2) \cup S_3 = S_3$ and $S_1 \cap (S_2 \cup S_3) = \emptyset$

6. $(S_1 \cup S_2) \cap S_3 = S_2$ and $S_1 \cup (S_2 \cap S_3) = S_2$

7. $(S_1 \cap S_2) \cap S_3 = S_1 \cap (S_2 \cap S_3) = \emptyset$

10.2 Sums and sequences

Exercises

What are the limits of the following infinite sequences?

- $a_i = 1$
- $b_i = i$
- $c_i = \frac{1}{i}$
- $d_i = (-1)^i$

What values have the following sums, based on the previous sequences?

- $\sum_{i=0}^{100} a_i$
- $\sum_{i=0}^{\infty} b_i$
- $\sum_{i=0}^{10} d_i$
- $\sum_{i=0}^{11} d_i$

Solutions

The limits of the sequences are

1. $\lim_{i\to\infty} 1 = 1$
2. $\lim_{i\to\infty} i = \infty$
3. $\lim_{i\to\infty} \frac{1}{i} = 0$
4. Does not have a limit.

The results for the sums are

1. 101
2. ∞
3. 1
4. 0

10.3 Function properties

Exercises

Determine

1. $f(x) = x^{+\sqrt{2}}$ for $x = 4$, 2 and $\sqrt{2}$
2. $g(x) = f(f(x))$
3. $h(x) = f(x)g(x) - g(x)/f(x)$
4. $l(x) = h(x)^{\frac{3}{2}} - (g(x)f(x)^{\sqrt{3}})^{-\frac{5}{2}} - \frac{g(x)}{f(x)}$
5. $\ln x$ for $x = 1/2$, 10^{-3}, 10^5, 1, and 10^{10}
6. $\ln x^4 + \ln x^5$
7. $\ln \frac{e^x}{e^{x^2}}$
8. $\ln x - \ln x^2 + \ln(2x)$

Determine the domain of the definition of the following functions.

1. $1/(x^2 - 1)$
2. $\sin^{-1}(x)$
3. x^2
4. $1/(x^4 + 1)$
5. x

Determine the image of the following functions.

1. $\frac{e^{-x} - e^x}{e^x + e^{-x}}$
2. x^2
3. $\frac{13}{2}(\sin x)^2$
4. $\sin \frac{1}{x}$
5. $\frac{e^{-x} + e^x}{2}$

6. x^3

7. $\exp(1/x)$

What are domain of definition and the image of the following functions, and what does this imply for the compositions? Note that in some cases some numbers will be hard to determine, and can be approximated, e. g., by drawing the function behavior.

1. $f(x) = 1/(x-1)$

2. $g(x) = x^2 \frac{1-x}{(1+x)(2-x)(9-x^2)}$

3. $h(x) = x^4 + x^2$

4. $l(x) = x^2 - x^3$

5. $a(x) = f(x)g(x)$

6. $b(x) = f(g(x))$

7. $c(x) = f(x)/g(x)$

8. $d(x) = l(x + h(x))$

Find the inverse function of the following functions, and give their domains of definition and image, and whether the solutions are multivalued.

1. $f(x) = x + 1 - x^2$

2. $g(x) = 1/(1-x)$

3. $h(x) = 1/x^2$

4. $l(x) = 2$

Solutions

The results are

1. $2^{\sqrt{2}}$ and $2^{\sqrt{2}}$, and $\sqrt{2}^{\sqrt{2}}$.

2. $g(x) = x^2$

3. $h(x) = x^{2+\sqrt{2}} - x^{2-\sqrt{2}}$

4. $l(x) = \frac{1}{2}\left(5x^{2+\sqrt{6}} - 2x^{2-\sqrt{2}} + 2\left(x^{2-\sqrt{2}}\left(x^{\sqrt{8}} - 1\right)\right)\right)$
5. $-\ln 2$, $-3\ln 10$, $5\ln 10$, 0, and $10\ln 10$
6. $9\ln x$
7. $x - x^2$
8. $\ln 2$

The domain of definitions are

1. $\mathbb{R}/\{\pm 1\}$
2. $[-1, 1]$
3. $\mathbb{R}$
4. $\mathbb{R}$
5. $\mathbb{R}$

The images are

1. $[-1, 1]$
2. $\mathbb{R}_+$
3. $\left[0, \frac{13}{2}\right]$
4. $[-1, 1]$
5. $\mathbb{R}_+$
6. $\mathbb{R}$
7. $\mathbb{R}_+$

The images and domain of definitions are

1. For $f(x)$ the domain of definition is $\mathbb{R}/\{1\}$ and the image is $\mathbb{R}$.
2. For $g(x)$ the domain of definition is $\mathbb{R}/\{-1, 2, \pm 3\}$ and the image is $\mathbb{R}$.
3. For $h(x)$ the domain of definition is $\mathbb{R}$ and the image is $\mathbb{R}_+$.
4. For $l(x)$ the domain of definition is $\mathbb{R}$ and the image is $\mathbb{R}$.

This implies for the chains

1. For $a(x)$ the domain of definition is $\mathbb{R}/\{-1,2,\pm 3\}$ and the image $\mathbb{R}$.

2. For $b(x)=\frac{(x-2)(1+x)(x^2-9)}{18+9x-12x^2+x^4}$ the domain of definition is given by $\mathbb{R}$ without the real roots of the polynomial (there are three, but need extended calculations). The image is $\mathbb{R}$.

3. For $c(x)=\frac{(2-x)(1+x)(x^2-9)}{x^2(x-1)^2}$ the domain of definition is $\mathbb{R}/\{0,1\}$ and the image is relatively involved and cna be obtained by drawing, but is $[-\infty,c_0]$, where c_0 is a constant of about 0.85.

4. For $d(x)=x+x^2+x^4$ the domain of definition is $\mathbb{R}$ and the image is $[d_0,\infty]$ with $d_0\approx -0.21$.

The inverse functions are given by

1. $f^{-1}(y)=\frac{1}{2}\left(1\pm\sqrt{5-4y}\right)$. It is multivalued.

2. $g^{-1}(y)=\frac{y-1}{y}$. It is singlevalued.

3. $h^{-1}(y)=\pm\frac{1}{\sqrt{y}}$. It is multivalued.

4. $l(x)$ has no well-defined inverse, as every variable is mapped to the same value, and thus it is infinitely mutlivalued.

10.4 Equations

Exercises

Solve the following equations for real x.

1. $ax+b=c$

2. $\exp(x)=4$

3. $x^2+x=0$

4. $x^4=16$

5. $\sqrt{x+4}=x$

6. $x^2-1=0$

7. $\sin(x) = \pi/4$
8. $x^4 + x^2 = 0$
9. $\frac{x+a}{x-d} = b$
10. $\frac{x^2+a}{x-d} = 0$
11. $ax - b = d$
12. $\exp(x+1) = 3$
13. $x^2 - x = 0$
14. $x^4 = 81$
15. $|\sqrt{x-2}| = x$
16. $x^2 + 1 = 0$
17. $\cos(x) = 0$
18. $x^4 - x^2 = 0$
19. $\frac{x^2-a}{2x-d} = 0$
20. $\exp(x+1) = 1$

Solutions

1. $x = \frac{c-b}{a}$
2. $x = \ln 4$
3. $x = 0, -1$
4. $x = \pm 2$
5. $(1+\sqrt{17})/2$. Note that the second sign of the square root does not solve the original equation
6. $x = \pm 1$
7. $x = 2n\pi + \sin^{-1}\frac{\pi}{4}$ and $x = (2n+1)\pi - \sin^{-1}\frac{\pi}{4}$ with n integer
8. The only real solution is 0

9. $x = (a + bd)/(b - 1)$

10. Has no solution

11. $x = \frac{b+d}{a}$

12. $x = -1 + \ln 3$

13. $x = 0, 1$

14. $x = \pm 3$

15. 1. Note that -2 solves the squared equation, but not the original one

16. No real solution

17. $x = (2n + 1)\pi/2$ with n integer

18. Rewriting it to $x^2(x^2 - 1) = 0$ yields $x = 0, \pm 1$

19. $x = \pm\sqrt{a}$, if $a \geq 0$ and $d \neq \pm 2\sqrt{a}$

20. $x = -1 + \ln 1 = -1$

10.5 Complex numbers

Exercises

Determine all possible sums and products of the following sets of numbers, and explicitly determine real part and imaginary part of the solutions.

1. $\{(1 + i), 3\exp(i\pi/2), \sqrt{i}\}$

2. $\{i - 1, \exp(i(\pi + i\pi)), (-1)^{\frac{1}{4}}, i + 1\}$

3. $\{1 - i, (1 - i)^*, \exp(-i\pi/2)\}$

4. $\{1/i, 2 - i, e^{i(\pi+i)}, 1^{1/4}\}$

Solutions

The numbers are labeled with letter a,... in the following.

1. $a+b=1+4i$, $a+c=(1+1/\sqrt{2})+(1+1/\sqrt{2})i$, $b+c=1/\sqrt{2}+i(3+1/\sqrt{2})$, $ab=-3+3i$, $ac=\sqrt{2}i$, $bc=(-3+3i)/\sqrt{2}$. It is useful to recognize that $\exp(i\pi/2)=i$ and $\sqrt{i}=\exp(i\pi/4)$. The latter can then be decomposed into real and imaginary part geometrically using the Pythagorean theorem.

2. Use that $\exp(i(\pi+i\pi))=-e^{-\pi}$ is real and $(-1)^{\frac{1}{4}}=\pm\frac{1}{\sqrt{2}}\pm\frac{1}{\sqrt{2}}$ with every possible combinations of + and −. This yields $a+b=-1-e^{-\pi}+i$, $a+c=(\pm 1-\sqrt{2})/\sqrt{2}+i(\pm 1+\sqrt{2})/\sqrt{2}$ (with all possible combinations of + and −), $a+d=2i$, $-e^{-\pi}\pm 1/\sqrt{2}\pm i/\sqrt{2}$, $b+d=1-e^{-\pi}+i$, $c+d=(\pm 1+\sqrt{2})/\sqrt{2}+i(\pm 1+\sqrt{2})/\sqrt{2}$ and $ab=e^{-\pi}-ie^{-\pi}$, $ac=\pm\sqrt{2}$ or $\pm i\sqrt{2}$, $ad=-2$, $bc=\pm(1\pm i)e^{-\pi}/\sqrt{2}$ with all combinations of + and −, $bd=-e^{-\pi}-ie^{-\pi}$, and cd yields the same as ca

3. Use that $(1-i)^*=1+i$ and $\exp(-i\pi/2)=-i$. This yields $a+b=2$, $a+c=1-2i$, $b+c=1$, $ab=2$, $ac=-1-i$, $bc=1-i$

4. Use that $1/i=-i$, and $\exp(i(\pi+i))=-1/e$, and $1^{1/4}$ is either ± 1 or $\pm i$. This yields $a+b=2-2i$, $a+c=-i-1/e$, $a+d=\pm 1-i$ or $-2i$ or 0, $b+c=(2-1/e)-i$, $b+d=(2\pm 1)-i$ or 2 or $2(1-i)$, $c+d=-1/e\pm 1$ or $-1/e\pm i$. $ab=-1-2i$, $ac=i/e$, $ad=\pm i$ or ± 1, $bc=-(2-i)/e$, $bd=\pm(2-i)$ or $\pm(1+2i)$, $cd=\pm 1/e$ or $\pm i/e$

Solutions

Domain of definition:

1. Real x: $x\neq\pm 1$, complex x: $x\neq\pm 1,\pm i$

2. Real x: $-1\leq x\leq 1$, complex x: $-1\leq\Re x\leq 1$ if $\Im x=0$

3. No restrictions

4. Real x: $\mathbb{R}$, complex x: $x\neq\pm\exp(\pm i\pi/4)$

5. Real x: $-1\leq x\leq 1$, complex x: No restrictions

6. No restrictions

Image:

1. Real x: $[-1,1]$. For imaginary x this can be rewritten as $y = ix \; -i\tan y$, yielding $[-i\infty, i\infty]$
2. Real x: $[0,\infty]$, imaginary x: $[0,-\infty]$
3. Real x: $[0, 13/2]$. For imaginary $x = iy$ this can be rewritten as $-\frac{13}{2}\sinh y = -\frac{13}{4}(e^x + e^{-x})^2$, yielding $[-\infty,\infty]$
4. Real x: $[-1,1]$. For imaginary $x = iy$ this can be rewritten as $\sinh\frac{1}{y}$ yielding $[-\infty,\infty]$
5. Real x: $[1,\infty]$. For imaginary x: $[-1,1]$
6. Real x: $[-\infty,\infty]$. For imaginary x: $[-i\infty, i\infty]$
7. Real x: $\{0\}$. For imaginary x: $[0,\infty]$
8. Real x: $[0,\infty]$. For imaginary x: A pure phase, but taking any value. Thus $\{\exp(i\theta)\}$ with $\theta \in [0, 2\pi)$

10.6 Derivatives

Exercises

Determine the first and second derivative of the following functions.

1. x^π
2. $\sin\exp(x)$
3. $c + x^4 - \cos^2 x - \sin^2 x$
4. $\ln(x\sqrt{x+1})$
5. $\exp(\pi x)$
6. $\sin\ln(x)$
7. $(x + \cos x)^a$
8. $\ln\left(x\exp(x^2)\right)$
9. $\frac{x+1}{x-1}$

10. $x^{1+\ln a}$

11. $\exp\cos(x)$

12. $x^3 - \cos^2(x+1) - \pi - \sin^2(x-1)$

13. $\exp(\sqrt{x^2+1})$

14. $\exp(2(x+a))$

15. $\ln\sin(x)$

16. $\exp(x^\pi)$

17. $\ln(x+1)$

18. $\exp(\sin^2(x))\exp(\cos^2(x))$

Solutions

1. $\pi x^{\pi-1}$, $(\pi-1)\pi x^{\pi-2}$

2. $\exp(x)\cos\exp(x)$, $\exp(x)\cos\exp(x) - \exp(2x)\sin\exp(x)$

3. $4x^3$, $12x^2$. Note that $\sin^2 x + \cos^2 x = 1$.

4. $(2+3x)/(2x+2x^2)$, $-(2+4x+3x^2)/(2x^2(1+x)^2)$

5. $\pi\exp(\pi x)$, $\pi^2\exp(\pi x)$

6. $\frac{\cos\ln x}{x}$, $-\frac{\sin\ln x+\cos\ln x}{x^2}$

7. $-a(\sin x-1)(x+\cos x)^{a-1}$, $a(a-1)(x+\cos x)^{a-2}(1-\sin x)(\sin x-1) - a\cos x(x+\cos x)^{a-1}$

8. Can be rewritten as $x^2+\ln x$, yielding $2x+\frac{1}{x}$, $2-\frac{1}{x^2}$

9. $-\frac{2}{(x-1)^2}$, $\frac{4}{(x-1)^3}$

10. $(1+\ln a)x^{\ln a}$, $x\ln a(1+\ln a)x^{\ln(a)-1}$

11. $-\sin(x)\exp(\cos(x))$, $\sin^2(x)\exp(\cos(x)) - \cos(x)\exp(\cos(x))$

12. $3x^2 + 2\cos(2x)\sin(2)$, $6x - 4\sin(2)\sin(2x)$

13. $\frac{x \exp \sqrt{1+x^2}}{\sqrt{1+x^2}}$, $\frac{\left(1+x^2\sqrt{1+x^2}\right)\exp\sqrt{x^2+1}}{(1+x^2)^{\frac{3}{2}}}$

14. $2\exp(2(x+a))$, $4\exp(2(x+a))$

15. $1/\tan(x)$, $-1/\sin^2(x)$

16. $\pi x^{\pi-1}\exp(x^\pi)$, $\pi x^{\pi-2}(\pi x^\pi+\pi-1)\exp(x^\pi)$

17. $1/(1+x)$, $-1/(1+x)^2$

18. 0, 0 because the original function evaluates to e

10.7 Extrema

Exercises

1. Determine for the function $f(x)=|\sin(x)|$ with x real all extrema and saddle points within the interval $[0,2\pi)$. Give a proof for every claim.

2. Determine for the function $f(x)=x^4-2x^2+x$ with x real all zeros, extrema, and saddle points within the interval $[0,2\pi)$. Give a proof for every claim. You can leave the numerical solutions x_1, x_2 and x_3 of the equation $x^3-x+1/4=0$ implicit.

3. Determine for the function $f(x)=+\sqrt{1+\sin(x)}$ with x real all extrema and saddle points within the interval $[0,2\pi)$. Give a proof for every claim.

4. Determine for the function $f(x)=x^3\exp(-x^2)+e^4/8$ with x real and $a>0$ all zeros, extrema, and saddle points. Give a proof for every claim.

Solutions

1. Because of the absolute value it is useful to rewrite the function as $(\sin(x)^2)^{1/2}$ for purpose of forming the derivatives. This gives as first derivative $\cos(x)\sin(x)/|\sin(x)|$ and $-|\sin(x)|$ as second derivative. The latter is thus either zero or negative. There are maxima at $x=\pi/2, 3\pi/2$. This can be seen either by using the rule of L'Hospitâl or because the function is bounded and takes its maximum value there. The points $x=0,\pi$ are special, since formally none of the derivatives are defined there, or always vanish when calculating them nonetheless. However, because $|\sin(x)|\geq 0$ and the function vanishes at $x=0,\pi$, these must be minima.

2. Rewriting the function as $x(x^3 - 2x + 1)$ yields a zero at $x = 0$. The second factor has a zero at $x = 1$. This allows to reduce the function to $x(x-1)(x^2 + ax + b)$ with $a = 1$ and $b = -1$. This yields finally two more zeros at $x = \frac{1}{2}\left(-1 \pm \sqrt{5}\right)$. The first derivative is $4x^3 - 4x + 1$ and the second $12x^2 - 4$. The last has only zeros at $x = \pm 1/\sqrt{3}$, for which the first derivative is not zero, and hence there are no saddle points. A direct determination of the maxima and minima is complicated, but an asymptotic analysis (or graphical representation) shows that the function has two minima and one maximum, and one of the minimum could be a global one. Formally, the first derivative's zeros are determined by the equation $x^3 - x + 1/4 = 0$, and thus the implicit values are the positions of the extrema. Because there are no saddle points and the function diverges to $+\infty$ for $x \to \pm\infty$ this implies a sequence of minimum, maximum, and minimum. Because of the zeros, the second minimum will be at negative values of the function. In this range the function is bounded by -1, since $x^4 < x^2 < x$ and hence $-2x^2$ has its largest negative value there. Because for $x = -1$ the function has the value -2 this implies that the first minimum is deeper, and thus the global one, while the second is only a local one.

3. The first and second derivatives are $\cos(x)/(2\sqrt{1+\sin x})$ and $\sqrt{1+\sin x}/4$. The first derivative has only a zero at $x = (2n+1)\pi$ and thus $\pi/2$ in the interval. The rule of L'Hospitâl yields that $x = 3\pi/2$ is not a zero. The second derivative is negative at $\pi/2$, it is thus a maximum. However, at $x = 3\pi/2$ the function takes on its minimum value, and it is defined there. But it cannot be detected using the standard methods, because the derivative is not continuous there.

4. The function has no zeros. The first derivative is $x^2(3 - 2x^2)\exp(-x^2)$. This yields as possible position for extrema $x = 0$ and $x = \pm\sqrt{3/2}$. The second derivative is $2x(3 - 7x^2 + 2x^4)\exp(-x^2)$. This remains zero at zero. At $\pm\sqrt{3/2}$ this yields $\mp 3\sqrt{6}/e^{\frac{3}{2}}$. Hence there is a minimum at the negative x value and a maximum at the positive x value. The third derivative is $(6 - 54x^2 + 48x^4 - 8x^6)\exp(-x^2)$, which differs from zero at $x = 0$. $x = 0$ is thus a saddle point.

10.8 Integration

Exercises

Determine the following definite integrals, but also always provide the indefinite integrals as well.

1. $\int_{-1}^{1} dx(x^2+1)$
2. $\int_{0}^{\infty} dx x^3 \exp(-x^4)$
3. $\int_{1}^{e} dx(4/x - \sin x \cos x)$
4. $\int_{0}^{1} dx \frac{1}{(1+x)^2}$
5. $\int_{\pi}^{2\pi} dx \ln(xa)$
6. $\int_{0}^{\pi} dx(1+2x)\cos(x+x^2)$
7. $\int_{a}^{b} \sin \frac{3x}{4}$
8. $\int_{-1}^{1} dx(x^3-x)$
9. $\int_{0}^{\infty} dx x^3 \exp(-x^4)$
10. $\int_{0}^{1} dx(1/((\cos(x))^2-1)$
11. $\int_{0}^{1} dx \frac{1}{(1+x)^3}$
12. $-\int_{0}^{\infty} dx x^2 \exp(-\pi x^3)$
13. $\int_{0}^{1-z} dx \frac{x+y}{x+z}$, $z>0$
14. $\int_{\pi}^{e} x^{\pi-e}$

Solutions

Here c always denotes the integration constant.

1. $x + x^3/3 + c$, $8/3$
2. $-\exp(-x^4)/4 + c$, $1/4$. The chain rule can be used
3. $\ln x + \cos(x)^2/2 + c$, $(16 - \cos(2) + \cos(2e))/4$. A trigonometric identify simplifies the integral prior to integration
4. $-\frac{1}{1+x} + c$, $1/2$. Substitution helps
5. $-x + x\log(ax) + c$, $2\pi\ln(2\pi a) - \pi\ln(a\pi) - \pi$. Here the decomposition of the logarithm helps
6. $\sin(x+x^2) + c$, $-\sin(\pi^2)$. Partial integration helps

7. $-\frac{4}{3}\cos\frac{3x}{4}+c$, $\frac{4}{3}\left(\cos\frac{3a}{4}-\cos\frac{3b}{4}\right)$
8. $x^2(x^2-2)/4+c$, 0
9. $-\exp(-x^4)/4+c$, $1/4$
10. $\tan x - x + c$, $\tan(1)-1$
11. $-\frac{1}{2(1+x)^2}+c$, $3/8$. Substitution helps
12. $\exp(-\pi x^3)/(3\pi)$, $-1/(3\pi)$. Partial integration helps
13. $x+(y-z)\ln(x+z)$, $1-z-(y-z)\ln z$. Substitution with $x\to x-z$ is useful
14. $x^{1-e+\pi}/(1-e+\pi)$, $(e^{1-e+\pi}-\pi^{1-e+\pi})/(1-e+\pi)$, using that $e<\pi$

10.9 Probability

Exercises

1. If one solves 8 exercises with probability 50% correctly, how probable is it to solve at least half of them correctly? How likely is it to solve all of them correctly? How many are expected to be solved correctly? Consider the case that the ability to solve the different exercises is not correlated, and ignore fatigue (if this only would be that easy).

2. If you have crossed 5 traffic lights on your way here, which are red with 40% probability, and you may have only two red to be on time, how likely is it that you will be on time? What would be the result at 30% probability for red, but three traffic lights margin? Assume that the traffic lights do not influence each other.

Solutions

1. Using $P^p_{nk}=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$ the results are $\frac{8!}{4!4!2^8}=\frac{35}{128}\approx 0.27$ and $\frac{8!}{8!0!2^8}=\frac{1}{256}\approx 0.0039$. The expectation value of correctly solved exercises is $\sum k\times\frac{1}{2^8}\times\frac{8!}{k!(8-k)!}=4$. This could also be argued because of the equal 50% probabilities.

2. Using $P^p_{nk}=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$ the probability for m to be red is $P^m_n=\sum_{k=0}^{k\leq m}P^p_{nk}$. For the calculation this requires $0!=1$, $1!=1$, $2!=2$, $3!=6$, $4!=24$, $5!=120$, $1-2/5=3/5$, and $1-1/5=4/5$. This yields for the first case $\frac{2133}{3125}\approx 0.68$ and $\frac{124}{125}\approx 0.99$ and thus 68% and 99% probability to be on time.

10.10 Linear algebra

Exercises

Determine for each pair of vectors from the following sets sum, scalar product and cross product, as well as the lengths of each of the vectors in the list.

1. $\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}a\\-a\\0\end{pmatrix},\begin{pmatrix}0\\0\\c\end{pmatrix}\right\}$

2. $\left\{\begin{pmatrix}0\\b\\0\end{pmatrix},\begin{pmatrix}2\\-2\\1\end{pmatrix},\begin{pmatrix}-b\\1\\c\end{pmatrix}\right\}$

3. $\left\{\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}a\\0\\a\end{pmatrix},\begin{pmatrix}0\\-c\\0\end{pmatrix}\right\}$

4. $\left\{\begin{pmatrix}a\\-1\\a\end{pmatrix},\begin{pmatrix}-a\\0\\-a\end{pmatrix},\begin{pmatrix}1\\-1\\0\end{pmatrix}\right\}$

Solutions

In the following, the results will be given as ordered combinations, i. e. first combing vectors 1 and 2, then 1 and 3, and then 2 and 3. For the cross product only one ordering is given, the other being the negative. For the other quantities, the order does not matter. For the lengths it should be noted that they are always positive, while named constants in the vectors may be either positive, negative, or zero, necessitating absolute values occasionally.

1. Lengths $\{1,\sqrt{2}|a|,|c|\}$, sums $\left\{\begin{pmatrix}1+a\\-a\\0\end{pmatrix},\begin{pmatrix}1\\0\\c\end{pmatrix},\begin{pmatrix}a\\-a\\c\end{pmatrix}\right\}$, scalar products $\{a,0,0\}$, and cross products $\left\{\begin{pmatrix}0\\0\\-a\end{pmatrix},\begin{pmatrix}0\\-c\\0\end{pmatrix},-ac\begin{pmatrix}1\\1\\0\end{pmatrix}\right\}$

2. Lengths $\{b|, 3, +\sqrt{1+b^2+c^2}\}$, sums $\left\{\begin{pmatrix}2\\-2+b\\1\end{pmatrix}, \begin{pmatrix}-b\\1+b\\c\end{pmatrix}, \begin{pmatrix}2-b\\-1\\1+c\end{pmatrix}\right\}$, scalar products $\{-2b, b, -2-2b+c\}$, and cross products $\left\{\begin{pmatrix}b\\0\\-2b\end{pmatrix}, \begin{pmatrix}bc\\0\\b^2\end{pmatrix}, \begin{pmatrix}-1-2c\\-b-2c\\2-2b\end{pmatrix}\right\}$

3. Lengths $\{1, \sqrt{2}|a|, |c|\}$, sums $\left\{\begin{pmatrix}a\\1\\a\end{pmatrix}, \begin{pmatrix}0\\1-c\\0\end{pmatrix}, \begin{pmatrix}a\\-c\\a\end{pmatrix}\right\}$, scalar products $\{0, -c, 0\}$, and cross products $\left\{\begin{pmatrix}a\\0\\-a\end{pmatrix}, \vec{0}, ac\begin{pmatrix}1\\0\\-1\end{pmatrix}\right\}$

4. Lengths $\{\sqrt{1+2a^2}, \sqrt{2}|a|, \sqrt{2}\}$, $\left\{\begin{pmatrix}0\\-1\\0\end{pmatrix}, \begin{pmatrix}1+a\\-2\\a\end{pmatrix}, \begin{pmatrix}1-a\\-1\\-a\end{pmatrix}\right\}$, scalar products $\{-2a^2, 1+a, -a\}$, and cross products $\left\{\begin{pmatrix}a\\0\\-a\end{pmatrix}, \begin{pmatrix}a\\a\\1-a\end{pmatrix}, \begin{pmatrix}-a\\-a\\a\end{pmatrix}\right\}$

Index