MATHEMATICAL PREPARATION COURSE before studying Physics

Accompanying Booklet to the Online Course:
www.thphys.uni-heidelberg.de/∼hefft/vk1

without Animations, Function Plotter
and Solutions of the Exercises

Klaus Hefft
Institute of Theoretical Physics
University of Heidelberg

Please send error messages to
k.hefft@thphys.uni-heidelberg.de

September 16, 2020

Contents

PREFACE

Johann Wolfgang von Goethe: FAUST, Part I
(transl. by Bayard Taylor)

WAGNER in Faust's study to Faust:

How hard it is to compass the assistance
Whereby one rises to the source!

FAUST on his Easter walk to Wagner:

That which one does not know, one needs to use;
And what one knows, one uses never.

O happy he, who still renews
The hope, from Error's deeps to rise forever!

From Knowledge to Skill

This course is intended to ease the **transition from school studies to university studies**. It is intended to diminish or compensate for the sometimes pronounced differences in mathematical preparation among incoming students, resulting from the differing standards of schools, courses and teachers. Forgotten and submerged material shall be recalled and repeated, scattered knowledge collected and organized, known material reformulated, with the goal of developing **common mathematical foundations**. No new mathematics is offered here, at any rate nothing that is not presented elsewhere, perhaps even in a more detailed, more exact or more beautiful form.

The main features of this course to emphasize are its selection of material, its compact presentation and modern format. Most of the material of an advanced mathematics school course is selected less for the development of practical math skills, and more for the purpose of intellectual training in logic and axiomatic theory. Here we shall organize much of the same material in a way appropriate for university studies, in some places supplementing and extending it a little.

It is well-known, that in the natural sciences you need to know mathematical terms and operations. You must also be able to work effectively with them. For this reason the many **exercises** are particularly important, because they allow you to determine your own location in the crucial transition region "from knowledge to technique" We shall especially stress practical aspects, even if thereby sometimes the mathematical sharpness (and possibly also the elegance) is diminished. This online course is not a replacement for mathematical lectures. It can, however, be a good preparation for these as well.

Repeating and training basic knowledge must start as early as possible, before gaps in this knowledge begin to impede the understanding of the basic lectures, and psychological barriers can develop. Therefore in Heidelberg the physics faculty has offered to physics beginners, since many years during the two weeks **prior to** the start of the first lectures, a crash course in form of an all-day block course. I have given this course several times since 84/85, with listeners also from other natural sciences and mathematics. We can well imagine that this course can make the beginning considerably easier for engineering students as well. In Heidelberg the online version shall by no means replace the proven crash courses for the beginners prior to their first semester. But it will usefully support and augment these courses. And perhaps it will help incoming students with their preparation, and later with solidifying their understanding of the material. This need might be especially acute for students beginning in summer semester, in particular when Easter holiday is unusually late. Over years this course may also help serve to standardize the material.

This electronic form of the course, free of charge available on the net, seems ideally suited for use during the relaxation time between school graduation and the beginning of lectures at university. A thoughtful student will have time to prepare, to cushion the unfortunately still frequent small shock of the first lectures, if not to avoid it altogether. It seems to us appropriate and meaningful to present this electronic form of the course (which is accessible to you always, and not only two weeks before the semester), to augment and deepen the treatment beyond what is normally possible in our block courses in intensive contact with the Heidelberg beginner students. Furthermore we have often noticed in practice that small excursions in "higher mathematics", historical reviews and physical applications beyond school knowledge energize and awaken a desire to learn more about what is coming. I shall therefore also address here some "higher things", especially toward the ends of the chapters. I will put these topics however in small or larger **inserts** or special exercises, so that they can be passed over without hesitation.

As you have seen from the **quotation** at the beginning from Goethe's (1749-1832) Faust we have to deal with an old problem. But **you** are now in the fortunate situation of having found this course, and you can hope. Don't hesitate! Begin! And have a little fun, too!

Acknowledgements

First of all I want to thank Prof. Dr. Jörg Hüfner for the suggestion and invitation to revise my old, proven "Vorkurs" manuscript. The idea was to redesign and reformat it attractively, making it accessible online, or in the form of the new medium of CD-ROM, to a larger number of interested people (before, during and after the actual preparation course). Thank you for many discussions including detailed questions, tips or formulations, and last not least for the continuous encouragement during the long labor on a project full of vicissitudes.

Then my special thanks go to Prof. Dr. Hans-Joachim Nastold who helped and encouraged me by answering a couple of mathematical questions nearly 50 years ago when I - coming from a law oriented home and a grammar school concentrating on classical languages, without knowing any scientist and lacking any access to mathematical textbooks or to a library, and

confronted with two young brilliant mathematics lecturers - was in a similar, but even more hopeless situation than you could possibly be in now. At that time I decided to someday do something really effective to reduce the math shock, if not to overcome it, if I survived this shock myself.

Prof. Dr. Dieter Heermann deserves my thanks for his competent advice, his influential support and active aid in an early stage of the project. I thank Dr. Thomas Fuhrmann cordially for his enthusiasm for the multimedia ideas, the first work on the electronic conversion of the manuscript, for the programming of the three Java applets, and in particular for the function plotter. To his wife Dr. Andrea Schafferhans-Fuhrmann I owe the correction of a detail important for the users of the plotter.

I also have to thank the following members of the Institute for numerous discussions, suggestions and help, especially Prof. F. Wegner for the attentive correction of the last chapters of the Word script in an early stage, Dr. E. Thommes for exceptionally careful aid during the error location in the HTML text, Prof. W. Wetzel for tireless advice and inestimable help in all sorts of computer questions, Dr. Peter John for relief with some illustrations, Mr. Ting Wang for computational assistance and many other members of the Institute for occasional support and ongoing encouragement.

My main thanks go to my immediate staff: firstly to Mrs. Melanie Steiert and then particularly to Mrs. Dipl.-Math. Katharina Schmock for the excellent transcription of the text into LATEX, Mrs. Birgitta Schiedt and Mr. Bernhard Zielbauer for their enthusiasm and skill in transferring the TEX formulae into the HTML version and finally to Olsen Technologies for the conception of the navigation and the fine organization of the HTML version. To the board of directors of the Institute, in particular Prof. C. Wetterich and Prof. F. Wegner, I owe a great dept of gratitude for providing the funds for this team in the decisive stage.

Furthermore I would like to thank the large number of interested students over the years who, through their rousing collaboration and their questions during the course, or via even later feedback, have contributed decisively to the quality and optimization of the compact form of my lecture script “Mathematical Methods of Physicists” of which the “Vorkurs” is the first part. As a representative of the many whose faces and voices I remember better than their names I want to name Björn Seidel. Many thanks also to all those users of the online course who spared no effort in reporting actual transference problems, or remaining typing and other errors to me, and thus helped me to asymptotically approach the ideal of a faultless text. My thanks go also to Prof. Dr. rer.nat.habil. L. Paditz for critical hints and suggestions for changes of the limits for the arguments of complex numbers.

My thanks especially go to my former student tutors Peter Nalbach, Rainer Tafelmayer, Steffen Weinstock and Carola von Saldern for their help in welcoming the beginners from near and far cheerfully, and motivating and encouraging them. They raised the course above the dull routine of everyday and helped to make it an experience which one may remember with pleasure even years later.

Finally I am full of sincere gratitude to my three children and my son-in-law Christoph Lübbe. Without their perpetual encouragement and untiring help at all times of the day or night I never would have been been able to get so deeply into the world of modern media. To them and to my grandchildren I want to dedicate this future-oriented project:

to **ANGELIKA, JOHANNES, BETTINA** and **CHRISTOPH**
as well as **CAROLINE, TOBIAS, FABIAN, NIKLAS** and **HENRI**.

After the online course resulted in a doubling of the number of German speaking beginners at the physics faculty in Heidelberg within two years, an English version was suggested by Prof. Dr. Karlheinz Meier. I am deeply grateful to cand. phil. transl. Aleksandra Ewa Dastych for her very careful, patient and indispensable help in composing this English version. Also I owe thanks to Prof. K. Meier, Prof. Dr. J. Kornelius and Mr. Andrew Jenkins, B.A. for managing contact to her. My thanks go also to the directors of the Institute, Prof. Dr. C. Wetterich and Prof. Dr. O. Nachtmann, for providing financial support for this task. Many special thanks go to my friend Prof. Dr. Alfred Actor (from Pennsylvania State University) for a very careful and critical expert reading of the English translation.

For the rapid and competent transfer of the English text to the LaTeX format in oder to allow easy printing my whole-hearted thanks go to cand. phys. Lisa Speyer. The support for this work was kindly provided by the relevant commission of our faculty under the chairman Prof. Dr. H.-Ch. Schultz-Coulon.

Chapter 1

MEASURING: Measured Value and Measuring Unit

1.1 The Empirical Method

All scientific insight begins when a curious and attentive person wonders about some phenomenon, and begins a detailed qualitative observation of this aspect of nature. This observing process then can become more and more quantitative, the object of interest increasingly idealized, until it becomes an experiment asking a well-defined question. The answers to this experiment, the measured data, are organized into tables, and can be graphically visualized in diagram form to facilitate the search for correlations and dependencies. After calculating or estimating the precision of the measurement, the so-called experimental error, one can interpolate and search for a **description or at least an approximation in terms of known mathematical curves or formulae** From such empirical connections, conformities to known laws may be discovered. These are mostly formulated in mathematical language (e.g. as differential equations). Once one has found such a connection, one wants to "understand" it. This means either one finds a theory (e.g. some known physical laws) from which one can derive the experimentally obtained data, or one tries using a "hypothesis" to guess the equation which underlies the phenomenon. Obviously also for doing this task a lot of mathematics is necessary. Finally mathematics is needed once again to make predictions which are intended to be checked against experiments, and so on. In such an upward spiral science is progressing.

1.2 Physical Quantities

In the development of physics it turned out again and again how difficult, but also important it was to develop the most suitable concepts and find the **relevant quantities** (e.g. force or energy) in terms of which nature can be described both simply and comprehensively.

> **Insert: History:** *It took more than 100 years for the discussion among the "natural philosophers" (especially D´Alembert, Bruno, Newton, Leibniz, Boskovic and Kant) to create our modern concepts of force and action from the old terms principium, substantia, materia, causa efficiente, causa formale, causa finale, effectum, actio, vis viva and vis insita.*

Every **physical quantity** consists of a **a measured value and a measuring unit**, i.e. a pure number and a dimension. All difficulties in conversations are avoided, if we treat both parts like a product "value **times** dimension".

Example: Velocity: In residential districts often a speed limit $v = 30\frac{km}{h}$ is imposed, which means 30 kilometers per hour. How many meters is that per second?
One kilometer contains 1000 meters: $1km = 1000m$, thus $v = 30 \cdot 1000\frac{m}{h}$.
Every hour consists of 60 minutes: $1h = 60min$, consequently $v = 30 \cdot 1000\frac{m}{60min}$.
One minute has 60 seconds: 1 min = 60 s , therefore $v = 30 \cdot 1000\frac{m}{60 \cdot 60s} = 8.33\frac{m}{s}$.
Even that may be too fast for a ball playing child.

> **Insert: Denotations:** *It is an accepted thing in international physics for long time past to abbreviate as many of the physical quantities as possible by the first letter of the corresponding* **English** *word, e.g. s(pace), t(ime), m(ass), v(elocity), a(cceleration), F(orce), E(nergy), p(ressure), R(esistance), C(apacity), V(oltage), T(emperature), etc..*
> *Of course there are some exceptions from this rule: e.g. momentum p, angular momentum l, electric current I or potential V*

Whenever the Latin alphabet is not sufficient, we use the **Greek** one:

alpha	α	A	iota	ι	I	rho	ρ	P
beta	β	B	kappa	κ	K	sigma	σ	Σ
gamma	γ	Γ	lambda	λ	Λ	tau	τ	T
delta	δ	Δ	my	μ	M	ypsilon	υ	Y
epsilon	ϵ	E	ny	ν	N	phi	ϕ	Φ
zeta	ζ	Z	xi	ξ	Ξ	chi	χ	X
eta	η	H	omikron	o	O	psi	ψ	Ψ
theta	θ	Θ	pi	π	Π	omega	ω	Ω

In addition the **Gothic** alphabet is at our disposal.

1.3 Units

The **units** are defined in terms of yardsticks. The search for suitable **yardsticks** and their definition, by as international a convention as possible, is an important part of science.

> **Insert: Standard units:** *What can be used as a standard unit? - The answers to this question have changed greatly through the centuries. Originally people everywhere used easily available comparative quantities like cubit or foot as units of length, and the human pulse beat as unit of time. (The Latin word tempora initially meant temple!) But not every foot has equal length, and the pulse can beat more quickly or slowly. Alone in Germany there have been more than 100 different cubit and foot units in use.*
>
> *Therefore, since 1795 people referred to the ten millionth part of the earth meridian quadrant as the "meter" and represented this length by the well-known rod made out of an alloy of platinum and iridium. The measurement of time was referred to the earth's rotation: for a long time the second was defined as the 86400th part of an average solar day.*
>
> *In the meantime more exact atomic standards have been introduced: One meter is now the distance light travels within the 1/299 792 485 part of a second. One second is now defined in terms of the period of a certain oscillation of cesium 133 atoms in "atomic clocks". Perhaps some day these standards will also be improved.*

Today, these questions are solved after many error ways by the conventions of the **SI-units** (Système International d'Unités) The following **fundamental quantities** are specified:

length measured in meters:	m
time in seconds:	s
mass in kilograms:	kg
electric current in ampere:	A
temperature in kelvin:	K
luminous intensity in candelas:	cd
even angle in radiant:	rad
solid angle in steradiant:	sr
amount of material in mol:	mol

All remaining physical quantities are to be regarded as **derived**, thus by laws, definitions or measuring regulations traced back to the fundamental quantities: e.g.

frequency measured in hertz:	Hz := 1/s
force in newton:	N := kg m/s^2
energy in joule:	J := Nm
power in watt:	W := J/s
pressure in pascal:	Pa := N/m^2
electric charge in coulomb:	C := As
electric potential in volt:	V := J/C
electric resistance in ohm:	Ω := V/A
capacitance in farad:	F := C/V
magnetic flux in weber:	Wb := Vs

Exercise 1.1 SI-units

a) What is the SI-unit of momentum?

b) From which law can we deduce the unit of force?

c) Who formulated this law first?

d) What is the dimension of work?

e) What is the unit of the electric field strength?

Insert: Old units: *Some examples of units which are still widely in use in spite of the SI-convention:*

grad:	$^\circ = (\pi/180) rad = 0.01745\ rad$
kilometer per hour:	*km/h = 0.277 m/s*
horse-power:	*PS = 735.499 W*
calorie:	*cal ≃ 4.185 J*
kilowatt-hour:	$kWh = 3.6 \cdot 10^6 J$
elektron volt:	$eV \simeq 1.6 \cdot 10^{-19} J$

Many **non-metric units** *are still used especially in England and the USA:*

inch = Zoll:	*in = " = 2.54 cm*
foot:	*ft = 12 in ≃ 0.30 m*
yard:	*yd = 3 ft ≃ 0.9144 m*
(amer.) mile:	*mil = 1760 yd ≃ 1609 m*
ounce:	*oz ≃ 28.35 g*
(engl.) pound:	*lb = 16 oz ≃ 0.454 kg*
(amer.) gallon:	*gal ≃ 3.785 l*
(amer.) barrel:	*bbl = 42 gal ≃ 158.984 l*

Exercise 1.2 Conversion of units

a) You are familiar with the conversion of angles from degrees to radiants using your pocket calculator: Calculate $30°$, $45°$, $60°$, *and* $180°$ *in radiant and 1 rad and 2 rad in degrees.*

b) How many seconds make up one sidereal year with 12 months, 5 days, 6 hours, 9 minutes and 9.5 seconds?

c) How much does it cost with an "electricity tariff" of 0.112 €/kWh, if you burn one night long a 60-Watt bulb for six hours and your PC runs needing approximately 200 watts?

d) Maria and Lucas measure their training distance with a stick, which is 5 feet and 2 inches long. The stick fits in 254 times. What is the run called in Europe?
How many rounds do Maria and Lucas have to run, until they put a mile back?

e) Bill Gates said: "If General Motors had kept up with technology like the computer industry has, we would all be driving twenty-five dollar cars that go 1000 miles per gallon." Did he mean the "3-litre car"?

1.4 Order of Magnitude

Natural phenomena are so various and cover so many **orders of magnitude**, that in relation to a standard unit, e.g. meter, tiny or enormous numbers often result. Just think of the diameter of an atom or the size of our Milky Way expressed in meters. In both cases "useless" zeros arise. One has therefore introduced powers of ten and as well as abbreviations and easily remembered names: e.g. the kilogram 1000 g = 103 g = kg. The **decimal prefixes**, too, are today internationally standardized. We indicate the most important ones:

tenth	10^{-1}	= d	dezi-	ten	10^{1}	= D	deka-
hundredth	10^{-2}	= c	centi-	hundred	10^{2}	= h	hecto-
thousandth	10^{-3}	= m	milli-	thousand	10^{3}	= k	kilo-
millionth	10^{-6}	= μ	mikro-	million	10^{6}	= M	mega-
billionth	10^{-9}	= n	nano-	billion	10^{9}	= G	giga-
trillionth	10^{-12}	= p	pico-	trillion	10^{12}	= T	tera-
quadrillionth	10^{-15}	= f	femto-	quadrillion	10^{15}	= P	peta-

Examples: In order to give you an idea of orders of magnitude, we give some examples from the field of length measurement:

- The diameter of the range, within which scattered electrons feel a proton, amounts to about 1.4 fm, atomic nuclei are between 3 and 20 fm thick.
- The wavelengths of gamma-rays lie within the range of pm. Atomic diameters reach from 100 pm to 1 nm.
- Important molecules are about 10 nm thick. 100 nm is the order of magnitude of viruses, and also the wavelengths of visible light lie between 300 and 800 nm.
- Bacteria have typical diameters of μm, our blood corpuscles of $10\mu m$, and protozoan measure some $100\mu m$.
- Thus we already come to your everyday life scale of pinheads: 1 mm, hazel-nuts: 1 cm and grapefruits: 1 dm.
- Electromagnetic short waves are 10 to 100 m long, medium waves 100 m to 1 km and oscillate with 1 MHz. The distance e.g. of the bridges over the Neckar river in Heidelberg amounts to 1 km. Flight altitudes of the large airliners are about 10 km.
- The diameter of the earth is to 12.7 Mm and that of the Jupiter is about 144 Mm. The sun's diameter is with 1.4 Gm, the average distance of the earth from the sun is approximately 150 Gm, and Saturn circles at a distance of approximately 1.4 Tm around the sun.
- Finally, light travels 9.46 Pm in one year.

Insert: Billion: *While these prefixes of the SI system are internationally fixed, this is by no means so with our familiar number words . The* **Anglo-American** *and also French expression "billion" in the above table means the German "Milliarde"* $= 10^9$ *and is different from the German Billion* $= 10^{12}$*. "The origin of our sun system 4,6 billion years ago..." must be translated as "die Entstehung unseres Sonnensystems vor 4,6 Milliarden Jahren...". Similar things apply to the Anglo-American "trillion"* $= 10^{12}$*, while the German "Trillion"* $= 10^{18}$*.*

Insert: Other unit names: *Special names are also still used for some metric units: You know perhaps* $10^2 m^2$ *as are,* $10^4 m^2$ *as hectare,* $10^{-3} m^3$ *as litre,* $10^2 kg$ *as quintal and* $10^3 kg$ *as ton.*

Do you also know $10^5 Pa$ *as bar,* $10^{-28} m^2 = bn$ *as barn,* $10^{-5} N = dyn$*,* $10^{-7} J = erg$*,* $10^{-15} m = fm$ *under the name of Fermi,* $10^{-10} m = 1\mathring{A}$ *after Ångström or* $10^{-8} Wb$ *under the name of Maxwell?*

Exercise 1.3 Decimal prefixes

a) Express the length of a stellar year (365 d + 6 h + 9 min + 9.5 s) in megaseconds.

b) The ideal duration of a scientific seminar talk amounts to one microcentury.

c) How long does a photon need, in order to fly with the speed of light $c = 2.997\,924\,58 \cdot 10^8$ *m/s 21 m far through the lecture-room?*

d) With the Planck energy of $E_p = 1.22 \cdot 10^{16}$ *TeV gravitation effects for the elementary particles are expected. Express the appropriate Planck mass* M_P *in grams.*

In the following we are only concerned with the **numerical values** of the examined physical quantities, which we read off usually in the form of lengths or angles from our measuring apparatuses, these being calibrated for the desired measuring range in appropriate units of the measured quantities.

Chapter 2

SIGNS AND NUMBERS and Their Linkages

The laws of numbers and their linkages are the main objects of mathematics. Although numbers have developed from basic needs of human social interaction, and natural science has inspired mathematics again and again, e.g. for differential and integral calculus, mathematics actually does not belong to natural sciences, but rather to humanities. Mathematics does not start from empirical (i.e. measured) facts. Instead, it investigates the logical structure of numbers and their generalizations within the human ability of thought. In many cases empirical facts can be well represented in terms of these logical structures. In this way mathematics became an indispensable tool for natural scientists and engineers.

2.1 Signs

Mathematics like every other science has developed its own language. This language includes among other things some mathematical and logical **signs**, which we would like to list here for quick, clear reference, because we will be using them continually:

Question game: Some *mathematical* signs

The meaning of the following mathematical signs is known to most of you. ONLINE you can challenge yourself and click directly on the symbols to check if you are right. If your browser does not support this, you will find a complete list of answers here:

$+$: plus	-: minus	$\pm$: plus or minus
$\cdot$: times	/: divided by	$\perp$: is perpendicular to
$<$: is smaller than	$\leq$: is smaller or equal to	$\ll$: is much smaller than
$=$: is equal to	$\neq$: is unequal to	$\equiv$: is identically equal to
$>$: is bigger than	$\geq$: is bigger or equal to	$\gg$: is much bigger than
$\angle$: angle between	$\simeq$: is approximately equal to	∞ : bigger than every number

Insert: Infinity: *Physicists often use the sign ∞, known as "infinity", rather casually. Assuming the meaning "bigger than every number" we avoid the problems mathematicians warn us about: thus $a < \infty$ means a is a finite number. Shortly we will use the combination of symbpols $\rightarrow \infty$ whenever we mean that a quantity is "growing beyond all limits".*

In addition, we use the

Sum Sign $\sum$: for example $\sum_{n=1}^{3} a_n := a_1 + a_2 + a_3$

A famous example is the sum of the first m natural numbers:

$$\sum_{n=1}^{m} n := 1 + 2 + \ldots + (m-1) + m = \frac{m}{2}(m+1),$$

just as the young Gauss has proved by skillful composition and clever use of brackets: $\sum_{n=1}^{m} n = (1+m) + (2+(m-1)) + (3+(m-2)) + \ldots = \frac{m}{2}(m+1)$.

Another example is the sum of the first m squares of natural numbers:

$$\sum_{n=1}^{m} n^2 := 1 + 4 + \ldots + (m-1)^2 + m^2 = \frac{m}{6}(m+1)(2m+1),$$

a formula we will later need for the calculation of integrals.

A further example is the sum of the first m powers of a number q:

$$\sum_{n=0}^{m} q^n := 1 + q + q^2 + \ldots + q^{m-1} + q^m = \frac{1-q^{m+1}}{1-q} \text{ for q} \neq 1,$$

which is known as the **"geometrical" sum**.

Insert: Geometric sum: *Just as an exception, we want to prove the formula for the geometric series which we will need several times. To do this we define the sum*

$$s_m := 1 + q + q^2 + \ldots + q^{m-1} + q^m,$$

then we subtract from this $$q \cdot s_m = q + q^2 + q^3 + \ldots + q^m + q^{m+1}$$

and obtain (since nearly everything cancels)
$s_m - q \cdot s_m = s_m(1-q) = 1 - q^{m+1}$, *from which we easily get for* $q \neq 1$ *dividing by* $(1-q)$ *the above formula for* s_m.

Much more important than the **product sign** $\prod$, defined analogously to the sum sign: for instance $\prod_{n=1}^{3} a_n := a_1 \cdot a_2 \cdot a_3$ is for us the

factorial sign ! : $m! := 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (m-1) \cdot m = \prod_{n=1}^{m} n$

(speak: "m factorial"), e.g. $3! = 1 \cdot 2 \cdot 3 = 6$ or $5! = 120$, augmented by the convention $0! = 1$.

Question game: Some *logical* signs

From the logical symbols which most of you are familiar with from math class, we use the following symbols to display logical connections in a simpler, more concise, and memorable way, as well as to make it easier for us to memorize them. ONLINE you can click directly on the symbols to get the answer. If your browser does not support this, you will find a complete list of answers :

$\in$: is an element of	$\ni$: contains as element	$\notin$: is no element of
$\subseteq$: is a subset of or equal	$\supseteq$: contains as a subset or is equal	$:=$: is defined by
$\exists$: there exists	$\exists!$: there exists exactly one	$\forall$: for all
$\cup$: union of sets	$\cap$: intersection of sets	$\emptyset$: empty set
$\Rightarrow$: from this it follows that, is a sufficient condition for	$\Leftarrow$: this holds when, is a necessary condition for	$\Leftrightarrow$: this holds exactly when, is a nec. and suff. cond. for

These symbols will be explained once more when they occur for the first time in the text.

2.2 Numbers

In order to display our measured data we need the numbers which you have been familiar with for a long time. In order to get an overview, we shall put together here their properties as a reminder. In addition we recall some selected concepts which mathematicians have formulated as rules for the combination of numbers, so that we can later on compare those rules with the ones for more complicated mathematical quantities.

2.2.1 Natural Numbers

We begin with the set of *natural numbers* $\{1, 2, 3, \ldots\}$, given the name $\mathbb{N}$ by number theoreticians and called "natural" because they have been used by mankind to count within living memory. Physicists think for instance of particle numbers, e.g. the number of atoms or molecules in a mole.

For long time now there have been *two different linkages*: the operation of addition and multiplication, assigning a new natural number to each pair of natural numbers $a, b \in \mathbb{N}$ ("the numbers a and b are elements of the set $\mathbb{N}$") and therefore called *internal* linkages:

the **ADDITION:**

internal linkage:	$a + b = x \in \mathbb{N}$	with the
Commutative Law:	$a + b = b + a$	and the
Associative Law:	$a + (b + c) = (a + b) + c$	and

the **MULTIPLICATION:**

internal linkage:	$a \cdot b$ or $ab = x \in \mathbb{N}$	also with a
Commutative Law:	$ab = ba$	and a
Associative Law:	$a(bc) = (ab)c$	and furthermore a
Neutral element: the one:	$1a = a$	

Both linkages, addition and multiplication, are connected through the

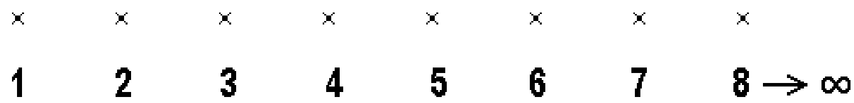
Distributive Law: $(a + b)c = ac + bc$

with each other.

> **Insert: Shorthand:** *If we want to express that in the set of natural numbers* ($n \in \mathbb{N}$) *there exists only exactly one* ($\exists!$) **element one** *which for all* ($\forall$) *natural numbers a fulfils the equation* $1a = a$ *, we could express this using the logical signs in the following manner:* $\exists!\ 1 \in \mathbb{N} : \forall a \in \mathbb{N} \quad 1a = a$. *Please appreciate this compact logical writing.*

> **Insert: Counter-examples:** *As an example of a linkage that leads out of a set, we will soon deal with the well known scalar product of two vectors, which combines their components into a simple number.*
> *Non-communicative are for example the rotations of the match box shown in Figure 9.10 in a Cartesian coordinate system: First turn it clockwise around the longitudinal symmetry axis parallel to the 3-axis and then around the shortest transversal axis parallel to the 1-axis and compare the result with the position of the box after you have performed the two rotations in the reversed order!*
> *Counter-examples of the bracket law for three elements of a set are very hard to find: From the home chemistry sector we remember the three ingredients for non-fat whipped cream for children: (sugar + egg-white) + juice = cream. If you try to whip first the egg-white together with juice as suggested by the instruction: sugar + (egg-white + juice) you will never get the cream whipped.*

We can clearly imagine the natural numbers as equally spaced points on a half line as shown in the next figure:



For physicists it is sometimes convenient to add the zero 0 as you would with a ruler, and so to extend $\mathbb{N}$ to $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Through this, the addition operation also obtains a uniquely defined

Neutral element: the zero: $\quad 0 + a = a$

In "logical shorthand": $\exists!\ 0 \in \mathbb{N}_0 : \forall a \in \mathbb{N}_0 \ : \ 0 + a = a$ in full analogy to the neutral element of multiplication.

Insert: History: *Even the ancient Greeks and Romans did not know numbers other than the natural ones:* $\mathbb{N} = \{I, II, III, IV, \ldots\}$. *The Chinese knew zero as "empty place" already in the 4th century BC. Not before the 12th century AD did the Arabs bring the number zero to Europe.*

2.2.2 Integers

Along with the progress in civilization and human culture it became necessary to extend the numbers. For example, when talking about money it is not sufficient to know the amount (e.g. the number of coins) we also need to be able to express whether we have or owe that amount. Sometimes this is expressed through the colour of the number ("black and red numbers") or through a preceding $+$ or $-$ sign. In the natural sciences such signs have been established.

Physicists can shift a marking on their ruler by an arbitrary number of points to the right, they will however encounter difficulties if they want to move it to the left. Mathematically speaking does not have for all natural numbers a and b the equation $a+x=b$ a solution x which is itself a natural number: e.g. the equation $2+x=1$. Such equations can then only be solved if we extend the natural numbers through the *negative* numbers $\{-a| \quad a \in \mathbb{N}\}$ to form the set of all **integers**:

To every positive element a there exists exactly one

Negative element $-a$ with: $\qquad a + (-a) = 0$

Even for 1 we get a -1, meaning owing a pound in contrast to possessing a pound. In "logical shorthand" : $\forall a \in \mathbb{Z}\ \exists! -a : a + (-a) = 0$.

Mathematicians refer to the set of integers, which consist of all natural numbers $a \in \mathbb{N}$, their negative partners $-a \in (-\mathbb{N})$ and zero as $\mathbb{Z} := \mathbb{N} \cup \{0\} \cup \{-a| \quad a \in \mathbb{N}\}$.

With this extension, the above equation $a + x = b$ has now, as desired, always a solution for all pairs of integers, namely the *difference* $x = b - a$ which once again is an integer $x \in \mathbb{Z}$. We also say that $\mathbb{Z}$ is "*closed*" concerning addition: i.e. addition does not lead out of the set. This brings us to a central concept in mathematics (and in physics), namely that of a *group :*

We call a set of objects (e.g. the integers) a **group**, if

1. it is closed concerning an internal linkage (like e.g. addition),
2. an Associative Law holds (like e.g.: $a + (b + c) = (a + b) + c$),
3. it encloses exactly one neutral element (like e.g. the number 0) and
4. if there exists exactly one reversal for each element (like e.g. the negative element).

If moreover the Commutative Law (like e.g. $a + b = b + a$) holds, mathematicians call the group **Abelian**.

> **Insert: Groups:** *Later on you will learn that groups play a very important role in the search for symmetries in physics, e.g. for crystals or the classification of elementary particles. The elements of a group are often operations, like e.g. rotations: the result of two rotations performed one after the other can also be reached by one single rotation. In performing three rotations the result does not depend on the brackets. The operation* no rotation *leaves the body unchanged. Each rotation can be cancelled. Usually these groups are not Abelian, e.g. two rotations performed in different order yield different results. Therefore mathematicians did not incorporate the Commutative Law into the properties of groups. The more specialized commutative groups are given the name Abelian after the Norwegian mathematician Niels Henrik Abel (1802-1829).*

We can imagine the integers geometrically as equidistant points on a *whole* straight line.

$$-\infty \leftarrow -3 \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \rightarrow \infty$$

> **Insert: Absolute value:** *If we, while viewing a number decide to ignore its sign, we use the term*
>
> **absolute value:** $|a| := a$ *for* $a \geq 0$ *and* $|a| := -a$ *for* $a < 0$,
>
> *so that* $|a| \geq 0\ \forall a \in \mathbb{Z}$.
> *For Instance for the number* $-5 : |-5| = 5$ *and for the number* $3 : |3| = 3 = 3$.
>
> *The multiplication rule for the product of absolute values:*

$$|a \cdot b| = |a| \cdot |b|$$

can easily be verified. For the absolute values of the sum and difference of integers there hold only inequalities *which we will meet later.*

$$||a| - |b|| \leq |a \pm b| \leq |a| + |b|.$$

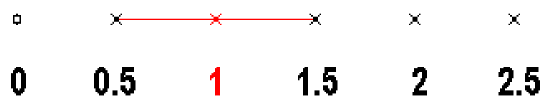
The second part is known as "Triangle Inequality".

The term $|a - b|$ *then gives the* distance *between the numbers* a *and* b *on the line of numbers.*
All points a *in the neighbourhood interval of a point* a_0*, having a distance from* a_0 *which is smaller than a positive number* ε *is called a* ε-neighbourhood $U_\varepsilon(a_0)$ *of* a_0 :

Insert: ε-neighbourhood: *You will often encounter the term of an* **ε-neighbourhood** *you will often meet in future mathematics lectures:*

$$a \in U_\varepsilon(a_0) \Longleftrightarrow |a - a_0| < \varepsilon \quad \textit{with } \varepsilon > 0.$$

We will use it only a few times here.



The Figure shows the ε-neighbourhood of the number 1 for $\varepsilon = 1/2$. It contains all numbers x with $0.5 < x < 1.5$. Realize that the borders (here 0.5 and 1.5) do not belong to the neighbourhood.

2.2.3 Rational Numbers

Whenever people have been forced to do division, they have noticed that integers are not enough. Mathematically speaking: to solve the equation $a \cdot x = b$ for $a \neq 0$ within a number set we are forced to extend the integers to **rational numbers** $\mathbb{Q}$ by adding the *inverse* numbers $\{\frac{1}{a}$ or $a^{-1}|a \in \mathbb{Z}\}$. We use the notation $\mathbb{Z} \setminus \{0\}$ for the set of integers without the zero. Then we have for each integer a different from 0 exactly one

$$\textit{inverse element } a^{-1} \text{ with:} \quad a \cdot a^{-1} = 1$$

In "logical shorthand": $\forall a \in \mathbb{Z} \setminus \{0\} \ \exists! \ a^{-1} : \ a \cdot a^{-1} = 1.$

We are familiar with this concept. The inverse to the number 3 is $\frac{1}{3}$, the inverse number to -7 is $-\frac{1}{7}$.

This way the *fraction* $x = \frac{b}{a}$ for $a \neq 0$ solves our starting equation $ax = b$ as desired. In general, a rational number is the quotient of two integers, consisting out of a numerator and a denominator (different from 0). Rational numbers are therefore mathematically speaking, *ordered pairs* of integers: $x = (b, a)$.

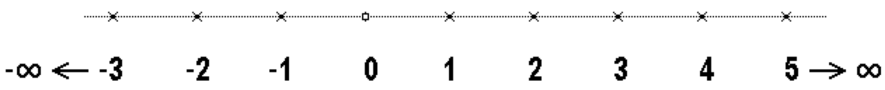
> **Insert: Class:** *Strictly speaking* one *rational number is always represented by a whole* class *of ordered pairs of integers, e.g.* $(1,2) = (2,4) = (3,6) = (1a, 2a)$ *for* $a \in \mathbb{Q}$ *and* $a \neq 0$ *should be taken as one single number:* $1/2 = 2/4 = 3/6 = 1a/2a$: *Cancelling should not change the number, as we know.*

When they are divided out, the rational numbers become finite, meaning breaking off or periodic decimal fractions: for example $\frac{1}{5} = 0.2$, $\frac{1}{3} = 0.3333333... = 0,\overline{3}$ and $\frac{1}{11} = 0.09090909... = 0.\overline{09}$, where the line over the last digits indicates the period.

With this definition of the inverse elements the rational numbers form a group not only relative to addition, but *also*, relative to multiplication (with the Associative Law, the one and the inverse elements). This group is, due to the Commutative Law of the factors $ab = ba$ **Abelian**.

> **Insert: Field:** *For sets which form groups subject to* two *internal linkages connected by a Distributive Law mathematicians have created a special name because of their importance: They call such a set a* ***field***.

The rational numbers lie *densely* on our number line, meaning in every interval we can find countable infinity of them:



Because of the finite accuracy of every physical measurement the rational numbers are in every practical aspect ***the working numbers of physics*** as well as in every other natural science. This is why we had paid such an attention to their rules.

By stating results as rational numbers, mostly in the form of decimal fractions, scientists worldwide have agreed on indicating only *as many decimal digits* as they have measured.

Along with every measured value the uncertainty should also be indicated. This for example is what we find in a table for Planck's quantum of action

$$\hbar = 1.054\ 571\ 68(18) \cdot 10^{-34}\ Js.$$

This statement can also be written in the following way:

$$\hbar = (1.054\ 571\ 68 \pm 0.000\ 000\ 18) \cdot 10^{-34}\ Js$$

meaning that the value of $\hbar$ (with a probability of 68 %) lies between the following two borders:

$$1.054\ 571\ 50 \cdot 10^{-34}\ Js \leq \hbar \leq 1.054\ 571\ 86 \cdot 10^{-34}\ Js.$$

Exercise 2.1

a) Show with the above indicated prescription of Gauss for even m, that the formula for the sum of the first m natural numbers $\sum_{n=1}^{m} n = \frac{m}{2}(m+1)$ holds also for odd m gilt.

b) Prove the above stated formula for the first m squares of natural numbers $\sum_{n=1}^{m} n^2 = \frac{m}{6}(m+1)(2m+1)$ by considering $\sum_{n=1}^{m}(n+1)^3$

c) What do the following statements out of the "particle properties data booklet" mean: $e = 1.602\ 176\ 53(14) \cdot 10^{-19}\ Cb$ and $m_e = 9.109\ 382\ 6(16) \cdot 10^{-31}\ kg$?

Insert: Powers: *Repeated application of the same factor we describe usually as* **power** *with the number of factors as*

exponent: $b^n := b \cdot b \cdot b \cdots b$ *in case of n factors b,*

where the known

calculation rules $b^n b^m = b^{n+m}$, $(b^n)^m = b^{n \cdot m}$ *and* $(ab)^n = a^n b^n$ *for* $n, m \in \mathbb{N}$

hold true. With the definitions $b^0 := 1$ and $b^{-n} := 1/b^n$ these calculation rules can be extended to all integer exponents: $n, m \in \mathbb{Z}$. Later we will generalize yet further.

As a first application of powers we mention the ***Pythagoras Theorem:*** *In a* right-angled *triangle the square over the hypotenuse c equals the sum of the squares over both catheti a and b:*

Pythagoras Theorem: $a^2 + b^2 = c^2$

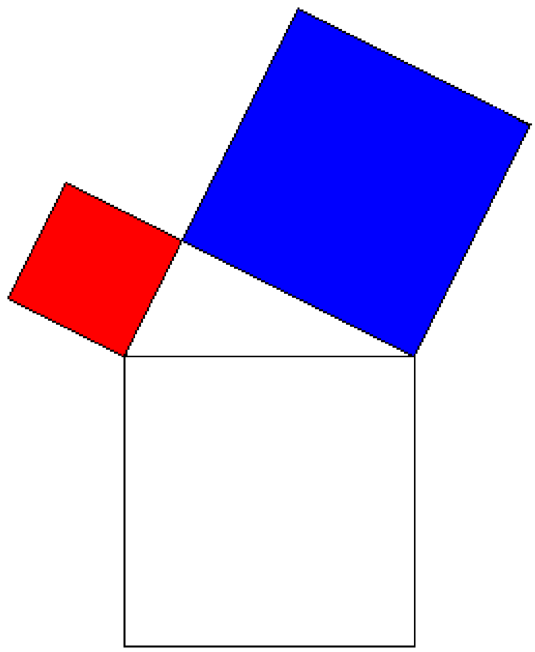


Figure 2.5 illustrates the Pythagoras Theorem, ONLY ONLINE with coloured parallelograms indicating the geometrical proof.

Very frequently we need the so-called

binomial formulas: $(a \pm b)^2 = a^2 \pm 2ab + b^2$ and $(a+b)(a-b) = a^2 - b^2$,

which can be easily derived, but need to be memorized.

The binomial formulas are a special case (for $n = 2$) of the more general formula

$$(a \pm b)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} a^{n-k} (\pm b)^k,$$

where $\frac{n!}{k!(n-k)!} =: \binom{n}{k}$ are the so-called **binomial coefficients**. We can calculate them either directly from the definition of the factorial, e.g.

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5}{1 \cdot 2 \cdot 3 \cdot 1 \cdot 2} = 10$$

or find them in the Pascal Triangle. This triangle is constructed in the following way:

$n = 0:$							1						
$n = 1:$						1		1					
$n = 2:$					1		2		1				
$n = 3:$				1		3		3		1			
$n = 4:$			1		4		6		4		1		
$n = 5:$		1		5		10		10		5		1	
$n = 6:$	1		6		15		20		15		6		1

We start with the number 1 in the line $n = 0$. In the next line ($n = 1$) we write two ones, one on each side. Then for $n = 2$ we add two ones to the left and right side once again, and in the gap between them a $2 = 1 + 1$ as the sum of the left and right "front man" (in each case a 1). In the framed box, we once again recognize the formation rule. The required binomial coefficient $\binom{5}{3}$ is then found in line $n = 5$ on position 3.

Exercise 2.2

a) Determine the length of the space diagonal in a cube with side length a.

b) Calculate $(a^4 - b^4)/(a - b)$.

c) Calculate $\binom{n}{0}$ and $\binom{n}{n}$.

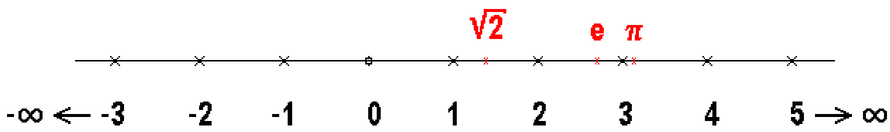
d) Calculate $\binom{7}{4}$ and $\binom{8}{3}$.

e) Show that $\binom{n}{n-k} = \binom{n}{k}$ holds true.

f) Prove the formation rule for the Pascal Triangle: $\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$.

2.2.4 Real Numbers

Mathematicians were however not fully satisfied with the rational numbers, seeing how for example something as important as the circumference π of a circle with the diameter of 1 is not a rational number: $\pi \notin \mathbb{Q}$. They also wanted the solution of the equation $x^2 = a$ at least for $a \neq 0$, as well as the *roots* $x = a^{1/2} =: \sqrt{a}$ to be included. This is why the rational numbers (by addition of infinite decimal fractions) have been extended to the *real* numbers $\mathbb{R}$ which can be mapped one-to-one onto a *straight line* $\mathbb{R}^1$ (meaning every point on the line corresponds to exactly one real number).



Insert: History: *Already in antiquity some mathematicians knew that there are numbers which cannot be represented as fractions. They showed this with a so-called indirect proof:*

If e.g. the diagonal of a square with side length 1 were a rational number, like $\sqrt{2} = b/a$, two natural numbers $b, a \in \mathbb{N}$ would exist with $b^2 = 2a^2$. Think now of the prime factor decompositions of b and a. On the left hand side of the equation there stands an even number of these factors, because of the square each factor appears twice. On the right hand side, however, an odd number of factors shows up, because in addition the factor 2 appears. Since the prime factor decomposition is unique, the equation cannot be right.

With this it is shown that the assumption, $\sqrt{2}$ can be represented as a fraction, leads to a contradiction and thus must be wrong.

With the real numbers, which have the same *calculation rules of a field* as the rational numbers, both solutions of the general

$$\textbf{quadratic equation:} \quad x^2 + ax + b = 0, \quad x_{1,2} = -\frac{a}{2} \pm \sqrt{\frac{a^2}{4} - b}$$

will then be real numbers, as long as the discriminant under the root is not negative: $a^2 \geq 4b$.

Insert: Preview: complex numbers: *Later in Chapter 8 we will go one step further by introducing the* **complex numbers** $\mathbb{C}$ *for which e.g. also $x^2 = a$ for $a < 0$ is always solvable and, amazingly enough, many other beautiful laws hold.*

Chapter 3

SEQUENCES AND SERIES and Their Limits

Direct mathematical study of sequences and series are, for natural scientists, less important than the fact that they greatly help us to understand and perform the limiting procedures which are of fundamental importance in physics. For this reason, we have combined in this chapter the most important facts of this part of mathematics. Later you will deal in greater detail with these things in your future mathematics lectures.

3.1 Sequences

The first important mathematical concept we have to inspect is that of a **sequence**. With this physicists think for instance of the sequence of the bounce heights of a steel ball on a plate, which due to the inevitable dissipation of energy decrease with time and tend more or less quickly to zero. After a while, the ball remains still. The resulting physical sequence of the jump heights has only a finite number of non-vanishing members in contrast to the ones that are of interest to mathematicians: Mathematically, a sequence is an infinite set of numbers which can be numbered consecutively, i.e. labelled by the set of the natural numbers: $(a_n)_{n\in\mathbb{N}}$. Because it is impossible to list all infinite many members $(a_1, a_2, a_3, a_4, a_5, a_6, \ldots)$, a sequence is mostly defined by the "general member" a_n, which is a law stating how to calculate the individual members of the sequence. Let us look at the following typical examples which already enable us to display all important concepts:

(F1) $1, 2, 3, 4, 5, 6, 7, \ldots = (n)_{n\in\mathbb{N}}$ the natural numbers themselves

(F2) $1, -1, 1, -1, 1, -1, \ldots = ((-1)^{n+1})_{n\in\mathbb{N}}$ a simple "alternating" sequence,

(F3) $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots = \left(\frac{1}{n}\right)_{n\in\mathbb{N}}$ the inverse natural numbers, the so-called "harmonic" sequence,

(F4) $1, \frac{1}{2}, \frac{1}{6}, \frac{1}{24}, \ldots = \left(\frac{1}{n!}\right)_{n\in\mathbb{N}}$ the inverse factorials,

(F5) $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots = \left(\frac{n}{n+1}\right)_{n\in\mathbb{N}}$ a sequence of proper fractions and

(F6) $q, q^2, q^3, q^4, q^5 \ldots = (q^n)_{n\in\mathbb{N}}\,, \quad q \in \mathbb{R}$ the "geometric" sequence.

> **Insert: Compound interest:** *Many of you know the geometrical sequence from school because it causes a capital K_0 at p% compound interest after n years to increase to $K_n = K_0 q^n$ with $q = 1 + \frac{p}{100}$.*

In order to give us a first clear idea of these sample sequences, we have plotted the sequence members a_n (in the 2-direction) over the equidistant natural numbers n (in the 1-direction) in the following Cartesian coordinate system in a plane:

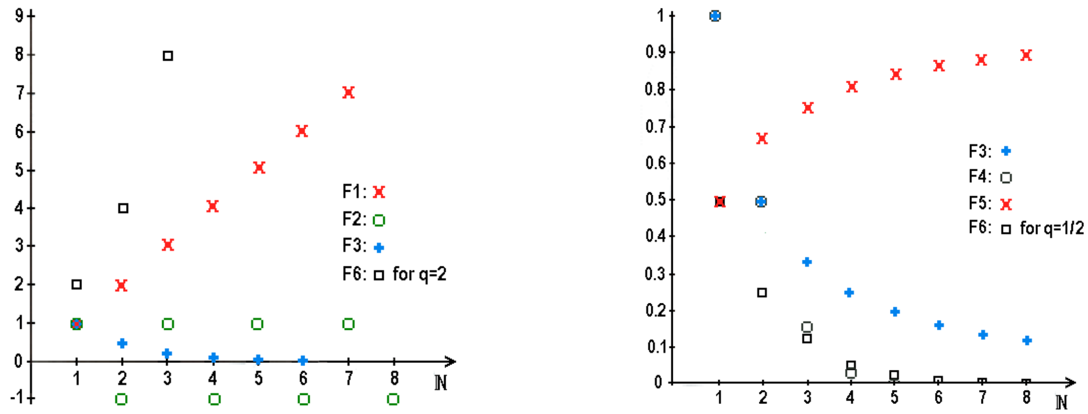


Figure 3.1: Visualization of our sample sequences over the natural numbers, in case of the geometrical sequence (F6) for $q = 2$ and $q = \frac{1}{2}$.

Also the sum, the difference or the product of two sequences are again a sequence. For example, the sample sequence (F5) with $a_n = \frac{n}{n+1} = \frac{n+1-1}{n+1} = 1 - \frac{1}{n+1}$ is the difference of the trivial sequence $(1)_{n\in\mathbb{N}} = 1, 1, 1, \ldots$, consisting purely of ones, and the harmonic sequence (F3) except for the first member.

The termwise product of the sample sequences (F2) and (F3) makes up a new sequence:

(F7) $1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \ldots = \left(\frac{(-1)^{n+1}}{n}\right)_{n\in\mathbb{N}}$ the "alternating" harmonic sequence.

Similarly the termwise product of the harmonic sequence (F3) with itself is once again a sequence:

(F8) $1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, ... = \left(\frac{1}{n^2}\right)_{n\in\mathbb{N}}$ the sequence of the inverse natural squares.

The termwise product of the sample sequences (F1) and (F6), too, gives a new sequence:
(F9) $q, 2q^2, 3q^3, 4q^4, 5q^5 \ldots = (nq^n)_{n\in\mathbb{N}}, \quad q \in \mathbb{R}$ a modified geometric sequence.

An other more complicated combined sequence will attract our attention later:

(F10) $2, (\frac{3}{2})^2, (\frac{4}{3})^3, \ldots = \left((1 + \frac{1}{n})^n\right)_{n\in\mathbb{N}}$ the so-called exponential sequence.

Exercise 3.1 *Illustrate these additional sample sequences graphically. Project the points on the 2-axis.*

There are **<u>three characteristics</u>** that are of special interest to us as far as sequences are concerned: boundedness, monotony and convergence:

3.2 Boundedness

A sequence is called **bounded above**, if there is an upper bound B for the members of the sequence: $a_n \leq B$: in shorthand notation this means:

$$(a_n)_{n\in\mathbb{N}} \ \textbf{bounded above} \iff \exists B : a_n \leq B \quad \forall n \in \mathbb{N}$$

Bounded below is defined in full analogy with a lower *lower* bound A:
$\exists A : \ A \leq a_n \ \forall n \in \mathbb{N}$.

For example, our first sample sequence (F1) consisting of the natural numbers is bounded only from below e.g. by 1: $A = 1$. The alternating sequence (F2) is obviously bounded from above and from below, e.g. by $A = -1$ and $B = 1$, respectively. For the harmonic sequence (F3) the first member, the 1, is an upper bound: $B = 1 \geq \frac{1}{n} \ \forall n \in \mathbb{N}$ and the zero a lower one: $A = 0$. The sample sequence (F4) of the inverse factorials has the lower bound $A = 0$ and the upper one $B = 1$.

Exercise 3.2 *Investigate the boundedness of the other two of our sample sequences.*

3.3 Monotony

A sequence is said to be **monotonically increasing**, if the successive members increase with increasing number: To memorize:

$(a_n)_{n\in\mathbb{N}}$ **monotonically increasing** $\iff a_n \leq a_{n+1}\ \forall n \in \mathbb{N}.$

If the stronger condition $a_n \lneqq a_{n+1}$ holds true, one calls the sequence *strictly* monotonic increasing.

In full analogy, *monotonically decreasing* is defined with $a_n \geq a_{n+1}$.

For example, the sequence (F1) of the natural numbers is strictly monotonic increasing, the alternating harmonic sequence (F2) is not monotonic at all and the harmonic sequence (F3) as well as the sequence (F4) of the inverse factorials are strictly monotonic decreasing.

Exercise 3.3 Monotonic sequences

Investigate the monotony of the other two of our sample sequences.

3.4 Convergence

Now we come to the central topic of the whole chapter: As you may have seen from the projection of the visualizing points onto the 2-axis there are sequences, whose members a_n accumulate around a number a on the number line, so that infinitely many members of the sequence lie in every ε-neighbourhood $U_\varepsilon(a)$ of this number a, which by the way needs not necessarily to be itself a member of the sequence. We call a in such a case a *cluster point* of the sequence.

In our examples we immediately realize that the sequence (F1) of the natural numbers has *none* and the harmonic sequence (F3) has *one* cluster point, namely the zero. The alternating sequence (F2) has even *two* cluster points: one at $+1$ and one at -1.

The **Theorem of Bolzano and Weierstrass** guarantees, that every sequence which is bounded above and below has to have **at least one cluster point**.

In the case that a sequence has only one single cluster point, it may occur that all sequence members from a certain number on, lie in the neighbourhood of that point. We then call this point the **limit** of the sequence and this situation turns out to be the central concept of analysis: Therefore mathematicians have several terms for it: They also say that the sequence converges or is convergent to a and write: $\lim\limits_{n\to\infty} a_n = a$, or sometimes more casually: $a_n \stackrel{n\to\infty}{\longrightarrow} a$.

$(a_n)_{n\in\mathbb{N}}$ **convergent:** $\exists a : \lim\limits_{n\to\infty} a_n = a$ $\Longleftrightarrow \forall\varepsilon > 0\ \exists N(\varepsilon) \in \mathbb{N} : \lvert a_n - a\rvert < \varepsilon\ \forall n > N(\varepsilon).$

The last shorthand reads: for every pre-set positive number ε which may be as tiny as you like, you can find a number $N(\varepsilon)$ so that the distance from the cluster point a for all sequence members with a number larger than $N(\varepsilon)$ is smaller than the pre-given small ε.

For many sequences we can recognize the convergence or even the limit value with some skill just by looking at it. But sometimes it is by no means easy to determine whether a sequence is convergent. This is why the Theorem of Bolzano and Weierstrass is so much appreciated: It shows us very generally when we can conclude the convergence of a sequence:

Theorem of Bolzano and Weierstrass: **Every monotonically increasing sequence which is bounded above is convergent, and**   **every monotonically decreasing sequence which is bounded below is convergent, respectively.**

In all cases where the limiting value is unknown or not easily identifiable mathematicians often make also use of the necessary ($\Leftarrow$) and sufficient ($\Rightarrow$)

Cauchy-Criterion: (a_n) convergent $\Longleftrightarrow \forall\varepsilon > 0\ \exists N(\varepsilon) \in \mathbb{N} : \lvert a_n - a_m\rvert < \varepsilon\ \forall n, m > N(\varepsilon)$

meaning, a sequence converges if and only if from a certain point onward the distances between the members of the sequence decrease more and more, i.e. the corresponding points on the number axis move closer and closer together. If that is not the case the sequence diverges. In addition, it can be shown that every subsequence of a convergent sequence and the sum and difference as well as the product and (provided the denominator is different from zero) also the quotient of two convergent sequences are convergent as well. This means that the limit is commutable with the rational arithmetic operations.

Many convergent sequences tend to zero as their cluster point, we call them *zero sequences.*

The harmonic sequence (F3) with $a_n = \frac{1}{n}$ is for example such a zero sequence.

> **Insert: Convergence proofs:** *For the **sequence F3**:*$(\frac{1}{n})_{n\in\mathbb{N}}$ *we want to test all convergence criteria:*
>
> *1. Most easily we check the Theorem of **Bolzano and Weierstrass**: the sequence (F3)*$(\frac{1}{n})_{n\in\mathbb{N}}$ *is monotonically decreasing and bounded below:* $0 < \frac{1}{n}$, *consequently it converges.*
>
> *2. The **cluster point** is apparently* $a = 0$: *We pre-set an* $\varepsilon > 0$ *arbitrarily, e.g.* $\varepsilon = \frac{1}{1000}$ *and look for a number* $N(\varepsilon)$, *so that* $|a_n - a| = |\frac{1}{n} - 0| = |\frac{1}{n}| = \frac{1}{n} < \varepsilon$ *for* $n > N(\varepsilon)$. *That is surely the case if we choose* $N(\varepsilon)$ *as the next natural number larger than* $\frac{1}{\varepsilon}$: $N(\varepsilon) > \frac{1}{\varepsilon}$ *(e.g. for* $\varepsilon = 0.001$ *we take* $N(\varepsilon) = 1001$*). Then there holds for all* $n > N(\varepsilon) : \frac{1}{n} < \frac{1}{N(\varepsilon)} < \varepsilon$.
>
> *3. Finally also the **Cauchy Criterion** can easily be checked here: If a certain* $\varepsilon > 0$ *is pre-given, it follows for the distance of two members* a_n *and* a_m *with* $n < m$:
> $|a_n - a_m| = |\frac{1}{n} - \frac{1}{m}| = |\frac{m-n}{nm}| < |\frac{m}{nm}| = \frac{1}{n} < \varepsilon$, *if* $n > N(\varepsilon) = \frac{1}{\varepsilon}$.

The sequences (F1) and (F2) obviously do not converge.

Exercise 3.4 Convergent sequences

a) Test the other three sample sequences for convergence.

b) Calculate - in order to become cautious - the first ten members of the sequence $a_n = n \cdot 0.9^n$, *the product of (F1) with (F6) for* $q = 0.9$, *and compare with* a_{60}, *as well as of* $a_n = \frac{n!}{10^n}$, *the quotient of (F6) for* $q = \frac{1}{10}$ *and (F4), and compare with the corresponding* a_{60}.

c) The sequence consisting alternately of the members of (F1) and (F3): i.e. 1,$\frac{1}{2}$*,3,*$\frac{1}{4}$*,5,*$\frac{1}{6}$*,... has only one single cluster point, namely 0. Does it converge to 0?*

3.5 Series

After having studied the limits of number sequences, we can apply our newly acquired knowledge to topics which occur more often in physics, for instance *infinite sums* $s = \sum\limits_{n=1}^{\infty} a_n$, called **series**:

These are often encountered sometimes in more interesting physical questions: For instance if we want to sum up the electrostatic energy of infinitely many equidistant alternating positive and negative point charges for one chain link (which gives a simple but surprisingly good one-dimensional model of a ion crystal) we come across the infinite sum over the members of the alternating harmonic sequence (F7): the series $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$. How do we calculate this?

Series are sequences whose members are finite sums of real numbers: The definition of a

$$\boxed{\textbf{series } \sum_{n=1}^{\infty} \boldsymbol{a_n} \text{ as } \textit{sequence of partial sums } s_m = \left(\sum_{n=1}^{m} a_n\right)_{m\in\mathbb{N}}}$$

reduces the series to sequences which we have been dealing with just above.

Especially, a series is exactly then **convergent** and has the value s, if the sequence of its partial sums s_m (*not* that of its *summands* a_n!!) converges: $\lim_{m\to\infty} s_m = s$:

$$\boxed{\text{series } s_m = \sum_{n=1}^{m} a_n \textbf{ convergent} \iff \lim_{m\to\infty} \sum_{n=1}^{m} a_n = s < \infty}$$

Also the multiple of a convergent series and the sum and difference of two convergent series are again convergent.

The few **sample** series that we need, to see the most important concepts, we derive simply through piecewise summing up our sample sequences:

(R1) The series of the partial sums of the sequence (F1) of the natural numbers: $s_m = \left(\sum_{n=1}^{m} n\right)_{m\in\mathbb{N}} = 1, 3, 6, 10, 15, \ldots$ is clearly divergent.

(R2) The series made out of the members of the alternating sequence (F2) always jumps between 1 and 0 and has therefore two cluster points and consequently no limit.

(R3) Also the "harmonic series" summed up out of the members of the harmonic sequence (F3), i.e. the sequence $\left(s_m = \sum_{n=1}^{m} \frac{1}{n}\right)_{m\in\mathbb{N}} = 1, \frac{3}{2}, \frac{11}{6}, \frac{25}{12}, \frac{137}{60}, \ldots$ is divergent. Because the (also necessary) Cauchy Criterion is not fulfilled: If we for instance choose $\varepsilon = \frac{1}{4} > 0$ and consider a piece of the sequence for $n = 2m$ consisting of m terms: $|s_{2m} - s_m| = \sum_{n=m+1}^{2m} \frac{1}{n} =$

$\frac{1}{m+1} + \frac{1}{m+2} + \ldots + \frac{1}{2m} > \underbrace{\frac{1}{2m} + \frac{1}{2m} + \ldots + \frac{1}{2m}}_{m \quad summands} = \frac{1}{2} > \varepsilon = \frac{1}{4}$ while for convergence $< \boldsymbol{\varepsilon}$ would have been necessary.

(R7) Their alternating variant however, created out of the sequence (F7), our physical example from above, converges $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$ $(= \ln 2$, as we will show later).

Because of this difference between series with purely positive summands and alternating ones, it is appropriate to introduce a new term: A series is said to be *absolutely convergent*, if already the series of the absolute values converges.

Series $s_m = \sum_{n=1}^{m} a_n$ **absolutely convergent** $\iff \lim_{m\to\infty} \sum_{n=1}^{m} \|a_n\| < \infty$

We can easily understand that within an absolutely convergent series the summands can be rearranged without any effect on the limiting value. Two absolutely convergent series can be multiplied termwise to create a new absolutely convergent series.

For absolute convergence the mathematicians have developed various *sufficient* criteria, the so-called *majorant criteria* which you will deal with more closely in the lecture about analysis:

Insert: Majorants: *If a convergent majorant sequence* $S = \lim_{m\to\infty} S_m = \sum_{n=1}^{\infty} M_n$ *exists with positive* $M_n > 0$*, whose members are larger than the corresponding absolute values of the sequence under examination* $M_n \geq |a_n|$*, then the series* $\lim_{m\to\infty} s_m = \sum_{n=1}^{\infty} a_n$ *is absolutely convergent, because from the Triangle Inequality it follows*

$$|s_m| = |\sum_{n=1}^{m} a_n| \leq \sum_{n=1}^{m} |a_n| \leq \sum_{n=1}^{m} M_n = S_m.$$

Very often the "geometric series"

(R6): $\sum_{n=0}^{\infty} q^n$, which follow from the geometric sequences (F6) $(q^n)_{n\in\mathbb{N}},\ q \in \mathbb{R}$, serve as majorants. To calculate them we benefit from the earlier for $q \neq 1$ derived geometric sum:

$$\lim_{m\to\infty} \sum_{n=0}^{m} q^n = \lim_{m\to\infty} \frac{1-q^{m+1}}{1-q} = \frac{1}{1-q} < \infty,$$

meaning convergent for $|q| < 1$ and divergent for $|q| \geq 1$.

Insert: Quotient criterion: *We present here as example for a majorant criterion only the* **quotient criterion** *which is obtained through comparison with the geometric series:*

$$\text{If } \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1, \text{ is } s_m = \sum_{n=1}^{m} a_n \text{ absolutely convergent.}$$

As an example we prove the absolute convergence of the series **(R9)** $\sum_{n=0}^{\infty} nq^n$ *for* $|q| < 1$, *which can be obtained from the for* $|q| < 1$ *convergent geometric series (R6) through termwise multiplication with the divergent sequence (F1) of the natural numbers. We calculate therefore*

$$\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| = \lim_{n\to\infty} \left|\frac{(n+1)q^{n+1}}{nq^n}\right| = |q| \lim_{n\to\infty} \frac{n+1}{n} = |q| < 1.$$

That the criterion is not necessary can be seen from the series **(R8)***, the summing up of the sample sequence (F8):*

$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$, *which is absolutely convergent, since all members are positive, but*

$\lim_{n\to\infty} \frac{n^2}{(n+1)^2} = \lim_{n\to\infty} \frac{1}{(1+n^{-1})^2} = 1.$

(R4) The series of the inverse natural factorials $\sum_{n=1}^{\infty} \frac{1}{n!}$ deserves to be examined in more detail:

First we realize that the sequence of the partial sums $\left(s_m = \sum_{n=1}^{m} \frac{1}{n!}\right)_{m\in\mathbb{N}}$ increases monotonically: $s_{m+1} - s_m = \frac{+1}{(m+1)!} > 0$. To get an upper bound B we estimate through the majorant geometric sum with $q = \frac{1}{2}$:

$$\begin{aligned}
|s_m| &= 1 + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{m!} \\
&< 1 + \frac{1}{2} + \frac{1}{2^2} + \ldots + \frac{1}{2^{m-1}} \\
&= \sum_{n=0}^{m-1} (\frac{1}{2})^n \\
&= \frac{1 - (\frac{1}{2})^m}{1 - \frac{1}{2}} \\
&< \frac{1}{1 - \frac{1}{2}} = 2.
\end{aligned}$$

Since the monotonically increasing sequence of the partial sums s_m is bounded from above by $B = 2$ the Theorem of Bolzano and Weierstrass guarantees us convergence. We just do not know the limiting value yet. This limit is indeed something fully new - namely an irrational number. We call it $= e - 1$, so that the number e after the supplementary convention $0! = 1$ is defined by the following series starting with $n = 0$:

Exponential series defined by: $\boldsymbol{e := \sum_{n=o}^{\infty} \frac{1}{n!}}$.

Insert: The number *e* is irrational: *we prove indirectly that the so defined number e is irrational, meaning it cannot be presented as quotient of two integers g and h:*

If e were writable in the form $e = \frac{g}{h}$ *with integers g and* $h \geq 2$*, then* $h!e = (h-1)!g$ *would be an integer:*
However, from definition it holds

$$\begin{aligned}(h-1)!g &= h!e = h!\sum_{n=0}^{\infty}\frac{1}{n!} = \sum_{n=0}^{h}\frac{h!}{n!} + \sum_{n=h+1}^{\infty}\frac{h!}{n!}\\ &= \left(h! + h! + \frac{h!}{2!} + \frac{h!}{3!} + \ldots + 1\right) +\\ &\quad + \lim_{n\to\infty}\left(\frac{1}{h+1} + \frac{1}{(h+1)(h+2)} + \ldots + \frac{1}{(h+1)(h+2)\ldots(h+n)}\right).\end{aligned}$$

While the first bracket is an integer if h is, this cannot be true for the second bracket, because

$$\begin{aligned}\frac{1}{h+1} &+ \frac{1}{(h+1)(h+2)} + \ldots + \frac{1}{(h+1)(h+2)\ldots(h+n)} = \ldots\\ &= \frac{1}{h+1}\left(1 + \frac{1}{h+2} + \ldots + \frac{1}{(h+2)\ldots(h+n)}\right),\end{aligned}$$

which can be estimated through the geometric series with $q = \frac{1}{2}$ *as follows,*

$$\begin{aligned}&< \frac{1}{h+1}\left(1 + \frac{1}{2} + \ldots + \frac{1}{2^{n-1}}\right) = \frac{1}{h+1}\cdot\frac{1-(\frac{1}{2})^n}{1-(\frac{1}{2})}\\ &< \frac{1}{h+1}\cdot\frac{1}{1-(\frac{1}{2})} = \frac{2}{h+1} \leq 2/3,\end{aligned}$$

Because h should be $h \geq 2$ *there is a contradiction. Consequently e must be irrational.*

To get the **numerical value** of $\boldsymbol{e}$ we first calculate the members of the zero sequence (F4) $a_n = \frac{1}{n!}$:

$a_1 = \frac{1}{1!} = 1, \quad a_2 = \frac{1}{2!} = \frac{1}{2} = 0.50, \qquad a_3 = \frac{1}{3!} = \frac{1}{6} = 0.1666,$
$a_4 = \frac{1}{4!} = \frac{1}{24} = 0.041\ 666, \qquad a_5 = \frac{1}{5!} = \frac{1}{120} = 0.008\ 33,$
$a_6 = \frac{1}{6!} = \frac{1}{720} = 0.001\ 388, \qquad a_7 = \frac{1}{7!} = \frac{1}{5\ 040} = 0.000\ 198,$
$a_8 = \frac{1}{8!} = \frac{1}{40\ 320} = 0.000\ 024, \qquad a_9 = \frac{1}{9!} = \frac{1}{362\ 880} = 0.000\ 002, \ldots$

then we sum up the partial sums: $s_m = \sum\limits_{n=1}^{m} \frac{1}{n!} = 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots + \frac{1}{m!}$

$s_1 = 1, \quad s_2 = 1.50, \quad s_3 = 1.666\ 666, \quad s_4 = 1.708\ 333,$
$s_5 = 1.716\ 666, \quad s_6 = 1.718\ 055, \quad s_7 = 1.718\ 253,$
$s_8 = 1.718\ 278, \quad s_9 = 1.718\ 281, \ldots.$

If we look at the rapid convergence, we can easily imagine that after a short calculation we receive the following result for the limiting value: $e = 2.718\ 281\ 828\ 459\ 045\ldots$

Insert: A sequence converging to $\boldsymbol{e}$: *Besides this exponential series which we used to define e there exists as earlier mentioned in addition a* sequence, *converging to the number e, the* exponential sequence*(F10):*
$\left((1 + \frac{1}{n})^n\right)_{n\in\mathbb{N}} = 2,\ (\frac{3}{2})^2,\ (\frac{4}{3})^3, \ldots,$ *which we will shortly deal with for comparison: According to the binomial formula we find firstly for the general sequence member:*

$$\begin{aligned} a_n &= (1 + \frac{1}{n})^n = \sum_{k=0}^{n} \frac{n!}{(n-k)!k!n^k} \\ &= 1 + \frac{n}{n} + \frac{n(n-1)}{n^2 2!} + \frac{n(n-1)(n-2)}{n^3 3!} + \ldots + \frac{n(n-1)(n-2)\ldots(n-(k-1))}{n^k k!} + \\ &\quad \ldots + \frac{n!}{n^n n!} \\ &= 1 + 1 + \frac{(1-\frac{1}{n})}{2!} + \frac{(1-\frac{1}{n})(1-\frac{2}{n})}{3!} + \ldots + \frac{(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{k-1}{n})}{k!} + \ldots \\ &\quad + \frac{(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{n-1}{n})}{n!} \end{aligned}$$

On the one hand we enlarge *this expression for a_n, by forgetting the subtraction of the multiples of $\frac{1}{n}$ within the brackets:*

$$a_n \le 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{n!} = 1 + s_n$$

and reach so (besides the term one) the corresponding partial sums of the exponential series s_n. Thus the exponential series is a majorant for the also monotonically increasing exponential sequence and ensures the convergence of the sequence through that of the series. For the limiting value we get:

$$\boxed{\lim_{n\to\infty} a_n \le e.}$$

On the other hand we diminish *the above expression for a_n by keeping only the first $(k+1)$ of the without exception positive summands and throwing away the other ones:*

$$a_n \ge 1 + 1 + \frac{(1-\frac{1}{n})}{2!} + \frac{(1-\frac{1}{n})(1-\frac{2}{n})}{3!} + \ldots + \frac{(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{(k-1)}{n})}{k!}.$$

When we now first let the larger n, of the two natural numbers tend to infinity, we get:

$$a := \lim_{n\to\infty} a_n \ge 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{k!} = 1 + s_k$$

and after letting also the smaller natural number k tend to infinity we reach:

$$\boxed{a \ge e.}$$

Consequently the limit $a := \lim_{n\to\infty} a_n$ of the exponential sequence a_n must be **equal** *to the number e defined by the exponential series:*

$$\boxed{\lim_{n\to\infty}(1+\tfrac{1}{n})^n = \sum_{n=0}^{\infty} \tfrac{1}{n!} = e}$$

When you, however, calculate the members of the sequence and compare them with the partial sums of the series, you will realize that the sequence converges much more slowly than the series.

Through these considerations we now have got a first overview over the limiting procedures and some of the sequences and series important for natural sciences with their limits, which will be of great use for us in the future.

Chapter 4

FUNCTIONS

4.1 The Function as Input-Output Relation or Mapping

We would like to remind you of the *empirical method* of physics discussed in Chapter 1, and take a look at the simplest, but common case: in an experiment we investigate the mutual dependency of two physical quantities: "y as a function of x" or $y = f(x)$: In our experiment one quantity x, called the *independent variable*, is measurably changed and the second quantity y, the *dependant variable,* is measured in each case. We may imagine the measuring apparatus in the way depicted below as a black box, into which the x are fed in as input, and from which the corresponding y come out as output.

Figure 4.1: Function as a black box with x as input and y as output

Physicists think for example of an electric circuit where the voltage is changed gradually by a potentiometer and the electric current is measured with a mirror galvanometer in order to investigate the characteristic curve. Also the time development of the amplitude of a pendulum or a radioactively decaying material as function of time are further candidates out of the huge number of physical examples.

The result of such a series of measurements is first of all a *value table* (x, y). The data can also be displayed in a *graphic illustration*, as shown below in our samples. Illustration of the functions as a picture, usually called by us *graph*, through plotting the measured values in a plane with a *Cartesian* (meaning right-angled) coordinate system (with the abscissa x on the 1-axis and the ordinate y on the 2-axis) is a matter of course for physicists.

In the following figures you will find examples for value tables, graphic illustrations and interpolating functions for a swinging spiral spring

$\frac{x}{cm}$	$\frac{F}{mN}$
1	−0.42
1.5	−0.55
2	−0.82
2.5	−1.03
3	−1.25
3.5	−1.45
4	−1.65
4.5	−1.80
5	−1.95
5.5	−2.20
6	−2.35
6.6	−2.60

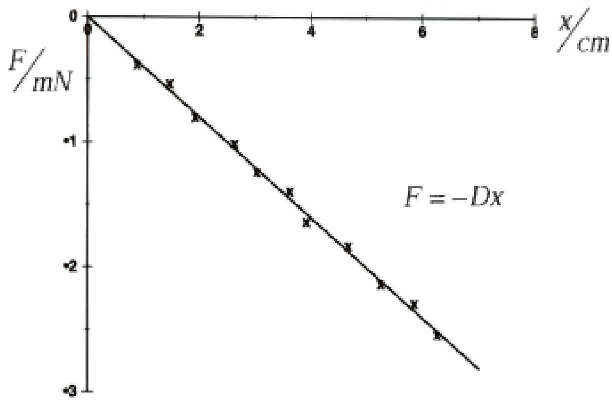


Figure 4.2 a: Reaction force F of the spring measured in mN in dependency on the amplitude x in cm.

$\frac{x}{cm}$	$\frac{E}{mJ}$
1	0.6
1.5	1.0
2.5	2.8
2.9	3.9
3.1	4.8
3.5	6.1

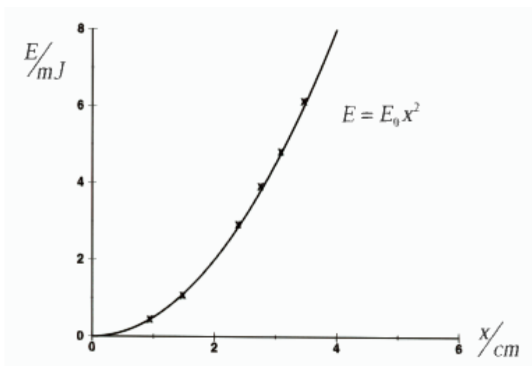


Figure 4.2 b: Potential energy E stored in the spring measured in mJ in dependency on the amplitude x in cm.

$\frac{t}{cs}$	$\frac{x}{cm}$
0.3	3.5
0.5	2.8
0.7	1.2
1.1	−1.8
1.7	−3.2
2.4	−0.8
2.6	1.5
3.2	2.4
3.6	1.4
4.3	−1.1
4.8	−1.8

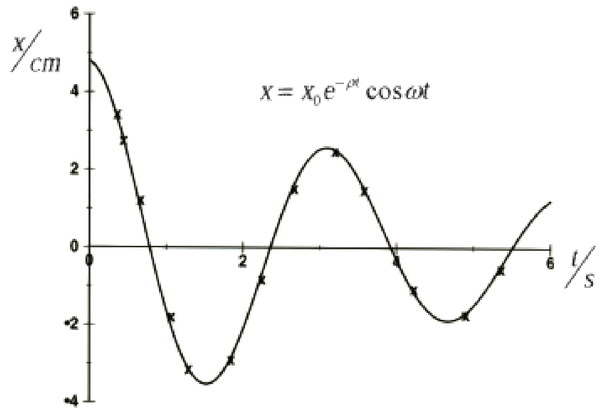


Figure 4.2 c: Deflection amplitude x of the spiral spring measured in cm in dependency on the time t in s.

$\frac{M}{g}$	$\frac{T}{s}$
2.5	0.75
10	1.63
14	1.91
20	2.23
25	2.46

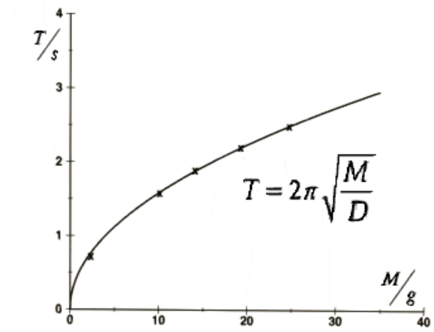


Figure 4.2 d: Oscillation time T of the spiral spring in s as a function of the mass M in g with unaltered spring constant D.

$\frac{D}{Nm^{-1}}$	$\frac{T}{s}$
3	3.25
4	2.72
5	2.16
7	1.75
8	1.71
10	1.59

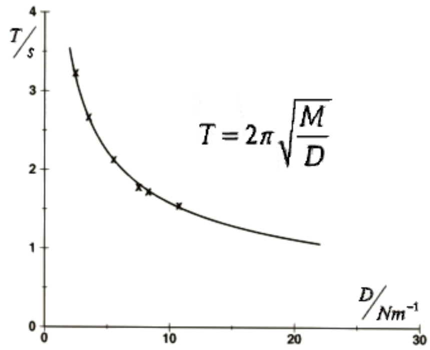


Figure 4.2 e: Oscillation time T of the spiral spring in s as function of the deflecting force D measured in Nm^{-1} with constant mass M.

After we have taken into account the inevitable measurement errors, we can start connecting the measured points by a curve or a mathematical calculation instruction, to look for a **function** which describes the dependence of the two quantities. If we succeed in finding such a function, we have achieved real progress: A mathematical formula is usually short and concise; it can be stapled, processed and conveyed to others much easier than extensive value tables. With its help we are able to interpolate more closely between the measurements and to extrapolate beyond the measured area, which suggests further experiments. Finally it is the first step towards a theory, and with it to the understanding of the experiment.

Insert: History: *T. Brahe measured in his laboratory the position of the planet Mars at different times. From that value table J. Kepler found the ellipse as an interpolating function for the orbit curve. This result influenced I. Newton in finding his gravitation law.*

Therefore, for **physical** reasons, we have to deal with **functions**, first with *real* functions of a *real* variable.

Mathematically, we can consider a function $y = f(x)$ as an *unambiguous mapping* $x \rightarrow f(x)$ of a point x of the area D_f, (the "*definition domain*" of f) of the independent variable x (also known as *abscissa* or argument) onto a point $f(x)$ of the area W_f (the "*value domain*" of f) of the dependent variable y (also known as *ordinate* or function value).

While the declaration of the definition domain in addition to the mapping prescription is absolutely necessary for a function, and often influences the properties of the function, the exact statement of the value domain $W_f := \{f(x)|x \in D_f\}$ is in most cases of less importance and sometimes takes much effort.

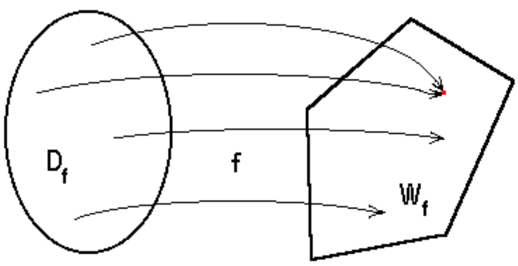


Figure 4.3: Function f as a mapping of the definition domain D_f into the value domain W_f (with two arrows, which lead from two pre-image points to *one* image point)

The pre-image set D_f is in most cases, just as is the set of images W_f, a part of the real number axis $\mathbb{R}^1$. The unambiguity included in the definition of a real function means that to each x there is one and only one $y = f(x)$. (It is however possible that two different pre-image points are mapped into one and the same image point.) To summarize in mathematical shorthand:

$$y = f(x) \textbf{ function: } \forall x \in D_f \subseteq \mathbb{R}^1 \quad \exists! y = f(x) : y \in W_f \subseteq \mathbb{R}^1$$

The **arithmetic for real functions** of a real variable follows according to the rules of the field $\mathbb{R}$ with both the Commutative and Associative Laws, as well as the connecting Distributive Law, which we have put together for the numbers in Chapter 2 : for example, the sum or the difference of two real functions $f_1(x) \pm f_2(x) = (f_1 \pm f_2)(x) =: g(x)$ gives a new real function, as well as the real multiple $r \cdot f(x) = (r \cdot f)(x) =: g(x)$ with $r \in \mathbb{R}$ and analogously also the product $f_1(x) \cdot f_2(x) = (f_1 \cdot f_2)(x) =: g(x)$ or, if $f_2(x) \neq 0$ all over the definition domain, the quotient, too. $\frac{f_1(x)}{f_2(x)} = \frac{f_1}{f_2}(x) =: g(x)$.

4.2 Basic Set of Functions

It is surprising that we can manage to go through daily physics with a basic set of very few functions which moreover you are mostly acquainted with from school. In this section we will introduce these basic set of functions as examples, then discuss some of their characteristics, and come back to them again and again.

4.2.1 Rational Functions

We start with the constant function $y = c$, independent of x. Afterwards we come to *linear* functions $y = s \cdot x + c$ with the graph of a straight line having a gradient s and the ordinate section c. We proceed to the *standard parabola* $y = x^2$ and the higher powers $y = x^n$ with $n \in \mathbb{N}$. Also the *standard hyperbola* $y = \frac{1}{x} = x^{-1}$ and $y = \frac{1}{x^2}$ which you are surely familiar with.

Straight line and parabola are for example defined over the whole real axis: $D_f = \mathbb{R}$. For the hyperbola we must omit the origin: $D_f = \mathbb{R} \setminus \{0\}$. Also in the image domain of the hyperbola the origin is missing: $W_f = \mathbb{R} \setminus \{0\}$. For the parabola the image domain is only the positive half-line including zero: $y \geq 0$. The following figure shows the graphs of these simple examples:

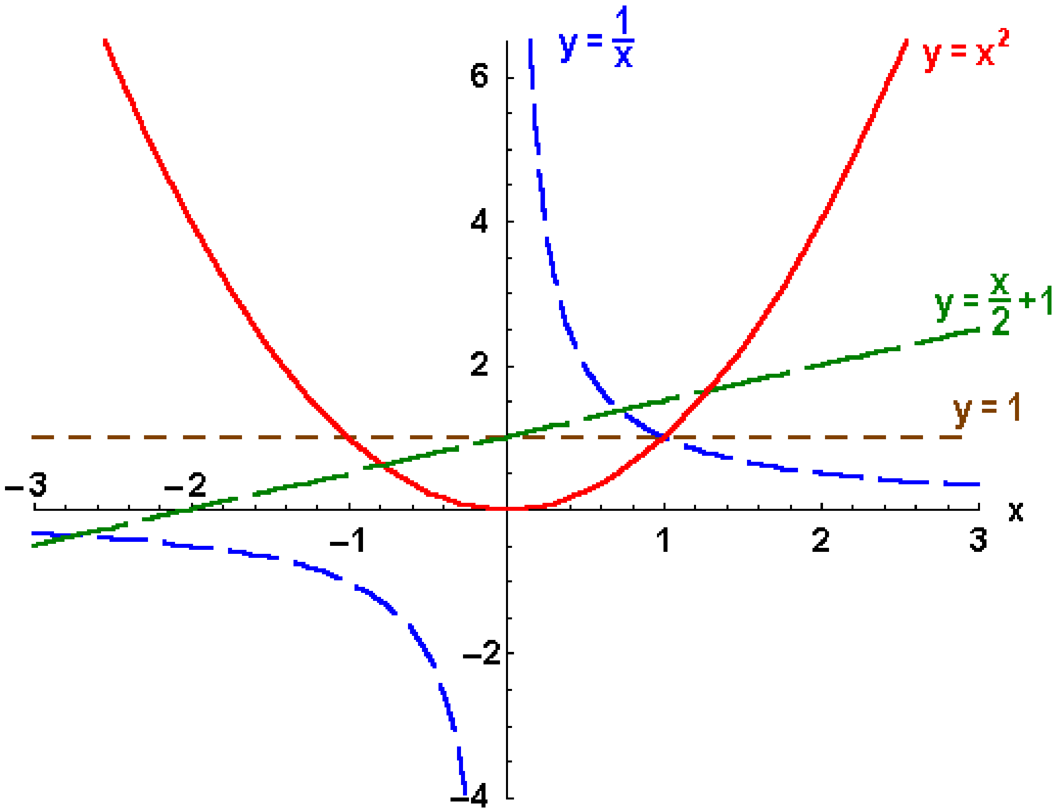


Figure 4.4: Graphs of simple functions

According to the calculation rules of the field of real numbers $\mathbb{R}$ we get from the straight line and the standard parabola $y = x^2$ all functions of second degree $y = ax^2 + bx + c$ as well as all further **polynomial functions** of higher, e.g. $m-$th degree:

$$y = P_m(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_m x^m = \sum_{k=0}^{m} a_k x^k.$$

Even the general rational function

$$y(x) = R(x) = \frac{P_m(x)}{Q_n(x)}$$

with a polynomial of $m-$th degree $P_m(x)$ in the numerator and a polynomial of the $n-$th degree $Q_n(x)$ in the denominator you are surely familiar with, for example $y = \frac{1}{x^2+1}$, the Lorentz distribution, which among other things describes the natural line width of a spectral line with $D = \mathbb{R}$ and $0 < y \leq 1$ or $y = \frac{x^2+1}{x-1}$. These rational functions are defined for all x except for those values x_m, where the denominator vanishes: $Q_n(x_m) = 0$.

Exercise 4.1 Graphs, definition domains and image domains

State the graphs and maximal definition domains of following functions and if possible also the image domains:

a) $f(x) = -2x - 2$; *b)* $f(x) = 2 - 2x^2$; *c)* $f(x) = x^2 - 2x - 3$; *d)* $f(x) = \frac{1}{3}x^3 - 3$;
e) $f(x) = x^4 - 4$; *f)* $f(x) = \frac{1}{1-x}$; *g)* $f(x) = \frac{2x-3}{x-1}$; *h)* $f(x) = \frac{1}{x^2-1}$;
i) $f(x) = \frac{1}{(x-1)^2}$; *j)* $f(x) = \frac{x+2}{x^2-4}$; *k)* $f(x) = \frac{x^2+5}{x-2}$.

4.2.2 Trigonometric Functions

A further group of fundamental functions for all natural sciences which you already know from school are the **trigonometric** functions. They play a central role in all periodic processes, whether it is in space or in time, for example during the oscillation of a pendulum, for the description of light or sound waves, and even for the vibration of a string. In the following figure a unit circle is pivoted rotatably around the centre carrying a virtual ink cartridge on its circumference at the end of the red radius. Please click with your mouse on the circular disc, pull the underlying sheet of paper out to the right under the uniformly rotating disc and look at the curve which the cartridge has drawn on the paper.

ONLINE ONLY	Figure 4.5 shows a virtually rotatable circle disc carrying an ink cartridge on its circumference, under which per mouse click a picture of the graph of $y = \sin x$ can be extracted.

With the help of the projection of the revolving pointer the cartridge has drawn for us onto the 2-axis the "length of the opposite leg" in the right-angled triangle built by the circulating radius of length one as hypotenuse, i.e. the graph of the function $y = \sin x$, the "sine" as function of the angle x.
Clearly, this construction rule gives a **periodic** function, meaning that in intervals of 2π of the independent variable the dependent variable takes on the same values: $\sin(x+2\pi) = \sin x$, generally:

$$y = f(x) \textbf{ periodic with } 2\pi: \quad f(x + 2\pi) = f(x)$$

Out of the sine function by simple operations we can build other trigonometric functions, which have received their own names due to their importance:
We get the "**cosine**-function" $y = \cos x$ analogously just like the sine function as the

"length of the **adjacent leg**" of the angle x in the right-angled triangle composed by the rotating radius and the sine, or as the projection of the circulating radius, that is now on the 1-axis. The fundamental connection:

$$\boxed{\cos^2 x + \sin^2 x = 1}$$

follows with the Pythagoras Theorem directly from the triangle marked in the figure. The ink cartridge would have obviously drawn the cosine immediately, if we had started with the angle $\frac{\pi}{2}$ instead of 0:

$$\boxed{\cos x = \sin(x + \frac{\pi}{2}).}$$

So the cosine function is in fact a sine function shifted to the left by the "Phase" $\frac{\pi}{2}$.

Also the cosine is periodic with the period 2π : $\cos(x + 2\pi) = \cos x$.

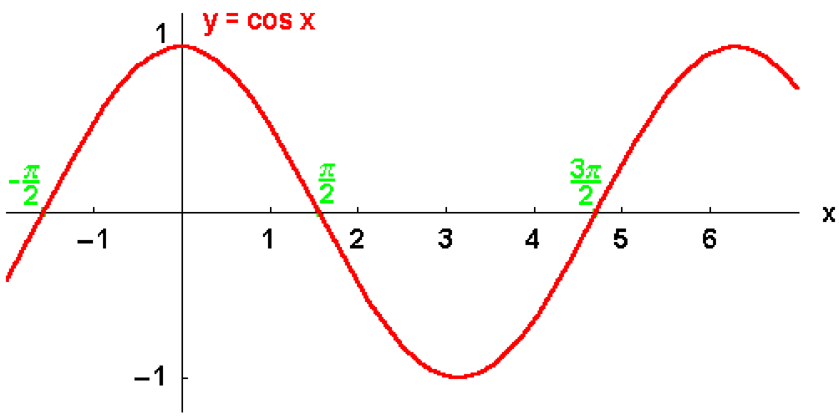


Figure 4.6: Graph of the cosine

From sine and cosine, through division we get two further important trigonometric functions: the

$$\boxed{\textbf{tangent: } y = \tan x = \frac{\sin x}{\cos x}}$$

and the

$$\textbf{cotangent: } y = \cot x = \frac{\cos x}{\sin x} = \frac{1}{\tan x}.$$

Insert: Notations: *In the German literature you may often find also* tg x *instead of* $\tan x$ *and* ctg x *instead of* $\cot x$.

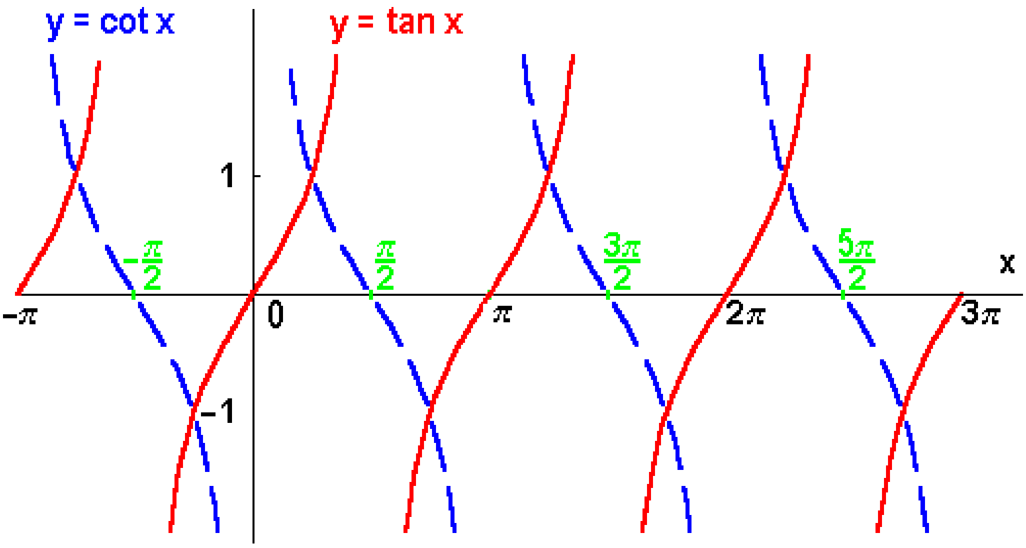


Figure 4.7: Tangent and cotangent

Tangent and cotangent are periodic with the period $\pi : \tan(x + \pi) = \tan x$.

In Chapter 6 we will learn how to calculate the functional value of even the trigonometric functions, e.g. of $y = \sin x$ for every value of the variable x through elementary calculations such as addition and multiplication.

Besides the Pythagoras Relation $\cos^2 x + sin^2 x = 1$ the

trigonometric addition theorems:

$$\begin{aligned} \cos(a \pm b) &= \cos a \cos b \mp \sin a \sin b \\ \sin(a \pm b) &= \sin a \cos b \pm \cos a \sin b \end{aligned}$$

are of major importance, and experience shows that we have to remind you of them and to recommend that they be learned by heart. In Chapter 8 we will learn to derive them in a much more elegant way than you did in school.

Exercise 4.2 Trigonometric Functions:

Sketch the graphs and the definition domains of the following functions, and also the value domains except for the last example:
a) $y = 1 + \sin x$, *b)* $y = \sin x + \cos x$, *c)* $y = \sin x - \cos x$, *d)* $y = x + \sin x$,
e) $y = x \sin x$, *f)* $y = \frac{1}{\sin x}$, *g)* $y = \frac{1}{\tan x}$ *und* *h)* $y = \frac{\sin x}{x}$.

4.2.3 Exponential Functions

While *raising to powers* b^n, we have until now introduced only *natural numbers* $n \in \mathbb{N}$ as exponents, which indicate how often a real base b occurs as a factor:

$$b^n := b \cdot b \cdot b \cdot \ldots \cdot b \text{ with } n \text{ factors } b$$

and we have got the calculation rules:

$$b^n b^m = b^{n+m} \text{ and } (b^n)^m = b^{n \cdot m} \text{ for } n, m \in \mathbb{N}.$$

We then have added negative exponents by the definition $b^{-n} := \frac{1}{b^n}$ and through the convention $b^0 := 1$ extended the set of exponents to *integers* $n \in \mathbb{Z}$.

In order to get to the **exponential functions** we have to allow *real* numbers x as exponents instead of taking only integers n (like with the bases b): $y = b^x$ with $x, b \in \mathbb{R}$ and to restrict ourselves to positive bases b, without changing the calculation rules for the powers, i.e. with the following

multiplication theorems for exponential functions:

$$b^{x+y} = b^x b^y, \quad (b^x)^y = b^{x \cdot y} \quad \text{with } x, y, b \in \mathbb{R}, b > 0$$

Of central importance for all natural sciences is the **natural** exponential function with the irrational number e defined in Section 3.5 as base:

$$y = e^x =: \exp x,$$

Its graph with its characteristically fast growth can be directly measured in the following figure:

ONLINE ONLY	Figure 4.8 illustrates the building of the exponential function e.g. during the increase of the number of biological cells with a fixed division rate.

For physicists the inverse function $y = \frac{1}{e^x} = e^{-x}$ is also of great importance, especially for all damping and decay processes. This function, too, is accessible to measurements, e.g. during a radioactive decay, in which the still available amount of matter determines the decay: $N(t) = N(0)e^{-\frac{t}{T}}$, where $N(t)$ is the number of nuclei at a time t and T the decay time:

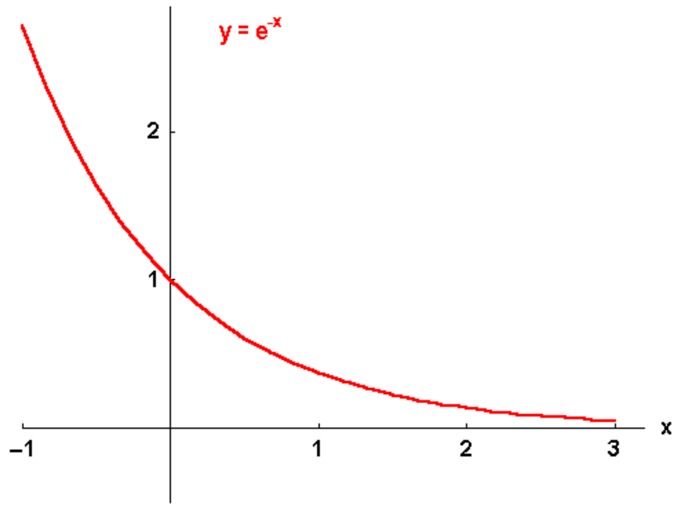


Figure 4.9: Inverse exponential function, e.g. during a radioactive decay

Even for the exponential functions we will get to know a method in Chapter 6 which will enable us to calculate the functional value $y = e^x$ for every value of the variable x by elementary calculation operations like addition and multiplication with every desired accuracy.

The following combinations of both the natural exponential functions have received special names due to their importance, which we will not understand until later: The

$$\textbf{hyperbolic cosine:} \quad y = \cosh x := \frac{e^x + e^{-x}}{2}$$

also known as catenary, because in the gravitational field of the earth a chain sags between two suspension points according to this functional curve, and the

$$\textbf{hyperbolic sine:} \quad y = \sinh x := \frac{e^x - e^{-x}}{2}$$

both connected by the easily verifiable relation:

$$\cosh^2 x - \sinh^2 x = 1.$$

In addition analogously to the trigonometric functions, we get the quotient of both, the

$$\textbf{hyperbolic tangent:} \quad y = \tanh x := \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

and the

$$\textbf{hyperbolic cotangent:} \quad y = \coth x := \frac{1}{\tanh x} = \frac{e^x + e^{-x}}{e^x - e^{-x}}.$$

The following figure shows the graphs of these functions, which are summarized under the term **hyperbolic functions**.

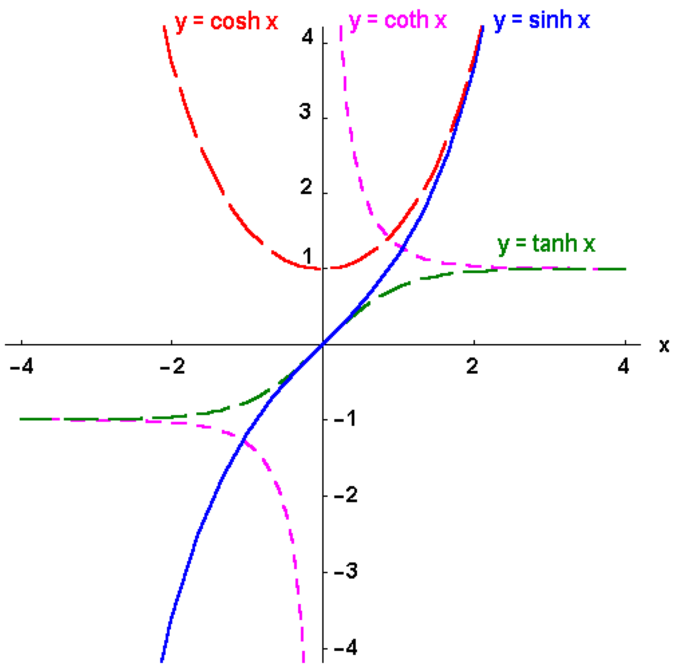


Figure 4.10: Hyperbolic functions

Insert: Notations: *The notation of the hyperbolic functions in the literature is not unique: also the following short hand notations are commonly used:* $\text{ch } x = \cosh x$, $\text{sh } x = \sinh x$ *and* $\text{th } x = \tanh x$.

Insert: Hyperbolic: *The name "hyperbolic" comes from the equation* $\cosh^2 z - \sinh^2 z = 1$*: With* $x = \cosh z$ *and* $y = \sinh z$ *in a Cartesian coordinate system this is the parameter representation* $x^2 - y^2 = 1$ *of a standard hyperbola which has the bisectors of the first and fourth quadrant as asymptotes and cuts the abscissa at* $x = \pm 1$*: Analogously with the unit circle we can draw the right branch of the hyperbola:* $\cosh x$ *is the projection of the moving point on the 1-axis and* $\sinh x$ *the projection on the 2-axis, as can be seen from the following figure.*

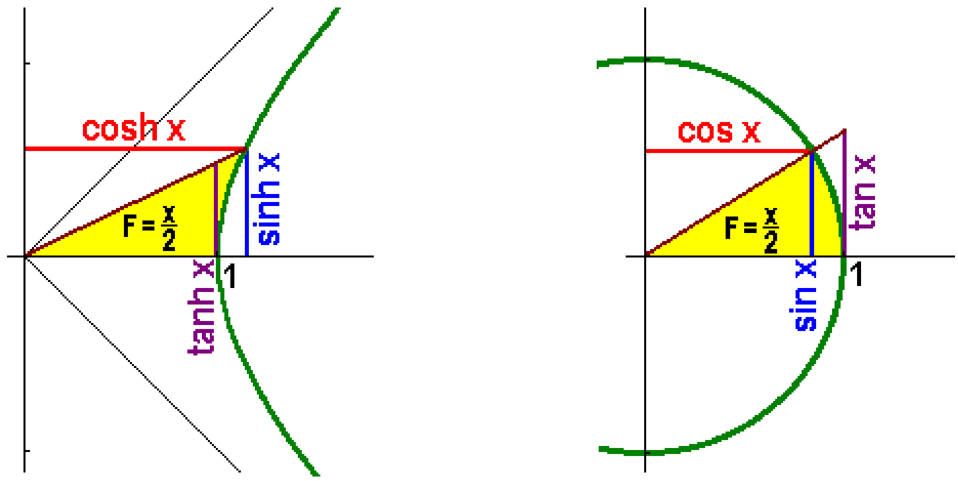


Figure 4.11: Right branch of the standard hyperbola with $\cosh x$ and $\sinh x$ to be compared with cos x and sin x in the unit circle

Exercise 4.3 Exponential functions:

Sketch the graphs for the following functions for $x \geq 0$*: a)* $y = 1 - e^{-x}$*, which describes e.g. the voltage during the charging of a capacitor*
b) $y = x + e^{-x}$,
c) the simple Poisson distribution $y = xe^{-x}$ *for totally independent statistic events,*
d) the quadratic Poisson distribution $y = x^2 e^{-x}$,
e) $y = \sin x + e^x$,
f) a damped oscillation $y = e^{-x} \sin x$,

g) the reciprocal chain line $y = \frac{1}{\cosh x}$
h) the Bose-Einstein distribution function of quantum statistics $y = \frac{1}{e^x - 1}$ *or*
i) the corresponding Fermi-Dirac distribution for particles with half-integer spin, e.g. conducting electrons $y = \frac{1}{e^x + 1}$,
j) the Planck formula for the spectral intensity distribution of the frequencies of a radiating cavity $y = \frac{x^3}{e^x - 1}$.

You may most easily check your sketches online with our function plotter or e.g. graph.tk or www.wolframalpha.com.

4.2.4 Functions with Kinks and Cracks

In addition to these sample functions, physicists use a few functions whose graphs show *kinks* (or corners) and *cracks* (or jumps). Among these, the following two are of special importance for us:

The first is the

absolute value function: $y = |x| := \begin{cases} x & \text{for} \quad x \geq 0 \\ -x & \text{for} \quad x < 0 \end{cases}$

This function is defined over the whole number axis, but as for the standard parabola the value domain covers only the non-negative half-line: $y \geq 0$. The following figure shows its graph with the "kink" at $x = 0$.

Exercise 4.4 Absolute value functions:

Sketch the graphs and the value domains of the following functions:
a) $y = 1 - |\frac{x}{a}|$, *b)* $y = x + |x|$, *c)* $y = \frac{1}{|x|}$ *and d)* $y = |x| \cos x$.

The second function is one you most likely have not encountered yet: the **Heaviside step function** $y = \theta(x)$, defined through:

Heaviside step function:

$$\begin{aligned} \theta(x) &:= 1 \quad \text{for} \quad x > 0, \\ \theta(x) &:= 0 \quad \text{for} \quad x < 0 \quad \text{and} \\ \theta(0) &:= \frac{1}{2} \quad . \end{aligned}$$

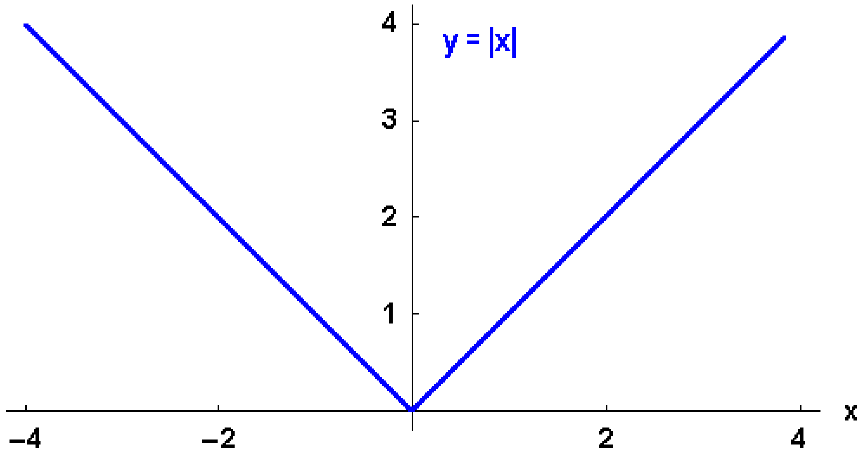


Figure 4.12: Graph of the absolute value function

The figure shows its graph with the characteristic two part step at $x = 0$.

We can easily imagine that the Heaviside function in physics is used among other things for start and stop situations and to describe steps and barriers.

> **Insert: Distributions:** *From the viewpoint of mathematics the step function is a sample of a discontinuous function. Thus it offers an access to the generalized functions, called distributions, of which the most important example in physics is the so-called Dirac δ-distribution.*

The calculation with the θ-function requires a little practice which we will gain further on: First we establish that

$$\boxed{\theta(ax) = \theta(x),}$$

if the argument is multiplied with a positive real number $a > 0$. Then we consider

$$\boxed{\theta(-x) = 1 - \theta(x).}$$

In order to get an idea of $\theta(x+a)$, we realize that the function vanishes where the argument is $x + a < 0$, thus $x < -a$, i.e. that the graph is "upstairs at $-a$". Analogously $\theta(x - a)$ means "upstairs at $+a$" and $\theta(a - x)$ "downstairs at $+a$".

Of further interest are the products of two step functions: for example $\theta(x)\theta(x+a) = \theta(x)$. With the same sign of the variables, the smaller argument gets its way. With different signs of variables in the argument, we receive either identically 0, as with $\theta(x)\theta(-x - a)$ or a *barrier* as for $\theta(x)\theta(-x + a) = \theta(x) - \theta(x - a)$ with the following graph: "upstairs at 0 and downstairs at $+a$":

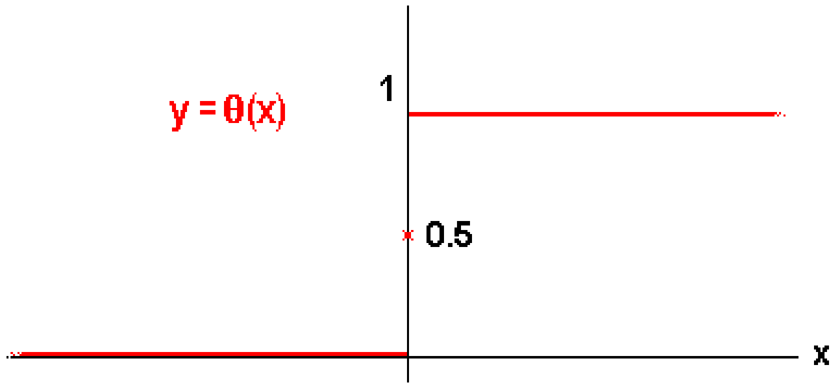


Figure 4.13: Heaviside function $\theta(x)$: "upstairs at 0".

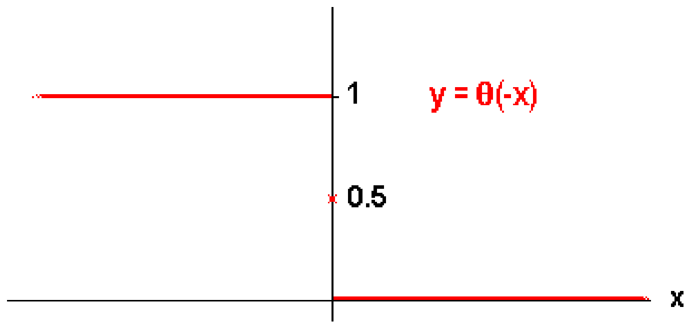


Figure 4.14: Graph of $\theta(-x)$: "downstairs at 0".

Exercise 4.5 Heaviside function: with a>0

a) Sketch $\theta(-x-a)$,

b) Sketch $\theta(x)\theta(x-a), \theta(-x)\theta(-x+a)$ *and* $\theta(-x)\theta(-x-a)$,

c) Visualize $\theta(-x)\theta(x+a) = \theta(x+a) - \theta(x), \theta(-x)\theta(x-a)$

and $\theta(x+a)\theta(a-x) = \theta(x+a) - \theta(x-a)$,

d) Draw the graph of $\theta(x)e^{-x}$,

e) Sketch the triangle function $(1 - |\frac{x}{a}|)\theta(x+a)\theta(a-x)$.

Insert: "δ-Function": *The family of functions* $\theta_a(x) = \frac{\theta(x+a)\theta(a-x)}{2a}$ *with the family parameter* a*, the "symmetrical box" of width* $2a$ *and height* $\frac{1}{2a}$ *(this means area 1), is one of the large number of function sets, whose limits (here the limit* $a \to 0$*)*

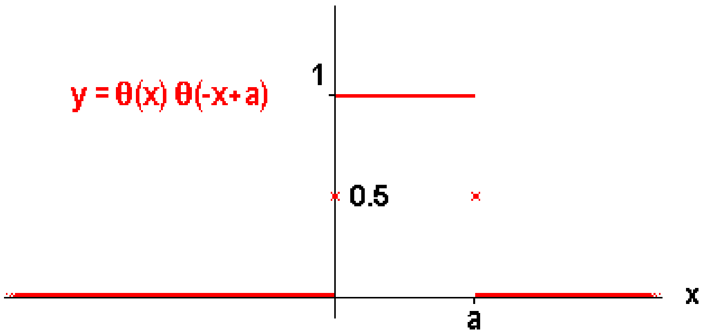


Figure 4.15: Graph of the product $\theta(x)\theta(-x+a)$

lead to the famous Dirac δ-distribution (casually also called Dirac's δ-function). We do not want to deal with them here any further since they are no more functions.

4.3 Nested Functions

Besides the possibilities which the field of real numbers offers to build new functions out of our basic set of functions with addition, subtraction, multiplication and division, there exists an important new operation to achieve that goal, namely the means of **nested functions**, sometimes also called *encapsulated functions*. It consists in "inserting one function into an other one": If for instance the value domain W_g of an ("inner") function $y = g(x)$ is lying in the definition domain D_f of an other ("outer") function $y = f(x)$: we get $y = f(g(x))$ with $x \in D_g$, i.e. a new functional dependency which is sometimes also written as $y = (f \circ g)(x)$. Since we are free in the notation of the independent and dependent variables, the nesting operation will become particularly clear if we write: $y = f(z)$ with $z = g(x)$ yields $y = f(g(x))$:

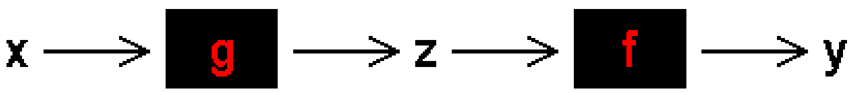


Figure 4.16: Diagram to visualize the nested function: $y = f(g(x))$

Simple examples are e.g.: $z = g(x) = 1 + x^2$ with $W_g : z \geq 1$ as inner function and $y = f(z) = \frac{1}{z}$ with $D_f = \mathbb{R}^1 \setminus \{0\}$ as outer one, which yields the Lorentz distribution function as nested function $y = \frac{1}{1+x^2}$, or $z = \sin x$ with $W_g : -1 \leq z \leq 1$ inserted into $y = |x|$ with $D_f = \mathbb{R}$ yields $y = |\sin x|$ to describe a rectified alternating current, or $z = -|2x|$ with $W_g = \mathbb{R}$ inserted into $y = e^z$ yields $y = \exp(-|2x|)$, an exponential top. Also the bell-shaped Gaussian function $y = \exp(-x^2)$ built out of $z = -x^2$ with $W_g : z \geq 0$ and $y = e^z$ is an interesting nested function which is widely used in all sciences.

Exercise 4.6 Nested Functions: *Sketch the graphs of the above mentioned examples and examine and sketch the following nested functions:*

a) $y = \sin 2x$,
b) $y = \sin x + \sin 2x + \sin 4x$,
c) $y = \cos^2 x - sin^2 x$,
d) $y = \sin(x^2)$,
e) $y = \sin\left(\frac{1}{x}\right)$,
f) $y = (\frac{\sin x}{x})^2$, *describing e.g. the intensity of light after diffraction,*
g) $y = \tan 2x$,
h) the classic Maxwell-Boltzmann velocity distribution of the colliding molecules of an ideal gas $y = x^2 e^{-x^2}$,
i) the Bose-Einstein distribution of the velocities of a gas according to quantum statistics $y = \frac{x^2}{e^{x^2}-1}$,
j) the Fermi-Dirac distribution of the velocities in an electron gas $y = \frac{\sqrt{x}}{e^{x-a}+1}$ *with the constant a depending on the temperature,*
k) Planck's formula for the spectral intensity of the wavelengths of the radiation of a cavity $y = \frac{1}{x^5[e^{\frac{1}{x}}-1]}$,
l) $y = e^{\sin x}$,
m) $y = 1 - |2x|$ *and*
n) $y = \frac{1}{|2x|}$.

You may easily check your sketches with our online function plotter or e.g. graph.tk or www.wolframalpha.com.

ONLINE ONLY	Figure 4.17 is a function plotter: It shows you in a Cartesian coordinate system the graphs of all the functions which you can build out of our basic set of functions as linear combinations, products or nested functions: You may type in the interesting function into the box above on the right using x as symbol for the independent variable and writing the function in *computer manner* (with a real number $r \in \mathbb{R}$): The plotter knows the number pi $:=\pi$, but it does not know the Euler number e. Addition, subtraction and division as usual: $x+r,\ x-r,\ x/r$ Multiplication with the star instead of the point symbol: $r^\star x := r{\cdot}x$, raising to a power with the hat: $x\hat{}r := x^r$ and $r\hat{}x := r^x$, square roots with sqrt(x) $:=\sqrt{x}$, other roots must be written as broken exponents, trigonometric functions with brackets: $\sin(x) := \sin x, \cos(x) := \cos x, \tan(x) := \tan x$, exponential functions with $\exp(x) := e^x$, because the plotter does not know the number e, hyperbolic functions also with brackets: $\sinh(x) := \sinh x, \cosh(x) := \cosh x, \tanh(x) := \tanh x$. The plotter knows only the three usual logarithms: $\ln(x) := \log_e x$, ld(x) $:= \log_2 x$ and lg$(x) := \log_{10} x$. The absolute value function and the Heaviside function must be synthesized by interval division. In any case only round brackets are allowed. You may change the scale in both directions independently within a wide range through a click at the magnifying glass symbol. If you are ready with the preparations, you should start the plotting by the return button. Of course this simple function plotter programmed by Thomas Fuhrmann computes the desired functions only at a few points and reproduces the graph only roughly. Especially in the neighbourhood of singularities the graphs must be taken with some caution.

Now, please play around with the plotter. I hope you will enjoy yourself!
If you are at the end of your wishes and fantasy, I would propose to study the building of interesting series: for instance
a) in the interval [-0.99,0.99]: first 1, then $1+x$, then $1+x+x\hat{}2$, and $+x\hat{}3,\ +x\hat{}4$, etc., and always compared with $\frac{1}{1-x}$,
b) in the interval [-0.1,0.1]: $1 - x\hat{}2/2 + x\hat{}4/2^\star 3^\star 4 - x\hat{}6/2^\star 3^\star 4^\star 5^\star 6^\star + -...$ etc., compared with $\cos(x)$,

c) in the interval [-pi,3*pi]: $\sin(x) - \sin(2{}^\star x)/2 + \sin(3{}^\star x)/3 - \sin(4{}^\star x)/4 + - \ldots$
What does this series yield?
d) in the interval [-pi,3*pi]: $\sin(x) + \sin(3{}^\star x)/3 + \sin(5{}^\star x)/5 + \sin(7{}^\star x)/7 + \ldots$
What do physicists need this series for?

4.4 Mirror Symmetry

Several **properties of functions** deserve to be considered in more detail:

Symmetry properties play an important role in all sciences: think for instance of crystals. A symmetric problem has mostly also a symmetric solution. Frequently this fact saves work. There are many kinds of symmetries. We want to select one of these, the mirror symmetry. Therefore we examine in this chapter the behavior of the functions $y = f(x)$, resp. of their graphs against reflections first in the 2-axis, i.e. in the straight line $x = 0$, if x is turned into $-x$.

In this case $y = f(x)$ is turned into $f(-x)$. In general there is no simple connection between $f(x)$ and $f(-x)$ for a given x. Take for example $f(x) = x+1$ for $x = 3 : f(3) = 4$, while $f(-3) = -2$. There exist however functions with a simple connection between the function values before and after the reflection. These functions are of special interest for physicists and mathematicians and have a special name:

A function which is symmetric against reflections in the 2-axis is called **even**:

$$\boxed{y = f(x) \textbf{ even} \iff f(-x) = f(x).}$$

For instance $y = x^2$, $y = \cos x$ and $y = |x|$ are even functions, their graphs turn into themselves through a reflection in the y-axis. The name "even" comes from the fact that all powers with even numbers as exponents are even functions.

On the other hand, a function is called ***odd***, if it is antisymmetric against a reflection in the 2-axis, i.e. it is turned into its negative or equivalently if the graph is unchanged through a point reflection in the origin:

$$\boxed{y = f(x) \textbf{ odd} \iff f(-x) = -f(x),}$$

for instance $y = \frac{1}{x}$, $y = x^3$ or $y = \sin x$.

The straight line function $y = s \cdot x + c$ is for $c \neq 0$ neither even nor odd. Every function can however be split into an *even* and an *odd part*:

$f(x) = \frac{f(x)+f(-x)}{2} + \frac{f(x)-f(-x)}{2} = f_+(x) + f_-(x)$ with the

$$\textbf{even part}: f_+(x) = \frac{f(x) + f(-x)}{2} = f_+(-x)$$

and the

$$\textbf{odd part}: f_-(x) = \frac{f(x) - f(-x)}{2} = -f_-(-x).$$

For instance c is the even part of the straight line function $y = s \cdot x + c$ and $s \cdot x$ is the odd part.

Exercise 4.7 Symmetry properties of functions:

1) Examine the following functions for mirror symmetry:
a) $y = x^4$, *b)* $y = x^5$, *c)* $y = \frac{\sin x}{x}$, *d)* $y = \tan x$, *e)* $y = \cot x$,
f) $y = \sinh x$, *g)* $y = \cosh x$ *and h)* $y = -|x|$.
2) Determine the even and odd part of e.g.:
a) $f(x) = x(x+1)$, *b)* $f(x) = x \sin x + \cos x$, *c)* $y = e^x$ *and d)* $y = \theta(x)$.

4.5 Boundedness

Our next step is to transfer the boundedness, known to us from sequences, onto functions. A function is said to be *bounded above* in an interval $[a, b]$, if there is an upper bound for the functional values in this interval:

$$y = f(x) \ \textbf{bounded above in} \ [a, b] \iff \exists B: \ B \geq f(x) \ \forall x \in [a, b]$$

Bounded below is defined analogously with the lower bound $A \leq f(x)$.
The standard parabola $y = x^2$ for example and the absolute value function $y = |x|$ are bound below through $A = 0$ and the step function $\theta(x)$ is bounded above through $B = 1$ and bounded below through $A = 0$.

Exercise 4.8 Boundedness:

Examine the following functions for boundedness within their definition domain in $\mathbb{R}$*:*
a) $y = 2 - 2x^2$, *b)* $y = x^2 - 2x - 3$, *c)* $y = \frac{2x-3}{x-1}$, *d)* $y = \sin x + \cos x$,
e) $y = x \sin x$, *f)* $y = 1 - e^{-x}$, *g)* $y = x + e^{-x}$, *h)* $y = xe^{-x}$,
i) $y = x^2 e^{-x}$, *j)* $y = e^{-x} \sin x$ *and* *k)* $y = \frac{1}{|x|}$.

4.6 Monotony

Also monotony can be transferred from sequences onto functions, since sequences can be understood as special functions over the definition domain $\mathbb{N}$:

A function is said to be **monotonically increasing** in the interval $[a, b]$, if with increasing argument the functional values also increase in the interval $[a, b]$:

$$y = f(x) \textbf{ monotonically increasing in } [a,b]$$
$$\iff x_1, x_2 \in [a,b] \in D_f : \; x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2)$$

For example is $y = sx + c$ for $s > 0$ monotonically increasing.

Monotonically decreasing is analogously defined with $f(x_1) \geq f(x_2)$, for instance $y = \frac{1}{x}$ in its definition domain D_f is monotonically decreasing.
If even $f(x_1) < f(x_2)$ holds true for all $x_1, x_2 \in D$ with $x_1 < x_2$, we call the function *strictly* monotonic increasing as in the case of sequences. In both previous examples the monotony is strict.

Exercise 4.9 Monotonic functions:

Examine the following functions for monotony:
a) $y = x^2$, *b)* $y = \frac{x^3}{3} - 3$, *c)* $y = \frac{2x-3}{x-1}$,
d) $y = \sin x$ *in* $[-\frac{\pi}{2}, \frac{\pi}{2}]$, *e)* $y = \tan x$ *in* $[-\frac{\pi}{2}, \frac{\pi}{2}]$,
f) $y = \cos x$ *in* $[0, \pi]$, *g)* $y = 1 - e^{-x}$, *h)* $y = \sinh x$,
i) $y = \cosh x$ *and* *j)* $y = \theta(x)$.

4.7 Bi-uniqueness

As we have stressed in the introduction of the term of a function, the definition contains the unambiguity of the mapping, i.e. to every pre-image point x exists **exactly** one image point $y = f(x)$. However, it is still possible that two different arguments yield the same functional value as image point, meaning $f(x_1) = f(x_2)$ for $x_1 \neq x_2$. Functions, where this is not the case anymore have got a special name: we call these functions **bi-unique** (also known as reversible or bijective) in an interval $[a, b]$, if also every functional value y out of the corresponding value domain emerges from **exactly** one argument:

$$y = f(x) \textbf{ bi-unique in } [a, b]: \iff \forall y \in W_f \; \exists! x \in [a, b] : \; y = f(x)$$

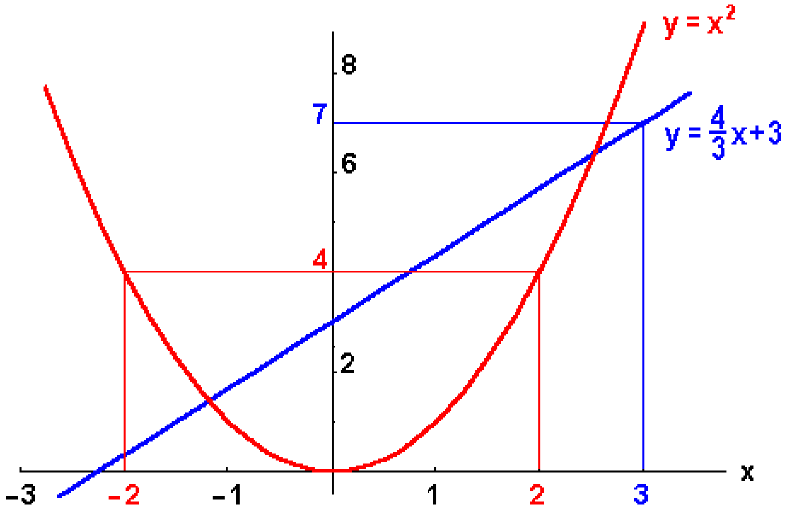


Figure 4.18: Graph of a straight line and the standard parabola

The figure shows as an example of a bi-unique function a straight line $y = sx + c$, especially $y = \frac{4}{3}x + 3$, where e.g. the functional value $x = 3$ corresponds exactly to the argument $y = 7$, and as a counter example the standard parabola $y = x^2$, for which we get the functional value $y = 4$ from both $x = 2$ and $x = -2$.

Insert: Bijective: *As the name "bijective" indicates, mathematicians approach the term "bijection" in two steps:*

*1) First they call a mapping **injective** ("one-to-one"), by which equal image points arise only from equal pre-image points:*

$$y = f(x) \textbf{ injective in } D_f \quad \Longleftrightarrow \forall x_1, x_2 \in D_f: \; f(x_1) = f(x_2) \Rightarrow x_1 = x_2,$$

or equivalently to this : if different pre-images always lead to different images:

$$y = f(x) \textbf{ injective in } D_f \; \Longleftrightarrow \forall x_1, x_2 \in D_f: \; x_1 \neq x_2 \Rightarrow f(x_1) \neq f(x_2).$$

In this case the equation $y = f(x)$ has for all y at most *one solution x, and every straight line parallel to the 1-axis hits the graph of the function at most once.*

2) Then in a second step they consider the area containing the image elements (which was not very interesting for us) and investigate whether it consists only of the image points or contains further points in addition. If the image area consists only of the image domain, they call the mapping **surjective ("onto"):**

$$y = f(x) \textbf{ surjective on } W_f \; \Longleftrightarrow \forall y \in W_f \; \exists x \in D_f: \; f(x) = y.$$

Then the equation $y = f(x)$ has for all y at least *one solution x.*

bijective := injective + surjective

Thus, in a bijective mapping the equation $y = f(x)$ has exactly *one solution x and the function is reversible.*

Exercise 4.10 Bi-unique Functions:

Examine the following functions for bi-uniqueness:
a) $y = x^2$, *b)* $y = x^3$, *c)* $y = \frac{2x-3}{x-1}$, *d)* $y = \sin x$,
e) $y = \tan x$, *f)* $y = \cos x$, *g)* $y = 1 - e^{-x}$, *h)* $y = \sinh x$,
i) $y = \cosh x$ *and* *j)* $y = \theta(x)$.

4.8 Inverse Functions

For all bi-unique functions, i.e. reversibly unambiguous mappings of a definition domain D_f onto a value domain W_f we can build *inverse functions* $x = f^{-1}(y) =: g(y)$ by resolving the equation $y = f(x)$ for x, where definition domain and value domain exchange their roles: $D_g = W_f$ and $W_g = D_f$. The inverse function describes the reversal of the original mapping. Since the original mapping $y = f(x)$ and the inverse mapping $x = g(y)$ cancel each other, for the nested function $f(g(y)) = y$ holds true. From this the above notation $g = f^{-1}$ can easily be understood: $f(f^{-1}(y)) = y = f^{-1}(f(y))$.

Insert: Function Symbol: *With a closer look, the symbol f has two different meanings:*
1) On the one hand f gives the operation, which creates the function values y out of the arguments x. If we write f^{-1}, we mean the reversal of this operation.
2) On the other hand the term $f(x)$ means the special function value, which emerges from a certain value x of the variable by this prescription. The inverse of this function value is denoted by $\frac{1}{f(x)} = (f(x))^{-1}$. $f^{-1}(x)$ is the function value of the inverse function. E.g. $\sin 1 = 0.84$, $(\sin 1)^{-1} = 1.188$, *but* $\sin^{-1} 1 = \arcsin 1 = 1.57$.

In the sense of our physical definition of the function as input-output relation, with the inverse function input and output are interchanged, i.e. the flow direction arrows in our figure are simply reversed.

Figure 4.19: Black boxes for $y = f(x)$ and the inverse function $x = g(y)$

Inverse functions occur very often in physics. As an example we consider the length L of the liquid column in a narrow glass pipe as function of temperature $T : L = f(T)$. If we use the glass pipe as thermometer to measure the temperature, we observe the length of the liquid column and infer from it the temperature $T = g(L)$.

Usually after the solution of the equation $y = f(x)$ for x has been found, the now independent variable y is renamed x and the dependent one y. For the graph the transition to the inverse function simply means the *reflection* in the straight line $y = x$, i.e. the *bisector* of the first and third quadrant. In this manner we receive a new function out of every bi-unique function.

An instructive example is the **standard parabola** $y = x^2$. Only through the limitation of the definition domain to $x \geq 0$ does it become a bi-unique function which is reversible: The inverse function is a new function for us, the **square root function** $x = +\sqrt{y}$ and after renaming the variables: $y = +\sqrt{x} = x^{\frac{1}{2}}$. Through the notation with the fraction in the exponent, powers were defined also for rational exponents, beyond our earlier considerations, without any change in the calculation rules for powers.

The possibility to create new functions by inversion out of bi-unique functions enlarges the treasure of our basic set of functions introduced in section 4.2 (rational, trigonometric, and exponential functions) nearly to twice as much. We want to devote ourselves now to this task:

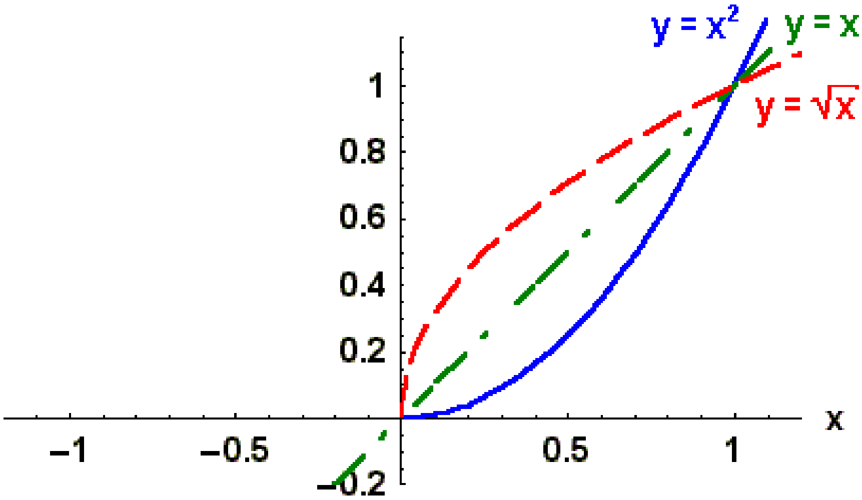


Figure 4.20: The standard parabola $y = x^2$ and its reflection in the bisector delivering the inverse function: the square root function $y = +\sqrt{x}$

4.8.1 Roots

The example of the parabola has already shown to us that the inverse functions of the powers $y = x^n$ with integer exponents $n \in \mathbb{Z}$ are the **root functions** with fractions as exponents: $x = y^{\frac{1}{n}}$, rewritten as: $y = \sqrt[n]{x} = x^{\frac{1}{n}}$ where however the even-numbered powers need to be made bi-unique before the reversal through limitation of the domain to $x \geq 0$.

As inverse functions of the polynomials we receive more complicated roots: for example from $y = x^2 + 1$ which is bi-unique for $x \geq 0$, we find $x = \sqrt{(y-1)}$ and after redefinition $y = \sqrt{(x-1)}$.

Exercise 4.11 Inverse functions:

Determine the inverse functions of the following functions:
a) $y = -2x - 2$, *b)* $y = 2 - 2x^2$, *c)* $y = x^2 - 2x - 3$, *d)* $y = \frac{x^3}{3} - 3$,
e) $y = \frac{1}{(1-x)}$ *and* *f)* $y = \frac{1}{x^2-1}$.

4.8.2 Cyclometric Functions

The trigonometric functions $y = \sin x$ or $y = \cos x$ are periodic in x with the period 2π and $y = \tan x$ or $y = \cot x$ with the period π. Thus they are by no means bi-unique functions. Only by limitation of the definition domain can reversibility be achieved.

For the odd functions $y = \sin x$, $y = \tan x$ and $\cot x$ we usually choose
$-\frac{\pi}{2} < x \leq \frac{\pi}{2}$, for the even one $y = \cos x$ $0 < x \leq \pi$. Due to the periodicity we

can also choose other intervals, shifted by the multiple of 2π. Here you should be very careful with the calculations: Especially before using a calculator, you should familiarize yourself with the domains of the reverse functions beforehand. The inverse functions of the trigonometric functions are called **cyclometric** or **arcus functions:**

to	$y = \sin x$	$y = \arcsin x$	and	$y = \cos x$	$y = \arccos x,$
to	$y = \tan x$	$y = \arctan x$	and	$y = \cot x$	$y = \text{arccot } x.$

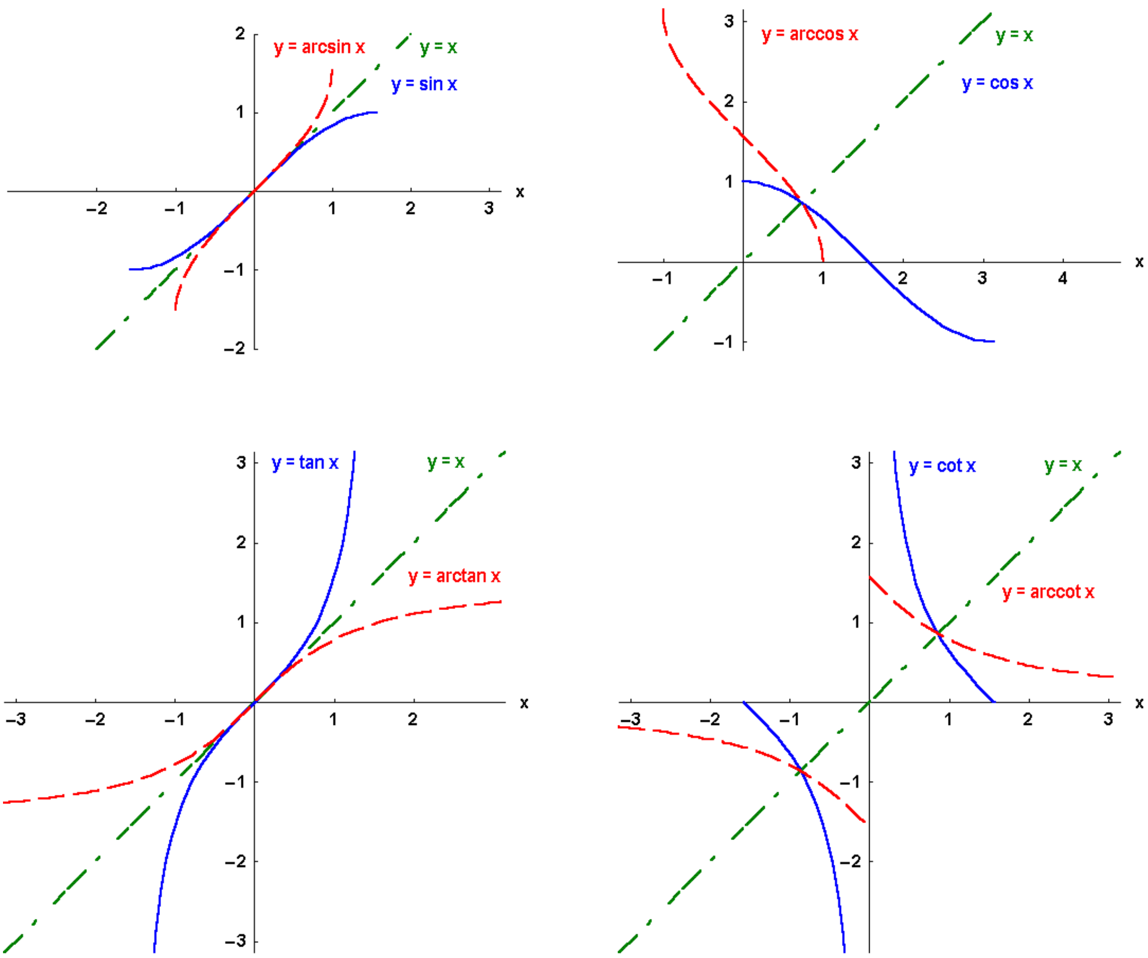


Figure 4.21: The trigonometric functions and their inverse functions, the cyclometric functions.

Insert: Arcus: *The term "*arcsin x*", pronounced: "arcus sine* x*", means the arc (lat.: arcus) in the unit circle, i.e. the angle whose sine has the value* x*.*

Insert: Notation: *Unfortunately the notation in the literature (especially in the Anglo-American) is not standardized. You may also find* $\arcsin x$ *or simply* asin x *instead of* $\sin^{-1} x$*. Particularly the last notation causes sometimes confusion, since it can easily be mixed up with the inverse of the sine function* $(\sin x)^{-1} = \frac{1}{\sin x}$.

4.8.3 Logarithms

Through reflection of the graph of the natural exponential function $y = e^x$ in the bisector for $x > 0$ we get the **natural logarithm** $y = \ln x$ which rises only very slowly:

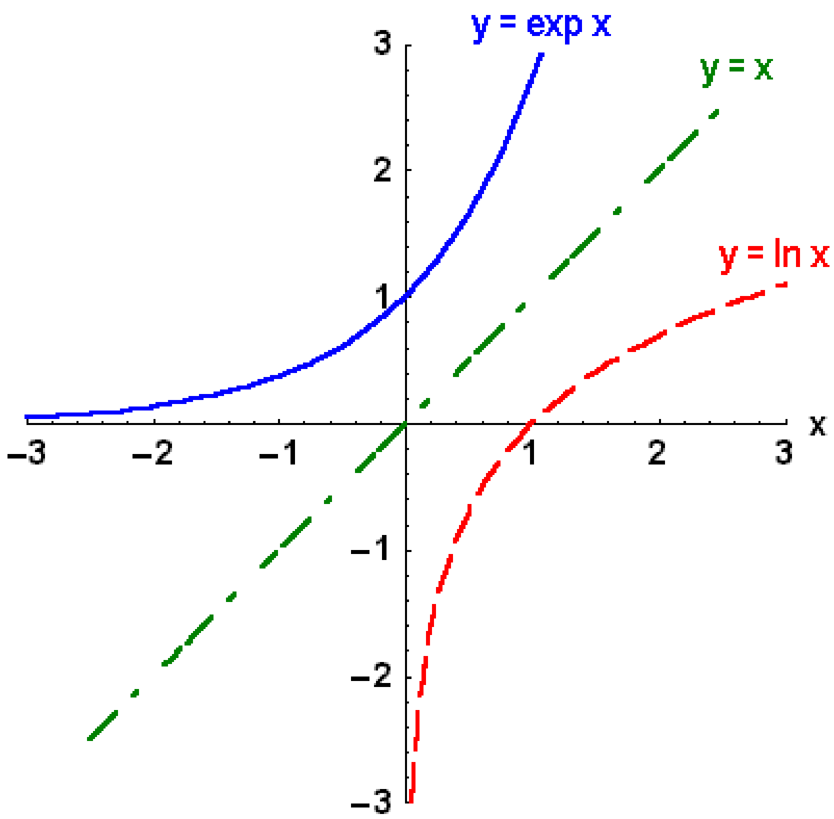


Figure 4.22: The exponential function and its inverse function, the natural logarithm

The characteristics of the strictly monotonic rising natural logarithm are readable from the graph above: $\ln 1 = 0$ and $\lim_{x \to 0} \ln x = -\infty$. From the calculation rules for powers we obtain the following calculation rules for natural logarithms:

$\ln y \cdot z = \ln y + \ln z,$ $\ln(\frac{y}{z}) = \ln y - \ln z$ and $\ln(z^y) = y \ln z$.

After having become familiar with especially the *natural* exponential function and the natural logarithm, we can use this knowledge to define the **general exponential function**:

general exponential function: $y = b^x := e^{x \ln b}$ for $b > 0$

and as its inverse function the ***general logarithm***:

general logarithm: $y = \log_b x := \frac{\ln x}{\ln b},$

which for $b > 1$ (just as for $\ln x$) rises strictly monotonic, for $b < 1$ however decreases monotonically.

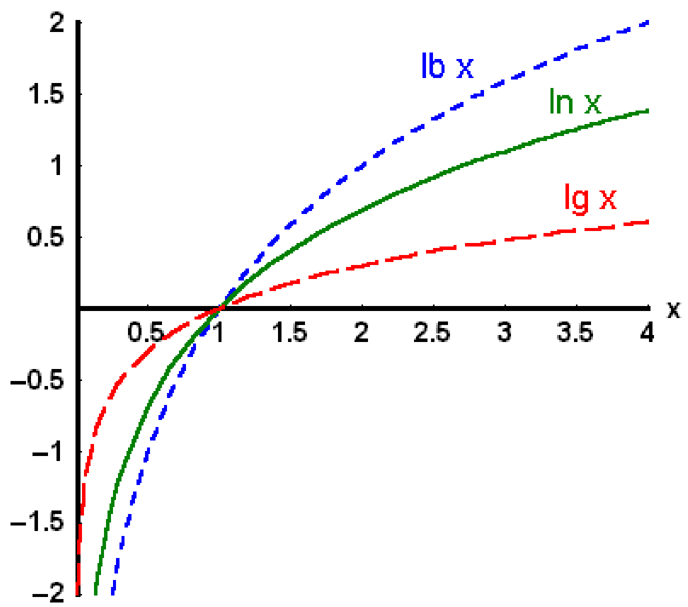


Figure 4.23: The three most important logarithms for the bases: 2, e, and 10

As calculation rules for the general logarithms with **unchanged base** b we get:

$\log_b yz = \log_b y + \log_b z,$ $\log_b(\frac{y}{z}) = \log_b y - \log_b z$ and $y \log_b z = \log_b(z^y)$.

Besides the very important natural logarithms $\boxed{\ln := \log_e}$ with the irrational number e = 2.7182... as base, there are simplified notations for two further widely used bases: the binary $\boxed{\text{lb} := \log_2}$ or dual $\boxed{\text{ld} := \log_2}$ and the common or Brigg's logarithms with the base 10 $\boxed{\text{lg} := \log_{10}}$.

The conversion between logarithms with **different** bases follows from the formula:

$$\boxed{\log_b y = \log_b z \cdot \log_z y}$$

especially, e.g. for $b = 10$ and $y = x$:

$$\boxed{\text{lg}x = \text{lg}\, e \ \ln x = 0.434 \ln x,}$$

since through the triple use of the identity

$$\boxed{b^{\log_b x} = 2^{\text{ld}\, x} = e^{\ln x} = 10^{\text{lg}\, x} = x}$$

we get $b^{\log_b y} = y = z^{\log_z y} = (b^{\log_b z})^{\log_z y} = b^{\log_b z \cdot \log_z y}$ and thus the exponents are equal.

Exercise 4.12 Logarithms:

a) What is $\log_b b$*?*
b) Show that $\ln 10 = \frac{1}{\text{lg}e}$ *or rather* $\ln 2 = \frac{1}{\text{lb}e}$.
c) Calculate $\text{lb}x$ *from* $\ln x$.
d) Calculate $25^{2.5}$.

Also the **general power function** is defined with the help of the natural exponential function and the natural logarithm for $x > 0$ and $r \in \mathbb{R}$ through:

$$\boxed{\textbf{general}\text{ power function: } \quad y = x^r = e^{r \ln x} \quad \text{for } x > 0 \text{ and } r \in \mathbb{R}.}$$

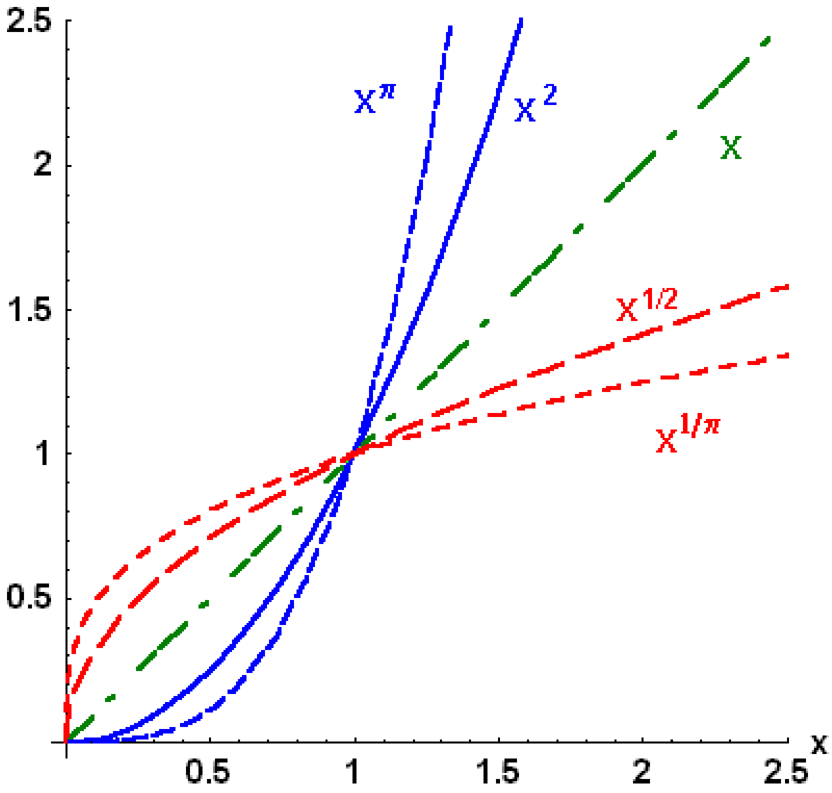


Figure 4.24: Power functions

We will not use it very often.
The inverse functions of the hyperbolic functions: $\cosh x$, $\sinh x$ and $\tanh x$, which we have built out of natural exponential functions are more important for physics: They are called **area functions**, can obviously be expressed through natural logarithms and are displayed in the following figure:

to $y = \cosh x := \dfrac{e^x + e^{-x}}{2}$	$y = \text{arcosh } x = \ln\left(x + \sqrt{x^2 - 1}\right)$,
to $y = \sinh x := \dfrac{e^x - e^{-x}}{2}$	$y = \text{arsinh } x = \ln\left(x + \sqrt{x^2 + 1}\right)$,
to $y = \tanh x := \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$	$y = \text{artanh } x = \frac{1}{2} \ln \dfrac{1 + x}{1 - x}$.

Insert: Notations: *Also in this case, there are notation difficulties like with the inverse functions of the trigonometric functions: arsinh* x = *arcsinh* x = *arsh* x = *sinh*^{-1}x*, etc.*

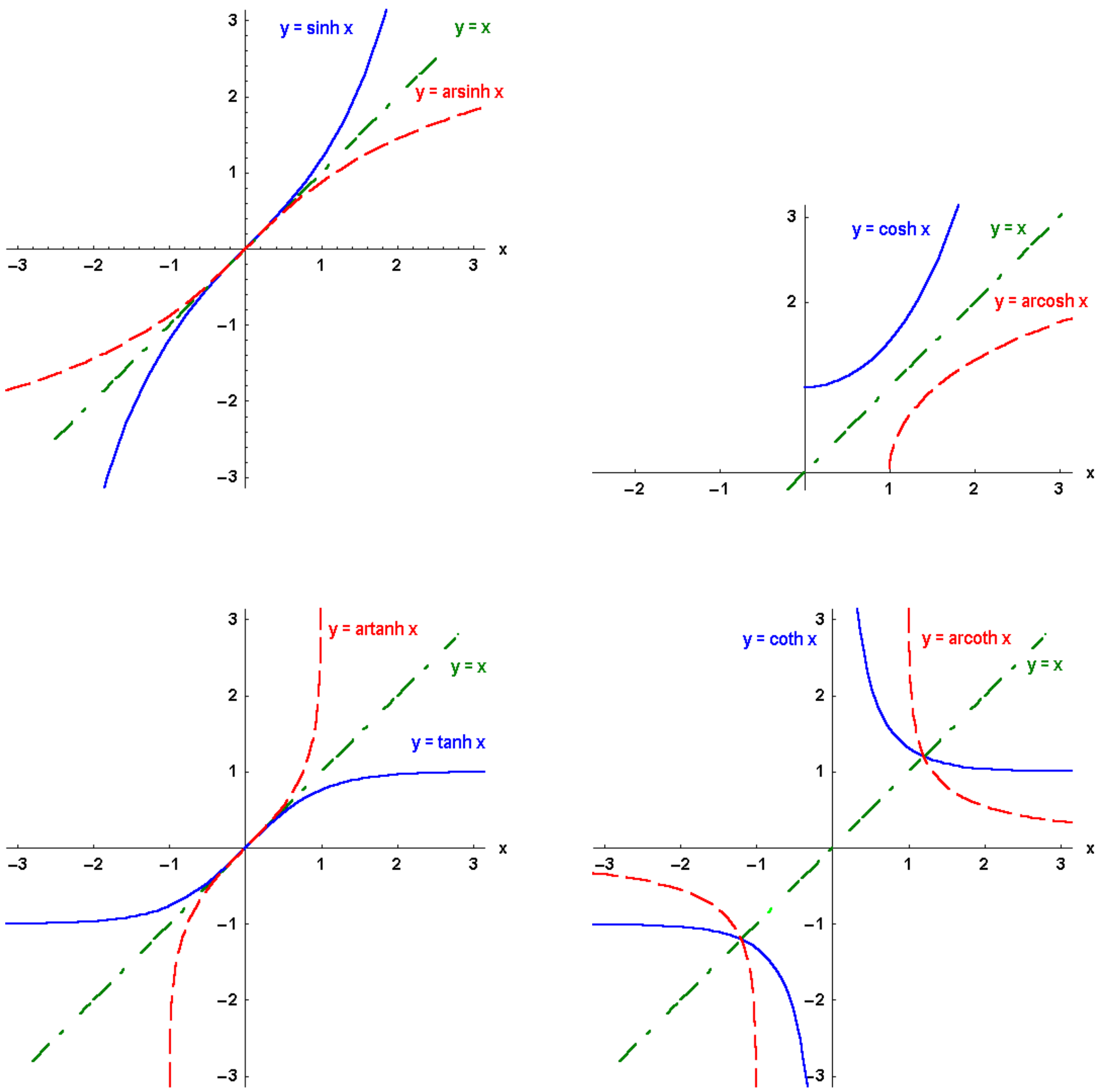


Figure 4.25: The hyperbolic functions and their inverse functions, the area functions

Insert: Area: *The name* **"arcosh** x**"**, *i.e.* **"area hyperbolic cosine** x**"**, *comes from the connection with the* area *(also lat.: area) of a sector of the standard hyperbola: It can be shown that y is the square measure of the area of the hyperbola sector which in Figure 4.11 is dyed (between the origin, the moving point, the vertex, and the moving point reflected at the 1-axis), if the 1-coordinate of the moving point (which means the hyperbolic cosine equals x).*

Exercise 4.13 Area functions:

a) Show that from $y = \ln\left(x + \sqrt{x^2+1}\right)$ *follows* $x = \sinh y$.
b) Show that from $y = \frac{1}{2}\ln\frac{1+x}{1-x}$ *follows* $x = \tanh y$.

4.9 Limits

The calculation of **limits for functions** is reduced to our limiting procedure for sequences by the following consideration: Whenever we want to know whether the functional values $f(x)$ of a real function f tend to a number y_0 when the arguments x_0 approach a real number x_0, we choose a sequence $(x_n)_{n\in\mathbb{N}} \subseteq D_f$ of real numbers within the definition domain D_f of the function f, which for $n \to \infty$ tends to the number $x_0 \in D_f$. Then we calculate the functional values $f(x_n)$ for these arguments, which once again form a sequence $(f(x_n))_{n\in\mathbb{N}}$, and check whether this sequence of the functional values converges to y_0. If this holds true for *every* sequence taken out of the domain and converging to x_0 then we call such sequence of functional values convergent to $y_0 : \lim_{x\to x_0} f(x) = y_0$:

$$\boxed{\lim_{x\to x_0} f(x) = y_0 \textbf{ convergent: } \iff \forall (x_n)_{n\in\mathbb{N}} : \lim_{n\to\infty} x_n = x \Rightarrow \lim_{n\to\infty} f(x_n) = y_0}$$

If we modify our definition for the convergence of sequences, we find:

$$\boxed{\lim_{x\to x_0} f(x) = y_0 \textbf{ convergent: } \iff \forall \varepsilon > 0\ \exists \delta > 0 : |f(x) - y_0| < \varepsilon \quad \forall x \in D_f \quad \mathrm{with} |\mathrm{x} - \mathrm{x_0}| < \delta}$$

To show this for **all** sequences is easier said than done! We do not need however to occupy ourselves with these partly difficult mathematical questions which you will have enough opportunity to deal with in the analysis lecture. Instead we shall be content with some examples which are important for physicists.

Already from the graphs we can see e.g. for the behaviour **at the origin** that for the powers $\boxed{\lim_{x\to 0} x^n = 0}$ while $\boxed{\lim_{x\to 0} x^{-n} \text{ divergent}}$ is divergent for $n \in \mathbb{N}$. Moreover we can clearly see that $\boxed{\lim_{x\to 0} \cos x = 1}$ and $\boxed{\lim_{x\to 0} \sin x = 0}$.

For the behaviour **at large values** of the variable $\boxed{\lim_{x\to\infty} x^{-n} = 0}$, and

$\boxed{\lim_{x\to\infty} x^{n} \text{ is divergent}}$. The same goes for $\boxed{\lim_{x\to\infty} x^{n}e^{-x} = 0}$ and

$\boxed{\lim_{x\to\infty} x^{-n}e^{x} \text{ is again divergent}}$.

We say therefore that the exponential function rises more strongly than any power function.

Insert: $\frac{\sin x}{x}$**:** *For the important limit* $\boxed{\lim_{x\to 0} \frac{\sin x}{x} = 1}$ *a nice vivid*

proof of de l'Hospital *exists according to the following figure:*

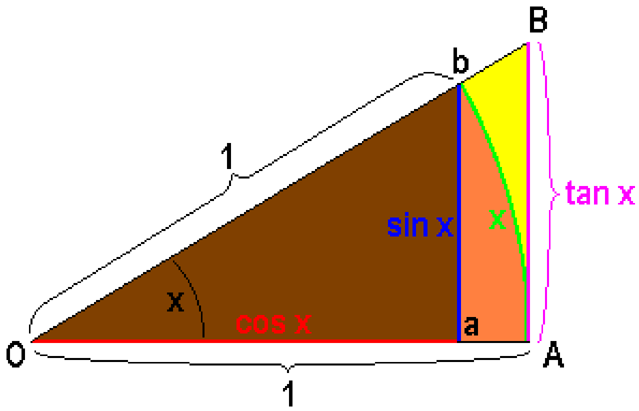


Figure 4.26: Concerning the proof of de l'Hospital

We consider the sector $A0b$ of the unit circle with the central angle x near 0, the line segments: $|0A| = 1$ and $|0b| = 1$ and the arc (Ab) over the angle x, as well as the point a on the line segment $|0A|$, the line segments $|0a| = \cos x$ and $|ab| = \sin x$, and the prolongation of the line segment $|0b|$ to B, so that the line segment is $|AB| = \tan x$.

Obviously the following inequalities hold for the areas of the triangles and the sector, respectively:

	F(triangle: 0ab)	$\leq$	*F(sector: 0Ab)*	$\leq$	*F(triangle: 0AB),*
meaning	$\frac{1}{2}\sin x\cos x$	$\leq$	$\frac{1^2\pi x}{2\pi}$	$\leq$	$\tan\frac{x}{2}$.
times $\frac{2}{\sin x}$ *gives:*	$\cos x$	$\leq$	$x/\sin x$	$\leq$	$1/\cos x$
reciprocal:	$\frac{1}{\cos x}$	$\geq$	$\frac{\sin x}{x}$	$\geq$	$\cos x$,
thus in the limit $x \to 0$*:*	*1*	$\geq$	$\lim_{x\to 0}\frac{\sin x}{x}$	$\geq$	*1.*

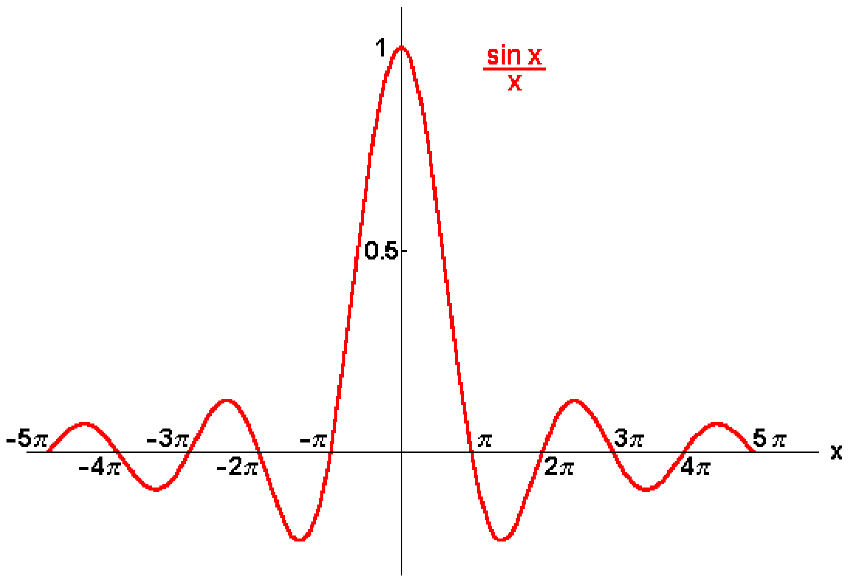


Figure 4.27: Graph of the function $\frac{\sin x}{x}$

Exercise 4.14 Limits of functions:

Calculate
a) $\lim\limits_{x\to\frac{1}{2}}\frac{1+x}{1-x}$,
b) $\lim\limits_{x\to\pi}\frac{\sin x}{x-\pi}$,
c) $\lim\limits_{x\to 0}(\tan x)^2$ *and*
d) examine the following limit $\lim\limits_{x\to 0}\frac{e^x-1}{x}$ *with the help of the inverse function and the exponential sequence. We will need this in the next chapter and will derive it in the chapter after the next one in a much more elegant way.*

4.10 Continuity

The last important property of functions which we need is **continuity:** Particularly in classical physics we often take the viewpoint: "natura non facit saltus" (nature does not jump), i.e. we deal with continuous functions. For many experiments continuity is unavoidable because of the finite accuracy of measurements. But there are also discontinuous processes in nature, e.g. switching on or off, and "quantum leaps".

Mathematicians define a function as continuous at a point x_0 if it maps points in the neighbourhood of x_0 to other neighbouring points, in shorthand:

$$\boxed{y = f(x) \textbf{ continuous} \text{ at } x_0 \iff \forall \varepsilon > 0\ \exists \delta(\varepsilon) > 0 : |f(x) - y_0| < \varepsilon\ \forall x \text{ with } |x - x_0| < \delta}$$

For limits this means that at the considered point x_0 the right-hand limit and the left-hand one are equal and are given by the function value $y_0 = f(x_0)$ of the limit x_0 of a sequence x_n out of the definition domain of the arguments:

$$\boxed{\lim_{x \to x_0 + 0} f(x) = \lim_{x \to x_0 - 0} f(x) =: \lim_{x \to x_0} f(x) = y_0 = f(x_0) = f(\lim_{n \to \infty} x_n).}$$

The graph of a continuous function "does *not jump*". The Heaviside function is the function with the "unit jump". With its help all discontinuities occurring in physics can be described. Sums, differences, products, quotients, and nested functions of continuous functions are again continuous. Therefore all functions considered until now *apart from* the Heaviside step function are continuous *within* their definition domains. The step function jumps at the point 0 by an amount of 1: $\lim_{x \to 0+0} \theta(x) = 1$, $\lim_{x \to 0-0} \theta(x) = 0$ while $\theta(0) = \frac{1}{2}$. The standard hyperbola $y = \frac{1}{x}$ is admittedly discontinuous at the point $x = 0$, but it is not defined there.

Exercise 4.15 Continuous functions:

Check the continuity of the following functions at the origin $x_0 = 0$*:*
a) $y = x$, *b)* $y = x^2$, *c)* $y = \frac{1}{1-x}$, *d)* $y = x \sin x$, *e)* $y = x + e^{-x}$,
f) $y = \frac{\sin x}{x}$, *g)* $y = \frac{e^x - 1}{x}$, *h)* $y = |x|$, *i)* $y = \theta(x + a)\theta(a - x)$,
j) $y = \theta(x)\theta(-x - a)$ *k)* $y = \theta(x)e^{-x}$ *and* *l)* $y = \theta(x)xe^{-x}$.

Exercise 4.16 Function quiz:

Suppose you have measured the functional dependence of a physical quantity $y = f(x)$ (e.g. the current strength) on another quantity x (e.g. the electric voltage) by repeated careful measurements between the values 0 and 3 and your measured values are within your accuracy well described by one of the 18 curves sketched in Figure 4.28. Which simple hypothesis about the functional dependency of the measured quantity $y = f(x)$ from the varied quantity x would you set up?

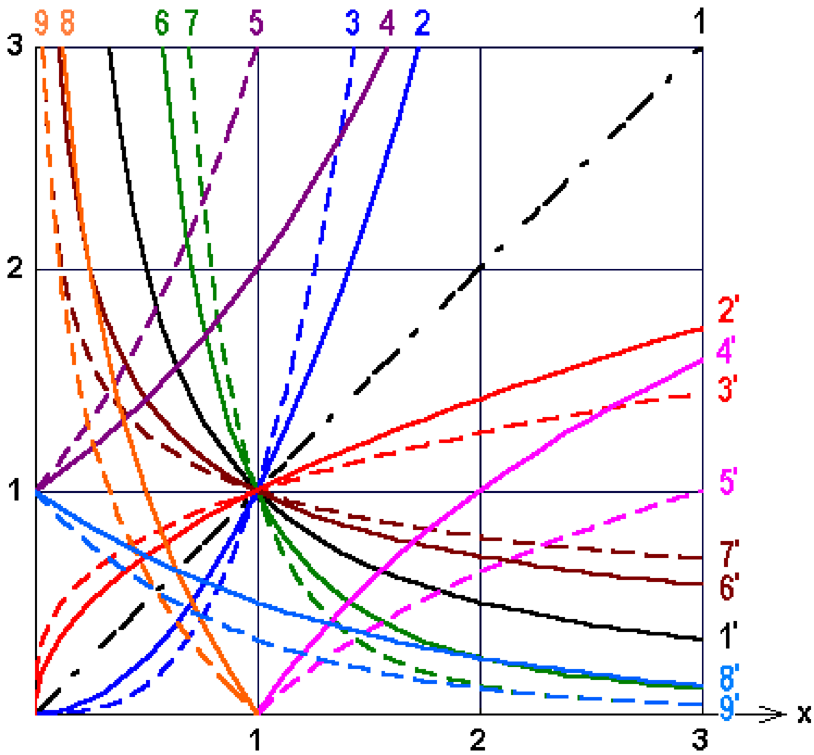


Figure 4.28: Function quiz

Chapter 5

DIFFERENTIATION

It is hardly an overstatement to claim that theoretical physics began with the concept of differentiation. The development of this branch of mathematics by Leibniz and Newton in the 17^{th} century made possible the formulation of the exact laws describing a wide range of very important physical phenomena: Newton's laws of mechanics, the Maxwell equations for electrodynamics and the Schrödinger equation of quantum mechanics are all expressed in the language of differential equations. The solutions of these require the entire apparatus of analysis, especially of the differential and integral calculus. For this reason the present chapter is of the utmost importance to us. Differentiation and integration are indispensable tools of every physicist.

5.1 Differential quotient

We first examine the uniform motion of a particle along a straight line. Thereby, the distance covered is: $x(t) = st + x_0$ as a function of time t, where $x_0 = x(0)$ means the position at the time $t = 0$.

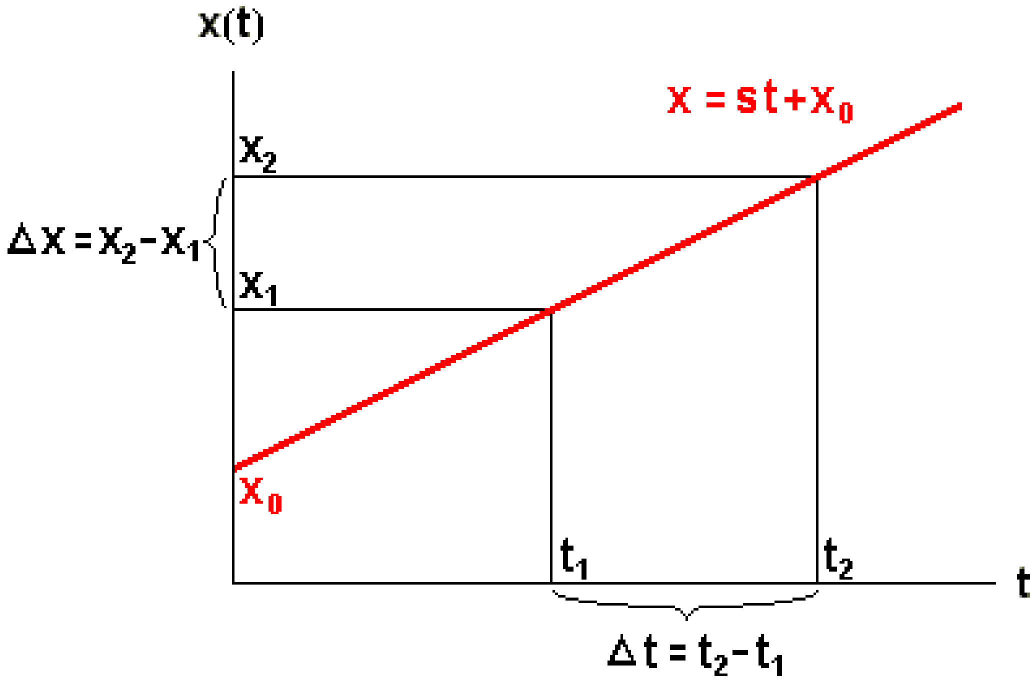


Figure 5.1: The straight line $x(t) = st + x_0$

Physicists are first of all interested in the **velocity of the motion**. From the graph of this linear function, an ascending straight line with the gradient s through the point $(0, x_0)$, we extract the velocity v as quotient of the covered distance $x(t) - x(0)$ divided by the time needed t, which gives us exactly

$$v = \frac{x(t) - x_0}{t} = s,$$

the gradient. It is therefore the **gradient of the graph**, which interests us. For the straight line we obviously could have taken an other time interval $t_2 - t_1$, too:

$$v = \frac{x(t_2) - x(t_1)}{t_2 - t_1}$$

and received the same result. Because with a uniform motion equal distances are covered within equal time intervals: the velocity is constant. We mark the *differences* in numerator and denominator of the quotient in the following way through upper-case Greek deltas: $(x(t_2) - x(t_1)) =: \Delta x(t)$ and $(t_2 - t_1) =: \Delta t$ and call the quotient of both difference quotient:

$$v = \frac{\Delta x(t)}{\Delta t}.$$

Generally, when the independent variable is called x again, we receive for the

$$\textbf{difference quotient:} \quad \frac{\Delta f(x)}{\Delta x} := \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

How will the situation change, if we consider in our physical example a general motion along our straight line with a time *varying* velocity, which is displayed through an arbitrary function of the distance from time $x(t)$?

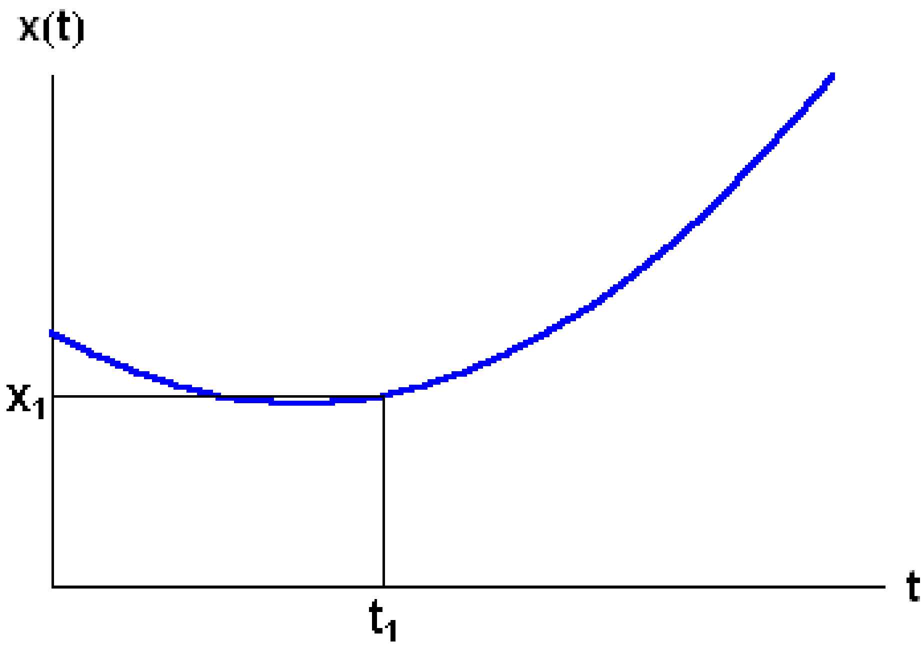


Figure 5.2: Graph of an arbitrary function $x(t)$ of time

Through the figure we recognize that the difference quotient gives the gradient of the secant, which connects the two points $(t_1, x(t_1))$ and $(t_2, x(t_2))$. The value of the difference quotient is then

$$v_m = \frac{\Delta x(t)}{\Delta t} = \frac{x(t_2) - x(t_1)}{t_2 - t_1},$$

the **middle** or **average velocity** during the time interval Δt. For many purposes the average velocity plays an important role. In physics however, the **instantaneous velocity**, say at the time t_1, is far more important. We get it from the average velocity between t_1 and t_2 by the limit in which t_2 approaches t_1. This limit is called the differential quotient.

5.2 Differential Quotient

In order to determine the **instantaneous velocity** for example at the time t_0, we choose an arbitrary point in time $t = t_0 + \Delta t$ near t_0 and draw the **secant** through the function values $x(t) = x(t_0) + \Delta x(t)$ and $x(t_0) = x_0$. Then we determine their gradient as the difference quotient $s = \frac{\Delta x(t)}{\Delta t} = \frac{x(t)-x(t_0)}{t-t_0}$ and let the time point t tend to t_0, i.e. let $\Delta t \to 0$. With this procedure for a continuous function also $x(t)$ tends to $x_0 = x(t_0)$ and the secant becomes the **tangent** to the graphs with the

$$\textbf{tangent gradient:} \quad \left.\frac{dx}{dt}\right|_{t_0} = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t}$$

In our usual mathematical terminology with x for the independent and $y = f(x)$ for the dependent variable, the results is

$$\textbf{differential quotient:} \quad \left.\frac{df(x)}{dx}\right|_{x_0} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

The difference quotient $\frac{\Delta y}{\Delta x} \equiv \frac{\Delta f(x)}{\Delta x} \equiv \frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}$ has turned into the so-called differential quotient through this limit $\Delta x \to 0$, denoted by its inventor Leibniz with $\left.\frac{df(x)}{dx}\right|_{x_0}$ or $\left.\frac{dy}{dx}\right|_{x_0}$, which is however, at first sight not quotient itself, but only the limit of a quotient. Just as for all with all fundamental terms in mathematics there are several notations for it: The alternative notation which most of you are familiar with from school, $f'(x_0)$, spoken "f prime at the point x_0", was proposed by Lagrange and serves as a reminder that the gradient of the graph, also called (first) derivative of a function, in general is itself again a new function varying along the curve of the independent variable x, here especially given at the point x_0. Also the term $\left(\frac{d}{dx}\right) f(x)\big|_{x_0}$ is used which emphasizes that the differentiation is an "operation" where the "differential operator" $\left(\frac{d}{dx}\right)$ is to act on the function $f(x)$ standing to the right of it, and the result should be evaluated especially at the point $x = x_0$. It makes sense to have all these notations available, and to use the notation that is most convenient for one's purpose.

Equivalent denotations: $f'(x_0) \equiv \left.\frac{df(x)}{dx}\right|_{x_0} \equiv \left(\frac{d}{dx}\right) f(x)\big|_{x_0} \equiv \left(\frac{dy}{dx}\right)\big|_{x_0}$

Hereby, we have to mention a curiosity of physicists: If the independent variable is the time t, as of course happens quite often, physicists write and speak a high-placed "point" $\dot{}$ instead of the "prime" $'$: $\dot{x}(t_0) \equiv \left.\frac{dx(t)}{dt}\right|_{t_0}$.

Insert: Linear Approximation: *We have seen that the differential quotient of a function $f(x)$ at a point x_0 has a nice vivid meaning: it gives us the gradient of the tangent to the graph of the function at the point x_0. We may understand this tangent property also in the following way:*

*We set ourselves the task of **approximating** the curve described by the function $y = f(x)$ in the neighborhood of the point x_0 as nearly as possible **by a straight line** $g(x) = sx + a$: For this purpose we demand therefore:*

1. At the point x_0, there should hold $f(x_0) = g(x_0) = sx_0 + a$, from which the absolute term $a = -sx_0 + f(x_0)$ can be determined. Inserted above this gives: $g(x) = s(x - x_0) + f(x_0)$. Thus we find for the deviation of the approximating line g from the curve f :

$$f(x) - g(x) = f(x) - f(x_0) - s\Delta x$$

with the distance $\Delta x := x - x_0$ of the independent variable from the approximation point.

2. This deviation of the approximation line g from the curve f, measured by Δx :

$$\frac{f(x) - g(x)}{\Delta x} = \frac{f(x) - f(x_0)}{\Delta x} - s$$

should vanish in the limit x tending to x_0, i.e. $\Delta x \to 0$. But this means exactly:

$$\lim_{\Delta x \to 0} \frac{f(x) - f(x_0)}{\Delta x} - s = f'(x_0) - s = 0 \quad \text{and thus } s = f'(x_0).$$

*Therefore, we get the best **linear approximation** to the graph of the function $f(x)$ in the neighbourhood of the point x_0, if we choose a straight line with the differential quotient as gradient, and this is of course exactly the definition of the tangent.*

Insert: Differentials: *To become even more acquainted with the important concept of the differential quotient, we want to study as a further alternative access the term "differential". Here nothing really new will be obtained. We only get new insight in the concept already obtained. This insight will be useful for later applications and extensions, since it can easily be transferred to several dimensions and will only there display its full power:*

We avoid *deliberately the* limit *and write*

the difference quotient as an equation: $\frac{\Delta f(x)}{\Delta x}\Big|_{x_0} = f'(x_0) + Rf(x_0, \Delta x)$

with a remainder $R_f(x_0, \Delta x)$, which depends on the function f, the point x_0 and the interval Δx, and vanishes with Δx. If we multiply this difference quotient with the increase of the variable Δx, we get

the **"real increase"** *of our function at x_0:* $\Delta f(x)\big|_{x_0} = f'(x_0)\Delta x + r_f(x_0, \Delta x)$

with the new remainder $r_f(x_0, \Delta x) = R_f(x_0, \Delta x)\Delta x$, which apparently vanishes even more strongly than R_f does with Δx. If we can ignore this remainder, we find for the real increase of the function $\Delta f(x)\big|_{x_0}$ a first approximation linear in Δx,

the **"linear part of the increase of the function"** $df(x)\big|_{x_0} = f'(x_0)\Delta x$,

*which is called the "***differential***".*

In particular for the function $y = f(x) = x$, the bisector line, we get because of $f'(x) = 1$

the linear part of the linear function of the independent variable: $dx = \Delta x$,

which is not necessarily infinitesimal and which we may insert above in order to get (using the Lagrange or Leibniz form of the differential quotient):

differential:

$df(x) = f'(x)dx \equiv \left(\frac{df}{dx}\right) dx$, *linear part of the increase of the function.*

Thus we have got an equation in which the symbols df and dx, having been defined only as quotient in the Leibniz form of the differential quotient, now appear as **singles***, as "linear parts of the increase" and are* ***defined*** *also as non-infinitesimal quantities. Because of this possibility we prefer the far sighted and suggestive writing form for the differential quotient by Leibniz to that of Lagrange known to most of you from school.*

5.3 Differentiability

From the above construction of the differential quotient as limit of the difference quotient and our knowledge about the formation of limits, it follows immediately that we cannot determine a gradient for every function at every point, meaning not every function is differentiable at every point of its definition domain:

$f(x)$ **differentiable** at x_0 :
$\Longleftrightarrow$ the limit of the difference quotient exists.

For this we have to require that both the "limit from the right" $\lim_{x\to x_0^+} \frac{f(x)-f(x_0)}{x-x_0}$, for which like above in Figure 5.2 we have chosen the second point for the secant to the right of x_0, and the "limit from the left" $\lim_{x\to x_0^-} \frac{f(x)-f(x_0)}{x-x_0}$ with the second secant point x to the left of x_0 exist, meaning that they both are finite and in addition agree with each other:

> $f(x)$ **differentiable** at x_0 :
> $$\iff -\infty < \lim_{x\to x_0^+} \frac{f(x)-f(x_0)}{x-x_0} = \lim_{x\to x_0^-} \frac{f(x)-f(x_0)}{x-x_0} < \infty$$

For the graph this obviously means that "kinks" or "corners" are not allowed. For example, the continuous absolute value function $y = |x|$, which is defined everywhere on the x axis, is not differentiable at $x = 0$, because $\lim_{x\to 0^+} \frac{|x|-|0|}{x-0} = +1$, while $\lim_{x\to 0^-} \frac{|x|}{x} = -1$. Even though both limits exist, they are not equal to each other.

As another example, the root function $y = f(x) = +\sqrt{x}$ defined only on the non-negative half-line is on the left border of its domain at $x = 0$ not differentiable, because the only possible "limit from the right" $\lim_{x\to 0^+} \frac{\sqrt{x}-\sqrt{0}}{x-0} = \lim_{x\to 0^+} \frac{1}{\sqrt{x}}$ does not exist, since the gradient becomes infinite.

From the definition above we can immediately see the differentiability of a function at a certain point implies its continuity there. This is because, for a sequence x_n from the definition domain which tends to x_0, $|f(x_n)-f(x_0)| = \left|\frac{f(x_n)-f(x_0)}{x_n-x_0}\right| |x_n-x_0| \longrightarrow f'(x_0)\cdot 0 = 0$. The reverse however is **not** true: Not every continuous function is differentiable, just as we have seen in the above example of the absolute value function $f(x) = |x|$.

Insert: Can a discontinuous function be differentiable?: *Can it happen that a discontinuous function is differentiable? To answer this question we investigate our sample discontinuous function the Heaviside function $\theta(x)$ which is discontinuous at the point $x = 0$. Thus we can illustrate that we were right in not assuming continuity for the differentiability:*

We see easily that neither the "limit from the right" $\lim_{x\to 0^+} \frac{\theta(x)-\theta(0)}{x-0} = (1-\frac{1}{2}) \lim_{x\to 0^+} \frac{1}{x}$ nor the "limit from the left" $\lim_{x\to 0^-} \frac{\theta(x)-\theta(0)}{x-0} = (0-\frac{1}{2}) \lim_{x\to 0^-} \frac{1}{x}$ exist.

This problem cannot be cured by redefining the value $\theta(0) = \frac{1}{2}$, which may at first sight look quite arbitrary. For example we could try to define it by 1. In this *case the limit from the right would become 0, but the left-side limit would continue being ∞. Thus our sample discontious function is incurably* **not** *differentiable at the discontinuity point, although it has a horizontal tangent at both sides of this point.*

Exercise 5.1 Differentiability

Examine the following functions $f(x)$ for differentiability at $x = 0$:
a) x^2, b) $\frac{1}{x}$, c) $\frac{\sin x}{x}$, d) $e^{-|x|}$ and e) $\theta(x + a)$.

It is only one single theorem on differentiable functions which we will use occasionally:

Mean Value Theorem of Differential Calculus:
If a function $f(x)$ is continuous in an closed interval $[a, b]$ and differentiable in the open interval $]a, b[$, then there exists at least one point $x_0 \in]a, b[$, called "mean value", such that the gradient $f'(x_0)$ of the tangent to the graph of the function at this point is equal to the gradient of the secant over the interval:

$$\frac{f(b) - f(a)}{b - a} = f'(x_0).$$

The proof follows vividly from the following figure:

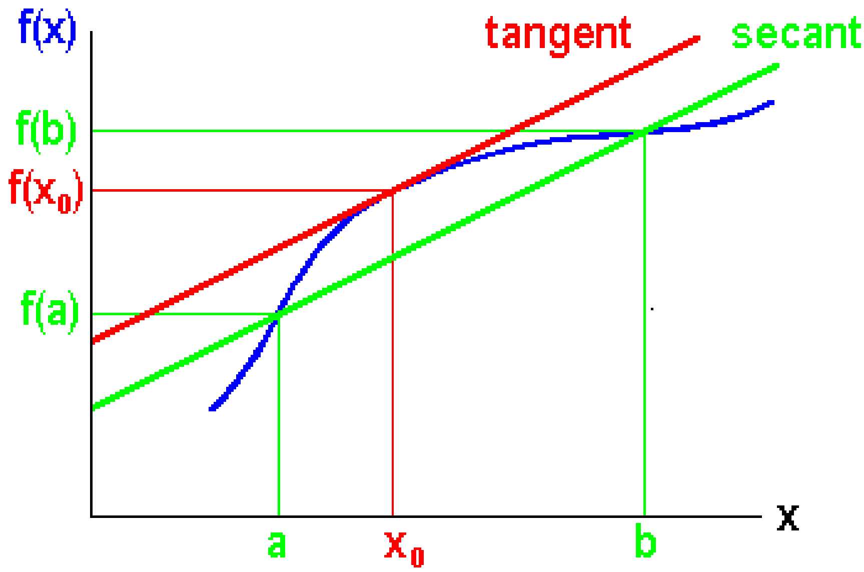


Figure 5.3: Proof of the Mean Value Theorem of Differential Calculus

5.4 Higher Derivatives

The differential quotient $f'(x)$ of a function $f(x)$ is itself a function of the independent variable x. If it too is differentiable, then we can pass from the differential quotient, the "**first derivative**" or gradient of a function, to the

> **second derivative:** $f''(x) := \frac{d}{dx}f'(x)$, i.e. $= \lim\limits_{\Delta x \to 0} \frac{f'(x_0+\Delta x)-f'(x_0)}{\Delta x}$

There are once again various ways of writing this: $f''(x) \equiv \frac{d^2 f}{dx^2} \equiv \left(\frac{d}{dx}\right)^2 f(x)$. The geometrical meaning of the second derivative as **curvature** results from the fact that the increase of the gradient, meaning a positive second derivative $f''(x) > 0$ (viewed in the positive direction of the independent variable) means a left curve, while a negative one corresponds to a right curve. If $f''(x) = 0$, we recognize that $f(x)$ is a straight line.

For physicists we get something well-known if time t is the independent variable, namely the **acceleration** as the first time derivative of the velocity or second time derivative of the space: $a = \dot{v}(t) = \ddot{x}(t)$.
Successively, we can also define for many functions even higher derivatives, in general the

> **n-th derivative:** $f^{(n)}(x) := \frac{d}{dx}f^{(n-1)}(x)$, mit $f^{(n)}(x) \equiv \frac{d^n f}{dx^n} \equiv \left(\frac{d}{dx}\right)^n f(x)$.

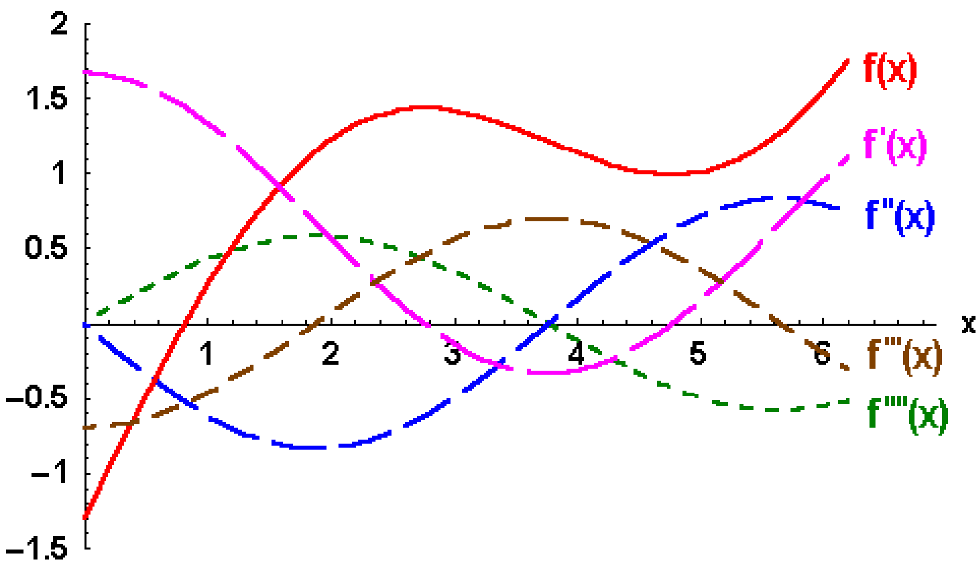


Figure 5.4: Graph of a function and its higher derivatives

Insert: Extrema: *From school many of you are familiar with this use of the first and second derivative of a function from the* ***discussion of curves****:*
The vanishing of the first derivative $f'(x_0) = 0$ at a point x_0, the criterion for a horizontal tangent in this point, is a ***necessary*** *condition for the existence of a local* ***extremum****. The condition cannot be sufficient because a horizontal turning tangent may occur.*
Sufficient *conditions for a local* ***maximum or minimum*** *can be obtained only by looking at the second derivative: $f''(x_0) > 0$ means a turn to the left, i.e. a local minimum, while $f''(x_0) < 0$ indicates a turn to the right, thus a maximum.*

Insert: Limits of Quotients: *For the calculation of limits of quotients of functions the* **Rule of de l'Hospital** *may be helpful, which states that the limit of the quotient of two differentiable functions is not changed by (also multiple) differentiation of numerator and denominator, if all concerned limits exist:*

$$\lim f(x)/g(x) = \lim f'(x)/g'(x) = \lim f^{(n)}(x)/g^{(n)}(x)$$

Z.B.: $\lim\limits_{x\to 0} \frac{\sin x}{x} = \lim\limits_{x\to 0} \frac{\cos x}{1} = 1,$
as was already shown geometrically in an insert of Section 4.9,
or $\lim\limits_{x\to 0} \frac{e^x - 1}{x} = \lim\limits_{x\to 0} \frac{e^x}{1} = 1,$
which took us much effort to prove in Exercise 4.14d,
or $\lim\limits_{n\to\infty} \frac{\ln x}{x} = \lim\limits_{n\to\infty} \frac{1/x}{1} = 0$
or $\lim\limits_{x\to 0} \frac{1-\cos x}{x^2} = \lim\limits_{x\to 0} \frac{\sin x}{2x} = \lim\limits_{x\to 0} \frac{\cos x}{2} = 1/2$
or $\lim\limits_{x\to 0} \frac{x-\sin x}{x^3} = \lim\limits_{x\to 0} \frac{1-\cos x}{3x^2} = \lim\limits_{x\to 0} \frac{\sin x}{6x} = \lim\limits_{x\to 0} \frac{\cos x}{6} = 1/6$

5.5 The Technique of Differentiation

Now we can start looking at examples, applying the general definition to specific functions which will lead us to the rules according to which the technique of differentiation works. We shall calculate the differential quotients of all important elementary functions and put them all together clearly in a **table** which will be of unexpected value for us later on.

5.5.1 Four Examples

Initially we calculate the differential quotients of **four** prominent examples out of our basic set of functions, from which we will then be able to find the derivatives of all other interesting functions with help of a few simple rules:

1. First of all we examine

the **powers** with natural number exponents $n \in \mathbb{N}$: $\quad (x^n)' = nx^{n-1} \quad (*)$

For the proof we use the binomial theorem:

$$\begin{aligned}(x^n)' &= \lim_{\Delta x \to 0} \frac{(x+\Delta x)^n - x^n}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{x^n + nx^{n-1}\Delta x + n(n-1)x^{n-2}(\Delta x)^2 + ... - x^n}{\Delta x} \\ &= \lim_{\Delta x \to 0} nx^{n-1} + n(n-1)x^{n-2}\Delta x + ... \\ &= nx^{n-1}\end{aligned}$$

The n-th powers are n-times differentiable, so that $x^{n(n)} = n(n-1)(n-2) \cdot \ldots \cdot 2 \cdot 1 = n!$.

2. Secondly we determine the differential quotient of

sine: $\quad (\sin x)' = \cos x$

The proof uses an addition theorem and a previously calculated limit:

$$\begin{aligned}\{\sin x\}' &= \lim_{\Delta x \to 0} \frac{sin(x+\Delta x) - sinx}{\Delta x} \quad \text{with help of the addition theorem} \\ &= \lim_{\Delta x \to 0} \frac{2\sin\left(\frac{x+\Delta x - x}{2}\right)\cos\left(\frac{x+\Delta x + x}{2}\right)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{\sin\left(\frac{\Delta x}{2}\right)\cos\left(x + \frac{\Delta x}{2}\right)}{\frac{\Delta x}{2}} \\ &= \lim_{\Delta x \to 0} \frac{\sin\left(\frac{\Delta x}{2}\right)}{\frac{\Delta x}{2}} \lim_{\Delta x \to 0} \cos\left(x + \frac{\Delta x}{2}\right) \\ &= \cos x.\end{aligned}$$

Entirely analogously it follows for the

cosine: $\quad (\cos x)' = -\sin x.$

3. Finally we differentiate the

exponential function: $\quad (e^x)' = e^x.$

The proof uses a limit previously calculated in Section 4.14d:

$$\begin{aligned}(e^x)' &= \lim_{\Delta x \to 0} \frac{e^{x+\Delta x} - e^x}{\Delta x} \\ &= e^x \lim_{\Delta x \to 0} \frac{e^{\Delta x} - 1}{\Delta x} \\ &= e^x\end{aligned}$$

This is the characteristic property of the exponential function and the deeper reason for its outstanding importance in science, that it is identical with its own differential quotient.

Exercise 5.2 *Derive the differential quotient of the cosine.*

From these examples we now obtain all desired differential quotients for all functions of our basic set and beyond, with the help of the following rules.

5.5.2 Simple Differentiation Rules: Basic Set of Functions

In practice we only very rarely have one of the four sample functions studied above to differentiate purely on its own. Usually we have more or less complicated expressions, composed of many various functions, like for example $f(x) = ax^n e^{-bx}(\cos cx + d \sin cx)$ with real constants a, b, c, d and n.

For this reason, we put together in this section the general rules which enable us to assemble the differential quotients of complicated expressions out of the well known derivatives of the individual parts. As sample applications we first of all check these rules using the functions of our basic set, and then beyond these further interesting functions important for science. We will arrange the results in a **TABLE** which we shall need later on also for integration.

In the following $f(x)$ and $g(x)$ are two differentiable functions and a, b, $c, \ldots$ represent real constants. You can attempt the proofs from the definition of the limit on your own, or you may have a look at the proofs in the inserts.

Due to the obvious homogeneity of the limit (a constant factor can be drawn out) $(c \cdot f(x))' = c \cdot f'(x)$, instead of the well-known **sum rule** $(f(x) \pm g(x))' = f'(x) \pm g'(x)$ we start immediately with

linearity: $(af(x) + bg(x))' = a \cdot f'(x) + b \cdot g'(x)$.

Here we see the differential quotient of a linear combination of functions is equal to the linear combination of the differential quotients.

Insert: Proof:

$$\begin{aligned}(af(x)+bg(x))' &\equiv (\tfrac{d}{dx})(af(x)+bg(x)) \\ &:= \lim_{\Delta x\to 0}\frac{af(x+\Delta x)+bg(x+\Delta x)-af(x)-bg(x)}{\Delta x} \\ &= a\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}+b\lim_{\Delta x\to 0}\frac{g(x+\Delta x)-g(x)}{\Delta x} \\ &= a(\tfrac{d}{dx})f(x)+b(\tfrac{d}{dx})g(x)=af'(x)+bg'(x)\end{aligned}$$

With this result, the differential quotient for every polynomial e.g. of m-th degree $P_m(x) = \sum_{n=0}^{m} a_n x^n$ as a polynomial of $(m-1)$-th degree follows from the power rule: $(\sum_{n=0}^{m} a_n x^n)' = \sum_{n=0}^{m} a_n n \cdot x^{n-1}$. Especially the (m+1)-th derivative vanishes: $P_m^{(m+1)}(x) = 0$.

Many of you are familiar also with the **product rule**:

product rule: $(f(x)\cdot g(x))' = f'(x)\cdot g(x) + f(x)\cdot g'(x).$

The differential quotient of the product of two differentiable functions $f(x)$ and $g(x)$ is the differential quotient of the first factor multiplied by the second factor, plus the differential quotient of the second factor multiplied by the first factor:

Insert: Proof:

$$\begin{aligned}(f(x)\cdot g(x))' &= (\tfrac{d}{dx})(f(x)g(x)) \\ &:= \lim_{\Delta x\to 0}\frac{f(x+\Delta x)g(x+\Delta x)-f(x)g(x)}{\Delta x} \\ &= \lim_{\Delta x\to 0}\frac{f(x+\Delta x)g(x+\Delta x)-f(x)g(x+\Delta x)+f(x)g(x+\Delta x)-f(x)g(x)}{\Delta x} \\ &= \lim_{\Delta x\to 0}\frac{(f(x+\Delta x)-f(x))g(x+\Delta x)}{\Delta x}+f(x)\lim_{\Delta x\to 0}\frac{g(x+\Delta x)-g(x)}{\Delta x} \\ &= (\tfrac{d}{dx})f(x)\cdot g(x)+f(x)\cdot(\tfrac{d}{dx})g(x) \\ &= f'(x)\cdot g(x)+f(x)\cdot g'(x)\end{aligned}$$

For example: $(x^2 \sin x)' = 2x \sin x + x^2 \cos x$

The next rule we need is the **inverse rule**:

$$\textbf{inverse rule:} \quad \left(\frac{1}{g(x)}\right)' = -\frac{g'(x)}{g^2(x)} \text{ for } g(x) \neq 0.$$

We obtain the differential quotient of the inverse of a differentiable non-vanishing function $g(x) \neq 0$ through division of the function's differential quotient through the negative of its square:

Insert: Proof:

$$\left(\frac{1}{g(x)}\right)' := \lim_{\Delta x \to 0} \frac{\frac{1}{g(x+\Delta x)} - \frac{1}{g(x)}}{\Delta x}$$

using the common denominator this gives

$$\begin{aligned} &= \lim_{\Delta x \to 0} \frac{g(x) - g(x+\Delta x)}{g(x)g(x+\Delta x)\Delta x} \\ &= -\lim_{\Delta x \to 0} \frac{g(x+\Delta x) - g(x)}{\Delta x g(x)g(x+\Delta x)} \\ &= -\frac{g'(x)}{g^2(x)} \end{aligned}$$

This way for example, it is possible to expand the power rule onto negative exponents, meaning integers
$(x^{-n})' = (1/x^n)' = -(x^n)'/x^{2n} = -nx^{n-1}/x^{2n} = -nx^{n-1-2n} = -nx^{-n-1}$ just like above, now however for $n \in \mathbb{Z}$.
Even the inverse exponential function can now be differentiated:
$(e^{-x})' = -e^x/(e^x)^2 = -e^{-x}$.

From this and linearity we obtain for the hyperbolic functions:
$(\sinh x)' = \frac{e^x - e^{-x}}{2}' = \frac{e^x + e^{-x}}{2} = \cosh x$ and analogously $(\cosh x)' = +\sinh x$.

The **quotient rule** follows from the product and inverse rule:

$$\textbf{quotient rule:} \quad \left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)} \text{ for } g(x) \neq 0.$$

The differential quotient of the quotient of two differentiable functions is the differential quotient of the numerator multiplied by the denominator function, minus the differential quotient of the denominator multiplied by the numerator function, both divided by the square of the denominator function which is not allowed to vanish.

Insert: Proof:

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

This is the way we can determine the differential quotients of all rational functions $R(x) = \frac{P_n(x)}{Q_m(x)}$, i.e. the quotient of two polynomials $P_n(x)$ and $Q_m(x)$.

We can now differentiate also tangent and cotangent:
$(\tan x)' = 1/\cos^2 x$ and $(\cot x)' = -1/\sin^2 x$,
and the corresponding hyperbolic functions:
$(\tanh x)' = 1/\cosh^2 x$ and $(\coth x)' = -1/\sinh^2 x$.

Insert: Proofs: *Using the quotient rule*

$$\begin{aligned}(\tan x)' &= \left(\frac{\sin x}{\cos x}\right)' = \frac{\cos x \cos x - \sin x(-\sin x)}{\cos^2 x} \\ &= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} \\ (\cot x)' &= \left(\frac{1}{\tan x}\right)' = -\frac{1}{\cos^2 x \tan^2 x} \\ &= -\frac{1}{\sin^2 x} \\ (\tanh x)' &= \left(\frac{\sinh x}{\cosh x}\right)' = \frac{\cosh^2 x - \sinh^2 x}{\cosh^2 x} \\ &= \frac{1}{\cosh^2 x} \\ (\coth x)' &= \left(\frac{1}{\tanh x}\right)' = -\frac{1}{\cosh^2 x \tanh^2 x} \\ &= -\frac{1}{\sinh^2 x} \quad \textit{for } x \neq 0.\end{aligned}$$

Now we have the differential quotients of all functions in our basic set, and enter these into a **TABLE** which summarizes all of our present knowledge:

DIFFERENTIATION TABLE			
Line	$F(x)$	$F'(x) \equiv (d/dx)F(x)$	Comments:
1	const	0	
2	x^r	rx^{r-1}	for the time being only $r \in \mathbb{Z}$
3			
4	$\sin x$	$\cos x$	-
5	$\cos x$	$-\sin x$	-
6	$\tan x$	$1/\cos^2 x$	$x \neq (z + 1/2)\pi,\ z \in \mathbb{Z}$
7	$\cot x$	$-1/\sin^2 x$	$x \neq z\pi,\ z \in \mathbb{Z}$
8	$\arcsin x$		
9	$\arccos x$		
10	$\arctan x$		
11	arccot x		
12	e^x	e^x	
13	r^x		
14	$\ln \lvert x \rvert$		
15	$\log_b \lvert x \rvert$		
16	$\sinh x$	$\cosh x$	
17	$\cosh x$	$\sinh x$	
18	$\tanh x$	$1/\cosh^2 x$	
19	$\coth x$	$-1/\sinh^2 x$	$x \neq 0$
20	arsinh x		
21	arcosh x		
22	artanh x		
23	arcoth x		

5.5.3 Chain and Inverse Function Rules

In order to differentiate the functions which occur in physics we need (according to the blanks in our preceding **TABLE**) in addition to these rules, often known from school, *two further* **differentiation rules**:

The **chain rule** helps us in the differentiation of nested functions: It gives us the differential quotient of an nested function $z = g(f(x))$ from the differential quotient of the inserted "inner" function $y = f(x)$ and the "outer" function $z = g(y)$ in which y was inserted with $W_f \subseteq D_g$. Using Leibniz's notations we obtain the product from the so-called "outer" $\frac{dz}{dy}$ and the "inner" derivative $\frac{dy}{dx}$:

chain rule: $\frac{dz(y(x))}{dx} = \frac{dz(y)}{dy} \cdot \frac{dy}{dx} = (\frac{dz}{dy}) \cdot (\frac{dy}{dx})$

in Leibniz's and in Lagrange's notation:

$$\boxed{(g(f(x)))' = g'(y)f'(x).}$$

Now, since we are familiar with the term *differential* this result may seem trivial to us, seeing how the fraction was simply reduced to higher terms by dy. Nevertheless, we want to quickly sketch the **proof** to demonstrate the advantages of the differentials, with whose help it is very simple:

First of all for the "inner" function $y = f(x) : dy = f'(x)dx + r_f$ with $\lim\limits_{\Delta x \to 0} r_f(x, \Delta x)/dx = 0$, then for the "outer" function $z = g(y) : dz = g'(y)dy + r_g$ with $\lim\limits_{\Delta y \to 0} r_g(y, \Delta y)/dy = 0$. After insertion it follows that: $dz = g'(y)(f'(x)dx + r_f) + r_g = g'(y)f'(x)dx + g'(y)r_f + r_g$, and after division by the differential dx in the limit it becomes:

$$\tfrac{dz}{dx} \equiv (\tfrac{d}{dx})g(f(x)) \equiv g'(y)f'(x) = (\tfrac{dg}{dy})(\tfrac{df}{dx}) \equiv (\tfrac{dz}{dy})(\tfrac{dy}{dx}).$$

The following **example** illustrates the advantages of the Leibniz notation: We are looking for the first derivative of $((x + 1/x)^4 - 1)^3$ for $x \neq 0$:

$$\begin{aligned}\left(((x + 1/x)^4 - 1)^3\right)' &= (\tfrac{d}{dx})w(z(y(x)))\\ &= (\tfrac{dw}{dz})(\tfrac{dz}{dy})(\tfrac{dy}{dx}) \quad \text{according to the chain rule,}\\ &= 12\left((x + 1/x)^4 - 1\right)^2 (x + 1/x)^3(1 - 1/x^2),\end{aligned}$$

$$\begin{aligned}\textit{since} \quad y = f(x) &= x + 1/x \quad \text{with } (\tfrac{dy}{dx}) = 1 - 1/x^2,\\ z = g(y) &= y^4 - 1 \quad \text{with } (\tfrac{dz}{dy}) = 4y^3 \quad \text{and}\\ w = h(z) &= z^3 \quad \text{with } (\tfrac{dw}{dz}) = 3z^2.\end{aligned}$$

A further example is the

$$\boxed{\textbf{general exponential function:} \quad (b^x)' = b^x \ln b}$$

Proof with $y := x \ln b : (b^x)' = (\frac{d}{dx})(e^{x \ln b}) = (\frac{d}{dx})e^y = (\frac{d}{dy})e^y(\frac{dy}{dx}) = e^y \ln b = b^x \ln b$.

Exercise 5.3 Chain rule:

Calculate the following differential quotients using the chain rule:
a) $(\cos x)' = (\sin(\frac{\pi}{2} - x))'$, *b)* $(\sin x^2)'$, *c)* $(\sin^2 x)'$, *d)* $(e^{-x})'$,
e) $(\exp(-x^2))'$ *and* *f)* $\left(\frac{1}{ax+b}\right)'$

Finally we need the **inverse function rule** for the differential quotient of the *inverse function* $x = f^{-1}(y)$ with $y \in W_f$ of a differentiable bi-unique function $y = f(x)$ with $x \in D_f$, whose differential quotient $f'(x) = dy/dx \neq 0$ is known and does not vanish in the whole D_f:

> **inverse function rule:** $\frac{dx}{dy} = \frac{1}{(\frac{dy}{dx})}$ for $(\frac{dy}{dx}) \neq 0.$

We want to derive this formula very simply in the Leibniz notation: In order to do this, we form the derivative of $x = f^{-1}(f(x))$ with respect to x according to the chain rule:

$$1 = (\tfrac{d}{dx})(f^{-1}(f(x))) = (\tfrac{d}{dy})(f^{-1}(y))(\tfrac{df(x)}{dx}) = \tfrac{dx}{dy}\tfrac{dy}{dx}$$

and after division by $\frac{dy}{dx} \neq 0$ we arrive at the stated result.

Equipped with these rules we are able now to calculate all desired derivatives. Most of the proofs you will find in inserts:

First of all the

> **roots:** $y = \sqrt[m]{x} = x^{\frac{1}{m}}$ for $x > 0: \ \sqrt[m]{x}\,' = (x^{\frac{1}{m}})' = (\frac{1}{m})x^{\frac{1}{m}-1}$

as inverse function of the exponential function $x = y^m$ for $y > 0$, for $\sqrt[m]{x}\,' = (x^{1/m})' = 1/(\frac{dx}{dy}) = 1/my^{m-1} = 1/m(x^{1/m})^{m-1} = (1/m)x^{1/m-1}$, i.e. our power rule (*) holds also for reciprocal integers in the exponents.

Even more generally for

> **rational powers:** $z = x^{\frac{n}{m}}$ for $x > 0 : (x^{\frac{n}{m}})' = \frac{n}{m}x^{\frac{n}{m}-1}$

Meaning, our **power rule** (*) holds true even for any rational exponents.

Insert: Proof: *with* $y = f(x) = x^{1/m}$ *in the chain rule:*

$$\begin{aligned}(x^{n/m})' &= (\frac{d}{dx})((x^{1/m})^n) = \frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx} \\ &= (\tfrac{d}{dy})y^n \cdot (\tfrac{d}{dx})x^{1/m} = ny^{n-1} \cdot (1/m)x^{1/m-1} \\ &= (n/m)(x^{1/m})^{n-1}x^{1/m-1} = (n/m)x^{n/m-1/m+1/m-1} \\ &= (n/m)x^{n/m-1}\end{aligned}$$

Then the

$$\boxed{\textbf{natural logarithm:} \quad y = \ln x \text{ for } x > 0 : (\ln x)' = \frac{1}{x} \text{ for } x \neq 0}$$

as inverse function to the exponential function $x = e^y$ for $y \in \mathbb{R}$.

Insert: Proof:
$(\ln x)' = \frac{dy}{dx} = 1/(\frac{dx}{dy}) = 1/(\frac{d}{dy})e^y = 1/e^y = 1/x$ *for $x \neq 0$. (Into the **TABLE!**)*

Even the following holds true: $(\ln |x|)' = \frac{1}{x}$.
Because $(\ln -x)' = dz/dx = dz/dy \cdot dy/dx = 1/y(-1) = -1/(-x) = 1/x$ for $x \neq 0$.

We then turn our attention to the

$$\boxed{\textbf{general power:} \quad z = x^r = e^{r \ln x} \text{ with } r \in \mathbb{R} : (x^r)' = rx^{r-1}}$$

i.e. our power rule (*) holds universally true even for any real exponent.

Insert: Proof: *With $z = e^y$ and $y = r \ln x$ in the chain rule we get: $(x^r)' = \frac{dz}{dx} = \frac{dz}{dy} \cdot \frac{dy}{dx} = e^y r/x = (r/x)e^{r \ln x} = (r/x)x^r = rx^{r-1}$. (Into the **TABLE**: L.2!)*

Even for the

$$\boxed{\textbf{general logarithm:} \quad y = \log_b x \text{ for } x > 0 : (\log_b x)' = \frac{1}{x \ln b}}$$

to any real base $b \in \mathbb{R}$, we now obtain the derivative, namely as inverse function of the general exponential function $x = b^y$:

Insert: Proof:
$(\log_b x)' = \frac{dy}{dx} = \frac{1}{(\frac{dx}{dy})} = \frac{1}{b^y \ln b} = \frac{1}{x \ln b}$. *(Into the **TABLE**: L.15!)*

We conclude this list of differential quotients, which is important also for he following chapters, with the cyclometric and the area functions:

For the cyclometric functions, the inverse functions to the trigonometric ones, we get

$$\textbf{arc tangent:} \text{ for } -\pi/2 < \arctan x < \pi/2 : (\arctan x)' = \frac{1}{1+x^2}$$

Insert: Proof: *With the inverse function* $x = \tan y$, *where from* $\frac{dx}{dy} = 1/\cos^2 y = (\cos^2 y + \sin^2 y)/\cos^2 y = 1 + \tan^2 y = 1 + x^2$ *follows:* $(\arctan x)' = \frac{dy}{dx} = 1/(\frac{dx}{dy}) = 1/(1 + x^2)$. *(Into the **TABLE**: L.10!)*

Analogously for the

$$\textbf{arc cotangent} \text{ for } 0 < \operatorname{arccot} x < \pi : \boldsymbol{(\operatorname{arccot} x)' = -\frac{1}{1+x^2}}$$

Exercise 5.4 *Prove this with the inverse function:* $x = \cot y$.

For the

$$\textbf{arc cosine} \text{ for } -\pi/2 < \arcsin x < \pi/2 : (\arcsin x)' = \frac{1}{\sqrt{1-x^2}} \text{ for } |x| < 1$$

Insert: Proof: *With the inverse function* $x = \sin y$, *where from* $\frac{dx}{dy} = \cos y = \sqrt{(1 - \sin^2 y)} = \sqrt{(1 - x^2)}$ *for* $|x| < 1$ *follows:* $(\arcsin x)' = \frac{dy}{dx} = 1/(\frac{dx}{dy}) = 1/\sqrt{(1 - x^2)}$. *(Into the **TABLE**: L.8!)*

Analogously for

$$\textbf{arc cosine} \text{ for } 0 < \arccos x < \pi : (\arccos x)' = -\frac{1}{\sqrt{(1-x^2)}} \text{ for } |x| < 1$$

Exercise 5.5 *Prove this with the inverse function:* $x = \cos y$.

The **area functions** the inverse functions of the hyperbolic functions, complete our differentiation table:

For the

$$\textbf{area hyperbolic tangent:} \quad (\text{artanh } x)' = \frac{1}{1-x^2} \text{for } |x| < 1$$

and the

$$\textbf{area hyperbolic cotangent:} \ (\text{arcoth } x)' = -\frac{1}{x^2-1} \text{ for } |x| > 1.$$

Exercise 5.6 *Prove this with the inverse function:* $x = \tanh y$, *respectively with* $x = \coth y$.

For the

$$\textbf{area hyperbolic sine:} \ (\text{arsinh } x)' = \frac{1}{\sqrt{(1+x^2)}} \text{ for } x \in \mathbb{R}.$$

and

area hyperbolic cosine: $0 < \text{arcosh } x$:

$$(\text{arcosh } x)' = \frac{1}{\sqrt{(x^2-1)}} \text{ for } x \geq 1.$$

Exercise 5.7 *Prove this with the inverse function:* $x = \sinh y$, *respectively with* $x = \cosh y \ \geq 1$, *bi-unique only for* $y > 0$.

You will find all the preceding results combined below results in the big **differentiation table**, to which we will return very often later on:

DIFFERENTIATION TABLE			
Line	$F(x)$	$F'(x) \equiv (d/dx)F(x)$	Comments:
1	const	0	
2	x^r	rx^{r-1}	$r \in \mathbb{R}$
3			
4	$\sin x$	$\cos x$	
5	$\cos x$	$-\sin x$	
6	$\tan x$	$1/\cos^2 x$	$x \neq (z+1/2)\pi,\ z \in \mathbb{Z}$
7	$\cot x$	$-1/\sin^2 x$	$x \neq z\pi,\ z \in \mathbb{Z}$
8	$-\pi/2 < \arcsin x < \pi/2$	$1/\sqrt{(1-x^2)}$	$\|x\| < 1$
9	$0 < \arccos x < \pi$	$-1/\sqrt{(1-x^2)}$	$\|x\| < 1$
10	$-\pi/2 < \arctan x < \pi/2$	$1/(1+x^2)$	
11	$0 < \text{arccot}\, x < \pi$	$-1/(1+x^2)$	
12	e^x	e^x	
13	r^x	$r^x \ln r$	$0 < r \in \mathbb{R}$
14	$\ln \|x\|$	$1/x$	$x \neq 0$
15	$\log_b \|x\|$	$1/x \ln b$	$x \neq 0,\ 0 < b \in \mathbb{R},\ b \neq 1$
16	$\sinh x$	$\cosh x$	
17	$\cosh x$	$\sinh x$	
18	$\tanh x$	$1/\cosh^2 x$	
19	$\coth x$	$-1/\sinh^2 x$	$x \neq 0$
20	$\text{arsinh}\, x$	$1/\sqrt{(x^2+1)}$	
21	$0 < \text{arcosh}\, x$	$1/\sqrt{(x^2-1)}$	$x > 1$
22	$\text{artanh}\, x$	$1/(1-x^2)$	$\|x\| < 1$
23	$\text{arcoth}\, x$	$-1/(x^2-1)$	$\|x\| > 1$

Exercise 5.8 Differentiation examples

Determine the differential quotients for the following functions $y = f(x)$ with constants a, b, c and d:
a) $y = \sin^3(4x)$, *b)* $y = \exp(-(x/a)^2)$, *c)* $y = \frac{1}{\sqrt{ax^2+b}}$, *d)* $y = \ln(3e^{2x})$,
e) $y = a\cosh\frac{x-b}{a}$, *f)* $y = ax^2\exp(-bx)$, *g)* $y = \cos(ax+b)\sin(cx+d)$,
h) $y = \frac{1}{1+(x/a)^2}$, *i)* $y = \left(\frac{\sin(x/a)}{(x/a)}\right)^2$, *j)* $y = \arctan(1/x) + (x/2)\left(\ln x^2 - \ln(x^2+1)\right)$

Calculate the first five derivatives of the following functions $f(x)$ which we will need in the next chapter:
k) $f(x) = \sin x$, *l)* $f(x) = \tan x$, *m)* $f(x) = e^x$ *and* *n)* $f(x) = \frac{1}{1-x^2}$

5.6 Numerical Differentiation

In some cases we cannot or do not want to calculate the derivative of a function analytically according to the rules from the last section. This is the case for example when we do not know an analytical form for the graph of a function. Then we have to rely on **numerical differentiation**.

Our definition formula of the differential quotient from Section 5.2:

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

can also be used for the numerical calculation of the derivative. For the point $(x_0, f(x_0))$, we look for neighbouring points $(x_0 + \Delta x, f(x_0 + \Delta x))$, calculate the difference quotient and let Δx tend to 0.

Since computer numbers are stored only with a certain accuracy (for example with 8 decimal places), rounding errors occur in forming the differences $f(x_0+\Delta x)-f(x_0)$ which, even though these rounding errors always remain of the same order, increase relative to the value of the difference more and more. Thus the difference quotient approximates $f'(x_0)$ increasingly accurately with decreasing Δx, but eventually the deviation begins to dominate with decreasing Δx due to the numerical roundings. Therefore, we must come to a compromise. We find a better approximation if

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0 - \Delta x)}{2\Delta x}$$

With the help of the Taylor series (which we will examine in the next chapter) we can show that the error is reduced through the symmetrization of $\Delta x f''(x_0)/2!$ to the order of $(\Delta x)^2 f'''(x_0)/3!$. Thus the error becomes smaller with Δx quadratically.

5.7 Preview of Differential Equations

To motivate you for this chapter about differentiation we mentioned at the outset that many laws of nature can be expressed in the form of differential equations. What are differential equations?

A **differential equation** is a relation between a **function** $f(x)$ **of interest and its differential quotients** $f'(x)$, $f''(x)$, ...

When e.g. we fasten a weight to a spiral spring hanging down from the ceiling and call the amount by which the spring is stretched $x(t)$, where t means the time, Newton's second Newtonian law yields, if we neglect air resistance, the following differential equation:

$$\ddot{x}(t) + \omega^2 x(t) = 0 \quad \text{with a constant } \omega.$$

Like most of the differential equations of physics this one is of "*second order*", i.e. the highest occurring differential quotient is the **second derivative** of the desired function. While in a "normal" equation for one unknown variable, for instance $x^2 - 1 = 0$, the numerical solutions x, like $x = \pm 1$ are the goal, in the differential equation above functions of the time variable $x(t)$ are what one seeks. We can easily see that $x(t) = \sin\omega t$ is a solution, since $\dot{x}(t) = \omega\cos\omega t$ and consequently $\ddot{x}(t) = -\omega^2\sin\omega t$. But is this the only solution? You surely will be dealing very extensively with questions of this kind.

Exercise 5.9 Physical differentiation

Determine the first $\dot{x}(t)$ derivative and second $\ddot{x}(t)$ derivative of the following functions $x(t)$ of time t with the constants x_0, v_0, g, ω, ω_0, γ, ρ, b_0, w, m_0 and μ:
The comparison of $\ddot{x}(t)$ with combinations of $x(t)$ and $\dot{x}(t)$ will lead you to "differential equations". Do you recognize the physical systems described by these differential equations? What is the physical meaning of the constants involved?

a) $x(t) = x_0 + v_0 t$

b) $x(t) = x_0 + v_0 t - gt^2/2$

c) $x(t) = x_0\cos\omega t + \frac{v_0}{\omega}\sin\omega t$

d) $x(t) = x_0 + \frac{v_0}{\rho}(1 - e^{-\rho t})$

e) $x(t) = x_0 - \frac{gt}{\rho} + \frac{v_0 + g/\rho}{\rho}(1 - e^{-\rho t})$

f) $x(t) = -\frac{1}{r}\ln\cosh(t\sqrt{gr})$

g) $x(t) = x_0\cosh\gamma t + (v_0/\gamma)\sinh\gamma t$

h) $x(t) = e^{-\rho t}\left(x_0\cos t\sqrt{\omega^2 - \rho^2} + \frac{v_0 + \rho x_0}{\sqrt{\omega^2 - \rho^2}}\sin t\sqrt{\omega^2 - \rho^2}\right)$

i) $x(t) = e^{-\rho t}\left(x_0\cosh t\sqrt{\rho^2 - \omega^2} + \frac{v_0 + \rho x_0}{\sqrt{\rho^2 - \omega^2}}\sinh t\sqrt{\rho^2 - \omega^2}\right)$

j) $x(t) = \frac{b_0}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\omega^2\rho^2}}\cos\left(\omega t - \arctan\frac{2\omega\rho}{\omega_0^2 - \omega^2}\right)$

k) $x(t) = x_0\tanh(\omega t)$

l) $x(t) = \frac{wm_0}{\mu}(1 - \frac{\mu t}{m_0})\ln(1 - \frac{\mu t}{m_0}) - gt^2/2 + wt$

Later on, in treating functions of several variables you will meet even more complicated differential operations: with help of the so-called "partial" derivatives you will form gradients of scalar fields as well as the divergence or rotation of vector fields. Whenever the calculation of numbers is concerned, you will need nothing else than what we have learned here.

Exercise 5.10 Partial derivatives

The change of functions of several variables, for instance fields of the three position variables x_1, x_2, x_3, you will later describe through so-called partial differential quotients, in which only one of the variables , e.g. x_1, is changed and all others are fixed: in this case $x_2 = const$ and $x_3 = const$. Without any further insight into the deeper background of the "partial" derivatives (for which a new notation Nabla: $\bigtriangledown_1$ must be introduced) given the things learned above you are already now able to calculate these "partial" derivatives. To do this you handle the fixed variables exactly as you do physical constants. Calculate for instance:

a) $\frac{d}{dx_1}(x_1 + x_2 + x_3)$ *b)* $\frac{d}{dx_1}(x_1^2 + x_2^2 + x_3^2)$ *c)* $\frac{d}{dx_1}(x_1 x_2 x_3)$

d) $\lim\limits_{x_1 \to 0} \dfrac{d}{dx_1}\left(\dfrac{2x_1 x_2}{x_1^2 + x_2^2}\right)$ *and* *e)* $\dfrac{d}{dx_1}\left(\dfrac{1}{\sqrt{x_1^2 + x_2^2 + x_3^2}}\right)$

Chapter 6

TAYLOR SERIES

Here we want to examine one of the principal applications of differential calculus, and this rather closely. Because far too little time is spent on this subject in schools, even though this technology is vital to natural scientists. We are going to talk about Taylor series which allow us to represent and calculate a large number of functions $f(x)$ needed in physics, in the neighbourhood of some value x_0 of the independent variable x, in terms of a power series.

6.1 Power Series

By far the simplest functions are the powers x^n with natural number exponents $n \in \mathbb{N}$, and the polynomials $P_m(x) = \sum_{n=0}^{m} a_n x^n$ formed through addition and multiplication of those, whose function values can quickly be calculated. Also the infinite **power series** $\sum_{n=0}^{\infty} a_n x^n$, being limits of polynomial partial sums, are relatively easy as long as the are absolutely convergent, compared to the huge variety of elementary functions studied by us. Power series can be added or subtracted and absolutely convergent ones can also be multiplied, divided and even differentiated term by term (and integrated as we will discover later). It would be great if we only needed to concern ourselves with such power series!

We shall see, this dream can become reality to a certain extent. Differential quotients are the key to this eldorado.

6.2 Geometric Series as Model

First of all let us examine once again the simplest of all polynomials, the geometric sum $G_m(x) := \sum_{n=0}^{m} x^n$ with coefficients $a_n = 1$ for all non-negative integers n and the accompanying powers series, the **geometric series**:

$$\textbf{geometric series:} \quad G_\infty(x) := \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + x^4 + ... = \frac{1}{1-x} \quad \text{for} \quad |x| < 1$$

Here we already have a prototype for our dream: the rational function $1/(1-x)$ is displayed in the open interval $]-1, 1[$ around the point $x_0 = 0$ in terms of a power series, namely the geometric series $\sum_{n=0}^{\infty} x^n$, meaning it is approximated by the sum of the constant "one", the bisector line of the first and third sector, the standard parabola, a function of third degree, and so on. Admittedly this series has an infinite number of terms, but it can be calculated using only multiplication and addition, and depending on the demand for accuracy, a few terms may be already sufficient. This representation however only exists in the interval $]-1, 1[$, while the function $1/(1-x)$ is defined everywhere other than for $x = 1$.

This example encourages us to ask the following **questions**:

1. Are there power series also for other functions which can represent these functions in certain intervals?
2. How can we obtain the coefficients in the series a_n ?
3. How many different series exist for a function?
4. How good is the convergence, respectively how large is the approximation error, if we break off the series?

6.3 Form and Non-ambiguity

Before we deal with the question of existence, we want to turn our attention to the questions 2 and 3: In order to get information about the characteristics of the desired series, we want to **assume** for the time being, that we already have found a suitable power series:

$$f(x) = \sum_{n=0}^{\infty} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + ...$$

with $f(0) = a_0$ which represents the function $f(x)$ of interest in an interval, e.g. for $|x| < R$ around the origin. Since all functions of our basic set are infinitely times differentiable, we are able to calculate the derivatives of the power series one after another:

$$\begin{aligned}
f'(x) &= \sum_{n=1}^{\infty} n a_n x^{n-1} = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \dots \\
&\text{with } f'(0) = a_1, \text{ thus } a_1 = \frac{f'(0)}{1!} \\
f''(x) &= \sum_{n=2}^{\infty} n(n-1) a_n x^{n-2} = 2a_2 + 3 \cdot 2a_3 x + 4 \cdot 3a_4 x^2 + \dots \\
&\text{with } f''(0) = 2a_2, \text{ thus } a_2 = \frac{f''(0)}{2!} \\
f'''(x) &= \sum_{n=3}^{\infty} n(n-1)(n-2) a_n x^{n-3} = 3! a_3 + 4! a_4 x + \dots \\
&\text{with } f'''(0) = 3!\, a_3, \text{ thus } a_3 = \frac{f'''(0)}{3!}
\end{aligned}$$

In general:

$$\begin{aligned}
f^{(n)}(x) &= n!\, a_n + (n+1)!\, a_{n+1} x + (n+2)! a_{n+2} x^2 + \dots \\
&\text{with } f^{(n)}(0) = n!\, a_n, \text{ thus } a_n = \frac{f^{(n)}(0)}{n!}.
\end{aligned}$$

In this way we find the desired coefficients a_n from the derivatives $f^{(n)}(0)$ of the function, evaluated at the expansion point $x_0 = 0$, to be represented. If then a power series representation of our function exists, it has the following **form** and we call it:

TAYLOR SERIES: $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n$

According to the given construction the coefficients are in addition **unambiguous**, so that we have answered question 3, too.

Our calculations also show us that the function to be represented **necessarily** must be **infinitely times differentiable** for the Taylor series to exist. That the necessary precondition is **not sufficient** for the existence of a Taylor series is shown through the following counter example: The function $f(x) = exp(-1/x^2)$ for $x \neq 0$ and $f(0) = 0$. Although this function is infinitely times differentiable, all its derivatives $f^{(n)}(0) = 0$ vanish at the point $x = 0$, so that no Taylor series around 0 can be constructed.

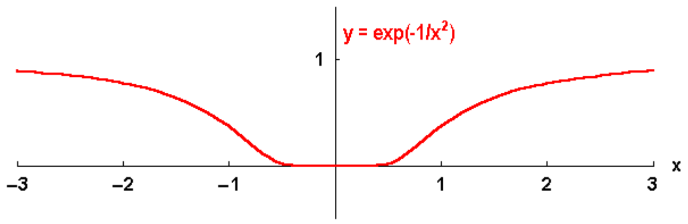


Figure 6.1: Graph of the function $f(x) = exp(-1/x^2)$ for $x \neq 0$ and $f(0) = 0$

Exercise 6.1 *As a consistency test, calculate the Taylor series of our model, the geometric series for* $|x| < 1$.

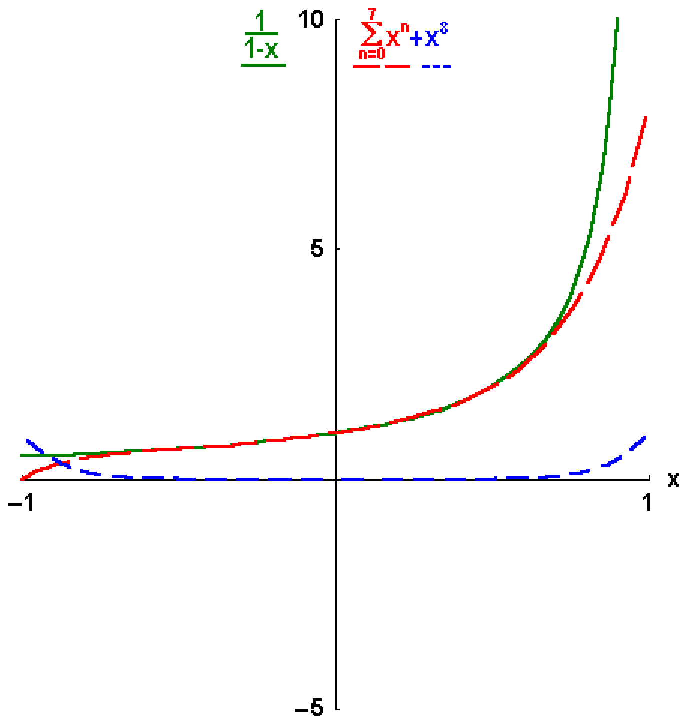


Figure 6.2: The geometric series as a Taylor series of the rational function

6.4 Examples from the Basic Set of Functions

First of all we consider some examples from our basic set of functions:

6.4.1 Rational Functions

The functions x^n with natural number exponents $n \in \mathbb{N}$ are special cases of power series with one single term only. Powers of x with negative exponents are not at all defined for $x = 0$.

However, the **general binomial series** with real exponents $r \in \mathbb{R}$ can be developed:

$$f(x) = (1+x)^r = \sum_{n=0}^{\infty} \binom{r}{n} x^n = 1 + \frac{r}{1!}x + \frac{r(r-1)}{2!}x^2 + \frac{r(r-1)(r-2)}{3!}x^3 + \dots$$

with the generalized binomial coefficient $\binom{r}{n} := \frac{r(r-1)(r-2)\dots(r-n+1)}{n!}$ and $\binom{r}{0} := 1$.

As proof:

$$f'(x) = r(1+x)^{r-1}, \qquad f''(x) = r(r-1)(1+x)^{r-2},$$

$$f'''(x) = r(r-1)(r-2)(1+x)^{r-3}, \qquad \dots \text{ etc.}$$

generally $f^{(n)}(x) = r(r-1)(r-2)\dots(r-n+1)(1+x)^{r-n}$

with

$$f(0) = 1, \qquad f'(0) = r,$$

$$f''(0) = r(r-1), \qquad f'''(0) = r(r-1)(r-2), \ \dots \text{ etc.}$$

generally $f^{(n)}(0) = r(r-1)...(r-n+1)$

altogether:

$$(1+x)^r = 1 + \frac{r}{1!}x + \frac{r(r-1)}{2!}x^2 + \frac{r(r-1)(r-2)}{3!}x^3 + \dots$$

$$= 1 + \sum_{n=1}^{\infty} r(r-1)(r-2)...(r-n+1)/n! \cdot x^n = \sum_{n=0}^{\infty} \binom{r}{n} x^n.$$

Some *special cases* are of particular importance:

First we recover for natural $r = n \in \mathbb{N}$ our previously derived *binomial formula* for the special case $a = 1$ and $b = x$, since the power series breaks off in case of natural number exponents:

$$(1+x)^n = \sum_{n=0}^{m} \binom{m}{n} x^n$$

For negative $r \in \mathbb{Z}$, e.g. for $r = -1$ the alternating *geometric series* results once more

$$\frac{1}{(1+x)} = 1 - x + x^2 - x^3 \pm \ldots = \sum_{n=0}^{\infty} (-1)^n x^n,$$

and for $r = -2$ its negative derivative:

$$\frac{1}{(1+x)^2} = 1 - 2x + 3x^2 - 4x^3 \pm \ldots = \sum_{n=0}^{\infty} (-1)^n (n+1) x^n.$$

For fractional $r \in \mathbb{Q}$, e.g. $r = 1/2$ or $-1/2$ we get the frequently needed series of the square root in the numerator or denominator

$$\sqrt{1+x} = 1 + (1/2)x - (1/8)x^2 + (1/16)x^3 - (5/128)x^4 \pm \ldots \text{ resp.}$$
$$1/\sqrt{(1+x)} = 1 - (1/2)x + (3/8)x^2 - (5/16)x^3 + (35/128)x^4 \mp \ldots$$

Exercise 6.2 *Calculate the Taylor series of* $(1+x)^r$ *for* $r = -3$, $1/3$ *and* $-1/3$.

6.4.2 Trigonometric Functions

As the next example we chose one of the trigonometric functions, namely the **sine**:

$$f(x) = \sin x = \sum_{n=0}^{\infty} (-1)^n x^{2n+1}/(2n+1)! = x - x^3/3! + x^5/5! - x^7/7! \pm \ldots$$

As proof:

$$f'(x) = \cos x, \qquad f''(x) = -\sin x,$$
$$f'''(x) = -\cos x, \qquad f^{(4)}(x) = \sin x, \ \ldots$$

with

$$f(0) = 0, \qquad f'(0) = 1,$$
$$f''(0) = 0, \qquad f'''(0) = -1,$$
$$f^{(4)}(0) = 0, \qquad \ldots$$

altogether: $\sin x = \sum_{n=0}^{\infty} (-1)^n x^{2n+1}/(2n+1)!$

Insert: : *From this we see once more that*

$$\lim_{x\to 0} \frac{\sin x}{x} = \lim_{x\to 0} \frac{x - x^3/\,3! + x^5/\,5! \mp \ldots}{x} = \lim_{x\to 0} (1 - x^2/\,3! + x^4/\,5! \mp \ldots) = 1$$

what we earlier have proved according to de l'Hospital.

Analogously, we find for the **cosine**:

$$f(x) = \cos x = \sum_{n=0}^{\infty} (-1)^n x^{2n}/(2n)! = 1 - x^2/\,2! + x^4/\,4! - x^6/\,6! \pm \ldots$$

Exercise 6.3 *Prove the above Taylor series for the cosine function.*

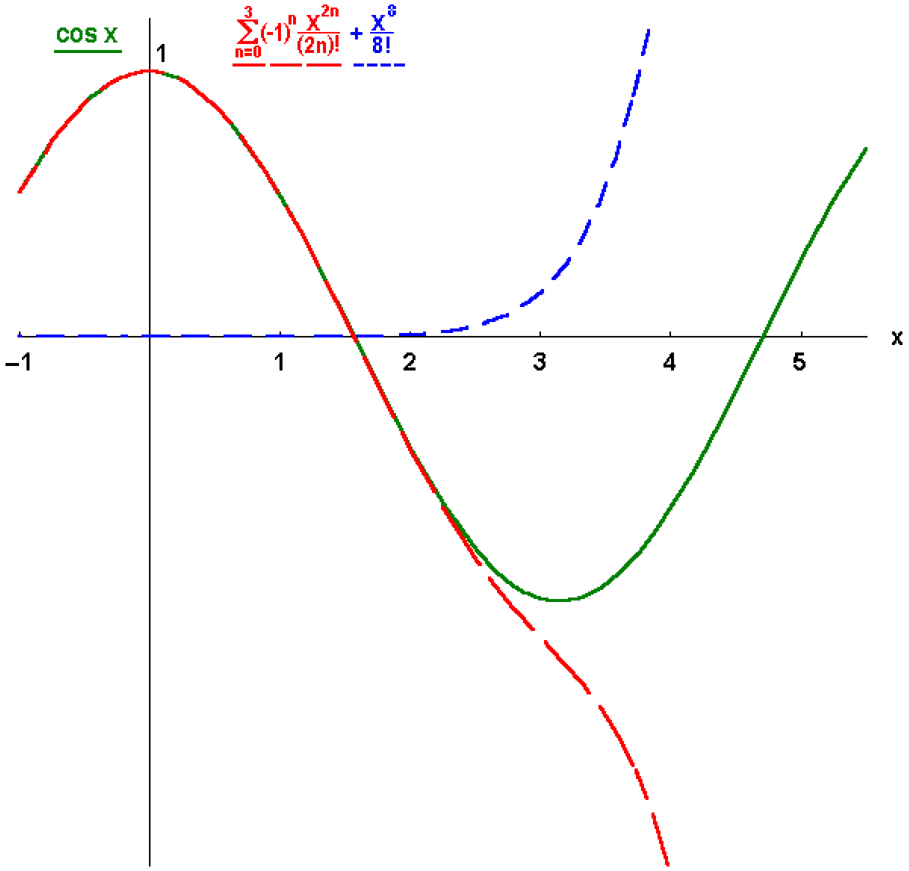


Figure 6.3: The Taylor series for the cosine function

6.4.3 Exponential Functions

An especially easy calculation is the series of the **natural exponential function:**

$$f(x) = e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + x^2/2 + x^3/6 + x^4/24 + \ldots$$

since $f'(x) = \ldots = f^{(n)}(x) = e^x$, mit $f(0) = f^{(n)}(0) = 1$.

In particular, for $x = 1$ we get $e = \sum_{n=0}^{\infty} \frac{1}{n!}$, the series through which we have defined the number e.

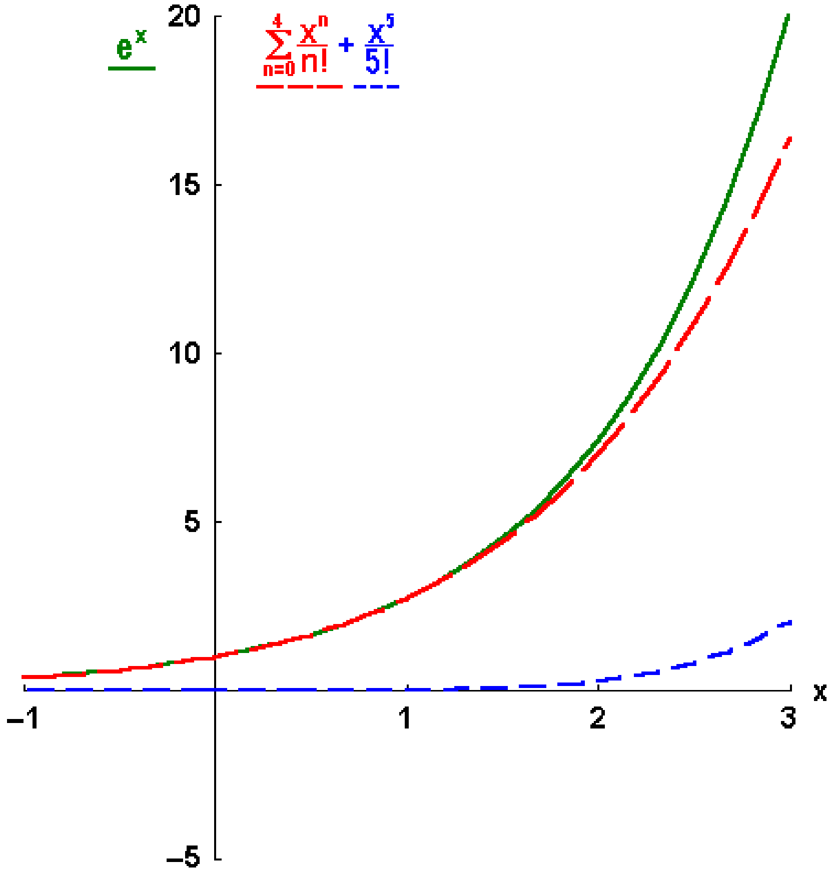


Figure 6.4: The Taylor series for the exponential function

Insert: : *From this series we can easily calculate once more the limit*

$$\lim_{x\to 0}(e^x - 1)/x = \lim_{x\to 0}(1 + x + x^2/2 + x^3/6 + \ldots - 1)/x$$
$$= \lim_{x\to 0}(1 + x/2 + x^2/6 + \ldots) = 1$$

which had made us some trouble earlier in Exercise 4.14d.

The inverse function of the natural exponential function, the **natural logarithm**, can not be expanded around $x = 0$, since $\lim_{x \to 0} \ln x = -\infty$. There is however a series for

$$\boxed{f(x) = \ln(x+1) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n} = x - x^2/2 + x^3/3 \mp \ldots .}$$

Exercise 6.4 *Prove this Taylor series for* $\ln(x+1)$.

In the following table we have put together for you the **first two terms** of the Taylor series for some common functions, to make memorization easier.

$f(x)$	$f(0) + xf'(0)$
$(1+x)^r$	$1 + r \cdot x$
$\sin x$	x
$\cos x$	1
$\exp x$	$1 + x$
$\ln(1+x)$	x

6.4.4 Further Taylor Series

From these few Taylor series we easily obtain a large number of further series, if we keep in mind what we have learned in the past about calculations with series. As an example of a **linear combination of two Taylor series** we calculate the series for the **hyperbolic sine**:

$$\begin{aligned} f(x) = \sinh x = \frac{e^x - e^{-x}}{2} &= \sum_{n=0}^{\infty} (1 - (-1)^n) \frac{x^n}{2n!} \\ &= \sum_{n=0}^{\infty} \frac{x^{2n+1}}{(2n+1)!} = x + x^3/3! + x^5/5! + x^7/7! + \ldots \end{aligned}$$

Surprisingly enough, this is exactly the Taylor series of the trigonometric sine, but **without** the change of signs, which sheds some light onto the nomenclature.

A further example shows how we can find the Taylor series of the **product of two functions** from the Taylor series of the factors, by simply multiplying the two series together, and sorting the result according to the powers:

$$\begin{aligned} f(x) = e^x \sin x &= \left(\sum_{n=0}^{\infty} \frac{x^n}{n!} \right) \left(\sum_{m=0}^{\infty} (-1)^m \frac{x^{2m+1}}{(2m+1)!} \right) \\ &= (1 + x + x^2/2 + x^3/6 + x^4/24 + \ldots)(x - x^3/3! + x^5/5! - x^7/7! \pm \ldots) \\ &= x + x^2 + (3-1)x^3/3! + (1-1)x^4/3! + (1 - 10 + 5)x^5/5! + \ldots \\ &= x + x^2 + x^3/3! - x^5/30 + \ldots \end{aligned}$$

Even with nested functions, whose Taylor series of inner and outer functions are known to us, it is often easier to insert these in each other than to directly calculate the differential quotients: For instance

$$\begin{aligned} f(x) = \exp(\sin x) &= \sum_{n=0}^{\infty}\left(\sum_{m=0}^{\infty}(-1)^m \frac{x^{2m+1}}{(2m+1)!}\right)^n /n! \\ &= 1 + (x - x^3/3! + \ldots) + (x - x^3/3! + \ldots)^2/2 + (x - x^3/3! + \ldots)^3/3! + \ldots \\ &= 1 + x + x^2/2 + (1-1)x^3/3! + (1-4)x^4/4! + \ldots \\ &= 1 + x + x^2/2 - x^4/8 + \ldots \end{aligned}$$

Exercise 6.5 *Calculate the Taylor series of the following functions*

a) the hyperbolic cosine: $\cosh x$,
b) the bell-shaped Gauss function : $\exp(-x^2)$,
c) of $1/(1-x)^2$ *through termwise differentiation of the geometric series.*

Exercise 6.6 *Calculate the first four terms of the Taylor series*

a) of $\tan x$ *through division of the series*
b) of the product $e^x \sin x$ *directly trough calculation of the derivatives*
c) of the nested function $\exp(\sin x)$ *likewise directly.*

6.5 Convergence Radius

Already with our model, the geometric series, the validity of the series development was limited to the interval $|x| < 1$ around the origin. Also for the other Taylor series, even if the function to be represented is infinitely often differentiable in a closed interval (i.e. including the boundary points), the convergence is generally limited to the inner of an interval symmetrical about the origin: $|x| < R$. The number R is called "**convergence radius**". The term "radius", however, is understandable only in the theory of functions, i.e. with power series of complex numbers. Within this symmetrical convergence region limited by R, all Taylor Series converge absolutely. Outside, i.e. for $|x| > R$, they are divergent. The convergence at both the boundary points must be examined separately in every individual case. Mathematicians provide us (through comparison with for example the geometric series) with methods for determining the convergence radius.

We would like to report here only one of these **sufficient conditions for the absolute convergence** of a series $f(x) = \sum_{n=0}^{\infty} a_n x^n$, namely D'Alembert's **quotient-criterion**, using which the radius can be written as follows:

$$R = \lim_{n\to\infty} \left|\frac{a_n}{a_{n+1}}\right| = \lim_{n\to\infty} \left|\frac{f^{(n)}(0)\,(n+1)!}{n!\,f^{(n+1)}(0)}\right|.$$

We find then for example for the general binomial series:

$$R = \lim_{n\to\infty} \left|\frac{r(r-1)\dots(r-n+1)(n+1)!}{n!r(r-1)\dots(r-n-1+1)}\right| = \lim_{n\to\infty} \left|\frac{(n+1)}{(r-n)}\right| = |-1| = 1,$$

however, for the trigonometric and the hyperbolic sine:

$$R = R[\sinh x; 0] = \lim_{n\to\infty} \left|\frac{(2(n+1)+1)!}{(2n+1)!}\right| = \lim_{n\to\infty} (2n+2)(2n+3) = \infty$$

as well as for the exponential function:

$$R = \lim_{n\to\infty} \left|\frac{(n+1)!}{n!}\right| = \lim_{n\to\infty} (n+1) = \infty$$

meaning the whole real axis is the convergence region.

Exercise 6.7 *What can you say about the convergence radius for the following Taylor Series around* $x_0 = 0$*: a)* $\cos x$ *and* $\cosh x$*, b)* $\frac{1}{1-3x}$*, c)* $\ln(1+x)$ *and d)* $\tan x$*?*

6.6 Accurate Rules for Inaccurate Calculations

Even if physics is the outstanding example of an exact science, approximations are an everyday occurrence. The crucial thing for accurate science is that we are able to give reasons for every approximation, we can put each approximation into practice and we can control its precision.

In physics often not meaningful to calculate more accurately than the experimental measurements. Also in mathematics it is sometimes sufficient to calculate the values of a function $f(x)$ from the Taylor series only to some order m, meaning to keep only the first m terms. We formulate this in the following manner:

$$f(x) = \sum_{n=0}^{m} \frac{f^{(n)}(0)}{n!} x^n + O(x^{m+1}).$$

Here $O(x^{m+1})$ means that the neglected terms are at least of order x^{m+1} , i.e. they contain $m + 1$ or more factors x.

In this section we want to put together the rules for doing approximate calculations. In this case the approximation consists in consistently taking into account all terms up to the order x^m. Which rules follow from this recipe can be displayed most easily for the frequent case $m = 1$, i.e. the series is broken off already after the second term:

$$f(x) = f(0) + x \cdot f'(0) + O(x^2) = f_0 + x \cdot f'_0 + O(x^2)$$

with the following abbreviations $f_0 = f(0)$ and $f'_0 = f'(0)$.

We consider a **second function** with analogous notations:

$$g(x) = g_0 + x \cdot g'_0 + O(x^2).$$

Please realize that we use the same term $O(x^2)$ in the series for both $f(x)$ and $g(x)$: $O(x^2)$ stands not for a certain numerical value, but is only a *symbolic way of writing* what has been omitted.

For the *product of the two functions* $f(x)$ and $g(x)$ we get then:

$$f(x) \cdot g(x) = f_0 g_0 + x(f'_0 g_0 + f_0 g'_0) + x^2 f'_0 g'_0 + (f_0 + x \cdot f'_0 + g_0 + x \cdot g'_0) O(x^2)$$

Here the first three terms are well defined, in the last term however concrete terms are multiplied with the symbol $O(x^2)$. What does this mean?

We may write $O(x^2) = \sum_{n=2}^{m} a_n x^n$ with some coefficients a_n. Then the first part of the term of interest can be written

$$f_0 O(x^2) = \sum_{n=2}^{m} (f_0 a_n) x^n = \sum_{n=2}^{m} b_n x^n = O(x^2),$$

since this series starts again with x^2, although with the coefficients $b_n = f_0 a_n$. Furthermore:

$$x f'_0 O(x^2) = \sum_{n=2}^{m} (f'_0 a_n) x^{n+1} = \sum_{n=3}^{m} c_n x^n = O(x^3),$$

because this series starts with a term proportional to x^3. For the sum we get

$$(f_0 + x f'_0) O(x^2) = O(x^2) + O(x^3) = O(x^2),$$

since the lowest (and in the neighbourhood of the origin dominant) power is x^2. If we identify also the third term in the expression for the product $f(x)g(x)$ as of the order x^2 and add it to the rest, we receive altogether:

$$\boxed{f(x) \cdot g(x) = f_0 g_0 + x(f_0' g_0 + f_0 g_0') + O(x^2).}$$

We would have gotten also the same expression if we had calculated the Taylor series for the product function $F(x) := f(x)g(x)$

$$F(x) = F(0) + xF'(0) + O(x^2),$$

where $F(0) = f_0 g_0$ and the product rule of differentiation would have lead us to $F'(0) = f_0' g_0 + f_0 g_0'$. One additional *warning*: Over-eager people might think that we should have included in the approximate expression for $f(x)g(x)$ at least the term $x^2 f_0' g_0'$ which could have easily been calculated. That would however be **inconsistent**, since other terms of the same order have been neglected.

For the **r-th power of an arbitrary function** $f(x)$ we get:

$$\boxed{f^r(x) = f_0^r (1 + x\tfrac{f_0'}{f_0})^r = f_0^r (1 + rx\tfrac{f_0'}{f_0}) + O(x^2)}$$

where we have taken the binomial series $(1+x)^r = 1 + r \cdot x + O(x^2)$ from our **small table**. Especially for $r = -1$, i.e. the **inverse of a function** $f(x)$ follows:

$$\boxed{f^{-1}(x) = f_0^{-1}(1 - x\tfrac{f_0'}{f_0}) + O(x^2).}$$

We find the same expression clearly also if we break off the Taylor series for the inverse function after the second term.

For **nested functions** $F(x) = f(g(x))$ we use most simply the Taylor series directly,

$$\boxed{f(g(x)) = f(g(0)) + xf'(g(0))\, g'(0) + O(x^2).}$$

As a numerical example let us consider the problem of calculating $(1.2)^{1/20} = (1 + 0.2)^{1/20} = 1 + 0.2/20 + O((0.2)^2) = 1.01 + O(0.04)$, while the exact value is 1.00915 and the error lies near 0.0008.

Exercise 6.8 *Develop around the point $x = 0$ up to first order:*

a) $(1+x)e^x$, *b)* $e^{-x}\sin x$, *c)* $\sqrt[3]{8+x}$, *d)* $\sin x \cos x$, *e)* $\frac{1}{\cosh x}$ *and* *f)* $\exp(\sin x)$

Exercise 6.9 *Calculate up to first order and compare with the exact values:*

a) $\sin 0.1$, *b)* $e^{-0.3}$, *c)* $\ln 0.8 = \ln(1-0.2)$ *and* *d)* $17^{1/4} = (16+1)^{1/4}$

6.7 Quality of Convergence: the Remainder Term

After the results of the last sections, only question 4) about the **quality of convergence** remains unanswered: Even if we are sure of the convergence of the series, it is clearly essential to know how large the error would be when we use only an

approximation polynomial of m-th degree: $P_m(x) = \sum_{n=0}^{m} \frac{f^{(n)}(0)}{n!} x^n$ instead of the entire infinite power series.

Instead of the exact calculation of the

$$\textbf{remainder term } r_m(x) := f(x) - P_m(x),$$

we shall only report here the formula which results from this calculation and can serve to estimate $r_m(x)$, the so-called

Lagrange form of the remainder term:
$$r_m(x) = \frac{f^{(m+1)}(\theta x)}{(m+1)!} x^{m+1}, \quad \text{where } 0 < \theta < 1.$$

At first sight this expression seems a bit astonishing, since it is expressed in terms of the $(m+1)$-th term of the series, i.e. the first neglected term. This remainder term should, however, replace the whole remainder of the series. The apparent contradiction is resolved by the observation that the $(m+1)$-th derivative in the remainder term is not to be evaluated at the point of development 0, but rather **at a unknown intermediate point** θx in between the expansion point 0 and the point of interest x, expressed by the unknown number θ with $0 < \theta < 1$. Because of the fact that θ is unknown, the remainder term can generally not be calculated, but must be

$$\textbf{estimated: } |r_m(x)| \leq \max_{0<\theta<1} \frac{|f^{(m+1)}(\theta x)||x^{m+1}|}{(m+1)!}.$$

From this estimate formula we can see that the error decreases as the $(m+1)$-th power of the distance from the point x to the expansion point 0. Thus it is favorable to go as near to the expansion point as possible.

As an example we calculate $\sin 100° = \sin(5\pi/9)$ up to $r_5(x)$ with help of the Taylor series around the point $x_0 = 0$, obtained in Section 6.4.2

$$\sin x = x - x^3/3! + x^5/5! - r_5(x) :$$

We get: $\sin(5\pi/9) = 0.9942 - r_5(5\pi/9)$, where the estimate of the remainder term with $\theta = 0$ yields:

$$|r_5(5\pi/9)| \leq (5\pi/9)^7/7! = 0.0098.$$

The correct value is 0.98480... differing from the approximation value by 0.0094.

Exercise 6.10 *Calculate the quadratic terms of the Taylor series and the remainder terms $r_2(x)$ in our small table from Section 6.4.3. Choose e.g. $r = 1/2$, $-1/2$ and -1 and estimate the errors: At what points x do the relative errors $r_2(x)/f$ amount to 1% or 10%, respectively?*

Exercise 6.11 *Calculate $\sqrt[4]{e^3}$ with the Taylor series around $x_0 = 0$ up to r_3.*

6.8 Taylor Series around an Arbitrary Point

In the last section we have seen how much depends on the proximity to the expansion point of development in the application of the Taylor series, when calculating to calculate function values in the neighbourhood of expansion points for which the function is well known. Therefore, we finally turn to the problem of optimizing the expansion point x_0 which till now we always have chosen to be 0.

We obtain the **general Taylor series around any point** x_0 from our present form very simply by replacing x everywhere by $y := x - x_0$ and expanding the resulting function of y in the neighborhood of $y = 0$. Thus we get for infinitely often differentiable functions $f(x)$ the form of the general

$$\textbf{Taylor series around the point } \boldsymbol{x_0} \textbf{ :} \; f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n,$$

for the **convergence radius R** the sufficient **quotient criterion**:

$$R = \lim_{n\to\infty} \left|\frac{a_n}{a_{n+1}}\right| = \lim_{n\to\infty} \left|\frac{f^{(n)}(x_0)\,(n+1)}{f^{n+1}(x_0)}\right|$$

and for the error resulting from breaking off after the nm-th term: the

Lagrange form of the remainder term

$$r_m(x) = \frac{f^{n+1}(x_0 + \theta(x - x_0))}{(n+1)!}(x - x_0)^{n+1}$$

again with $0 < \theta < 1$.

As an example we expand the **sine function** $f(x) = \sin x$ now around the point $x_0 = \frac{\pi}{2}$:

$$f(x) = \sin x = \sum_{n=0}^{\infty}(-1)^n(x - \frac{\pi}{2})^{2n}/(2n)!$$
$$= 1 - (x - \frac{\pi}{2})^2/2! + (x - \frac{\pi}{2})^4/4! \pm \ldots$$

To the proof:

$$f'(x) = \cos x, \qquad f''(x) = -\sin x,$$
$$f'''(x) = -\cos x, \qquad f^{(4)}(x) = \sin x, \ldots$$

with

$$f(\frac{\pi}{2}) = 1, \qquad f'(\frac{\pi}{2}) = 0,$$
$$f''(\frac{\pi}{2}) = -1, \qquad f'''(\frac{\pi}{2}) = 0,$$
$$f^{(4)}(\frac{\pi}{2}) = 1, \qquad \ldots$$

altogether: $\sin x = 1 - (x - \frac{\pi}{2})^2/2! + (x - \frac{\pi}{2})^4/4! \pm \ldots = \sum_{n=0}^{\infty}(-1)^n(x - \frac{\pi}{2})^{2n}/(2n)!$
with the convergence radius $R = \lim_{n\to\infty} |(2(n+1))!/(2n)!| = \lim_{n\to\infty} |(2n+1)(2n+2)| = \infty$.

We want to check immediately how the centre of expansion $x_0 = \frac{\pi}{2}$, lying nearer to 100°, improves our earlier calculation of $\sin 100° = \sin(10\pi/18)$ which we performed in Section 6.7 with the development point $x_0 = 0$:

$$\sin x = 1 - (x - \pi/2)^2/2! + (x - \pi/2)^4/4! - r_4(x) :$$

We get: $\sin(10\pi/18) = \sin(\pi/2 + \pi/18) = 0.984807773 - r_3(10\pi/18)$,
where the estimation of the rest term with $\theta = 0$ yields

$$|r_4(10\pi/18)| \leq (\pi/18)^6/6! = 3.93 \cdot 10^{-8}.$$

This result must be compared with the earlier error estimation of $9.38 \cdot 10^{-3}$, i.e. through the better centre of expansion the error can be reduced by more than four orders of magnitude, with comparable effort of calculation.

For other desired values in between we may develop the **sine also around** $x_0 = \pi/4$:

$$f(x) = \sin x = (1/\sqrt{2})[1 + (x - \pi/4) - (x - \pi/4)^2/2! - (x - \pi/4)^3/3! + + - - \dots]$$

since $f'(x) = \cos x$, $f''(x) = -\sin x$, $f'''(x) = -\cos x$, $f^{(4)}(x) = \sin x, \dots$
with $f(\pi/4) = 1/\sqrt{2}$, $f'(\pi/4) = 1/\sqrt{2}$, $f''(\pi/4) = -1/\sqrt{2}$, $f'''(\pi/4) = -1/\sqrt{2}$, $f^{(4)}(\pi/4) = 1/\sqrt{2}$,
altogether: $\sin x = (1/\sqrt{2})[1 + (x - \pi/4) - (x - \pi/4)^2/2! - (x - \pi/4)^3/3! + + - - \dots$
with the convergence radius $R = \lim_{n\to\infty} |(n + 1)!/n!| = \lim_{n\to\infty} |n + 1| = \infty$.

Now we are able to present a series also for the **natural logarithm**, e.g. around $x_0 = 1$:

$$f(x) = \ln x = \sum_{n=1}^{\infty} (-1)^{n+1}(x - 1)^n/n = (x - 1) - (x - 1)^2/2 + (x - 1)^3/3 + \dots$$

If we replace here x by $x + 1$, we arrive again at our earlier Taylor series for $\ln(x + 1)$ around the point 0.

Exercise 6.12 *Determine the Taylor series and the convergence regions for:*

1) $\sin x$ *around point* $x_0 = \pi$,
2) e^x *around point* $x_0 = 1$,
3) e^x *around point* $x_0 = 2$ *and*
4) prove the series for $\ln x$ *around the point* $x_0 = 1$ *given above.*

Exercise 6.13 *Calculate once more* $\sqrt[4]{e^3}$ *up to* r_3, *but now using the Taylor series around* $x_0 = 1$ *and compare with our earlier calculation around* $x_0 = 0$ *in Exercise 6.11.*

You will get to know and carry out many more series expansions during your studies. There are expansion series also for fields, for instance the famous multipole series. You will expand the frequently used periodic functions "after Fourier in the quadratic mean" with cosine and sine as basis functions, and later on also transform non-periodic functions "after Fourier" with help of the exponential functions. Complex functions can be developed "after Laurent" even in the neighbourhood of certain singularities. Finally in quantum mechanics, you will perform several perturbation theoretical expansions around the few solvable systems like the harmonic oscillator. Theoretical physics is in large part the high art of dealing with series expansions.

Chapter 7

INTEGRATION

The second main pillar in mathematical methods for all natural sciences is the integral calculus. In some sense integration is the inverse of differentiation. While differentiation assigns to a function $f(x)$ its gradient $f'(x)$ the integral calculus deals with problems in which something is known about the gradient of a function and other functions having this gradient are sought. This assignment is much more difficult than differentiation, it has however a central meaning for natural science. It turns out that the basic laws of mechanics, electrodynamics and quantum mechanics can be formulated as **differential equations** which make statements about the derivatives of functions. For example, Newton's second law makes a statement about the path or trajectory $x(t)$ of a particle under the influence of a force K, and makes this statement using the second derivative of $x(t)$, i.e. the acceleration:

$$m(\tfrac{d}{dt})^2 x(t) = K.$$

For a given force, e.g. gravity or the electric or magnetic force, a particle trajectory is to be determined. To integrate in this context means to solve the basic equations, applying the theory to the various cases which are encountered in practice.

In previous chapters you have learned and practiced the technique of differentiation. It is actually not that difficult, if you follow some rules. Integration however is an "**art**", as you will soon see. However, since no artist ever appeared out of nowhere, but rather had to develop his talent through learning, gaining experience and practice, the same goes for integration. The following chapters will give you plenty of opportunity to do this.

7.1 Work

First we take another look at the **uniform motion** of a mass point on a straight line, for example the centre of mass of a car on a highway. After we have answered our first question for the velocity with help of the differential calculus, we look as physicists to one level deeper for the reason: Why does the car move on this straight plane highway section with the observed uniform speed? Obviously, the reason for the uniform straight motion according to Newton is a force, namely the force K, with which the engine pushes the car against inertia, wind and other frictional loss. Latest at the gas station the **question arises** how much **work** A has the engine done over the travelled distance Δx. The work done is proportional to the needed force was, and proportional to the travelled distance Δx. In fact it is exactly equal to the product of both quantities: $A = K\Delta x$, geometrically equal to the **rectangular area** K times Δx, if we draw in a Cartesian (i.e. right angled) coordinate system with the travelled distance x in the direction of the 1-axis and the operating force K in direction of the 2-axis.

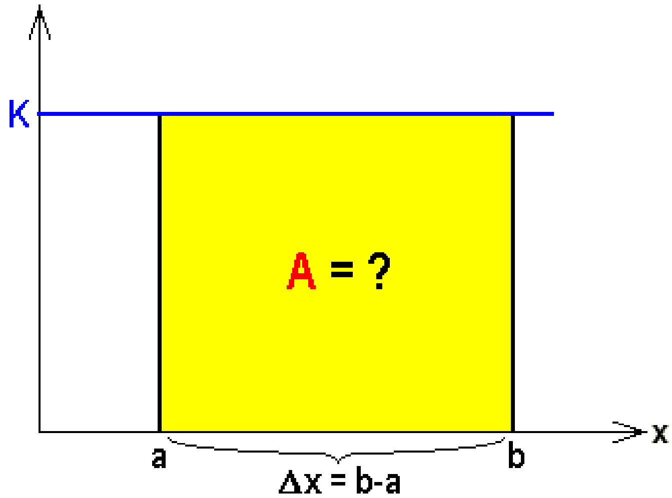


Figure 7.1: Constant force K as a function of the travelled distance from a to $b = a + \Delta x$. The yellow rectangular area is proportional to the work $A = K\Delta x$.

Following this idealized case of a constant operating force we turn our attention to a more realistic case, where the force over the distance is **increased linearly** from an initial value K_0 by giving more gas: $K(x) = sx + K_0$:

From the figure we immediately see how we can help ourselves, if the question arises (while filling up the gas tank) how much work was done over the entire distance. We take the mean value $\xi = (a + b)/2$ between the starting point a and the end point b, read out the corresponding function value of the force $K(\xi) = K((a + b)/2)$, and multiply this with the travelled distance $\Delta x = b - a$:

$$A = K(\xi)\Delta x = K(\tfrac{a+b}{2})(b - a).$$

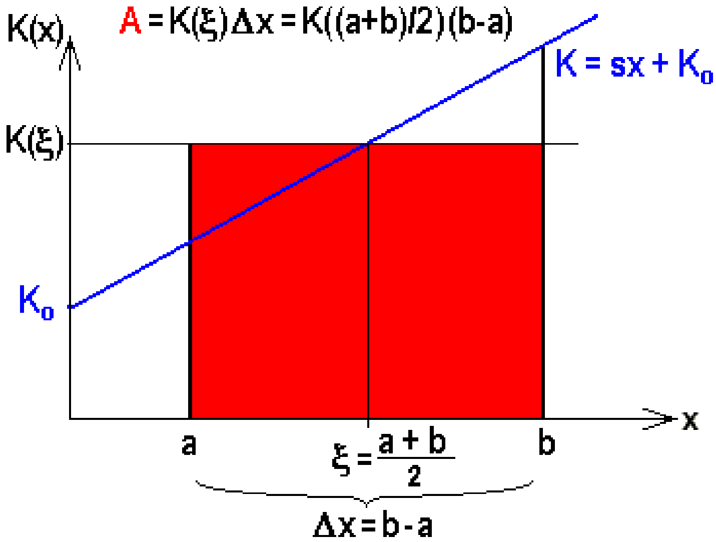


Figure 7.2: Force rising linearly with distance from the initial value K_0 to $K(x) = sx + K_0$ between the points a and b

In this way we reduce the physical question for the work to the geometrical question for the "**area under the force line over the interval $\boldsymbol{\Delta x}$**", more precisely for the content of the area which is bordered on the top by a known function, to the left by the straight line $x = a$, to the right by the line $x = b$, and on the bottom by the 1-axis $K(x) = 0$. The advanced question for the work done by a force varying according to an **arbitrary** function is herewith traced back to the mathematical problem of the determination of area of a rectangle, in which one side (in this case the upper one) is replaced by a curve. In the following we would like to explore this more general mathematical question and to give an answer which will fulfill all wishes of physicists concerning work, and many other wishes beyond.

7.2 Area under a Function over an Interval

We calculate the **area** $F(f(x); a, b)$ "under" any bounded continuous, however for the moment positive function $f(x) > 0$ "over" the finite distance "between" the starting point a and the final point b, also often called $\boldsymbol{F_a(b)}$following the recipe indicated above for the straight line: We divide the desired area in many narrow vertical strips, whose upper sides are nearly straight lines, calculate the areas of the strips as indicated above, sum up the single parts and finally let the number of strips grow without limit in the hope that this way we will find the wanted area as the limiting value.
We want to document this **limiting procedure** just this once and then let it run in a

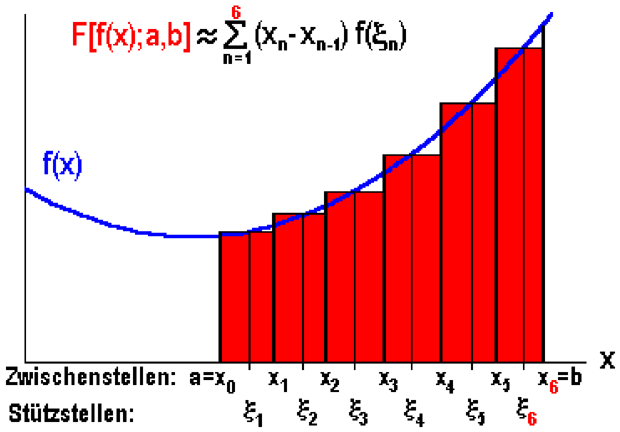


Figure 7.3: Interval dissection and strips

standardized way: the interval $[a, b]$ between the starting point, called $x = a =: x_0$, and the final point, called $x = b =: x_m$, will be divided through the choice of e.g. $m - 1$ **intermediate points** x_n with $a = x_0 < x_1 < x_2 < \ldots < x_{m-1} < x_m = b$ in m partial intervals $[x_{n-1}, x_n]$ of the respective length $\Delta x_n = x_n - x_{n-1}$. The partial intervals must not be of equal length, but they may be: $\Delta x_n = (b - a)/m$. In the interior of each of the small partial intervals we chose a **node** $\xi_n \in [x_{n-1}, x_n]$, which does not necessarily has to be the arithmetical midpoint $\xi_n = (x_n + x_{n-1})/2$, but it can be. Then we determine the functional values above these nodes and we approximate the actual area of the single strips through $\Delta x_n f(\xi_n)$, that of the corresponding rectangles under the horizontal lines through $f(\xi_n)$. We sum up these m rectangular areas and call them

Riemann sum: $S_m := \sum_{n=1}^{m} \Delta x_n f(\xi_n).$

The announced limiting procedure consists now in the refinement of the interval dissection by increasing the number $m - 1$ of the intermediate points without limit, where we have to pay attention that in case of a non-equidistant dissection, the width of the thickest stripe $\max \Delta x_n$ approaches zero. If the sequence of the Riemann sums S_m converges to a limiting value which is independent of the dissection of the interval and the selection of the nodes ξ_n in the single stripes, we call this limit the "definite" (or Riemann) integral of the function $f(x)$ from a to b, respectively the area "under" the function $f(x)$ "over" the interval $[a, b]$ and write it as proposed by Leibniz with a stylized S for sum:

(Riemann) integral: $F(f(x); a, b) \equiv F_a(b) = \int_a^b dx f(x) := \lim_{m \to \infty} S_m$

The integrand $f(x)$ may also be placed between the integral sign and the differential.

Mathematicians guarantee that the considered limit exists, if the integrand function $f(x)$ is continuous and bounded, and the interval is finite and closed:

> **A bounded continuous function is (Riemann) integrable over a closed interval.**

In contrast to this, differentiability did not at all follow from continuity.

Insert: Calculation of an Integral According to the Definition: *After this complicated definition of the limiting procedure, clearly an example is needed: We choose the **function** $f(x) = x^2$ and the interval $[0, b]$:*

For the sake of simplicity, we divide the interval of length b through $m-1$ equidistant intermediate points $x_n = nb/m$ in m intervals of the uniform length $\Delta x_n = b/m$. As nodes in the intervals we choose the arithmetic means $\xi_n = (x_n + x_{n-1})/2 = (2n-1)b/2m$. With this we form the Riemann sums:

$$S_m := \sum_{n=1}^{m} \Delta x_n f(\xi_n) = \sum_{n=1}^{m} \left(\frac{b}{m}\right)\left(\frac{(2n-1)b}{2m}\right)^2 = \left(\frac{b}{m}\right)^3 \left(\frac{1}{4}\right) \sum_{n=1}^{m} (2n-1)^2$$

The final sum causes a little trouble:

$$\begin{aligned}\sum_{n=1}^{m}(2n-1)^2 &= 4\sum_{n=1}^{m} n^2 - 4\sum_{n=1}^{m} n + \sum_{n=1}^{m} n^0 \\ &= 4m(m+1)(2m+1)/6 - 4m(m+1)/2 + m \\ &= m(4m^2-1)/3.\end{aligned}$$

Here we have used the sum of the first m numbers $\sum_{n=1}^{m} n = m(m+1)/2$ and the sum of the first m squares $\sum_{n=1}^{m} n^2 = m(m+1)(2m+1)/6$ from Chapter 2.1 and $\sum_{n=1}^{m} n^0 = m$. Thus we obtain for the sequence of the Riemann sums:

$$S_m := (b/m)^3(1/4)m(4m^2-1)/3 = \frac{b^3}{3}\left(1 - \frac{1}{4m^2}\right)$$

and in the limit $m \to \infty$:

$$\boxed{\int_0^b dx x^2 = \lim_{m\to\infty} S_m = \frac{b^3}{3}.}$$

The still-needed examination of whether the result is independent of the choice of the interval dissection and the choice of nodes we omitted.

This simple example shows us that on one hand the definition leads in fact to the expected result, but on the other hand also how much effort is needed to calculate such a simple example according to the definition.

Therefore we will have to search for other methods to calculate integrals, and this will be done in what follows.

7.3 Properties of the Riemann Integral

Before we go on to calculate further integrals, we want to get to know the most important properties of the derived integral concepts, and put them together. Also we want to free ourselves from some of the preconditions which until now we have accepted out of sheer idleness:

But first let us emphasize that the **denotation of the integration variable** x is of course **completely arbitrary**: we also could have named it y: $\int_a^b dx f(x) = \int_a^b dy f(y)$.

7.3.1 Linearity

From our knowledge of the characteristics of sums and the limits of sequences we can see, that the integral is **linear**, i.e. that the integral over a linear combination of functions is equal to the corresponding linear combination of the integrals of the single functions: for example with two functions $f(x)$ and $g(x)$ and real constants c respectively d :

Linearity: $$\int_a^b dx(cf(x) + dg(x)) = c\int_a^b dx f(x) + d\int_a^b dx g(x).$$

The special case $c = d = 1$ is known as **additivity** of the integral:

$$\int_a^b dx(f(x) + g(x)) = \int_a^b dx f(x) + \int_a^b dx g(x).$$

An other special case is called **homogeneity**; when $d = 0$, it follows:

$$\int_a^b dx c f(x) = c\int_a^b dx f(x).$$

In particular for $c = -1$ that means:

$$\int_a^b dx(-f(x)) = F(-f(x); a, b) = -F(f(x); a, b) = -\int_a^b dx f(x).$$

If $f(x) \geq 0$, as assumed, is a positive function and $F(f(x); a, b) = \int_a^b dx f(x)$ the area "under" the function, we see from the above relation that the integral over the negative, i.e. through the x-axis reflected function $-f(x) \leq 0$, yields exactly $-F(f(x); a, b)$, meaning a **negative** area. The area "above" the function $-f(x)$ (running in the fourth quadrant) receives in the integral automatically a negative sign. We can therefore give up our beginner assumption $f(x) \geq 0$, if we interpret the integral as **area with sign**: positive: "under" a function and above the x-axis, and negative: "above" a negative function and below the x-axis.

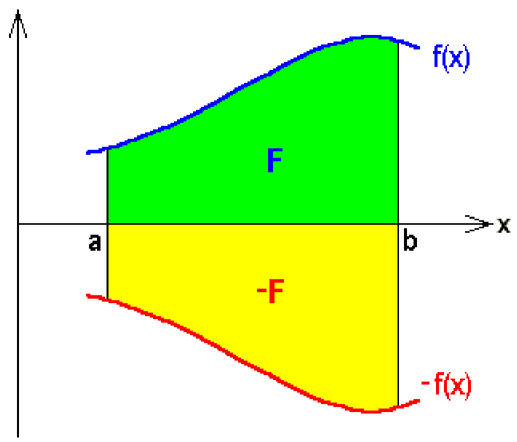


Figure 7.4: Functions f(x) and -f(x) with the coloured integral areas: F and -F.

If the integrand changes sign within the integration interval, the interval has to be divided into two parts, which must be treated separately and afterwards the results subtracted from each other with the proper sign.

7.3.2 Interval Addition

Next, we consider two adjoining integration intervals: i.e. two integrals, in which the upper limit of the first one coincides with the lower limit of the second integral while the integrand stays the same: From the meaning of the integral as area, there follows immediately the so-called:

$$\textbf{interval addition:} \quad \int_a^b dx f(x) + \int_b^c dx f(x) = \int_a^c dx f(x)$$

Thus we can for instance see without any calculation that $\int_0^{2\pi} dx \sin x = 0$.
With this knowledge we can free ourselves from a further beginner precondition, namely that the upper limit $b \geq a$ has to be larger than or equal to the lower one, if we define:

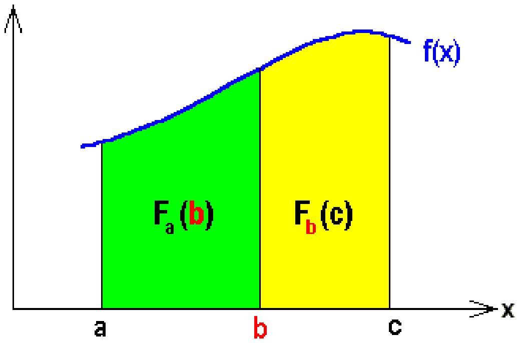


Figure 7.5: Interval addition

$$\int_a^b dx f(x) =: -\int_b^a dx f(x)$$

This definition is by no means in contradiction to our limiting procedure introducing the integral by the Riemann sum, since in case of exchanged integration limits all Δx_n and consequently all dx become negative.

With the help of the interval addition it now follows: $\int_a^a dx f(x) = 0$, as it should be.

7.3.3 Inequalities

Certain inequalities expand our understanding for the integral concept and are helpful later on in the calculation of integrals:

If for example a function $g(x)$ in the whole interval $[a, b]$ is larger than another function $f(x) \leq g(x)\ \forall x \in [a, b]$, there follows a corresponding relation for the integrals, the so-called

monotony: $f(x) \leq g(x)\ \forall x \in [a, b] \Rightarrow \int_a^b dx f(x) \leq \int_a^b dx g(x),$

because an analog relation for sums holds true.
Even the "triangle inequality" known from the sums over absolute values translates simply onto the integrals:

triangle inequality: $\left| \int_a^b dx f(x) \right| \leq \int_a^b dx\ |f(x)|$ for $a < b$

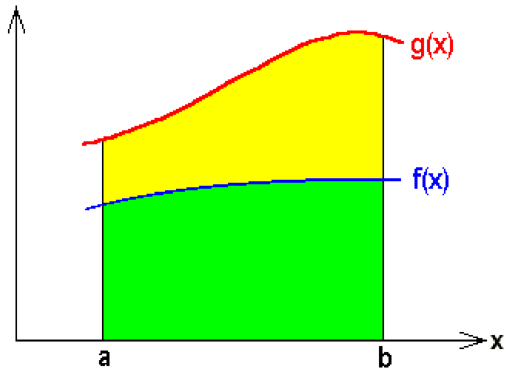


Figure 7.6: Illustration of the monotony of the integral

Finally we sometimes need the following estimate of an integral area under a function $f(x)$ being continuous in an interval $[a, b]$ in terms of square areas with its minimal m resp. maximal M function value in the interval:

estimate: $(b-a)m \leq \int_a^b dx f(x) \leq (b-a)M$, if $m \leq f(x) \leq M \; \forall x \in [a, b]$

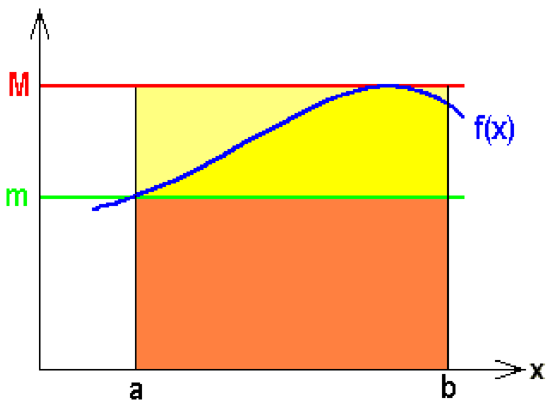


Figure 7.7: The three differently coloured areas of the estimate

7.3.4 Mean Value Theorem of the Integral Calculus

Just as in the differential calculus there holds a **mean value theorem** also for the integrals which sometimes helps to estimate an integral:

For an Integral over a function $f(x)$ which is continuous and bounded in an interval $[a, b]$ there exists always a **mean value** ξ, so that $\int_a^b dx f(x) = f(\xi)(b-a)$ for $\xi \in (a, b)$.

Because a continuous and bounded function in an interval passes through its extrema and all the values in between.

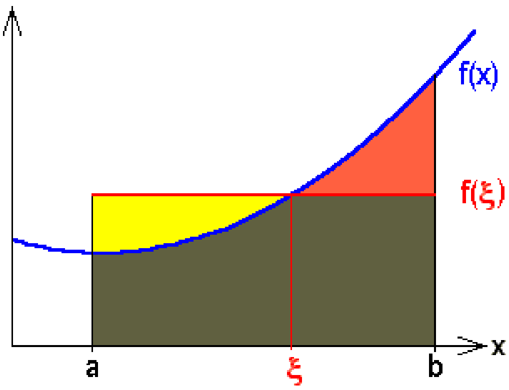


Figure 7.8: Mean value theorem of the integral calculus

7.4 Fundamental Theorem of Differential and Integral Calculus

7.4.1 Indefinite Integral

With this general knowledge about the integral concept we want to turn to the all-deciding question of the calculation of Riemann integrals, in which the Fundamental Theorem of Differential and Integral Calculus helps us decisively, as its name promises.

At first we want to expand the Riemann integral concept: the "**definite**" or Riemann integral $\int_a^b dx f(x)$ had assigned to a bounded continuous function $f(x)$, called the integrand, by a given lower interval border a and upper border b, the (signed) area $F(f(x); a, b) = F_a(b) \in \mathbb{R}$ under the function, meaning a real **number**. Mathematicians call such an entity a functional. Now we are interested in how this area will change, when we move the upper border. Thus we replace in the upper border the constant b by a **variable** y and treat the **integral as a function of its upper border**. This function of y is called

$$\textbf{indefinite integral:} \quad F(f(x); a, y) \equiv F_a(y) := \int_a^y dx f(x).$$

The procedure is fully analogous to the expansion step of the differential calculus from the gradient $f'(x_0)$ of a function $f(x)$ at a certain point x_0 to the first derivative $f'(x)$ as a function of the variable x.

7.4.2 Differentiation with Respect to the Upper Border

In order to study the functional dependence of the indefinite integral from the variable upper border, we are at first interested in the gradient of the function $F_a(y)$:

$$(\frac{d}{dy})F_a(y) := (\frac{d}{dy})\int_a^y dx f(x) = \lim_{\Delta y \to 0} \frac{\int_a^{y+\Delta y} dx f(x) - \int_a^y dx f(x)}{\Delta y} =$$

if we insert the definition of the derivative. Because of the interval addition we get:

$$= \lim_{\Delta y \to 0} \frac{\int_a^y dx f(x) + \int_y^{y+\Delta y} dx f(x) - \int_a^y dx f(x)}{\Delta y} = \lim_{\Delta y \to 0} \frac{\int_y^{y+\Delta y} dx f(x)}{\Delta y} =$$

According to the mean value theorem of the integral calculus there exists in the interval $[y, y + \Delta y]$ a mean value $y + \theta\Delta y$ with $0 \leq \theta \leq 1$, so that the following holds true:

$$= \lim_{\Delta y \to 0} \frac{f(y + \theta\Delta y)\Delta y}{\Delta y} = \lim_{\Delta y \to 0} f(y + \theta\Delta y) = f(y).$$

Altogether we get the

First part of the Fundamental Theorem: $F_a'(y) := (\frac{d}{dy})\int_a^y dx f(x) = f(y)$,
i.e. the differential quotient of a indefinite integral with respect to its upper border is the integrand taken at the upper border.

Exactly in this sense is differentiation the reversal of integration. For example $(\frac{d}{dy})\int_0^y dt \sin(\omega t + \alpha) = \sin(\omega y + \alpha)$ holds.

7.4.3 Integration of a Differential Quotient

After having learned to differentiate an integral, we are inquisitive about the reverse process, namely about the integral of a differential quotient: We start from the well-known continuous differential quotient $F'(x) = f(x)$ which may be given to us as a continuous function $f(x)$, i.e. really from a differential equation of the first order for $F(x)$. We want to integrate this differential quotient over the interval $[a, b]$:

$$\int_a^b dx F'(x) := \lim_{m\to\infty} \sum_{n=1}^{m} \Delta x_n F'(\xi_n)$$

with the nodes $\xi_n \in [x_{n-1}, x_n]$ within the interval of length $\Delta x_n = x_n - x_{n-1}$, if we insert the definition of the integral. According to the *Mean Value Theorem* of *Differential Calculus* the gradient at the nodes ξ_n can be replaced by the gradient of the secant, meaning the replacement of the differential quotient by the difference quotient:

$$F'(\xi_n) = \frac{F(x_n) - F(x_{n-1})}{x_n - x_{n-1}}.$$

Written in detail this yields:

$$\lim_{m\to\infty} \big((F(x_1) - \mathbf{F(x_0)}) + (F(x_2) - F(x_1)) + \ldots + (F(x_{m-1}) - F(x_{m-2})) + (\mathbf{F(x_m)} - F(x_{m-1}))\big).$$

We easily see that all terms are cancelling in pairs except the second and the second last one, which do not at all depend on m and thus will not be affected by the limiting process:

$$... = \mathbf{F(x_m)} - \mathbf{F(x_0)} = F(b) - F(a) =: F(x)\big|_a^b.$$

We therefore get altogether as

> **Second Part of the FUNDAMENTAL THEOREM:**
> $\int_a^b dx F'(x) = F(b) - F(a) =: F(x)\big|_a^b$,
> i.e. the definite integral of the differential quotient of a continuous differentiable function over an interval is equal to the difference of the function values at the upper and lower border of the interval.

Also in this sense integration is the reversal of differentiation.

For instance we get again the result: $F(x^2; 0, b) = \int_0^b dx x^2 = b^3/3$, which we were hard put to derive from the definition of the integral, but now effortlessly from the differentiation $(\frac{d}{dx})x^3 = 3x^2$.

This second part of the Fundamental Theorem is the crucial step to the solution of our integration problem: Because we are now able to calculate all definite integrals of all the functions which we find in the second column of our Differentiation Table in Chapter 5. We simply have to read the **TABLE** backward from right to left and complete the heading in the following accordingly:

TABLE FOR DIFFERENTIATION AND INTEGRATION			
Line	$F(x) = \int dx f(x)$	$F'(x) \equiv (\frac{d}{dx})F(x) = f(x)$	Comments:
1	const	0	
2	x^r	rx^{r-1}	$r \in \mathbb{R}$
3	$x^{r+1}/(r+1)$	x^r	$-1 \neq r \in \mathbb{R}$
4	$\sin x$	$\cos x$	
5	$\cos x$	$-\sin x$	
6	$\tan x$	$1/\cos^2 x$	$x \neq (z+1/2)\pi,\ z \in \mathbb{Z}$
7	$\cot x$	$-1/sin^2 x$	$x \neq z\pi,\ z \in \mathbb{Z}$
8	$-\pi/2 < \arcsin x < \pi/2$	$1/\sqrt{(1-x^2)}$	$\lvert x\rvert < 1$
9	$0 < \arccos x < \pi$	$-1/\sqrt{(1-x^2)}$	$\lvert x\rvert < 1$
10	$-\pi/2 < \arctan x < \pi/2$	$1/(1+x^2)$	
11	$0 < \mathrm{arccot}\, x < \pi$	$-1/(1+x^2)$	
12	e^x	e^x	
13	r^x	$r^x \ln r$	$0 < r \in \mathbb{R}$
14	$\ln \lvert x\rvert$	$1/x$	$x \neq 0$
15	$\log_b \lvert x\rvert$	$1/x \ln b$	$x \neq 0,\ 0 < b \in \mathbb{R},\ b \neq 1$
16	$\sinh x$	$\cosh x$	
17	$\cosh x$	$\sinh x$	
18	$\tanh x$	$1/\cosh^2 x$	
19	$\coth x$	$-1/\sinh^2 x$	$x \neq 0$
20	$\mathrm{arsinh}\, x$	$1/\sqrt{(x^2+1)}$	
21	$0 < \mathrm{arcosh}\, x$	$1/\sqrt{(x^2-1)}$	$x > 1$
22	$\mathrm{artanh}\, x$	$1/(1-x^2)$	$\lvert x\rvert < 1$
23	$\mathrm{arcoth}\, x$	$-1/(x^2-1)$	$\lvert x\rvert > 1$

Our example out of the insert above, can be obtained e.g. from line two for $r = 2$.

A further example from this line is $\int_a^b dx\ x^3 = (b^4 - a^4)/4$ with the borders a and b, generally for an arbitrary real $r \in \mathbb{R}$ follows:

$$F(x^r; a, b) = \int_a^b dx\ x^r = \frac{b^{r+1} - a^{r+1}}{r+1},$$

which we have filled in the third line of the **TABLE** being empty until now, because it occurs very often.

From the fourth line we find for instance:

$$F(\sin x; a, b) = \int_a^b dx \sin x = -\cos b + \cos a,$$

from the twelfth line: $$F(e^x; a, b) = \int_a^b dx\ e^x = e^b - e^a$$

and analogously many further integrals.

Exercise 7.1 *Calculate the following examples of integrals:*

a) $\int_1^3 \frac{1}{x}dx,$ *b)* $\int_{-1}^1 \frac{dx}{1+x^2},$ *c)* $\int_0^b \frac{dx}{\sqrt{1+x^2}}$, *d)* $\int_{-1/\sqrt{2}}^{1/\sqrt{2}} \frac{dx}{\sqrt{1-x^2}}$,

e) $\int_{-a}^a dx\ \cosh x$, *f)* $\int_0^{\pi/4} \frac{dx}{\cos^2 x}$, *g)* $\int_1^2 \frac{dx}{x^{1+a}}$, *h)* $\int_{-a}^a dx\ x^{2n+1}$ *for* $n \in \mathbb{Z}$

7.4.4 Primitive Function

Although we are able to calculate a considerable number of definite integrals over a large number of intervals, the question for the **indefinite integral of a differential quotient** $F'(x) = f(x)$ remains. We once again replace the constant upper border b of the definite integral through a **variable** y and as above we get

$$\int_a^y dx f(x) = \int_a^y dx F'(x) = F(y) - F(a).$$

This we rewrite in the following way:

$$F(y) = \int_a^y dx f(x) + F(a) =: \int_a^y dx f(x) + c,$$

because with respect to the variable y $F(a)$ is indeed a constant, although it depends on the starting point of the interval a. Since we want to have the usual letter x as symbol for the independent variable in the function $F(y)$, an extraordinary sloppy way of writing has sneaked into the above equation world-wide: symbolically it is written as:

> $F(x) = \int dx f(x) + c$ and $F(x)$ is called the **primitive function** of $f(x)$.

The x on the left side serves only as a hint, that it is a function of an independent variable, and has obviously nothing to do with the really arbitrarily denotable integration variable x on the right hand side, which obviously does not occur anymore after the integration on the right side. Once we have cleared up this sloppy mess, it is quite a comfortable matter and acknowledged world-wide.

Written in such a sloppy way, the primitive function is actually a whole family of functions with the *family parameter* c. The primitive function $F(x)$ of $f(x)$ is exactly the family of functions which solves our original differential equation $F'(x) = f(x)$ and that is the reason why it is so important for physicists. Out of this family with pre-given gradient $f(x)$, the physicist has to only pick out the solution by choosing the constant c so that the function fulfils the correct boundary condition $c = F(a)$, and the problem is solved.

For example, the primitive function $F(x)$ of $f(x) = 3x$ is searched for, which fulfils the boundary condition $F(1) = 2$. Out of the function family $F(x) = 3x^2/2 + c$ we have to choose the function which fulfills $F(1) = 3/2 + c = 2$, meaning $c = 1/2$: consequently $F(x) = (3x^2 + 1)/2$ is the desired solution.

Exercise 7.2 *Determine the primitive functions of the following functions:*
a) $f(x) = x^3$, *b)* $f(x) = \frac{1}{\sqrt{x^2-1}}$, *c)* $f(x) = \sinh x$ *and* *d)* $f(x) = 2^x$

Exercise 7.3 *Determine the primitive functions of the following functions with the given boundary conditions:*
a) $f(x) = \sin x$ *with* $F(\pi) = 1$, *b)* $f(x) = \frac{1}{\sqrt{x}}$ *with* $F(4) = 1$ *and*
c) $f(x) = \frac{1}{\cosh^2 x}$ *with* $F(a) = \frac{1}{2}$

7.5 The Art of Integration

In this chapter we want to put together what we need for the integration of the functions from our basic set and those assembled from it.

Initially after the discoveries of the Fundamental Theorem, we have the recipe: read the differentiation table backwards.

7.5.1 Differentiation Table Backwards

For example from line 14 we get:	$\int_1^e dx/x = \ln x\big\vert_1^e = \ln e - \ln 1 = 1 - 0 = 1$
or from line 4:	$\int_0^{\pi/2} dt \cos t = \sin t\big\vert_0^{\pi/2} = \sin(\pi/2) - \sin 0 = 1 - 0 = 1$
or from line 10:	$\int_a^b dx/(x^2+1) = \arctan x\big\vert_a^b$
or from line 8 indefinitely:	$\int dx/\sqrt{(1-x^2)} = \arcsin x + c.$

In our first enthusiasm for the consequences of the Fundamental Theorem for integration, by reading the differentiation table backwards, we have slightly overrated our success. Because upon closer inspection, we see that there are only relatively few functions, that occur in the second column of our **TABLE**. Even such simple functions out of our basic set as $f(x) = x^2 + 1$ cannot be found. All the more $f(x) = \sqrt{(1 - x^2)}$ is nowhere to be found, although we can easily integrate the reciprocal of this root. Also with the Gaussian bell-shaped $f(x) = \exp(-x^2)$ we are pretty helpless with out **TABLE**.

Differently as in differentiation there is no procedure of integration which automatically after a sufficient amount of work leads to the goal. This is the reason why **integration is an art** while differentiation can be called **atechnique**.

Nevertheless, there are a lot of rules and tricks which follow from the characteristics of the integral and can make our lives considerably easier. Therefore we want to turn out attention now to those: the simplest one is the linear decomposition.

7.5.2 Linear Decomposition

We have seen in Chapter 7.3.1, that the integral is a linear operation:

$$\int_a^b dx(cf(x) + dg(x)) = c\int_a^b dxf(x) + d\int_a^b dxg(x).$$

We will use this recipe in a large number of integrals, e.g. in determining the area of the yellow "leaf" in the following figure:

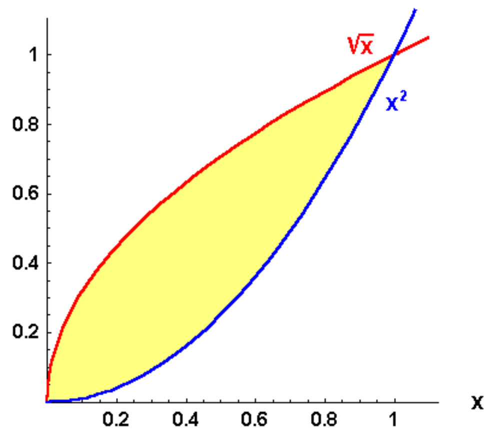


Figure 7.9a: Integral over the "leaf"

$$\int_0^1 dx(\sqrt{x} - x^2) = \int_0^1 dx\sqrt{x} - \int_0^1 dx x^2 = \frac{x^{1/2+1}}{(1/2+1)}\Big|_0^1 - \frac{x^3}{3}\Big|_0^1 = 2/3 - 1/3 = 1/3$$

Or in this integral over "half a hill cross section":

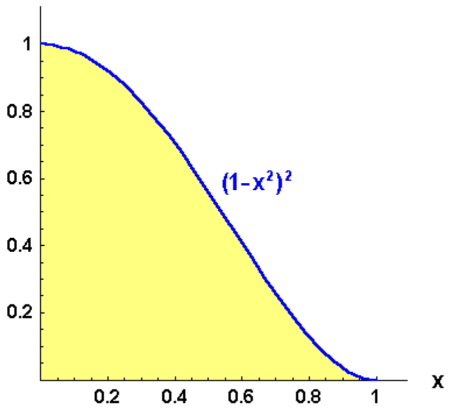


Figure 7.9b: Integral over "half a hill cross section"

$$\int_0^1 dx(1 - x^2)^2 = \int_0^1 dx(1 - 2x^2 + x^4) = \int_0^1 dx 1 - 2\int_0^1 dx\, x^2 + \int_0^1 dx\, x^4$$
$$= (x - 2x^3/3 + x^5/5)\Big|_0^1 = 1 - 2/3 + 1/5 = 8/15.$$

Exercise 7.4 *Integrate through linear partition* $\int_{-1}^1 dx(1 + 2x^3)^3$.

The most frequently used and successful integration method is substitution.

7.5.3 Substitution

Substitution is always recommended when the integrand $f(x)$ depends continuously in a more simple or more appropriate way on **another variable** y, which is connected reversibly and continuously differentiably with $x = g(y)$, whereas $W_g \subset D_f$.

In order to make this clearer we rename the integration borders of the wanted integrals $\int_a^b dx f(x)$ now $x_a := a$ and $x_b := b$, respectively: $\int_{x_a}^{x_b} dx f(x)$. Because of the bi-unique relation of y and x an inverse function $y = g^{-1}(x)$ exists, in particular $y_a = g^{-1}(x_a)$ and $y_b = g^{-1}(x_b)$. Furthermore, due to continuous differentiability the derivative $\frac{dx}{dy} = g'(y)$ exists. Then there holds (in the suggestive Leibniz shorthand almost trivially) the:

$$\textbf{substitution formula} \quad \int_{x_a}^{x_b} dx f(x) = \int_{y_a}^{y_b} dy (\frac{dx}{dy}) f(g(y)) = \int_{y_a}^{y_b} dy g'(y) f(g(y)).$$

Insert: Proof of the Substitution Formula: *If the primitive function $F(x)$ of the integrand $f(x)$, i.e. the solution of the differential equation $(\frac{d}{dx})F(x) = f(x)$ were known, then from the second part of the Fundamental Theorem for the* left *side of the equation would follow*

$$\int_{x_a}^{x_b} dx f(x) = F(x_b) - F(x_a).$$

*If we assume that the solution of the differential equation $(\frac{d}{dy})F(g(y)) = f(g(y))g'(y)$ is known, i.e. $F(g(y))$ is the primitive function of $f(g(y))g'(y)$ as function of the other variable y, then for the **right** hand side of the equation follows from the Fundamental Theorem:*

$$\int_{y_a}^{y_b} dy\, g'(y) f(g(y)) = F(g(y_b)) - F(g(y_a)) = F(x_b) - F(x_a)$$

with $x = g(y)$ as stated above.

From experience, the explanation of the substitution procedure is more complicated than this method in practice. Here are some typical examples:

$$\int_1^5 dx \sqrt{(2x-1)} = ... \quad \text{with } x_a = 1 \text{ and } x_b = 5.$$

As new variable $y := 2x - 1$ is pretty obvious with $y_a = 2 \cdot 1 - 1 = 1$, $y_b = 2 \cdot 5 - 1 = 9$ and $\frac{dy}{dx} = 2$, so $\frac{dx}{dy} = 1/2$. Thus it follows:

$$... = \int_1^9 dy \frac{dx}{dy} y^{1/2} = (1/2) y^{3/2}/(3/2)\Big|_1^9 = (9 \cdot 3 - 1)/3 = 26/3.$$

An important example for physics is:

$$\int_0^\pi dt \cos \omega t = ... \quad \text{with } t_a = 0 \text{ and } t_b = \pi.$$

We chose the substitution $y := \omega t$ with $y_a = 0$, $y_b = \omega\pi$ and $\frac{dy}{dt} = \omega$, so $\frac{dt}{dy} = 1/\omega$. Thus we get:

$$... = \int_0^{\omega\pi} dy (\tfrac{dt}{dy}) \cos y = (1/\omega) \sin y\Big|_0^{\omega\pi} = (1/\omega) \sin \omega\pi.$$

A third example from physics:

$$\int_0^b dt\, t \exp(-\alpha t^2) = ... \text{ with } t_a = 0 \text{ and } t_b = b.$$

We substitute $y := -\alpha t^2$ with $y_a = 0$, $y_b = -\alpha b^2$ and $\frac{dy}{dt} = -2\alpha t$, so $\frac{dt}{dy} = -1/2\alpha t$. From this we get:

$$... = \int_0^{-\alpha b^2} dy (\tfrac{dt}{dy}) t \exp y = (-1/2\alpha) \exp y\Big|_0^{-\alpha b^2} = (1 - \exp(-\alpha b^2))/2\alpha.$$

Exercise 7.5 *Substitution*

a) $\int dx/(ax+b)$, *b)* $\int_0^t dx\, e^{-2x/a}$, *c)* $\int_0^1 dx\, \sqrt{(1-x^2)}$,

d) $\int_0^r dx\, \sqrt{r^2 - x^2}$, *e)* $\int dt\, \dot{x}(t)$, *f)* $\int_{-a}^{a} dx\, \cosh(x/A)$

We obtain even classes of integrals in the following way by using the "**substitution formula backward**". What we mean by this is best explained through the following examples:

Supposing we have to calculate an integral of the following type, where we have suggestively chosen y for the arbitrary integration variable:

$$\int_{y_a}^{y_b} dy\, g'(y)/g(y) = ...$$

This is obviously the right side of our substitution formula, especially for the function $f(x) = 1/x$ and we can apply the substitution formula immediately with $x = g(y)$ from right to left, i.e. "backwards", in order to find:

$$... = \int_{x_a}^{x_b} \frac{dx}{x}.$$ This however, according to line 14 of our **TABLE** is equal to

$$... = \ln|x|\Big|_{x_a}^{x_b}$$ and with $x = g(y)$ it follows altogether:

$$\boxed{\int_{y_a}^{y_b} dy\, g'(y)/g(y) = \ln|g(y)|\Big|_{y_a}^{y_b}.}$$

As examples we have:

for $g(y) = ay \pm b$ with $g'(y) = a$: $\int dy\ a/(ay \pm b) = \ln|ay \pm b| + c$,
for $g(y) = \sin y$ with $g'(y) = \cos y$: $\int dy\ \cos y/\sin y = \ln|siny| + c$,
for $g(y) = y^2 \pm b$ with $g'(y) = 2y$: $\int dy\ 2y/(y^2 \pm b) = \ln|y^2 \pm b| + c$,
and so on.

In full analogy we show with $f(x) = x^n$ for $1 \leq n \in \mathbb{N}$:

$$\int_{y_a}^{y_b} dy g'(y) g^n(y) = \int_{x_a}^{x_b} dx x^n = x^{n+1}/(n+1)\Big|_{x_a}^{x_b} = \frac{g^{n+1}(y)}{n+1}\Big|_{y_a}^{y_b}.$$

Exercise 7.6 *Derive from this formula further ones by specifying $g(y)$:*

a) $g(y) = ay \pm b$, *b)* $g(y) = \sin y$, *c)* $g(y) = y^2 \pm b$, *d)* $g(y) = \ln y$

Exercise 7.7 *Prove analogously as above the formula:*

$$\int_{y_a}^{y_b} dy\ g'(y) \sqrt[n]{g(y)} = ng(y) \sqrt[n]{g(y)}/(n+1)\Big|_{y_a}^{y_b}$$

for $1 < n \in \mathbb{Z}$ and specify in it $g(y)$.

Exercise 7.8 *What do we obtain analogously for $\int_{y_a}^{y_b} dy\ g'(y)/g^n(y)$?*

Exercise 7.9 *Further mixed examples for substitution:*

a) $\int_a^{a+2\pi} dt\ \cos t$, *b)* $\int dx\ \sqrt{1+x^2}$, *c)* $\int_{-1}^{1} dz/\sqrt{az+b}$, *d)* $\int dt\ \dot{x}(t)x(t)$,
e) $\int_{-a}^{a} dx\ \sinh(2x/b)$, *f)* $\int dx\ \sqrt{x \pm b}$, *g)* $\int_{-a}^{a} dx/x^{2n+1}$, *h)* $\int dx/x^{1-a}$,
i) $\int_{-\pi}^{\pi} d\varphi\ \sin\varphi/(\cos^2\varphi + 1)$, *j)* $\int dx\ x\sqrt{x^2 \pm a}$,
k) $\int dx\ (x + b/2a)/(ax^2 + bx + c)^3$, *l)* $\int dx\ x/(1 + x^4)$

Based on these examples you can get a feeling for the enormous amount and diversity of integrals which can be calculated through application of the substitution formula. Nevertheless, this is not enough for our physics needs. In cases of simple integrands such as $\ln x$, $x \cos x$ or $\sin^2 x$ we are still helpless.

For these and similar cases we have a method, which will not fully give us the integral, but will permit at least a partial calculation and sometimes will lead to the solution in multiple steps. This method is sensibly called "partial integration".

7.5.4 Partial Integration

Whenever an integrand can be decomposed into a product $f'(x) \cdot g(x)$, so that for one of the factors, e.g. $f'(x)$, the primitive function $f(x)$ is known, you should not at any rate leave partial integration untried. We remember the product rule of differentiation from Section 5.5.2:

$$(f(x) \cdot g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x),$$

integrate this and use the second part of the Fundamental Theorem to integrate the product:

$$\int_a^b dx (f(x) \cdot g(x))' = f(x)g(x)\Big|_a^b = \int_a^b dx f'(x) \cdot g(x) + \int_a^b dx f(x) \cdot g'(x).$$

After we have resolved this equation for the first term on the right side, we get the formula for the:

$$\textbf{partial integration:} \int_a^b dx f'(x)g(x) = f(x)g(x)\Big|_a^b - \int_a^b dx\ f(x)g'(x).$$

Because of the remaining integral with the characteristic minus sign on the right side, this is not a finished solution of our problem. But sometimes this integral is easier to calculate than the original one.

We consider immediately the typical examples mentioned above, e.g. with $f'(x) = x$ and $g(x) = \ln x$:

$$\int_a^b dx\ x \ln x = (x^2/2) \ln x\Big|_a^b - \int_a^b dx\ (x^2/2)(1/x) = (x^2/2)(\ln x - 1/2)\Big|_a^b$$
$$= (b^2/2)(\ln b - 1/2) - (a^2/2)(\ln a - 1/2).$$

Or another unexpected, however, not rare example with $f'(x) = 1$ and $g(x) = \ln x$, by which a trivial product is produced through insertion of a number 1:

$$\int_1^y dx\ \ln x = \int_1^y dx\ 1 \ln x = x \ln x\Big|_1^y - \int_1^y dx\ x(1/x)$$
$$= x(\ln x - 1)\Big|_1^y = y \ln y - y + 1.$$

It is not always easy to see which factor of the product we appropriately consider as the derivative f'. To illustrate this here we study the example $x \sin x$, first obviously with $f'(x) = x$:

$$\int_0^y dx\; x \sin x = (x^2/2) \sin x\Big|_0^y - \int_0^y dx (x^2/2) \cos x.$$

We see immediately that this choice will not get us further. Instead, it makes the remaining integrand even more complicated. The other possibility $f'(x) = \sin x$ leads to success:

$$\int_0^y dx\; x \sin x = -x \cos x\Big|_0^y - \int_0^y dx\; 1(-\cos x) = -y \cos y + \sin x\Big|_0^y = -y \cos y + \sin y.$$

As a further example we consider the integrand $x^2 \cos x$ and choose based on our experience from just above $f'(x) = \cos x$:

$$\int_0^y dx\; x^2 \cos x = x^2 \sin x\Big|_0^y - \int_0^y dx\; 2x \sin x = y^2 \sin y - 2 \int_0^y dx\; x \sin x.$$

Although this is not the solution to our problem, it is a step in the right direction, since we have just calculated the remaining integral in our last example. With this result it follows:

$$\int_0^y dx\; x^2 \cos x = y^2 \sin y + 2y \cos y - 2 \sin y.$$

Altogether, **a double partial integration** has lead us here to the solution.

A further interesting example is (for a change an indefinite integral):

$$\int dx \cos x \sin x = \sin x \sin x - \int dx \sin x \cos x + c.$$

The remaining integral is equal to the original one, hence we get with a new c :

$$\int dx \cos x \sin x = \frac{\sin^2 x}{2} + c.$$

The next example with $f'(x) = g(x) = \sin x$ is also remarkable:

$$\int dx \sin^2 x = -\cos x \sin x - \int dx (-\cos x) \cos x = -\cos x \sin x + \int dx \cos^2 x = ...$$

After inserting the relation $\cos^2 x + \sin^2 x = 1$, we get:

$$... = -\cos x \sin x + \int dx (1 - sin^2 x) = x - \cos x \sin x - \int dx \sin^2 x = (x - \cos x \sin x)/2.$$

An entire series of examples of this kind follows with $f'(x) = e^{-x}$, for which a special notation was created because of their importance in physics:

$$E_1(y) = \int_0^y dx\ xe^{-x} = -xe^{-x}\big|_0^y + \int_0^y dx\ 1e^{-x} = -ye^{-x} + 0 - e^{-x}\big|_0^y = -ye^{-y} + e^{-y} - 1$$

$$E_2(y) = \int_0^y dx\ x^2e^{-x} = -x^2e^{-x}\big|_0^y + \int_0^y dx\ 2xe^{-x} = -y^2e^{-y} + 2E_1(y)$$

$$\vdots$$

$$E_n(y) = \int_0^y dx\ x^ne^{-x} = -y^ne^{-y} + nE_{n-1}(y).$$

Exercise 7.10 *Integrate the following integrals by partial integration:*

a) $\int_0^y dx\ \sin x\ e^{-x}$, *b)* $\int_0^y dx\ \cos x\ e^{-x}$, *c)* $\int dx\ \arcsin x$, *d)* $\int dx\ x\sqrt{1+x}$,
e) $\int dx\ x^3 \exp(x^2)$, *f)* $\int dx\ x^2 \ln x$, *g)* $\int dx\ \ln(x^2+1)$

and prove the following useful recursion formulae for $n \in \mathbb{N}$*:*

h) $\int dx\ f'(x)x^n = f(x)x^n - n\int dx\ f(x)x^{n-1}$,

i) $\int dx\ g(x)/x^n = -g(x)/(n-1)x^{n-1} + \int dx\ g'(x)/(n-1)x^{n-1}$ *for* $n \neq 1$,

j) $\int dx\ \sin^n x = -\frac{1}{n}\cos x \sin^{n-1} x + \frac{n-1}{n}\int dx\ \sin^{n-2} x$,

k) $\int dx\ (1 \pm x^2)^n = x\frac{(1 \pm x^2)^n}{2n+1} + 2n\int dx\ \frac{(1 \pm x^2)^{n-1}}{2n+1}$

Exercise 7.11 *Show that for the motion of a mass point on a straight line the traveled distance* $x(t)$ *can be calculated from a pre-given acceleration function* $a(t)$ *with starting velocity* v_0 *and position* x_0 *by partial integration in the following form:* $x(t) = \int_0^t dy(t - y)a(y) + v_0t + x_0$.

7.5.5 Further Integration Tricks

Apart from substitution and partial integration with their broad field of application there are a whole range of further integration tricks, mostly for a limited classes of integrals. As representative for these, we want to take a closer look at the **Hermite ansatz**, which expresses integrals of type $\int dxP_m(x)/\Gamma(x)$ with a polynomial $P_m(x)$ of the m-th degree over a root $\Gamma(x) := \sqrt{(ax^2+bx+c)}$ of a quadratic expression, in terms of the corresponding integral with the polynomial 1, because these kinds of integrands are very common in physics. Hermite made the following ansatz:

Hermite ansatz:
$\int dxP_m(x)/\Gamma(x) = Q_{m-1}(x)\Gamma(x) + a_m \int dx/\Gamma(x)$ with $Q_{m-1}(x) := \sum_{n=0}^{m-1} a_n x^n$

with a polynomial of the $(m-1)$-th degree $Q_{m-1}(x)$ and real numbers a_m. We get the required $(m+1)$ numbers a_n with $n = 0, 1, 2, \ldots, m$ through coefficient comparison of the following two polynomials of m-th degree:

$$P_m(x) = Q'_{m-1}(x)\Gamma^2(x) + (ax + b/2)Q_{m-1}(x) + a_m$$

The differentiation of the Hermite ansatz yields

$$\begin{aligned} P_m(x)/\Gamma(x) &= Q'_{m-1}(x)\Gamma(x) + Q_{m-1}(x)\Gamma'(x) + a_m/\Gamma(x) \\ &= Q'_{m-1}(x)\Gamma(x) + Q_{m-1}(x)(2ax+b)/2\Gamma(x) + a_m/\Gamma(x) \end{aligned}$$

and multiplication with $\Gamma(x)$ gives the above determination equation for the $(m+1)$ coefficients a_n.

Exercise 7.12 *Show that the coefficient comparison*

a) for $P_3(x) = 3x^3 + 5x^2 + 3x$ *and* $\Gamma(x) = \sqrt{(x^2 + 2x + 2)}$ *leads to* $a_3 = a_2 = 1$, $a_1 = 0$ *and* $a_0 = -1$, *so that* $Q_2(x) = x^2 - 1$ *follows and*

b) for $P_2(x) = x^2$ *and* $\Gamma(x) = \sqrt{1 - x^2}$ *leads to* $a_2 = 1/2$, $a_1 = -1/2$ *and* $a_0 = 0$ *with* $Q_1(x) = -x/2$ *as solution.*

The remaining **integrals of type $\int dx/\Gamma(x)$** we solve by a trick, known to many of you from school, namely the **quadratic completion**, i.e. addition and subtraction of the same term (here bold printed): if e.g. $a > 0$

$$\begin{aligned} \int dx/\Gamma(x) &= \int dx/\sqrt{(ax^2 + bx + c)} \\ &= (1/\sqrt{a}) \int dx/\sqrt{x^2 + bx/a + \mathbf{(b/2a)^2} + c/a - \mathbf{(b/2a)^2}} \\ &= (1/\sqrt{a}) \int dx/\sqrt{(x + b/2a)^2 + \Delta/4a^2} \end{aligned}$$

with the discriminant $\Delta := 4ac - b^2$.

Finally the substitution $\boldsymbol{y := x + b/2a}$ brings us to

$$= (1/\sqrt{a}) \int dy/\sqrt{y^2 + \Delta/4a^2}$$

and then, if $\Delta > 0$, the further substitution $\boldsymbol{z := 2ay/\sqrt{\Delta}}$ with $\frac{dz}{dy} = 2a/\sqrt{\Delta}$ to

$$= (1/\sqrt{a}) \int dz/\sqrt{z^2 + 1}$$

which we find in line 20 of our **TABLE**

$$= (1/\sqrt{a})\text{arsinh}\, z + d = (1/\sqrt{a}) \ln(z + \sqrt{(z^2 + 1)}) + d$$

as we have shown earlier in Exercise 4.13. Altogether with other constants c :

$$\begin{aligned}
&= (1/\sqrt{a}) \ln d(z + \sqrt{(z^2 + 1)}) \\
&= (1/\sqrt{a}) \ln d(y + \sqrt{(y^2 + \Delta/4a^2)}) \\
&= (1/\sqrt{a}) \ln d(x + b/2a + \sqrt{(x^2 + bx/a + c/a)}).
\end{aligned}$$

Exercise 7.13 *Solve the integral* $\int dx/\Gamma(x)$ *for the case* $a < 0$ *and* $\Delta < 0$.

All integrals, which were solved over the years with different formulae and tricks, are collected in integral tables, some of which we have listed below: We start our collection with the smaller German books for daily use of every physics student and continue on to the large comprehensive works stored for you in the libraries:

1. K.ROTTMANN: Mathematische Formelsammlung, BI,
2. I.N.BRONSTEIN + K.A.SEMENDJAJEW: Taschenbuch der Mathematik, H.Deutsch,
3. W.GRÖBNER + N.HOFREITER: Integraltafel I + II, Springer,
4. M.ABRAMOWITZ + I.A.STEGUN: Handbook of Mathematical Functions, Wiley,
5. I.S.GRADSHTEYN + I.M.RYZHIK: Tables of Integrals, Series and Products, Academic,
6. A.P.PRUDNIKOV + Yu.A.BRYCHKOV + O.I.MARICHEV: Integrals and Series I + II, Gordon + Breach.

Nowadays the information collected in these tables is also incorporated in **programs** such as MATHEMATICA or MAPLE . These find the primitive functions, if they exist, even in difficult cases. You should not be afraid of using these tables and programs in your daily work. The use of these programs though requires in general some knowledge of the different integration techniques and tricks, because otherwise the given comments and constraints on the regions of validity can not be taken into account appropriately and correctly. This is the reason why we have studied these problems here.

7.5.6 Integral Functions

Despite all formulae and tricks some integrals, even ones needed in science remain unsolvable. Examples are the integrands $\exp(-x^2)$, e^x/x, $1/\sqrt{(1+x^4)}$, $\sin x/x$ or $1/\ln x$. Mathematicians can prove that the corresponding limit exists, but it cannot be expressed in a closed form using **elementary** functions.

In this situation we would like to remind you of the indefinite integral and our **TABLE**:

e.g. line 14: $\ln y = \int_1^y dx/x$
or line 10: $\arctan y = \int_0^y dx/(1+x^2)$
or line 8: $\arcsin y = \int_0^y dx/\sqrt{(1-x^2)}$.

If we had not gotten to know the functions on the left side as inverse functions or calculated them as Taylor series, we could consider them as defined by these equations.

According to these models we can treat other non-elementary integrals: We give a name to the analytic unsolvable integral and look for an other procedure for calculating the functional values. Here we give only two examples: The

error function: $\mathrm{erf}(y) := (2/\sqrt{\pi}) \int_0^y dx \exp(-x^2)$,

which plays a role in the calculus of accidental error, and the

elliptic integrals: $F(k;y) := \int_0^y dx/\sqrt{(1-x^2)(1-k^2x^2)}$,

which are needed for pendulum oscillations.

Exercise 7.14 *Show through an suitable substitution, that also the integrals*

$\int_0^y dx/\sqrt{(1-k^2\sin^2 x)}$ *and* $\int_0^y dx/\sqrt{\cos 2x}$ *are elliptic integrals.*

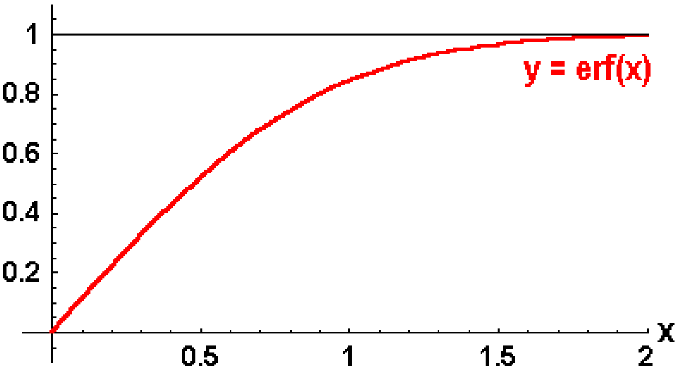


Figure 7.11: Error function

Obviously the beautiful name alone does not help. A method must be found which allows the calculation of the functional values. When all other means fail, we have no choice but the numerical integration.

7.5.7 Numerical Integration

The oldest and most primitive way to determine a definite integral, i.e. to calculate the area "under" a function "over" an interval, is of course to draw the integrand onto graph paper and then count the squares.

With today's electronic calculators we can do this more elegantly and much faster, by considering the defining limit and summing up the areas of the strips after the dissection of the interval. In doing so, the fewer strips and steps of iteration are needed, the closer we fit the upper edge of the strip to the actual function to be integrated.

The simplest choice are the horizontal straight lines as we have done in the definition, so that the strips become rectangles, whose areas are easy to calculate. Instead of the nodes ξ_n in the interior of the partial intervals just as with the Riemann sum, we can also take the minima and maxima of the function in the partial intervals. Then we obtain an under and over sum, which enclose the integral in the limit.

The next elegant method is the **chord-trapezium rule**, by which the function in each partial interval is approximated through the secant, and the real areas of the stripes through the areas of trapezia. For each partial interval we take instead of $\Delta x_n f(\xi_n)$:

$$(\Delta x_n)(f(x_{n-1}) + f(x_n))/2.$$

An even better approximation gives the **Simpson rule**, in which the integrand function in the partial intervals is approximated through a parabola according to **Kepler's barrel rule**. Here the function is taken at three points in each strip:

$$(\Delta x_n)(f(x_{n-1}) + 4f(\tfrac{x_{n-1}+x_n}{2}) + f(x_n))/6.$$

To find even more sophisticated rules through which the desired integral with a pre-given demand for accuracy can be calculated in the fewest number of steps is one of the tasks of numerical mathematics.

7.6 Improper Integrals

All that we have learned about integration is still not enough for physicists. There are two further conditions to which they need answers:

Firstly the **finite interval**, since in physics we often have to integrate up to infinity, and secondly the **bounded integrand**, because the integrands of physicists sometimes become infinite within the integration interval, and often that is the place where something interesting happens. Here, we would just like to indicate the considerations with which both of these challenges can be overcome by adding one further limit:

7.6.1 Infinite Integration Interval

If for example the integral over a function being continuous in the interval $[a, \infty)$ should extend on the upper border up to infinity, we calculate this only up to a large finite value y and let afterwards in the result of the integration the large value y in a further limiting procedure grow over all borders. If also this limit exists, we call it an **improper integral of the first kind** and write:

$$\boxed{F_a(\infty) \equiv \int_a^\infty dx f(x) := \lim_{y\to\infty} \int_a^y dx f(x) \equiv \lim_{y\to\infty} F_a(y).}$$

As an example we calculate for $a > 0$ and a tiny positive $\varepsilon > 0$:

$$\int_a^\infty \frac{dx}{x^{1+\varepsilon}} := \lim_{y\to\infty} \int_a^y dx\, x^{-1-\varepsilon} = \lim_{y\to\infty} \frac{x^{-1-\varepsilon+1}}{-1-\varepsilon+1}\Big|_a^y = \frac{-1}{\varepsilon}\Big(\lim_{y\to\infty} \frac{1}{y^\varepsilon} - \frac{1}{a^\varepsilon}\Big) = \frac{1}{\varepsilon a^\varepsilon}.$$

From this calculation we see that the improper integral exists, if the function for the growing x drops down even a little bit stronger than $1/x$, i.e. for instance $1/x^2$. However for $\varepsilon \to 0$ the coloured area in the following figure
under the function $1/x$ possesses **no** finite area and that the same holds for all less strongly dropping functions like e.g. $1/\sqrt{x}$.

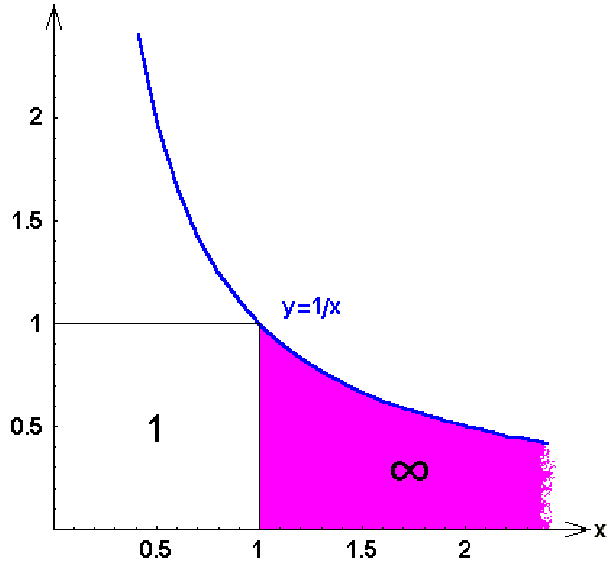


Figure 7.12 Improper integration interval

Exercise 7.15 *Try to calculate the following improper integrals of the first kind:*

a) $\int_a^\infty dx/x^2$, *b)* $\int_0^\infty dx\ \exp(-x)$, *c)* $\int_0^\infty dx/(1+x)$,

d) $\int_0^\infty dx\ \cos x$, *e)* $\int_0^\infty dx\ \cos x\ e^{-x}$

Analogously we proceed at the lower border with a function which is continuous and bounded in $(-\infty, b]$:

$$F_{-\infty}(b) \equiv \int_{-\infty}^b dx f(x) := \lim_{a\to-\infty} \int_a^y dx f(x) \equiv \lim_{a\to-\infty} F_a(y)$$

or for both borders with a function which is continuous and bounded on the whole real number axis:

$$\int_{-\infty}^\infty dx f(x) := \lim_{a\to-\infty} \int_a^c dx f(x) + \lim_{y\to\infty} \int_c^y dx f(x)$$

with an arbitrary division point c.

Exercise 7.16 *Calculate:* $\int_{-\infty}^{-2/\pi} dx\ \sin(1/x)/x^2$ *and* $\int_{-\infty}^{0} dx\, x/(1+x^4)$.

Insert: Cauchy Principal Value: *It is possible that both the limits exist only if the huge finite borders grow* ***in the same manner****. In this case the result is called*

$$\textbf{Cauchy principal value:}\quad P\int_{-\infty}^{\infty} dx f(x) := \lim_{c\to\infty}\int_{-c}^{c} dx f(x).$$

In order to illustrate this we calculate the following example for $n \in \mathbb{N}$*:*

$$\begin{aligned}\int_{-\infty}^{\infty} dx\, x^{2n-1} &:= \lim_{a\to-\infty}\int_{a}^{c} dx\, x^{2n-1} + \lim_{y\to\infty}\int_{c}^{y} dx\, x^{2n-1}\\ &= \lim_{a\to-\infty}(-a^{2n}/2n) + \lim_{y\to\infty} y^{2n}/2n.\end{aligned}$$

Apparently both the limits do not exist. If we, however, form the principal value, there follows:

$$P\int_{-\infty}^{\infty} dx\, x^{2n-1} := \lim_{c\to\infty}\int_{-c}^{c} dx\, x^{2n-1} = \lim_{c\to\infty} 0 = 0,$$

since once again the integral of an odd function over an interval symmetric to the origin vanishes.

Exercise 7.17 *Calculate:* $\int_{-\infty}^{\infty} dx/(1+x^2)$ *and* $\int_{-\infty}^{\infty} dx\, x/(1+x^4)$.

7.6.2 Unbounded Integrand

We now take a quick look at the second case, namely when the integrand is unbounded at a point x_0 of the finite integration interval $[a, b]$, e.g. on the **lower** border: $x_0 = a$. We calculate in this case the integral starting from the value $x_0 + \eta$, which lies only a tiny distance $\eta > 0$ above the critical point x_0 and let only in the result of the integration this tiny distance approach zero. If this limit exists, we call it an **improper integral of the second kind** and write:

$$F_{x_0}(b) \equiv \int_{x_0}^{b} dx f(x) := \lim_{\eta\to 0} \int_{x_0}^{b} dx f(x) \equiv \lim_{\eta\to 0} F_{x_0+\eta}(b).$$

As an example we calculate for $b > 0$ and a tiny positive $\varepsilon > 0$:

$$\int_0^b \frac{dx}{x^{1-\varepsilon}} := \lim_{\eta\to 0} \int_\eta^b dx\, x^{\varepsilon-1} = \lim_{\eta\to 0} \frac{x^{\varepsilon-1+1}}{\varepsilon-1+1}\Big|_\eta^b = \frac{1}{\varepsilon}\big(b^\varepsilon - \lim_{\eta\to 0}\eta^\varepsilon\big) = \frac{b^\varepsilon}{\varepsilon}.$$

From this we see that the improper integral does in fact exist, if the function for decreasing x rises slightly weaker than $1/x$, hence e.g. for $1/\sqrt{x}$. However for $\varepsilon \to 0$ in the following figure

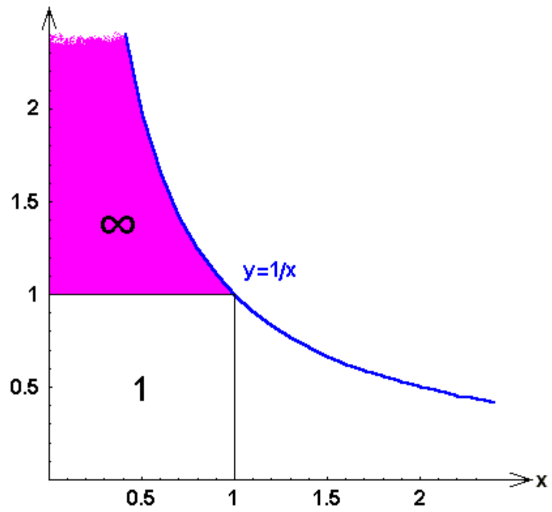


Figure 7.13 Unbounded Integrand

the colored area under the function $1/x$ (which is exactly the image of the earlier considered area reflected through the bisector line) does **not** possess a finite area **any more** and this holds true also for all stronger increasing functions like e.g. $1/x^2$.

Exercise 7.18 *Calculate the following improper integrals of the second kind:*

a) $\int_0^b dx/\sqrt{x}$, *b)* $\int_1^2 dx/\sqrt{x-1}$, *c)* $\int_0^b dx/x^3$

Again we proceed analogously with a function which is unbounded at the **upper** border:

$$F_a(x_0) \equiv \int_a^{x_0} dx f(x) := \lim_{\varepsilon \to 0} \int_a^{x_0 - \varepsilon} dx f(x) \equiv \lim_{\varepsilon \to 0} F_a(x_0 - \varepsilon)$$

or in case the function is unbounded within the integration interval at a point x_0:

$$\int_a^b dx f(x) := \lim_{\varepsilon \to 0} \int_a^{x_0 - \varepsilon} dx f(x) + \lim_{\eta \to 0} \int_{x_0 + \eta}^b dx f(x).$$

Insert: Cauchy Principal Value: *Here it is possible that both the limits exist only, if the tiny distances ε and η off the critical point vanish* ***in the same manner****. Also in this case the result is called*

$$\textbf{Cauchy principal value:} \quad P \int_a^b dx f(x) := \lim_{\varepsilon \to 0} \big(\int_a^{x_0 - \varepsilon} dx f(x) + \int_{x_0 + \varepsilon}^b dx f(x) \big).$$

The following example may explain this for $n \in \mathbb{N}$:

$$\begin{aligned} \int_{a<0}^{b>0} dx/x^{2n+1} &:= \lim_{\varepsilon \to 0} \int_a^{-\varepsilon} dx/x^{2n+1} + \lim_{\eta \to 0} \int_\eta^b dx/x^{2n+1} \\ &= \lim_{\varepsilon \to 0} x^{-2n}/(-2n)\big|_a^{-\varepsilon} + \lim_{\eta \to 0} x^{-2n}/(-2n)\big|_\eta^b \end{aligned}$$

Also here both the limits do not exist. If we, however, form the principal value e.g. for $a = -b$, it follows:

$$P \int_{-b}^b dx/x^{2n+1} := \lim_{\eta \to 0} \int_{-b}^b dx/x^{2n+1} = \lim_{\eta \to 0} 0 = 0,$$

since once again the integral of an odd function over an interval symmetric to the origin vanishes.

Exercise 7.19 *Calculate:* $\int_0^1 dx/\sqrt{1 - x^2}$ *and* $\int_0^{\pi/2} dx \, \tan x$.

Exercise 7.20 *Calculate:* $P \int_{-1}^2 dx/x$ *and* $P \int_0^\pi dx \, \tan x$.

Exercise 7.21 *Show that the improper integral of the second kind:* $\int_0^1 dx/\sqrt{x}$ *through the substitution* $x = 1/y^2$ *becomes an improper integral of the first kind.*

We do not want to occupy ourselves with integrals which are of first as well as of second kind, although we could master even these through cleanly separated limits.

Exercise 7.22 Examples of physical integrals

a) If you would calculate the ***falling time of the moon****, which is needed to fall vertically onto the earth surface, if the moon suddenly would stop in its path, you would need the following integral:* $J = \int dx\, \sqrt{x}/\sqrt{1-x}$.

b) In order to calculate the ***electrical potential*** $V(x)$ *of a* ***homogeneously charged spherical shell*** *(between the inner radius* r *and the outer radius* R*) at the distance* x *from the centre point, besides the integral* $\int_{-1}^{1} dz/\sqrt{az+b}$ *in the numerator of the Coulomb factor* $1/x$ *you also need the following more complicated integral* $Z(x) = \int_r^R dy\, y(|y-x| - (y+x))$*, with which you can very well practice the putting together of integrals.*

Later on, especially treating functions of several variables, you will learn about a variety of further integrals: curve or line integrals, surface and volume integrals in spaces of various dimensions. But whenever it comes to calculating numbers to compare them with measurements, you will not do anything other than calculating Riemann integrals as we have learned together here.

Exercise 7.23 A multiple integral

At the end of this chapter we take a look at multiple integrals, so that you can see based on a simple example that you are able to calculate even much more complicated kinds of integrals with the techniques learned here: For instance the area of a ***circular disc*** *with radius* r *(located in the first quadrant and touching the coordinate axes) as a double integral: The equation of the circle is:* $(x_1 - r)^2 + (x_2 - r)^2 = r^2$*, thus* $x_1 = r \pm \Gamma$ *with* $\Gamma = \sqrt{r^2 - (x_2 - r)^2}$.
Later you will learn that the following double integral then describes the area content: $F = \int_0^{2r} dx_2 \int_{r-\Gamma}^{r+\Gamma} dx_1$.

Chapter 8

COMPLEX NUMBERS

8.1 Imaginary Unit and Illustrations

8.1.1 Motivation

In Section 2.2.4 we decided, mainly for fundamental and mathematical reasons, to use the **field of real numbers** as the basis of our physical considerations, although the finite precision of all physical measurements would have allowed us to use only the rational numbers. In this chapter, we want to once more extend the field of numbers without any compelling physical need, namely to the **complex numbers** $\mathbb{C}$.

Although there are many mathematical reasons to extend the field of numbers so that it contains also the solution of the equation $x^2 = -1$, and with this extension many unexpected and deep insights become possible, the complex numbers are for physics only a typical **tool**. At no point in physics there is a true need to use complex numbers. Nevertheless they are a very convenient way to describe various phenomena in many areas of physics. You can not yet imagine how convenient something unnecessary may be! Complex numbers are not only helpful everyday tools for all oscillation processes, and the wide field of electrical engineering, but also especially quantum mechanics makes heavy use of complex quantities. The decisive difference between the Schrödinger equation and the heat conduction equation is the imaginary unit "i". Without the use of complex numbers we would have to deal with two complicated coupled real differential equations instead of solving a single simple complex one. Therefore, this chapter is also a direct preparation for your lecture on quantum mechanics.

The surprising results of complex numbers in mathematics are interesting in themselves and well worth the effort: with the extension from $\mathbb{R}$ to $\mathbb{C}$ not only $x^2 = -1$ becomes solvable, but every equation of second degree. And every algebraic equation of n-th degree has according to the Fundamental Theorem of Algebra exactly n solutions in $\mathbb{C}$. Each rational function has an expansion into partial fractions. Only now does the convergence radius of power series become understandable from the definition domain, and a lot of other things besides.

So we ask ourselves: What do we have to do to get to this paradise?

8.1.2 Imaginary Unit

We have seen that with real numbers every equation $x^2 = a > 0$ is solvable: $x = \pm\sqrt{a}$, but no real number $x \in \mathbb{R}$ exists, for which $x^2 + 1 = 0$. You see this indirectly: If such a real number $x \neq 0$ would exist, then this number on the real number line should be either $x < 0$ or $x > 0$, so the square of it should be $x^2 = x \cdot x = (-x)(-x) > 0$. Consequently, $x^2 + 1 > x^2 > 0$ should also hold, and that would be a contradiction to the equation: $x^2 + 1 = 0$ from which we started.

An extension of real numbers calls for at least the addition of the solution $z \in \mathbb{C}$ of the equation $z^2 + 1 = 0$. This problem was solved by Euler in 1777 with an ingenious trick by simply giving a name to the unknown new number:

$$\textbf{imaginary unit: } \mathbf{i^2 := -1}.$$

With this, we can write the solution of the equation $z^2 + 1 = 0$ simply in the following way: $z = \pm\mathrm{i}$.

Next, we draw some direct conclusions from this definition for the powers of i, where we try to **maintain** all known **calculation rules** from the field of real numbers:

$$\mathrm{i} := +\sqrt{-1}, \quad \mathrm{i}^2 = -1, \quad \mathrm{i}^3 = \mathrm{i}^2\mathrm{i} = -\mathrm{i}, \quad \mathrm{i}^4 = \mathrm{i}^2\mathrm{i}^2 = (-1)(-1) = +1, \quad \textit{etc.}$$

$$\mathrm{i}^{4n+1} = \mathrm{i}, \quad \mathrm{i}^{4n+2} = -1, \quad \mathrm{i}^{4n+3} = -\mathrm{i}, \quad \mathrm{i}^{4n+4} = \mathrm{i}^{4n} = +1 \text{ with } n \in \mathbb{N}.$$

Also for the negative powers we can easily derive rules: first it follows for i, the inverse number of i^{-1} from

$$\mathrm{i}\mathrm{i}^{-1} = 1 = \mathrm{i}^4 = \mathrm{i}\mathrm{i}^3 \quad \Rightarrow \quad \mathrm{i}^{-1} = \mathrm{i}^3 = -\mathrm{i}$$

and then:

$$\mathrm{i}^{-2} = \frac{1}{\mathrm{i}^2} = \frac{1}{-1} = -1, \quad \mathrm{i}^{-3} = \mathrm{i}, \quad \text{and so on.}$$

This means that the framed results above even hold true for all integers $n \in \mathbb{Z}$.

Exercise 8.1 Imaginary Unit *Calculate:* $\mathrm{i}^{15}, \mathrm{i}^{45}$ *and* $(-\mathrm{i})^{-20}$.

8.1.3 Definition of complex numbers

With this first success we go on to the general equation $z^2 = -b^2 < 0$ and derive the solution $z = \pm\sqrt{-b^2} = \pm b\sqrt{-1} = \pm \mathrm{i}b$. We call a real number multiplied with the imaginary unit i an

imaginary number: $z := \mathrm{i}y \quad$ with $y \in \mathbb{R}$.

If we continue looking forward toward even more general equations, namely $(z-a)^2+b^2 = 0$, we obtain the solution $z = a \pm \mathrm{i}b$, as a linear combination of a real and an imaginary number. This we call a general

complex number: $z := x + \mathrm{i}y \quad$ with $x, y \in \mathbb{R}$.

Thus, a complex number is an **ordered pair** of unambiguously defined real numbers: the purely real first part is called the

real part of the complex number: $x = \mathrm{Re}\, z \ \in \mathbb{R}$,

and the second part equipped by the factor i is called the

imaginary part of the complex number: $y = \mathrm{Im}\, z \ \in \mathbb{R}$.

This decomposition into real and imaginary parts is unambiguous in contrast to the rational numbers, which in the past we introduced also as "ordered pairs" of integers. Then however, we identified whole equivalence classes: (1,2) = (2,4) = (3,6) = ... since cancelling should be possible without changing the number:
$\frac{1}{2} = \frac{2}{4} = \frac{3}{6} = \ldots$.

The **equality** $z = w$ of two complex numbers $z = x + \mathrm{i}y$ and $w = u + \mathrm{i}v$ means the equality of both the real and imaginary parts: $x = u$ and $y = v$, i.e.

a complex equation $\boldsymbol{z = w}$ includes two real equations: $\mathbf{Re}\,\boldsymbol{z} = \mathbf{Re}\,\boldsymbol{w}$ and $\mathbf{Im}\,\boldsymbol{z} = \mathbf{Im}\,\boldsymbol{w}$.

Especially, a complex number vanishes if and only if the real- and imaginary parts both are equal to zero:

$$z = x + \mathrm{i}y = 0 \iff \mathrm{Re}\,z = x = 0 \textbf{ and } \mathrm{Im}\,z = y = 0.$$

The real numbers $\mathbb{R} : z = x$ are a subset of the set of complex numbers: $\mathbb{R} \subset \mathbb{C}$, namely all those with $\mathrm{Im}\,z = y = 0$. To these we just added the purely imaginary numbers $z = \mathrm{i}y$ as new elements.

Before we turn our attention to calculation rules, we want to get an overview over the methods that enable us to visualize the complex numbers:

8.1.4 Gauss Number Plane

To represent an ordered pair of real numbers the plane is very useful. We have already used planes to illustrate variable and functional value of a real function. Now we want to introduce the **Gauss number plane**: to every point (or "pointer" as electronics engineers say) of the Gauss number plane there corresponds exactly one complex number.

As an orientation guide we draw in the plane two real number lines standing perpendicular to each other, the real axis $\mathbb{R}_x$ and the imaginary axis $\mathbb{R}_y$, meaning we choose a **Cartesian coordinate system**: the real part x of a complex number z, i.e. of a point (or pointer) z, is then the projection of its distance from the origin (or its length) on the real 1-axis, and the imaginary part y accordingly on the imaginary 2-axis, just as it is shown in the following figure:

As an alternative to Cartesian coordinates in the plane we can of course also use **planar polar coordinates**, by writing

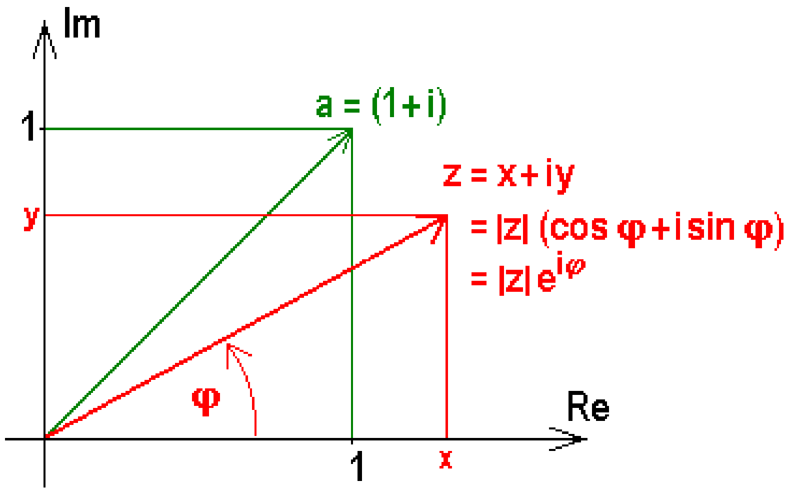


Figure 8.1: Gauss number plane with Cartesian coordinate system

$$z = |z|(\cos\varphi + \mathrm{i}\sin\varphi)$$

with the

real part $\mathrm{Re}\, z = x = |z|\cos\varphi$

and the

imaginary part $\mathrm{Im}\, z = y = |z|\sin\varphi$.

From this it follows according to the Pythagoras theorem for the

absolute value of a complex number: $0 \leq |z| := +\sqrt{(x^2 + y^2)} < \infty$.

The polar angle which you get from $\frac{y}{x} = \tan\varphi$ is only determined up to additive terms 2π and is called

argument of a complex numbers: **e.g.** $0 \leq \varphi = \arg(z) < 2\pi$.

Determining the argument we come across a small difficulty, an **ambiguity** which occurs because of $\frac{y}{x} = \frac{-y}{-x}$:
When we for example look at the complex number $a = 1+i$ with the real part Re $a = 1$ and the imaginary part Im $a = 1$, we obtain, for the absolute value unambiguously $|a| = \sqrt{2}$ for the argument, however, at first glance **two** values $\alpha = \arg a = \arctan 1 = \frac{\pi}{4}$ **or** $\frac{5}{4}\pi$, which both lie in the interval $[0, 2\pi)$. By inserting both values in question in Re a we can, however, find the right argument: for $\alpha = \frac{\pi}{4}$ we correctly find $\mathrm{Re}\, a = |a| \cos\alpha = \sqrt{2}\cos\frac{\pi}{4} = \frac{\sqrt{2}}{\sqrt{2}} = +1$, while $\alpha = \frac{5}{4}\pi$ gives us the wrong result $\mathrm{Re}\, a = \sqrt{2}\cos\frac{5}{4}\pi = \frac{\sqrt{2}}{-\sqrt{2}} = -1$.

Insert: Precise calculation of the arguments: *For precise calculation of the argument of a complex number from its real and imaginary part within our chosen borders you can use the following equations:*
for $x > 0$ *and* $y \geq 0$ *use* $\varphi = \arg(x+iy) = \arctan\frac{y}{x}$,
for $x > 0$ *and* $y < 0$ *use* $\varphi = \arg(x+iy) = \arctan\frac{y}{x} + 2\pi$,
for $x < 0$ *and* y *arb. use* $\varphi = \arg(x+iy) = \arctan\frac{y}{x} + \pi$,
for $x = 0$ *and* $y > 0$ *use* $\varphi = \arg(x+iy) = \pi/2$,
for $x = 0$ *and* $y < 0$ *use* $\varphi = \arg(x+iy) = 3\pi/2$ *and*
for $x = 0$ *and* $y = 0$ *the argument is indeterminate.*

Insert: Alternative choice of the borders of the arguments: *Looking forward to future applications, many textbooks, computer programs and ISO- or DIN-standardization conventions choose for the arguments of the complex numbers instead of the polar angle interval (you are accustomed to from school) an interval of length* 2π *which is symmetrical around the real axis. They call these* **main arguments of a complex number** *and sometimes signify this choice by a capital A:* $-\pi < \varphi = Arg(z) \leq \pi$.

In the case of these symmetrical borders you get the main argument of a complex number from its real and imaginary part out of the following equations:
for $x > 0$ *and* y *arb. out of* $\varphi = Arg(x+iy) = \arctan\frac{y}{x}$,
for $x < 0$ *and* $y \geq 0$ *out of* $\varphi = Arg(x+iy) = \arctan\frac{y}{x} + \pi$,
for $x < 0$ *and* $y < 0$ *out of* $\varphi = Arg(x+iy) = \arctan\frac{y}{x} - \pi$,
for $x = 0$ *and* $y > 0$ *out of* $\varphi = Arg(x+iy) = \pi/2$,
for $x = 0$ *and* $y < 0$ *out of* $\varphi = Arg(x+iy) = -\pi/2$ *and*
for $x = 0$ *and* $y = 0$ *once more the main argument is indeterminate.*

Exercise 8.2 Argument of a complex number:

Determine the argument of the complex number $b = 1 - \mathrm{i}$.

8.1.5 Euler's Formula

When we insert for cosine and sine the Taylor series into the representation of complex numbers in planar polar coordinates, we come across an interesting relation:

$$\begin{aligned}
\frac{z}{|z|} &= \cos\varphi + \mathrm{i}\sin\varphi \\
&= \sum_{n=0}^{\infty}\frac{(-1)^n\varphi^{2n}}{(2n)!} + \mathrm{i}\sum_{n=0}^{\infty}\frac{(-1)^n\varphi^{2n+1}}{(2n+1)!} \quad \text{If in this we insert the relation } -1 = \mathrm{i}^2 \text{ we obtain:} \\
&= \sum_{n=0}^{\infty}\frac{(\mathrm{i}^2)^n\varphi^{2n}}{(2n)!} + \mathrm{i}\sum_{n=0}^{\infty}\frac{(\mathrm{i}^2)^n\varphi^{2n+1}}{(2n+1)!} \quad \text{and contracted:} \\
&= \sum_{n=0}^{\infty}\frac{(\mathrm{i}\varphi)^{2n}}{(2n)!} + \sum_{n=0}^{\infty}\frac{(\mathrm{i}\varphi)^{2n+1}}{(2n+1)!}. \quad \text{Written out, we get:} \\
&= 1 + \frac{(\mathrm{i}\varphi)^2}{2!} + \frac{(\mathrm{i}\varphi)^4}{4!} + \frac{(\mathrm{i}\varphi)^6}{6!} + \ldots + \frac{(\mathrm{i}\varphi)}{1!} + \frac{(\mathrm{i}\varphi)^3}{3!} + \frac{(\mathrm{i}\varphi)^5}{5!} + \frac{(\mathrm{i}\varphi)^7}{7!} + \ldots \\
&= \sum_{n=0}^{\infty}\frac{(\mathrm{i}\varphi)^n}{n!}.
\end{aligned}$$

This series looks familiar to us: We recognize the **Taylor series** which enabled us to calculate the functional values of the exponential function. The difference is, an "*i*" is now preceding the real variable φ. Through this discovery we feel encouraged to define the exponential function for an imaginary variable through the above mentioned series:

exponential function for an imaginary variable: $e^{\mathrm{i}\varphi} := \sum_{n=0}^{\infty}\frac{(\mathrm{i}\varphi)^n}{n!}$.

With this definition the derived relation takes a very simple form, which has become famous as:

Euler's formula: $e^{\mathrm{i}\varphi} := \cos\varphi + \mathrm{i}\sin\varphi$.

From this, we extract $Re\, e^{\mathrm{i}\varphi} = \cos\varphi$ and $\mathrm{Im}\ e^{\mathrm{i}\varphi} = \sin\varphi$. Since cosine and sine are periodic functions with the period 2π, accordingly the exponential function of an imaginary variables must be a 2π-periodic function, too:

$$\textbf{2}\pi\textbf{i-periodic: } e^{i(\varphi+2\pi k)} = e^{i\varphi} \textbf{ with } k \in \mathbb{Z}.$$

With Euler's formula we have derived a third very popular representation of complex numbers beside the ones in Cartesian and polar coordinates:

$$\textbf{exponential representation of a complex number: } z = |z|e^{i\varphi}.$$

Especially for the following complex numbers on the unit circle: $|z| = 1$ we put together some important relations for you to learn by heart:

$$1 = e^{0i} = e^{2\pi i}, \quad -1 = e^{\pm i\pi}, \quad i = e^{\frac{i\pi}{2}}, \quad -i = e^{\frac{-i\pi}{2}}$$

8.1.6 Complex Conjugation

Here we want to illustrate a very often used transformation of complex numbers in the Gauss number plane: The complex conjugation assigns to every complex number z its complex conjugate z^* by reversing the sign of the Euler "i" wherever it appears (Mathematicians often use instead of the star a line just above the symbol, which is not available to us here). In the Gauss number plane, complex conjugation means obviously the reflection of complex numbers through the real axis: all imaginary parts and arguments suffer a sign reversal, the real part and the absolute value remain unchanged:

complex conjugation:

$$z = x + iy = |z|e^{i\varphi} \quad \Rightarrow \quad z^* = x - iy = |z|(\cos\varphi - i\sin\varphi) = |z|e^{-i\varphi}$$

i.e. $\mathrm{Re}\, z^* = \mathrm{Re}\, z, \quad \mathrm{Im}\, z^* = -\,\mathrm{Im}\, z, \quad |z^*| = |z|$ and $\arg z^* = -\arg z$.

For example for $a = 1 + i$ follows $a^* = (1 + i)^* = 1 - i$.
Through a double reflection we receive again the original number back: $(z^*)^* = z$.

Exercise 8.3 Complex conjugation:

Calculate for the following complex number $c := 3 + 3i\sqrt{3}$:
$\mathrm{Re}\, c, \quad \mathrm{Im}\, c, \quad |c|, \quad \arg c, \quad c^*, \quad c + c^*, \quad c - c^*.$

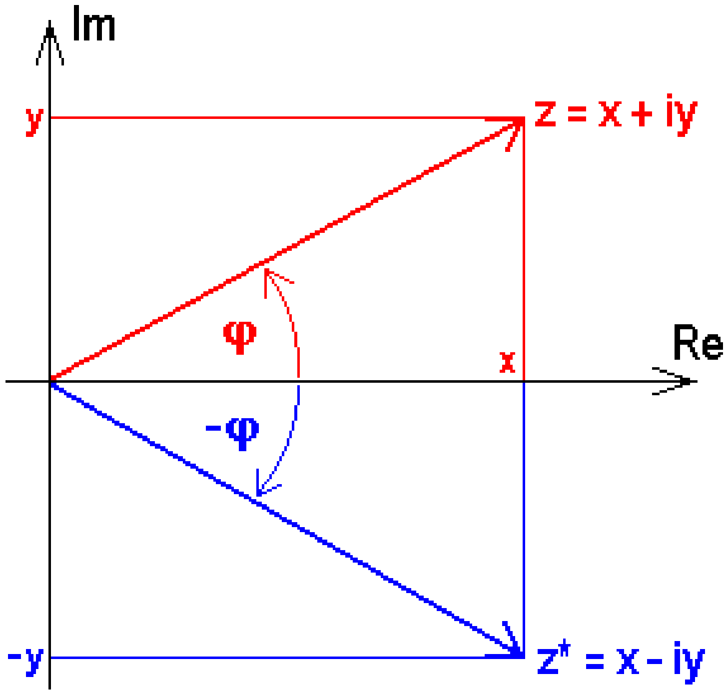


Figure 8.2: Complex conjugation

Insert: Number Sphere: *Finally we mention an alternative to the representation of the complex numbers in the Gauss number plane: the Riemann* ***number sphere****:*

In order to do this we imagine a sphere of diameter 1 positioned with its South Pole touching the origin of the Gauss plane, and all points of the plane connected by straight lines with the North Pole of the sphere, as sketched in the following figure:

By means of this "stereographic projection" (as it is called by mathematicians) each point of the plane is unambiguously assigned to one point on the surface of the sphere which can serve as alternative representative for the corresponding complex number: You see immediately that the origin is mapped onto the "South Pole", the interior of the unit circle onto the "southern hemisphere", the unit circle onto the "equator" and the region outside the unit circle onto the "northern hemisphere". The most interesting feature of this illustration, however, is the "North Pole" which turns out to be the continuous image of all infinitely far-off points of the plane. This expresses the fact that from the viewpoint of complex numbers there exists only one "number" ∞ *and that in its neighbourhood the same arguments go through as for every other complex number. At some later time in your studies you will return to this representation and consider the neighbourhood of the point* ∞ *on the Riemann*

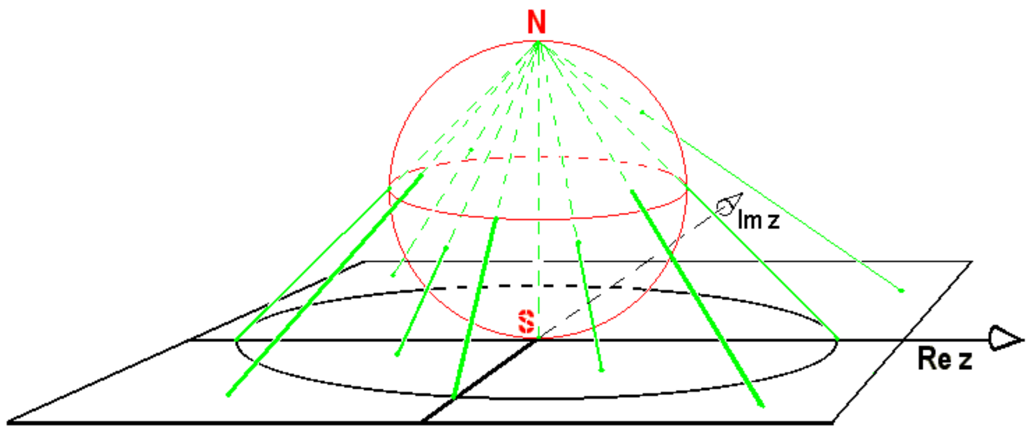


Figure 8.3: Riemann number sphere on top of the Gauss plane

sphere. We will not further use it in this course.

8.2 Calculation Rules of Complex Numbers

The ingenious invention of the number “i” solves automatically all calculational problems for complex numbers. In the following we observe how the field characteristics of real numbers $\mathbb{R}$ transfer onto the complex ones $\mathbb{C}$ and endeavour to obtain their meaning and illustration in the Gauss plane:

8.2.1 Abelian Group of Addition

The complex numbers form an Abelian group of addition like the real ones: If $z = x + iy$ and $w = u + iv$ are two complex numbers, their sum is:

$$\textbf{sum:} \quad \boldsymbol{z + w = x + iy + u + iv = (x + u) + i(y + v),}$$
$$\text{i.e. } \mathbf{Re}(\boldsymbol{z+w}) = \mathbf{Re}\,\boldsymbol{z} + \mathbf{Re}\,\boldsymbol{w} \text{ and } \mathbf{Im}(\boldsymbol{z+w}) = \mathbf{Im}\,\boldsymbol{z} + \mathbf{Im}\,\boldsymbol{w}.$$

The visualization is most excellently achieved in the Gaussian plane with help of a Cartesian coordinate system: The “pointer” of the sum is the one physicists know as the resultant force from the parallelogram of forces, as you can see in the next figure:

The **group laws** arise simply from the corresponding relations of the real numbers for real and imaginary parts.

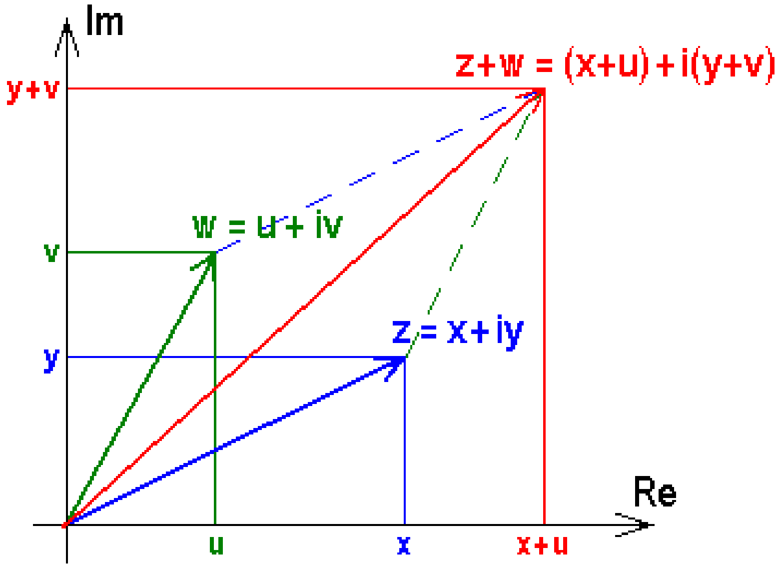


Figure 8.4: Addition of complex numbers

Insert: Group Laws:

Commutative Law:	$a + b = b + a$
Associative Law:	$(a + b) + c = a + (b + c)$
Zero element:	$\exists!\, 0 := 0 + \mathrm{i}0 \quad \forall z \in \mathbb{C} : \quad z + 0 = z$
Negative elements:	$\forall z \in \mathbb{C} \quad \exists!\, -z := -x - \mathrm{i}y : \quad z + (-z) = 0.$

To the unambiguously determined **negative** $-z$ of every complex number z, there corresponds in the Gaussian plane the mirror point (or pointer) which you obtain by reflection through the origin. With that, **subtraction** of complex numbers is possible just as you are familiar with from the real numbers: The **difference** $a-b$ is the unambiguous solution of the equation $z + b = a$.

The **absolute value** of the difference $|a-b| = \sqrt{(\mathrm{Re}\, a - \mathrm{Re}\, b)^2 + (\mathrm{Im}\, a - \mathrm{Im}\, b)^2} = |b-a|$ is the distance between the corresponding points or pointer tips, respectively in the plane.

In particular, the triangle inequality holds true:

triangle inequality: $|a + b| \leq |a| + |b|.$

Insert: Triangle Inequality: *The proof here anticipates the multiplication rule, but may serve as practice in the calculation with complex numbers: We consider*

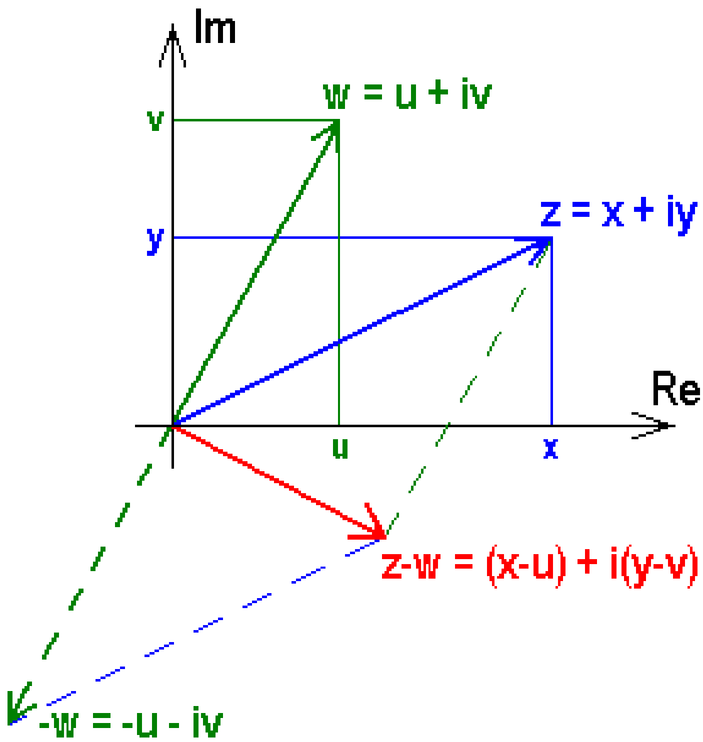


Figure 8.5: Subtraction of complex numbers

the square:

$$
\begin{aligned}
|a+b|^2 &= (a+b)(a+b)^* \\
&= (a+b)(a^*+b^*) \\
&= |a|^2+|b|^2+ab^*+a^*b \\
&= |a|^2+|b|^2+ab^*+(ab^*)^* \\
&= |a|^2+|b|^2+2\,\mathrm{Re}(ab^*) \\
&\leq |a|^2+|b|^2+2|ab^*|, \ \textit{because for every complex number } \mathrm{Re}\, z \leq |z| \\
&= |a|^2+|b|^2+2|a||b^*| \\
&= |a|^2+|b|^2+2|a||b| \\
&= (|a|+|b|)^2.
\end{aligned}
$$

From this the triangle inequality follows as the positive square root of both sides.

For the complex numbers themselves there exist no more inequalities. Apparently it is impossible to decide for any two complex numbers which of them is the larger one. This is the important difference from the real numbers which can be arranged along the number line and the price we have to pay for the extension. There is however yet a "**memory of ordering**", namely from $a \neq b$ follows $a + c \neq b + c$ just as before.

The complex conjugate of a sum is the sum of the conjugate summands:

$$(z + w)^* = (z^* + w^*)$$

meaning, the star can be drawn into the brackets. From the sum resp. difference of a complex number and its complex conjugate we can easily obtain the real resp. imaginary part:

$$\mathrm{Re}\, z = \frac{z + z^*}{2} \text{ resp. } \mathrm{Im}\, z = \frac{z - z^*}{2\mathrm{i}}.$$

The complex conjugate $z^* = |z|(\cos\varphi - \mathrm{i}\sin\varphi) = |z|e^{-\mathrm{i}\varphi}$ also allows us the reversal of the Euler formula:

$$\cos\varphi = \frac{z + z^*}{2|z|} \quad \text{and} \quad \sin\varphi = \frac{z - z^*}{2\mathrm{i}|z|}.$$

8.2.2 Abelian Group of Multiplication

Also for the multiplication rule Euler's "i" regulates everything automatically. We can simply multiply both the complex numbers $z = x + \mathrm{i}y$ and $w = u + \mathrm{i}v$ with each other applying the laws known from the real numbers and taking into account that $\mathrm{i}^2 = -1$:

$$zw = (x + \mathrm{i}y)(u + \mathrm{i}v) = xu + \mathrm{i}(yu + xv) + \mathrm{i}^2 yv = (xu - yv) + \mathrm{i}(yu + xv)$$

Thus, we get a fairly complicated expression for the

$$\textbf{product: } zw = (xu - yv) + \mathrm{i}(yu + xv).$$

This expression cannot easily be visualized in Cartesian coordinates, even if we write it in polar coordinates using $z = |z|(\cos\varphi + \mathrm{i}\sin\varphi)$ resp. $w = |w|(\cos\omega + \mathrm{i}\sin\omega)$:

$$\begin{aligned} zw &= |z|(\cos\varphi + \mathrm{i}\sin\varphi)|w|(\cos\omega + \mathrm{i}\sin\omega) \\ &= |z||w|((\cos\varphi\cos\omega - \sin\varphi\sin\omega) + \mathrm{i}(\sin\varphi\cos\omega + \cos\varphi\sin\omega)). \end{aligned}$$

Therefore we go over to the exponential representation by the Euler formula $z = |z|e^{\mathrm{i}\varphi}$ and $w = |w|e^{i\omega}$, which satisfies our need for visualization:

product: $zw = |z|e^{\mathrm{i}\varphi}|w|e^{i\omega} = |z||w|e^{\mathrm{i}(\varphi+\omega)} = |z||w|(\cos(\varphi+\omega) + \mathrm{i}\sin(\varphi+\omega))$.

This means for the

absolute value of the product: $|zw| = |z||w|$

and for the

argument of the product: $\arg(zw) = \arg z + \arg w$.

From the equation for the absolute values we receive

$$\frac{|zw|}{|w|} = \frac{|z|}{1},$$

meaning that the length $|zw|$ of the product pointer is to the length $|w|$ of the pointer of one factor, as the length $|z|$ of the other pointer is 1. For **visualization** we have to draw in addition to the pointer of one factor, e.g. w, the argument φ of the other factor to get the product pointer zw exactly, if the triangle $\Delta 0w(zw)$ is similar to the triangle $\Delta 01z$. We illustrate this in the next figure:

As a side result of our above effort for a visualization in polar coordinates, and because of the unambiguity of complex numbers, we have derived the trigonometric addition theorems for the **sums** of angles which we had a lot of trouble deriving and learning by heart earlier:

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

The **laws of the Abelian group of multiplication** follow once more simply from the corresponding relations for the real numbers.

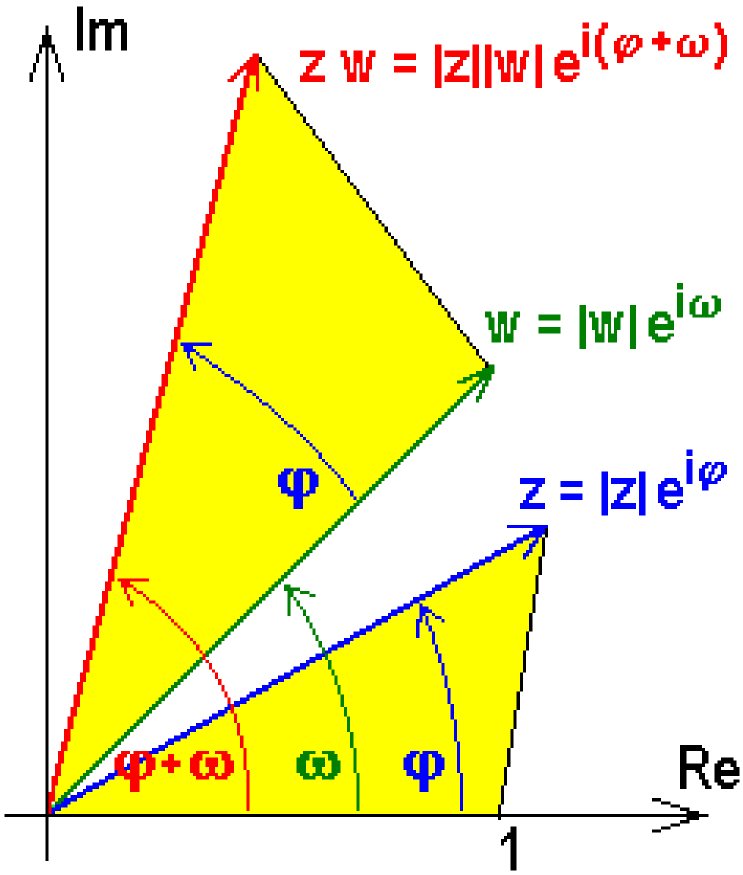


Figure 8.6: Multiplication of complex numbers

Insert: Group Laws:

Commutative Law:	$ab = ba$
Associative Law:	$(ab)c = a(bc)$
Unity element:	$\exists!\, 1 := 1 + \mathrm{i}0 \quad \forall z \in \mathbb{C}: \quad 1 \cdot z = z$
Inverse elements:	$\forall z = re^{\mathrm{i}\varphi} \quad \exists! z^{-1} := r^{-1}e^{-\mathrm{i}\varphi}$ *with* $z \cdot z^{-1} = 1$

The existence of an unambiguous **inverse** enables **division** by complex numbers: the **quotient** $a \cdot b^{-1} =: \frac{a}{b}$ solves the equation $z \cdot b = a$ for $b \neq 0$. To visualize the quotient we calculate

$$\textbf{quotient:}\ \frac{z}{w} = \frac{|z|e^{\mathrm{i}\varphi}}{|w|e^{\mathrm{i}\omega}} = \frac{|z|}{|w|}e^{\mathrm{i}(\varphi-\omega)} = \frac{|z|}{|w|}(\cos(\varphi-\omega) + \mathrm{i}\sin(\varphi-\omega)).$$

This means for the

$$\boxed{\textbf{absolute value of the quotient: } \left|\frac{z}{w}\right| = \frac{|z|}{|w|}}$$

and for the

$$\boxed{\textbf{argument of the quotient: } \arg(\tfrac{z}{w}) = \arg z - \arg w.}$$

From the equation for the absolute values we get

$$\boxed{\frac{\left|\frac{z}{w}\right|}{|z|} = \frac{1}{|w|},}$$

i.e. the length $\left|\frac{z}{w}\right|$ of the quotient pointer is to the length $|z|$ of the pointer of the numerator, as 1 to the length $|w|$ of the denominator. To get a good **visualization** we have just to subtract the argument ω of the denominator from the argument φ of the numerator z to get the quotient pointer $\frac{z}{w}$ exactly, if the triangle $\Delta 0(\frac{z}{w})z$ is similar to the triangle $\Delta 01w$. We have a look at this in more detail in the next figure:

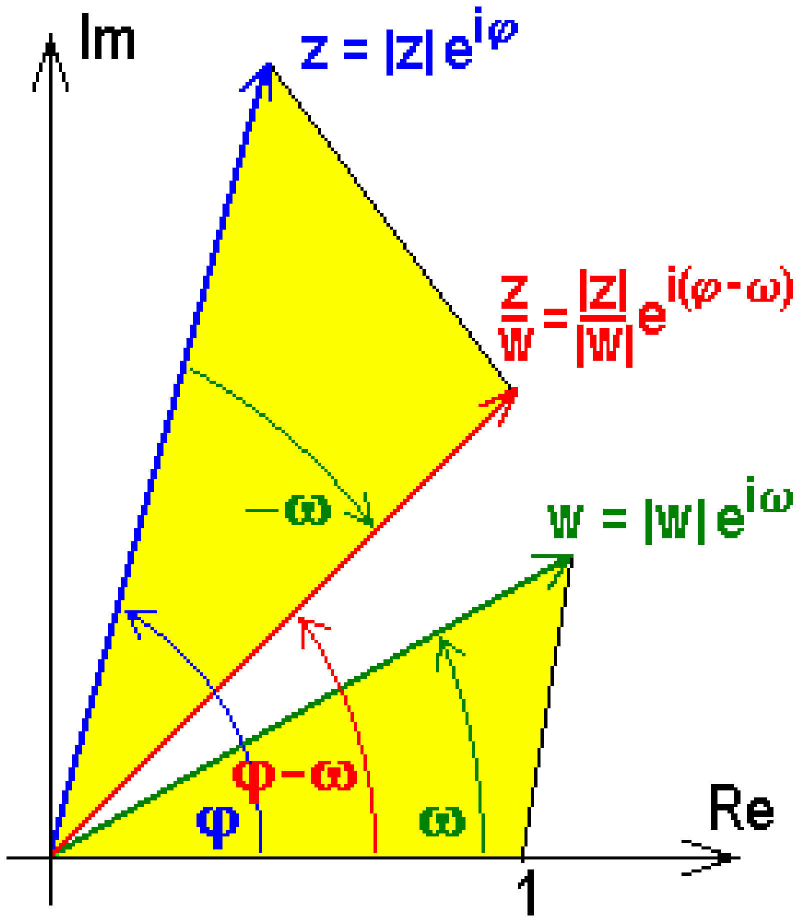


Figure 8.7: Division of two complex numbers

In order to calculate the quotient in Cartesian or planar polar coordinates, it is the best to use the relation

$$\boxed{\frac{z}{w} = \frac{zw^*}{|w|^2},}$$

which you should memorize because it is frequently used.
To make the laws of the field complete, there holds as earlier in addition a

Distributive Law: $(a + b)c = ac + bc.$

The complex conjugate of a product is the product of the conjugate of its factors:

$$\boxed{(zw)^* = z^* w^*.}$$

The star may be drawn into the brackets as in the case of the sum.

In calculations with complex numbers an often-used fact is that the product of a complex number with its own complex conjugate is always a real number:

$$zz^* = (x + \mathrm{i}y)(x - \mathrm{i}y) = x^2 + y^2 = |z|^2 \in \mathbb{R}.$$

This relation may also be helpful, if you want to keep the denominator real: $\frac{1}{z} = \frac{z^*}{|z|^2}$.

Also in multiplication, there is a modest "memory" of the **order** which was lost in the extension of the real number into the complex ones: From $a \neq b$ and $c \neq 0$ there follows $ac \neq bc$.

Exercise 8.4 Multiplication and division of a complex number:

Calculate for the complex number $c := 3 + 3\mathrm{i}\sqrt{3}$:
cc^*, c^2, c/c^*, $1/c$, $1/c^*$, $1/c + 1/c^*$, $1/c - 1/c^*$ *and* c^3.

Exercise 8.5 Multiplication and division of complex numbers:

Calculate for a general complex number $z = re^{\mathrm{i}\varphi}$:
zz^*, z^2, z/z^*, $|z/z^*|$, $1/z + 1/z^*$, $1/z - 1/z^*$.

Exercise 8.6 Simple mappings:

Choose a complex number z *and calculate for it: a)* $\mathrm{i}z$, *b)* $1/z$ *and c)* $1/z^*$.

Exercise 8.7 Other product definitions:

Show by a counterexample that the "memory" of order mentioned above: $a \neq b$, $c \neq 0$ $\Rightarrow ac \neq bc$ *would not hold, if we had chosen the simpler definition* $a \times b := a_1b_1 + ia_2b_2$ *instead of the one suggested by Euler's "i".*

8.3 Functions of a Complex Variable

8.3.1 Definition

In full analogy with the real case we define **complex functions of a complex variable** again as an input-output relation or mapping, **however with one very important difference**: For the complex functions we do **not at all want to require** the uniqueness with $\exists! y = f(x)$ which was deliberately incorporated into the definition of a real function:

$w = f(z)$ **complex function:** $\forall z \in D_f \subset \mathbb{C} \quad \exists w = f(z) : w \in W_f \subset \mathbb{C}$

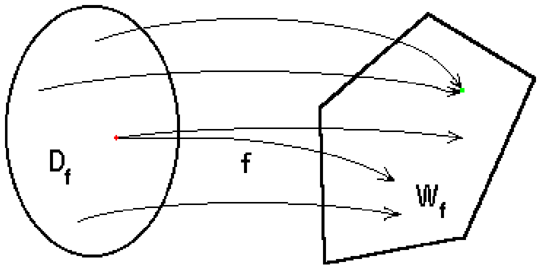


Figure 8.8: Complex function with the possibility of polyvalence

It will be possible and even common that for one value of the independent complex variable z in the definition domain $D_f \subset \mathbb{C}$ there exist several function values $f(z)$ in the value domain $W_f \subset \mathbb{C}$. We will have to acquaint ourselves with one-, two-, three- etc., i.e. **multi-valued** functions and even with ∞-valued ones.

In Section 8.1.5 we have already met an important complex function, the exponential function.

Since for the complex numbers we have no order any more, we do **not** have **anything analogous to the monotony** which was very important for the real functions.

The **calculation with complex functions of a complex variable** is regulated by the rules of the field $\mathbb{C}$ which we have put together in the last section with both the Commutative and Associative Laws as well as the connecting Distributive Law: For instance the sum resp. the difference of two complex functions $f_1(z) \pm f_2(z) = g(z)$ is a new complex function, the same holds for the complex multiple $cf(z) = g(z)$ with $c \in \mathbb{C}$ and analogously for the product $f_1(z) \cdot f_2(z) = g(z)$ or, if $f_2(z) \neq 0$ in the definition domain, also for the quotient $f_1(z)/f_2(z) = g(z)$.

8.3.2 Limits and Continuity

Also the transfer of the concept of the limit being central for real series and functions presents no serious difficulties, since it was based on the distance between points which we have at our disposal also for complex numbers.

We say a **sequence of complex numbers** $(z_n)_{n\in\mathbb{N}}$ has a complex number z_0 as **limit** and write: $\lim\limits_{n\to\infty} z_n = z_0$ (sometimes more casually: $z_{n\to\infty} \longrightarrow z_0$), or call the sequence:

> $(z_n)_{n\in\mathbb{N}}$ **convergent to** z_0: $\exists z_0 : \lim\limits_{n\to\infty} z_n = z_0 \iff$
> $\forall\varepsilon > 0\, \exists N(\epsilon) \in \mathbb{N} : |z_n - z_0| < \varepsilon\, \forall n > N(\epsilon).$

The last shorthand again means: for every pre-given tiny positive number ε a number $N(\varepsilon)$ can be found, such that the distance from the cluster point z_0 for all members of the sequence with a larger number than N is smaller than the given tiny ε.

With this definition of the limit of complex numbers, all considerations about convergence of complex numbers are reduced to the examination of the corresponding distances.

For complex functions we choose again a sequence $(z_n)_{n\in\mathbb{N}} \subset D_f$ of complex numbers in the definition domain D_f of the function f, which for $n \to \infty$ tends to the number $z_0 \in D_f$. Then we calculate the function values at these points $f(z_n)$ which form a sequence again $(f(x_n))_{n\in\mathbb{N}}$ and check whether the sequence of these function values converges. If this can be shown for **every** sequence, out of the definition domain converging to z_0, and always with the same limit w_0, we call the sequence of the function values convergent to w_0 : $\lim\limits_{z\to z_0} f(z) = w_0$:

> $\lim\limits_{z\to z_0} f(z) = w_0$ ***convergent:*** $\iff \forall (z_n)_{n\in\mathbb{N}} : \quad \lim\limits_{n\to\infty} z_n = z_0 \Longrightarrow \lim\limits_{n\to\infty} f(z_n) = w_0$

If we fill in our definition of the convergence of sequences, we receive:

> $\lim\limits_{z\to z_0} f(z) = w_0$ ***convergent:***
> $\iff \forall\varepsilon > 0\, \exists\delta > 0 : \quad |f(z) - w_0| < \varepsilon \quad \forall z \in D_f$ with $|z - z_0| < \delta$

To show this for all sequences is of course again easier said than done! We leave this problem to the mathematicians as earlier in the real number case and restrict ourselves to some interesting cases which are clear anyway.

With this limit definition we can easily also define **continuity** for our complex functions analogously to our earlier definition:

$$w = f(z) \textbf{ continuous} \text{ at } x_0 \iff \forall \varepsilon > 0 \ \exists \delta(\varepsilon) > 0 : \ |f(z) - w_0| < \varepsilon \ \ \forall z \text{ with } |z - z_0| < \delta$$

For limits this means again that at the considered point z_0 the limit is given through the function value $w_0 = f(z_0)$ of the limit z_0 of a sequence (z_n) in the definition domain of the argument: $\lim_{z \to z_0} f(z) = w_0 = f(z_0) = f(\lim_{n \to \infty} z_n)$. Visually this means that the function $f(z)$ maps pre-image points in the neighbourhood of z_0 again in neighbouring image points.

8.3.3 Graphic Illustration

Unfortunately, the **graphic illustration** of complex functions turns out to be significantly more difficult than in the real case, since a complex function correlates *four real quantities*. Instead of mapping one real pre-image straight line, namely the x-axis, onto an other one, the image straight line or y-axis, we now have to map the complete pre-image complex plane, the z-plane, onto another one, the image or w-plane. In the graphic illustration of a real function of a real variable, we have become used to placing the pre-image and image number lines orthogonal to one another and illustrating the mapping by means of a curve in this plane. For complex functions of a complex variable, we have to find new ways of illustration. Mostly we will use the **pre-image plane and the image plane** drawn side by side and characterize selected points or curves in the z-plane and their images in the w-plane by using the same symbols or colours.

Furthermore, a net of level curves of the **real part u and imaginary part v over the z-plane** or of the absolute value $|w|$ **and the argument** $arg \quad w$ **of the functional values over an xy-net of the z-plane** can give us a more exact idea of the mapping.

The best impression of the effect of a function we get through a **perspective** relief of mountains for example with a xy-net over a certain range of the z-plane.

Because of these difficulties, we shall study only the most important functions chosen from our basis of real functions in the complex field:

8.3.4 Powers

First of all, as before, we consider the **powers** z^n with natural exponents $n \in \mathbb{N}$:

$$w = z^n = (x + \mathrm{i}y)^n = |z|^n(\cos\varphi + \mathrm{i}\sin\varphi)^n = |z|^n e^{in\varphi} = |z|^n(\cos n\varphi + \mathrm{i}\sin n\varphi),$$

where at the end we have used the Euler formula and thus written down its extension: the

Moivre formula: $z^n = |z|^n(\cos n\varphi + \mathrm{i}\sin n\varphi)$.

This implies for the

absolute value of the n-th power: $|z^n| = |z|^n$

and for the

argument of the n-th power: $\arg(z^n) = n\arg(z)$.

We want to discuss **two examples** in more detail:

1) As the first example we choose the **quadratic function**, i.e. $n = 2$:

quadratic function: $w = u + \mathrm{i}v = z^2 = (x + \mathrm{i}y)^2 = (x^2 - y^2) + \mathrm{i}2xy = |z|^2 e^{2\mathrm{i}\varphi}$,

meaning for the real part: $u = x^2 - y^2$ and for the imaginary part: $v = 2xy$, respectively for the absolute value: $|w| = |z|^2$ and for the argument: $\arg(w) = 2\arg(z)$.

Firstly we determine **some image points**:

$$\begin{aligned} w(\pm 1) &= (\pm 1)^2 = 1, \\ w(\pm \mathrm{i}) &= (\pm \mathrm{i})^2 = e^{\pm \mathrm{i}\frac{\pi}{2}2} = e^{\pm \mathrm{i}\pi} = -1 \text{ and} \\ w(1 \pm \mathrm{i}) &= (1 \pm \mathrm{i})^2 = (\sqrt{2}e^{\pm\frac{\mathrm{i}\pi}{4}})^2 = \pm 2\mathrm{i}. \end{aligned}$$

Then we look at the **vertical straight line** $x = 1 : u = x^2 - y^2 = 1 - y^2$ and $v = 2xy = 2y$. Thus $1 - u = y^2 = \frac{v^2}{4}$, i.e. $u = -\frac{v^2}{4} + 1$, which is the equation of a parabola open to the left side.

Analogously you can show that the **horizontal straight line** $y = 1$ becomes the parabola open to the right side $u = \frac{v^2}{4} - 1$.

Apparently the **unit circle** $|z| = 1$ is mapped onto itself by the quadratic function: $|w| = 1$.

From $u = x^2 - y^2 = const.$ you can see that the **hyperbolae with the bisectors as asymptotes** become vertical straight lines, and from $v = 2xy = const.$, that the **hyperbolae with the axes as asymptotes** become horizontal straight lines.

The following figure gives an overview over the whole mapping. Note that the left-side half of the pre-image plane is omitted, since the **image of the right half alone** covers the entire w-plane.

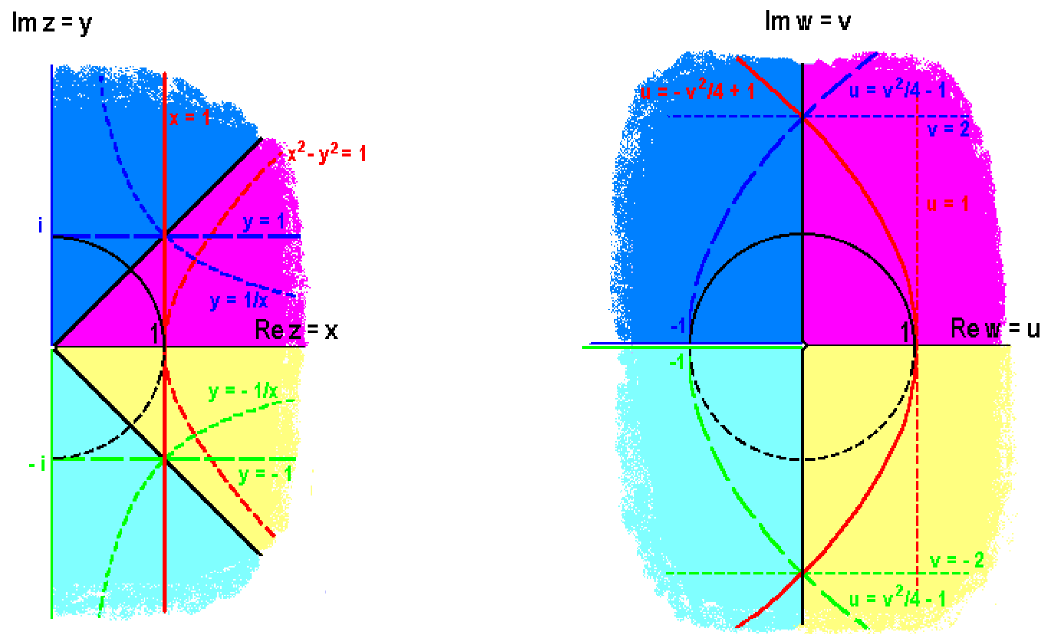


Figure 8.9: Right half of the z-plane and the entire (upper) w-plane for the quadratic function

Insert: Rubber sheet: *You can imagine the mapping procedure approximately in the following way: think of the right-half of the z-plane as made of an elastic sheet and then rotate the positive and negative halves of the imaginary axis in opposite direction by* $90°$ *around the origin until they meet each other along the negative real axis.*

The image of the left half of the Gauss z-plane leads to a **second** cover of the whole w-plane. We already encountered part of this in connection with the **real** quadratic function, where the image of the negative pre-image half-axis covered the positive image half-axis a second time, so that the square root function could only be defined over the positive half line. In order to allow now for the **complex** quadratic function an inverse function over the whole plane, mathematicians cut the two picture planes, imagined lying one over the other, (the cut being e.g. along the negative real axis), and connect the upper edge of the cut of the upper sheet with the lower edge of the cut in the lower sheet and think of the lower edge of the upper sheet as penetrating "straight through the other connection", stuck together with the upper edge of the lower sheet. The whole construction consisting of the two planes connected crosswise along the negative real axis is called a **Riemannian surface with two sheets**. Thus we can say: The complex quadratic function maps the z-plane bi-uniquely onto the two-sheeted Riemannian surface, where the special position of the cut is arbitrary. Decisive is, that the cut runs between the two branching points 0 and ∞. The next figure tries to visualize this situation.

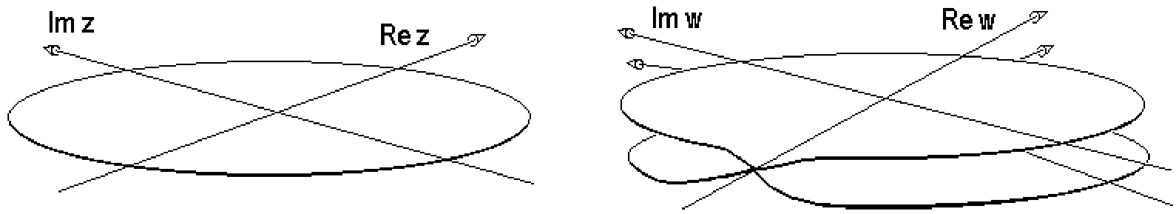


Figure 8.10: Riemannian surface of the quadratic function

During the motion of a mass point e.g. on the unit circle in the z-plane with the starting point $z = 1$ the image point w runs along the unit circle in the upper w-plane, however, with twice the velocity until it dives near $z = \mathrm{i}$, i.e. $w = -1$ into the lower sheet of the Riemannian w-sheet. It goes on running on the unit circle in the lower sheet, reaches for $z = -1$ the point $w = +1$ in the lower sheet and appears only for $z = -i$ at the diving point $w = -1$ again in the upper sheet, to reach finally on the upper unit circle for $z = 1$ the starting point $w = 1$.

2) A similar construction holds for the **cubic function** with $n = 3$:

cubic function: $w = z^3 = |z|^3 e^{3\mathrm{i}\varphi} = |z|^3(\cos 3\varphi + \mathrm{i}\sin 3\varphi)$,

meaning for the absolute value: $|z^3| = |z|^3$ and for the argument: $\arg(z^3) = 3\arg(z)$.
We determine only a few **image points**:

$$\begin{aligned} w(\pm 1) &= (\pm 1)^3 = \pm 1, \\ w(\mathrm{i}) &= i^3 = e^{\frac{3\pi\mathrm{i}}{2}} = -\mathrm{i} \text{ and} \\ w(1+\mathrm{i}) &= (1+\mathrm{i})^3 = -2(1-\mathrm{i}). \end{aligned}$$

We see that already **one third** of the z-plane is mapped onto the entire w-plane, and that the entire pre-image plane is mapped onto a Riemannian surface consisting of **three** sheets which are cut between 0 and ∞ connected with each other. The following figure sketches the situation:

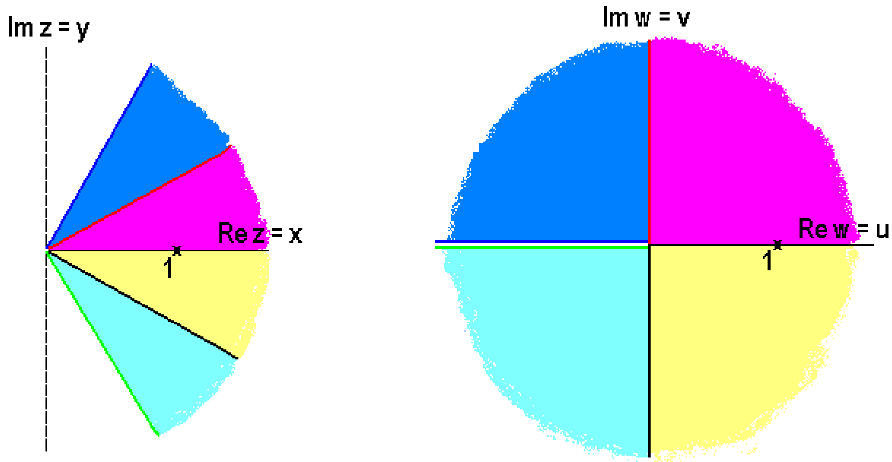


Figure 8.11: One third of the z-plane and the upper sheet of the w-plane for $w = z^3$

Continuing in this manner an overview over all **power functions** $w = z^n$ can be reached. In particular, one n-th of the z-plane is bi-uniquely mapped onto the whole w-plane or the whole z-plane onto a n-sheeted Riemannian surface. At least in principle this gives a feeling for the mapping action of complex **polynomials**: $P_m(z) = \sum_{n=0}^{m} a_n z^n$.

For every complex polynomial of m-th degree the **Fundamental Theorem of Algebra** guaranties the **existence of m complex numbers** z_n, such that the sum can be represented as a product of m factors:

$$P_m(z) = \sum_{n=0}^{m} a_n z^n = a_m(z - z_1)(z - z_2)(z - z_3)\ldots(z - z_{m-1})(z - z_m) :$$

Fundamental Theorem of Algebra:
$\exists z_n \in \mathbb{C}, n = 1, 2, 3, \ldots, m : \quad P_m(z) = \sum_{n=0}^{m} a_n z^n = a_m \prod_{n=1}^{m} (z - z_n).$

Exercise 8.8 Concerning the fundamental theorem of algebra:

Show with help of the Fundamental Theorem of Algebra, that the sum respectively the product of the m zero points w_n of a polynomial $P_m(w) = 0$ holds: $\sum\limits_{n=1}^{m} w_n = -\frac{a_{m-1}}{a_m}$ *respectively* $\prod\limits_{n=1}^{m} w_n = (-1)^m \frac{a_0}{a_m}$.

For the complex infinite power series $\sum\limits_{n=0}^{\infty} a_n(z - z_0)^n$, the ones the mathematicians call also "analytic functions", we report without proof that all these series converge absolutely inside a circle domain $|z-z_0| < R$ with the radius R around the development centre z_0 and diverge outside that region. Only now can we really understand the "convergence **radius**" R, that can be determined according to the **criteria of convergence** we have explained earlier. For instance for the complex geometric series $\sum\limits_{n=0}^{\infty} z^n = \frac{1}{1-z}$ the singularity at $z = 1$ restricts the radius of convergence to $R = 1$, as we have seen in Section 6.5 with help of the **quotient criterion**. Now we want to examine in more detail three very important power series as examples: the natural exponential function, which we already met, and the complex sine and cosine functions.

8.3.5 Exponential Function

The most important complex function by far is the **natural exponential function**. Already in Section 8.1.5 through Euler's formula we have been lead to its definition for purely imaginary variables, and of course we can easily complete it for general complex variables:

exponential function:

$$w = e^z := exp(x + \mathrm{i}y) = exp(x)(\cos y + \mathrm{i}\sin y) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

i.e. for the absolute value:

$$|w| = |e^z| = exp(\mathrm{Re}\, z) = exp(x)$$

and for the argument:

$$\arg(w) = \arg(e^z) = \mathrm{Im}\, z = y.$$

While the exponential function still rises rapidly as a function of its real part as we have seen in the past, it is 2π-periodic in its dependence on the imaginary part of its variable.

$$\textbf{2}\boldsymbol{\pi}\textbf{i-periodic:} \quad e^{\mathrm{i}(\varphi+2k)} = e^{\mathrm{i}\varphi} \text{ with } k \in \mathbb{Z}.$$

The convergence radius of the Taylor series is infinite as we have seen earlier.

The **functional equation:**

$$\boldsymbol{e^z e^w = e^{z+w}} \text{for } z,\, w \in \mathbb{C} \text{ still holds true.}$$

In order to visualize the function we first calculate once again **some image points**:

$$\begin{aligned} w(1) &= e, \\ w(0) &= 1, \\ w(-1) &= \frac{1}{e}, \\ w(\mathrm{i}\pi) &= -1 \text{ and} \\ w(\frac{\mathrm{i}\pi}{2}) &= \mathrm{i}. \end{aligned}$$

Then we see that **vertical straight lines** $\operatorname{Re} z = x = const.$ are mapped into circles $|w| = |e^z| = e^{\operatorname{Re} z} = exp(x) = const.$: the straight line $x = 0$ into the circle $|w| = 1$, the line $x = 1$ into the circle $|w| = e$ and the straight line $x = -1$ into the circle $|w| = \frac{1}{e}$.

The **horizontal straight lines** $\operatorname{Im} z = y = const.$ are mapped onto spokes $\arg w = \arg(e^z) = \operatorname{Im} z = const.$: to be specific the straight line $y = \pi$ onto the spoke $\arg w = \pi$, the line $y = \frac{\pi}{2}$ onto $\arg w = \frac{\pi}{2}$, etc...

From these results we see that the exponential function maps a **horizontal stripe** of the z-plane with height 2π, for example the so-called **fundamental area** with $-\pi < \operatorname{Im} z \leq \pi$ onto the w-plane which is cut open between the branching points 0 and ∞ (e.g. along the negative real axis). The entire z-plane therefore goes over into a Riemannian surface with **infinitely many** sheets. For each sheet the upper edge is continuously connected along the cut with the lower edge of the underlying sheet, and the upper edge of the last sheet "through all other connections" with the lower edge of the first one. The following figure can help you to make a mental image of the effect of the function.

8.3.6 Trigonometric Functions

Having now studied the exponential function we will quickly take a look at the **trigonometric functions**, cosine and sine, which we can easily derive with help of the Euler formula from the exponential function, or define through their power series:

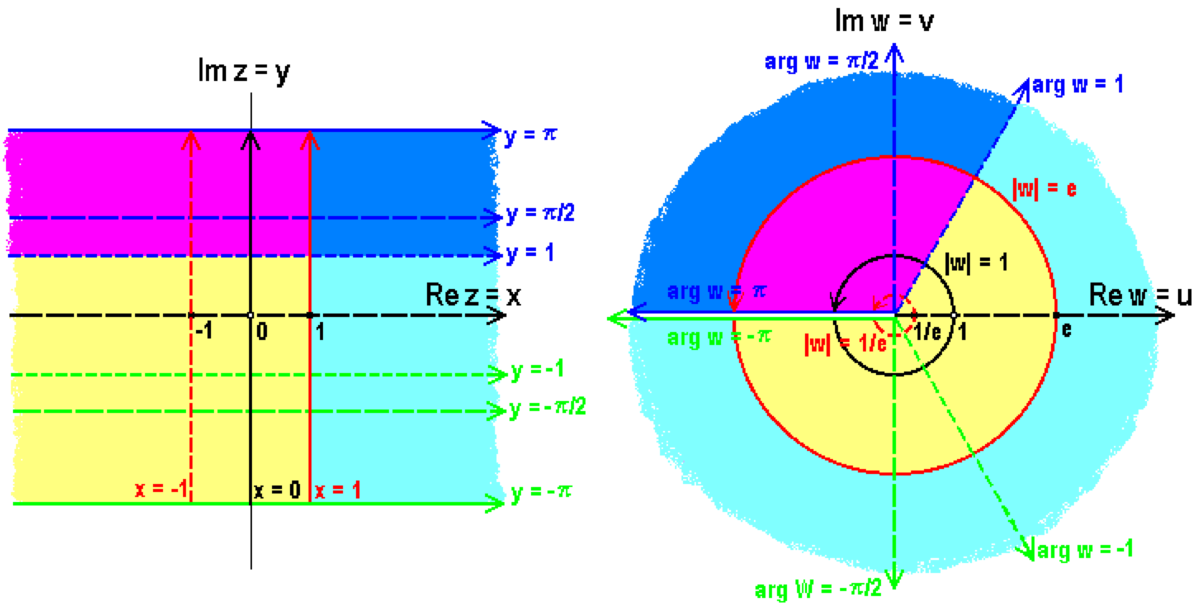


Figure 8.12: Illustration of the horizontal fundamental stripe in the z-plane and the cut w-plane for the exponential function

cosine: $w = \cos z = \dfrac{e^{\mathrm{i}z} + e^{-\mathrm{i}z}}{2} = \sum_{n=0}^{\infty} (-1)^n \dfrac{z^{2n}}{(2n)!}$ and

sine: $w = \sin z = \dfrac{\mathrm{i}(e^{-\mathrm{i}z} - e^{\mathrm{i}z})}{2} = \sum_{n=0}^{\infty} (-1)^n \dfrac{z^{2n+1}}{(2n+1)!}$.

Both series converge in the entire complex plane. As we know, cosine and sine are

2π -periodic: $\cos(z + 2\pi) = \cos z$ and $\sin(z + 2\pi) = \sin z$.

Like our old trigonometric addition theorems from Section 4.2.2:

$$\cos(z \pm w) = \cos z \cos w \mp \sin z \sin w$$
$$\sin(z \pm w) = \sin z \cos w \pm \cos z \sin w$$

and moreover

$$\cos^2 z + \sin^2 z = 1, \quad \cos^2 z - \sin^2 z = \cos 2z \text{ and } 2 \cos z \sin z = \sin 2z$$

hold true for general complex variables z, $w \in \mathbb{C}$.

Especially for $z + w = x + \mathrm{i}y$ we obtain from these equations

$$\boxed{\cos \mathrm{i}z = \cosh z \quad \text{resp.} \quad \sin \mathrm{i}z = \mathrm{i} \sinh z :}$$

$$\boxed{\begin{aligned} \cos(x + \mathrm{i}y) &= \cos x \cos \mathrm{i}y - \sin x \sin \mathrm{i}y = \cos x \cosh y - \mathrm{i} \sin x \sinh y \\ \sin(x + \mathrm{i}y) &= \sin x \cos \mathrm{i}y + \cos x \sin \mathrm{i}y = \sin x \cosh y + i \cos x \sinh y. \end{aligned}}$$

The definition of the hyperbolic functions is just as before

hyperbolic cosine: $w = \cosh z = \dfrac{e^z + e^{-z}}{2} = \sum_{n=0}^{\infty} \dfrac{z^{2n}}{(2n)!}$ and

hyperbolic sine: $w = \sinh z = \dfrac{e^z - e^{-z}}{2} = \sum_{n=0}^{\infty} \dfrac{z^{2n+1}}{(2n+1)!}$.

From this relations we see that, for complex arguments, $\cos z$ and $\sin z$ are by no means bounded as they are for the real numbers, but rather increase for growing imaginary parts like the hyperbolic functions. Differently from the exponential function, **vertical stripes** of the z-plane with width 2π, e.g. the **fundamental area** with $-\pi < \operatorname{Re} z \leq \pi$, is here mapped on the two-sheeted w-plane cut between -1 and $+1$.

Exercise 8.9 Addition theorems:

Prove one of the addition theorems, e.g. $\cos(z - w) = \cos z \cos w + \sin z \sin w$ *with help of the exponential functions and then show that* $\cos^2 z + \sin^2 z = 1$.

Exercise 8.10 Connection with the hyperbolic functions

Show that: a) $\cos \mathrm{i}z = \cosh z$, *b)* $\sin \mathrm{i}z = \mathrm{i} \sinh z$ *and c)* $4 \sin^3 \alpha = 3 \sin \alpha - \sin 3\alpha$.

Exercise 8.11 Functional values of the cosine

Calculate the following functional values of the cosine function: $\cos \pm \frac{\pi}{2}$, $\cos \pm \pi$, $\cos \pm \mathrm{i}\frac{\pi}{2}$, $\cos \pm \mathrm{i}\pi$, $\cos(\frac{\pi}{2} \pm \mathrm{i}\frac{\pi}{2})$, $\cos(\frac{\pi}{2} \pm \mathrm{i}\pi)$, *and* $\cos(\pi \pm \mathrm{i}\pi)$.

In the case of the complex sine function we will demonstrate the wide variety of **representation possibilities** which are at our disposal. Because of the symmetry properties it is sufficient to visualize $\sin z$ over the **square** $0 < x < \pi$ **and** $0 < y < \pi$:

The following figures show the **level curves** for the real part $\mathrm{Re}\sin z$, the imaginary part $\mathrm{Im}\sin z$ (dashed), the absolute value $|\sin z|$ and the argument $\arg\sin z$ (also dashed) of the mapped function $w = \sin z$ over the square.

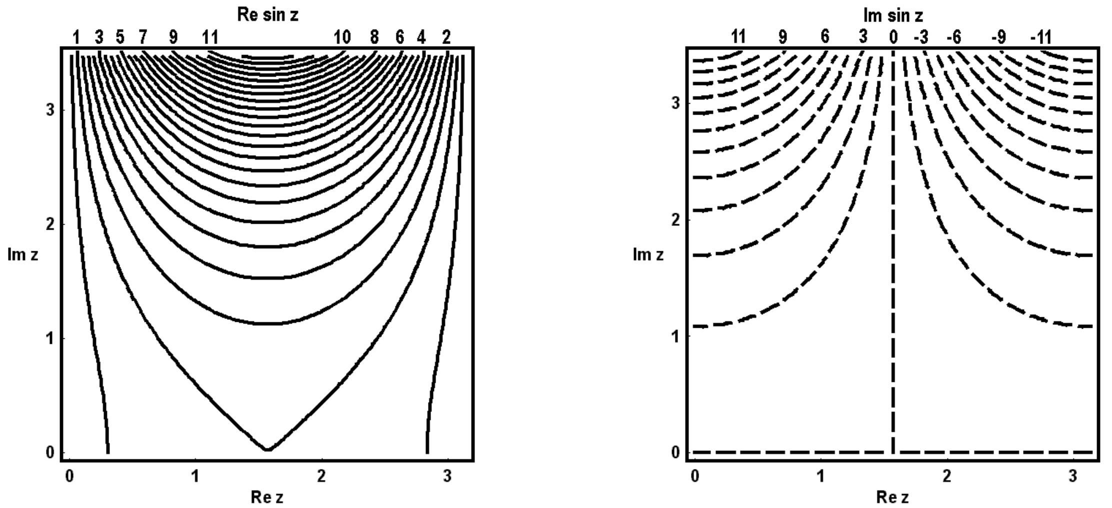


Figure 8.13 a + b: Level curve representation for $\mathrm{Re}\sin z$ and $\mathrm{Im}\sin z$ over the selected square $0 < \mathrm{Re}\, z < \pi$ and $0 < \mathrm{Im}\, z < \pi$.

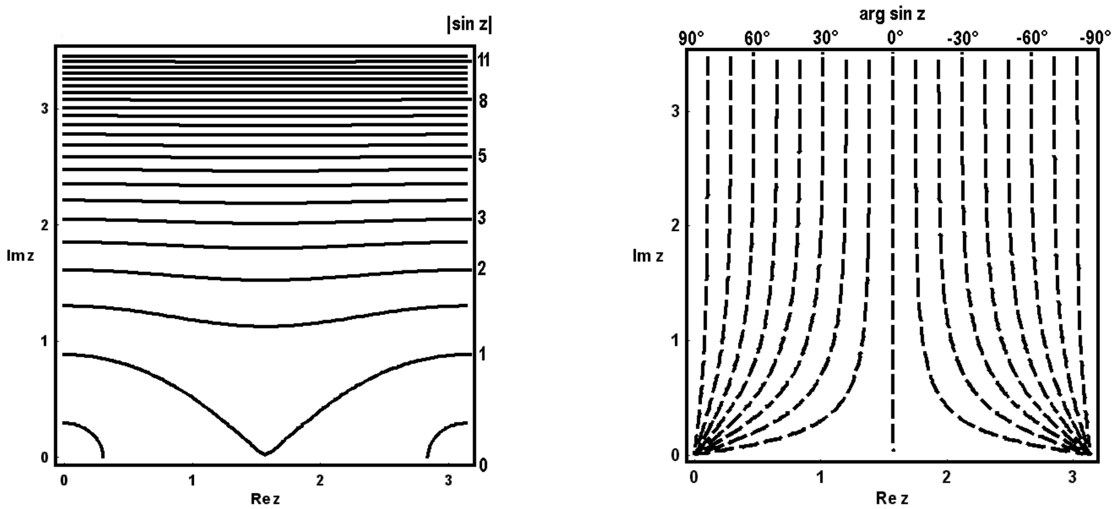


Figure 8.13 c + d: Level curve representation for $|\sin z|$ and $\arg\sin z$ over the selected square $0 < \mathrm{Re}\, z < \pi$ and $0 < \mathrm{Im}\, z < \pi$.

Usually these representations are put together in pairs into one single diagram to form a **level net** as we have done in the next two figures:

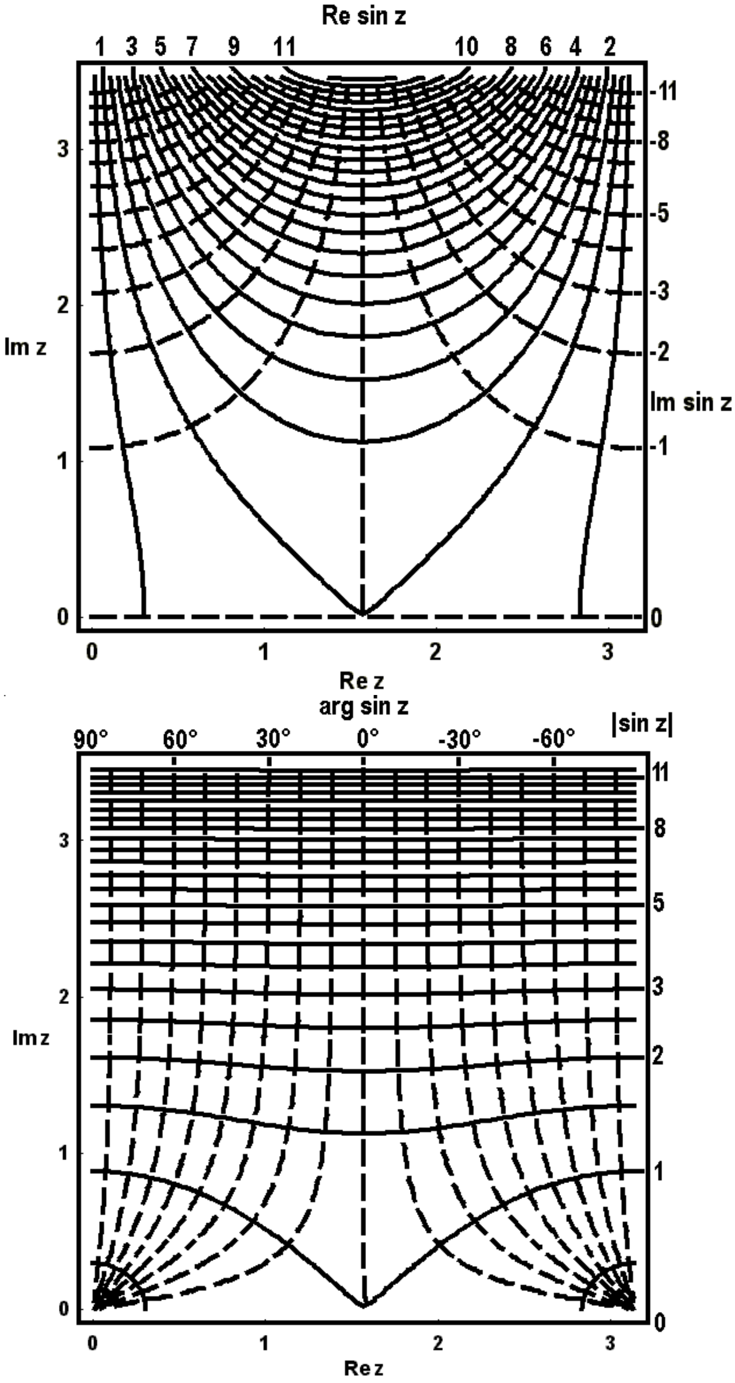


Figure 8.13 e + f: Level net for $\mathrm{Re} \sin z$ and $\mathrm{Im} \sin z$ and resp. $|\sin z|$ and $\arg \sin z$ over the square

It requires some effort to get an impression of the represented function from the level curves of the image points. We succeed a little better if correlated with the mean value of the function in that region, the areas between the curves are tinted in **grey shades** over a scale which extends from deep black for small values to white for large values. This kind of representation is demonstrated by the Figures g) till j). Now, we can more easily imagine how the values of the imaginary part $\operatorname{Im}\sin z$ with growing $\operatorname{Im} z$ one the one hand increase near $\operatorname{Re} z = 0$ and on the other hand decrease near $\operatorname{Re} z = \pi$. Also the tremendous increase of $\operatorname{Re}\sin z$ and $|\sin z|$ with growing distance from the real axis shows up clearly.

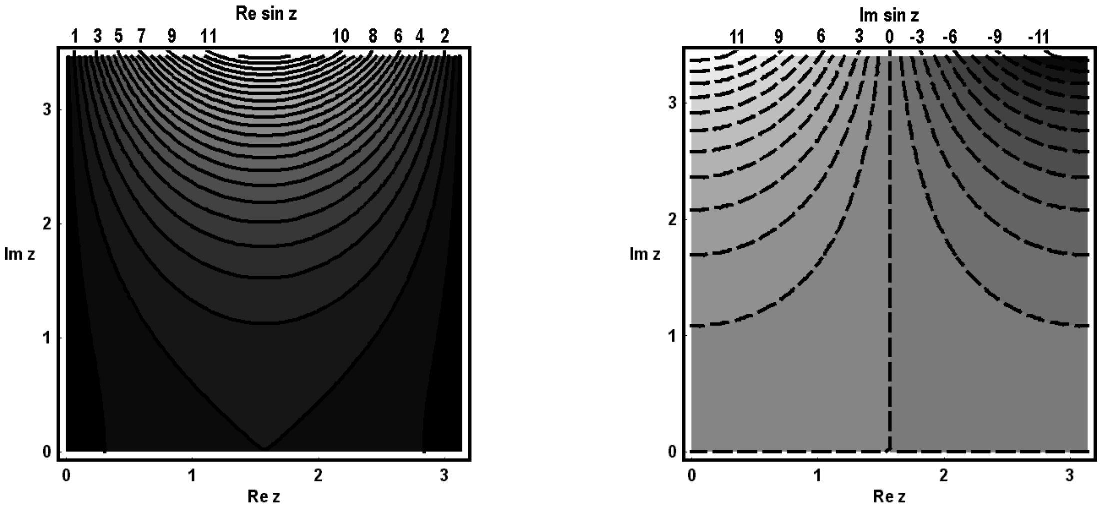


Figure 8.13 g + h: Grey tinted level curve representation for $\operatorname{Re}\sin z$ and $\operatorname{Im}\sin z$ over the square.

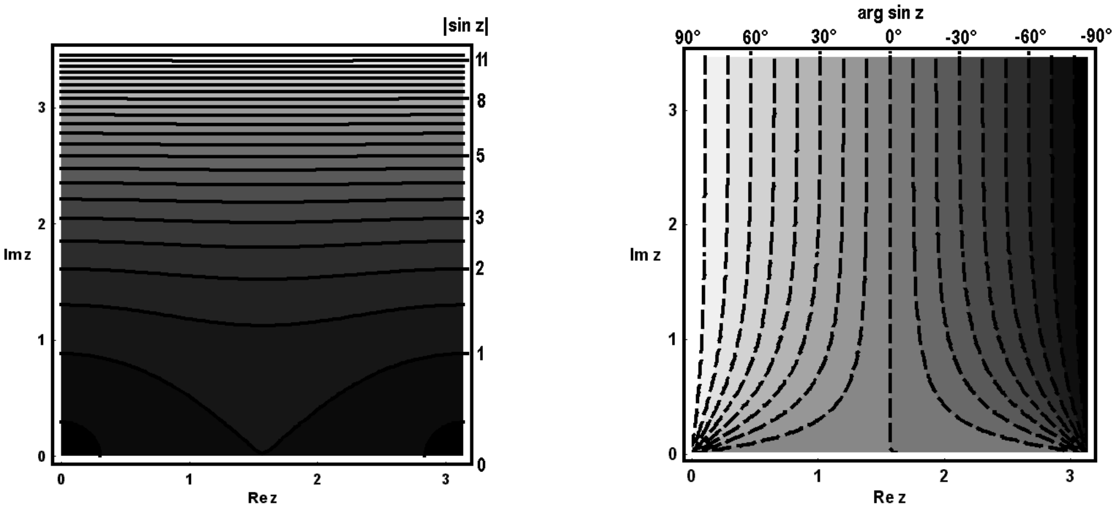


Figure 8.13 i + j:Grey tinted level curve representation for $|\sin z|$ and $\arg\sin z$ over the square.

We achieve even more impressive pictures, if we use a **colour scale** to characterize the relative heights, for instance as done in **geographical maps** with their deep blue of the deep oceans, different shades of green representing low lying land, the beige, and finally the darker and darker browns of the mountains. In the computer program MATHEMATICA used here in the Figures k) till n) the **colours of the rainbow** are used to represent function height according to the frequency of the light. (Magma-)Red colours represent smaller values and (sky-)blue colours stand for higher functional values. Through these figures we get a significant impression of the structure of the "function value mountains".

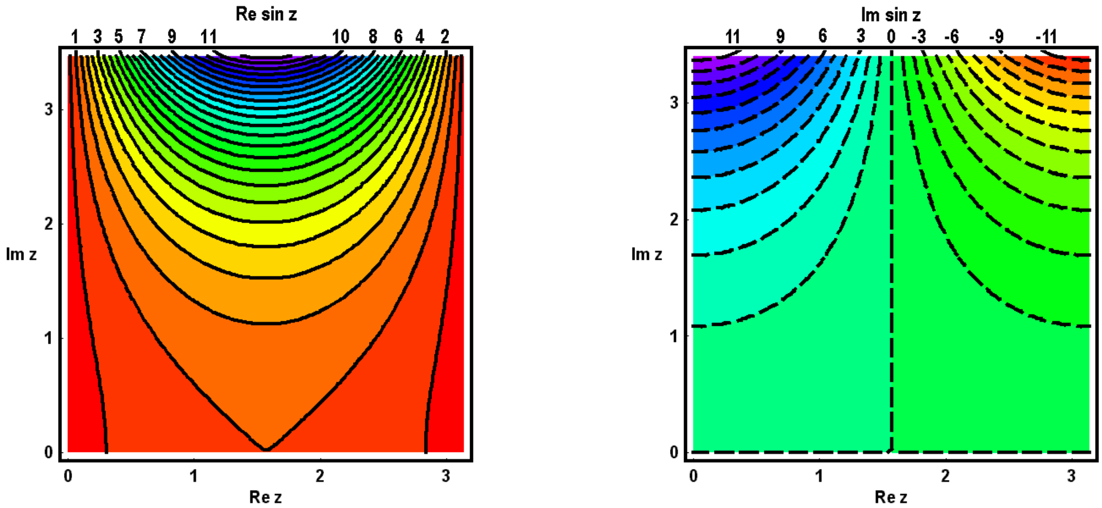


Figure 8.13 k + l: Rain-bow-like coloured level curve representation for $\mathrm{Re}\sin z$ and $\mathrm{Im}\sin z$ over the square.

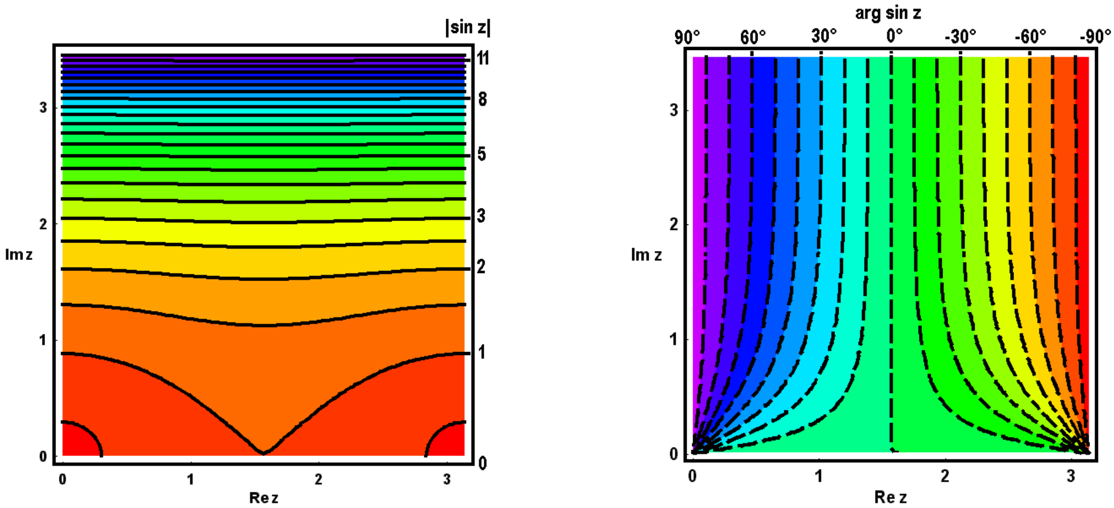


Figure 8.13 m + n: Rain-bow-like coloured level curve representation for $|\sin z|$ and $\arg\sin z$ over the square.

In figure 8.13 n we see particularly well the linear increase of the phase from $-90°$ at $\mathrm{Re}\, z = \pi$ to $+90°$ at $\mathrm{Re}\, z = 0$.

Also with this kind of representation, the coloured visualized level curves of a variable can be upgraded to a **net** through the marking in of the (dashed) level curves of a second variable which however **cannot** yet be represented by colours. This fact is illustrated in the next two figures:

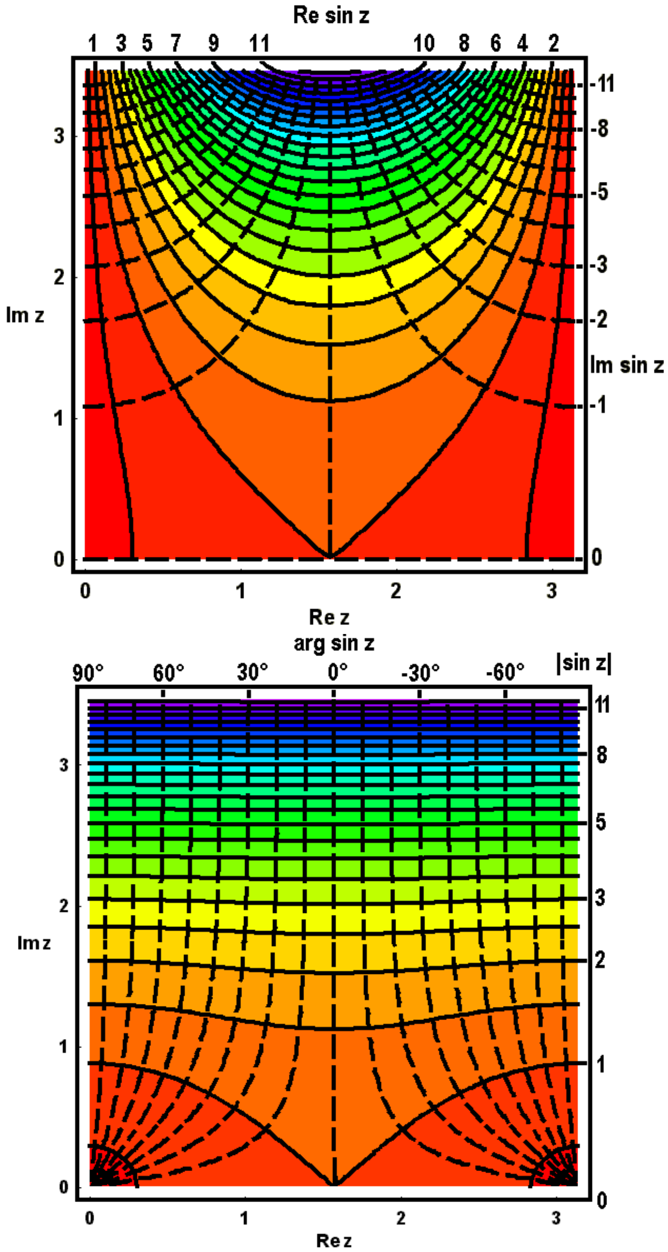


Figure 8.13 o + p: Rain-bow-like coloured level net representation for $\text{Re}\sin z$ and $\text{Im}\sin z$ and resp. $|\sin z|$ and $\arg\sin z$ over the square.

We get, however, a more vivid impression than these two-dimensional projections are able to provide, if we look at the **pictures** of the function values **in perspective** which are offered by the drawing programs of the modern computers as shown in the next figures:

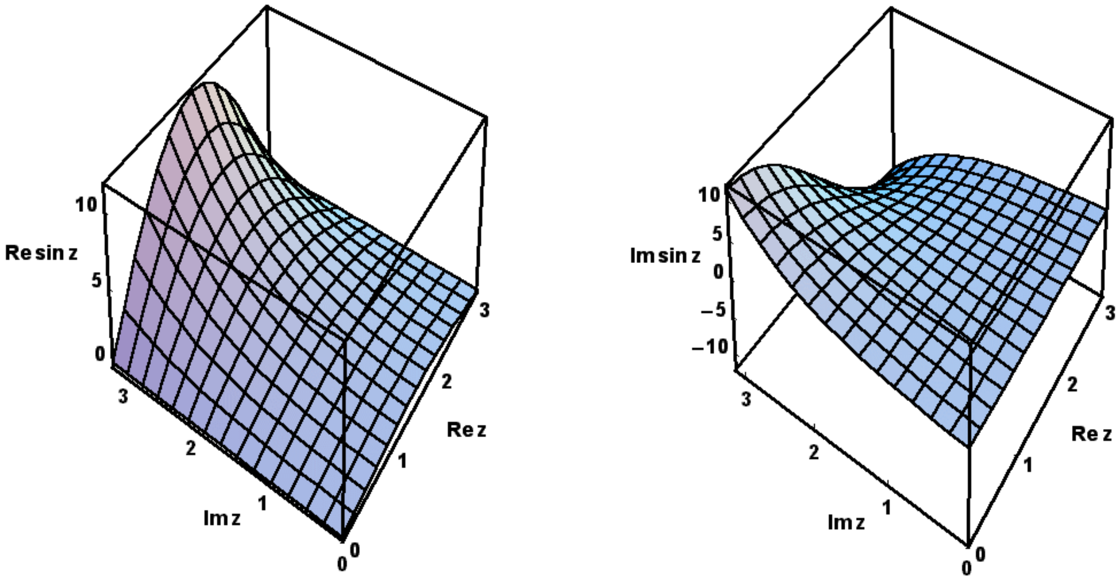


Figure 8.14 a + b: Relief picture in perspective of the function values of Re $sin z$ and $\operatorname{Im}\sin z$ with an x-y net over the selected square $0 < \operatorname{Re} z < \pi$ and $0 < \operatorname{Im} z < \pi$.

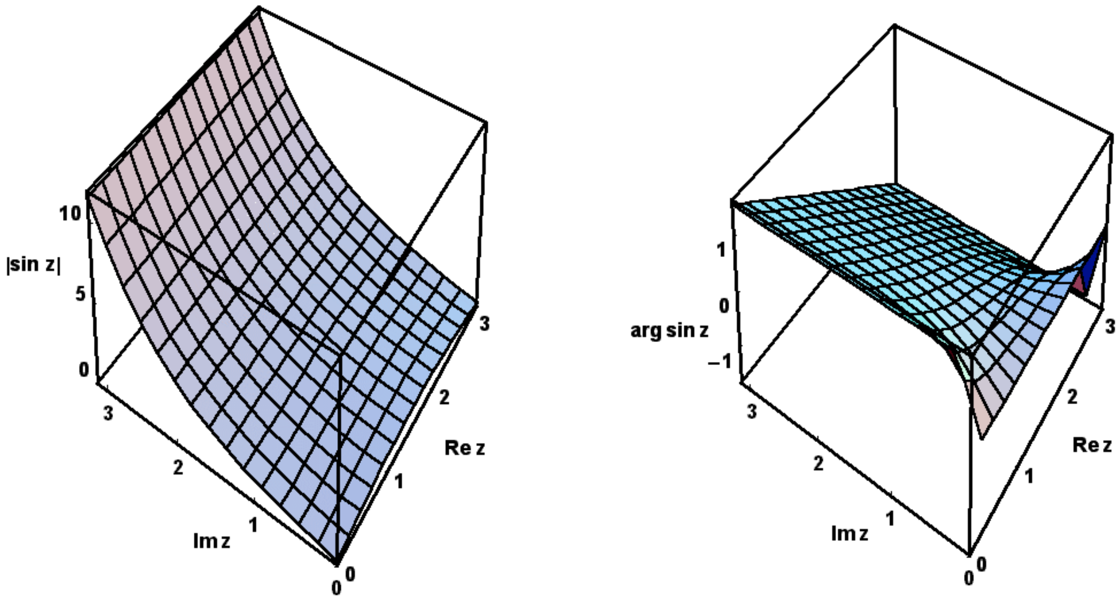


Figure 8.14 c + d: Relief picture in perspective of the function values of $|\sin z|$ and $\arg\sin z$ with an x-y net over the selected square $0 < \operatorname{Re} z < \pi$ and $0 < \operatorname{Im} z < \pi$.

To demonstrate the influence of the changes in sign, we have displayed for you finally the four interesting variables with the help of the program MATHEMATICA **over the larger rectangle region** $0 < \text{Re}\, z < \pi$ and $-\pi < \text{Im}\, z < \pi$, using **revolving** pictures:

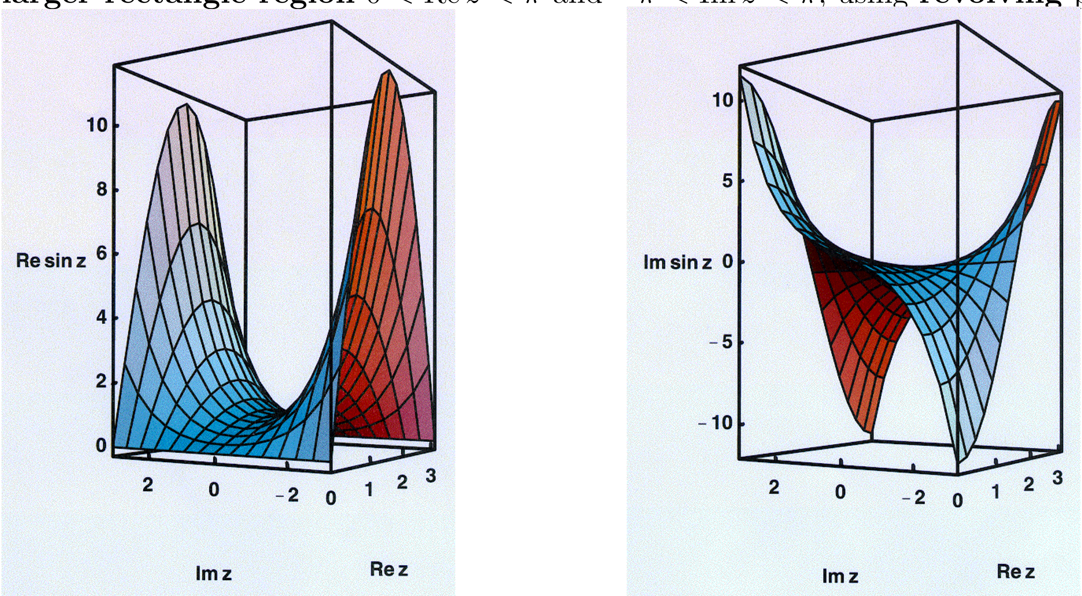


Figure 8.15 a + b: Revolving relief picture in perspective of the function values of Re *sinz* and $\text{Im}\sin z$ with an x-y net over the selected rectangle region: $0 < \text{Re}\, z < \pi$ and $-\pi < \text{Im}\, z < \pi$.

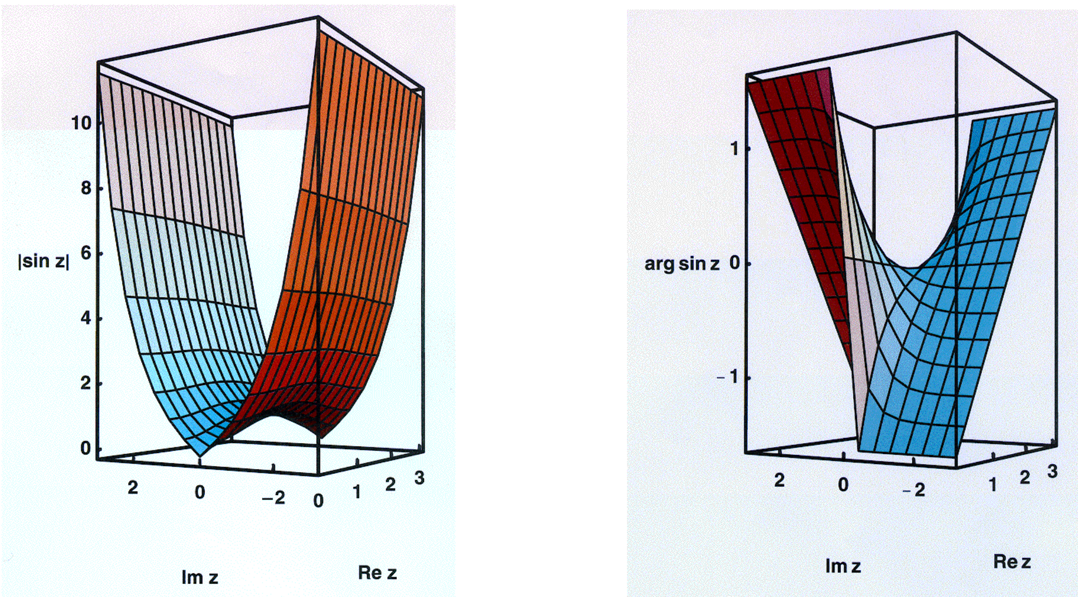


Figure 8.15 c + d: Revolving relief picture in perspective of the function values of $|\sin z|$ and $\arg\sin z$ with an x-y net over the selected rectangle region: $0 < \text{Re}\, z < \pi$ and $-\pi < \text{Im}\, z < \pi$.

If you view this valley along the positive imaginary axis, you will see clearly the **real**

function $|\sin z|$ at a vertical cut over the real axis $\operatorname{Im} z = 0$. Looking along the positive real axis, you recognize the **real function** $\sinh y$ over the imaginary axis, and even the **real catenary function** $\cosh y$ is visible as envelope above the straight line $\operatorname{Re} z = \frac{\pi}{2}$. After shifting the origin by $\frac{\pi}{2}$ in direction of the real axis, these figures describe the relief of the **complex cosine function**.

8.3.7 Roots

At the end of this chapter we take a look at some **inverse functions** in the complex field, where once again characteristic differences occur compared to the real case: First of all we examine the **root functions**.

After having seen how the n-th power maps an n-th sector of the complex z-plane in the whole w-plane, we expect inversely that the n-th root will map the whole z-plane in an n-th sector of the w-plane, meaning, it is an n-valued function, as we have admitted in the complex field:

$$w = \sqrt[n]{z} = z^{\frac{1}{n}} = (re^{i\varphi})^{\frac{1}{n}} = (re^{\mathrm{i}(\varphi+2\pi k)})^{\frac{1}{n}} = \sqrt[n]{r}e^{\frac{\mathrm{i}(\varphi+2\pi k)}{n}} \quad \text{for} \quad n \in \mathbb{N} \quad \text{and} \quad k \in \mathbb{N}_0.$$

Here we have taken into consideration that $e^{2\pi i} = 1$.

For the argument φ of the independent variable z there is the restriction (agreed to earlier) $0 \leq \varphi < 2\pi$. But, for which non-negative integers k does the corresponding relation for the argument of the image hold?

$$\frac{\varphi + 2\pi k}{n} < 2\pi \Longleftarrow k + \frac{\varphi}{2\pi} < n \Longleftarrow k \leq n-1 < n - \frac{\varphi}{2\pi}, \quad \text{i.e. } \mathbf{k = 0, 1, 2, 3, ..., n-1}.$$

Thus there are exactly n n-th roots w_k, which we shall label by the index k:

$$\textbf{n n-th roots: } w_k = \sqrt[n]{z} = \sqrt[n]{r}e^{\frac{\mathrm{i}(\varphi+2\pi k)}{n}} \text{ for } n \in \mathbb{N}.$$

The complex number w_0 is called **principal value**. Furthermore we see that the n roots lie on a circle around the origin with radius $\sqrt[n]{r}$ at the corners of a regular n-polygon: $\sum\limits_{k=0}^{n-1} w_k = 0$.

Exercise 8.12 Roots:

Prove that $\sum\limits_{k=0}^{n-1} w_k = 0$ *with the help of the result of Exercise 8.8.*

As an example we calculate first $w_k = \sqrt{\mathrm{i}} = e^{\mathrm{i}(\frac{\pi}{2}+2\pi g)/2} = e^{\mathrm{i}(\frac{\pi}{4}+k\pi)}$ with $k = 0$ and 1, also

$$w_0 = e^{\mathrm{i}\frac{\pi}{4}} = \frac{1+\mathrm{i}}{\sqrt{2}} \text{ and}$$
$$w_1 = e^{\mathrm{i}\frac{5}{4}\pi} = -\frac{1+\mathrm{i}}{\sqrt{2}} = -w_0.$$

A further example is: $w_k = \sqrt[3]{1} = e^{\mathrm{i}(2\pi+2\pi k)/3} = e^{\mathrm{i}(\frac{2}{3}\pi+k\frac{2}{3}\pi)}$ with $k = 0, 1, 2$, thus

$$w_0 = e^{i2\pi/3} = \cos\frac{2}{3}\pi + \mathrm{i}\sin\frac{2}{3}\pi = \frac{-1+i\sqrt{3}}{2},$$
$$w_1 = e^{\mathrm{i}(2\pi/3+2\pi/3)} = e^{4\mathrm{i}\pi/3} = \frac{-1-i\sqrt{3}}{2} = w_0^* \text{ and}$$
$$w_2 = e^{\mathrm{i}(2\pi/3+2\cdot 2\pi/3)} = e^{2\pi i} = +1.$$

A last example is: $w_k = \sqrt[3]{8\mathrm{i}} = \sqrt[3]{8}e^{\mathrm{i}(\pi/2+2\pi k)/3} = 2e^{\mathrm{i}(\pi/6+k2\pi/3)}$ with $k = 0, 1, 2$, thus

$$w_0 = 2e^{\mathrm{i}\pi/6} = 2(\cos\frac{\pi}{6} + \mathrm{i}\sin\frac{\pi}{6}) = 2(\frac{\sqrt{3}}{2} + \frac{i}{2}) = \sqrt{3} + i,$$
$$w_1 = 2e^{\mathrm{i}(\pi/6+2\pi/3)} = 2e^{5\mathrm{i}\pi/6} = 2(\cos\frac{5\pi}{6} + \mathrm{i}\sin\frac{5\pi}{6}) = -\sqrt{3} + i \text{ and}$$
$$w_2 = 2e^{\mathrm{i}(\pi/6+4\pi/3)} = 2e^{3\mathrm{i}\pi/2} = -2i.$$

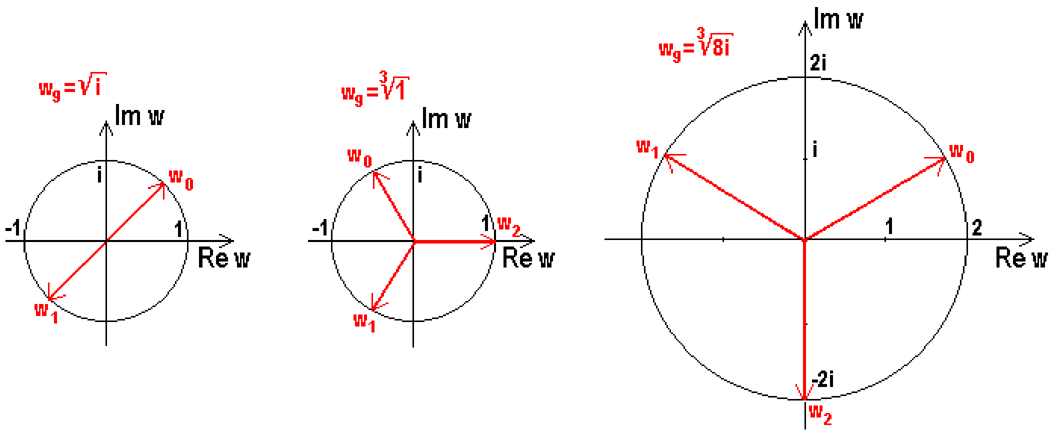


Figure 8.16: Illustration of the roots in the w-plane for these three examples: a) $w = \sqrt{\mathrm{i}}$, b) $w = \sqrt[3]{1}$ and c) $w = \sqrt[3]{(8\mathrm{i})}$.

Exercise 8.13 Roots:

Calculate and sketch the following roots: a) $w = \sqrt[3]{\mathrm{i}}$, *b)* $w = \sqrt[4]{(-1)}$, *c)* $w = \sqrt[8]{1}$, *d)* $w = \sqrt[4]{-\mathrm{i}}$ *and e)* $w = \sqrt[2]{8\mathrm{i}}$.

8.3.8 Logarithms

Finally we take a look at an infinitely multiple valued function, the natural

logarithm: $w = \ln z$ for $z \neq 0$.

Since the logarithm is the inverse function of the exponential function, the following has to hold true:

$$z = |z|e^{\mathrm{i}\arg z} = e^{w+2\pi \mathrm{i}k} = e^{\mathrm{Re}\, w+\mathrm{i}(\mathrm{Im}\, w+2\pi k)} \quad \text{with } k \in \mathbb{Z}$$

meaning, we have for the

real part of the logarithm: $\mathrm{Re}(\ln z) = \ln |z|$

and the

imaginary part of the logarithm: $\mathrm{Im}(\ln z) = \arg z + 2\pi k$ with $k \in \mathbb{Z}$

has infinitely many values, which differ by integer multiples of 2π. We choose for instance the region $-\pi < \mathrm{Im}(\mathrm{Ln}\, z) \leq \pi$ as the **principal value** and write it with the capital letter: Ln, so that we obtain:

$$\ln z = \mathrm{Ln}\, z + 2\pi \mathrm{i}k \text{ with } k \in \mathbb{Z}$$

For example it follows

1)	from	$\mathrm{e}^0 = 1:$	$\mathrm{Ln}\, 1 = 0,$	thus	$\ln 1 = 2\pi \mathrm{ik},$
2)	from	$\mathrm{e}^{\mathrm{i}\pi} = -1:$	$\mathrm{Ln}(-1) = \mathrm{i}\pi,$	thus	$\ln(-1) = \mathrm{i}\pi(1+2\mathrm{k})$and
3)	from	$3\mathrm{e}^{\mathrm{i}\pi} = -3:$	$\mathrm{Ln}(-3) = \ln|-3| + \mathrm{i}\pi,$	thus	$\ln(-3) = 1.098 + \mathrm{i}\pi(1+2\mathrm{k}).$

Exercise 8.14 Logarithms:

Calculate $\ln \mathrm{i}$.

8.3.9 General Power

We need the logarithm here just as for the real field to define the **general power** function, which has therefore infinitely many values in the complex field:

General Power: $w = b^z := e^{z \ln b}$, where $\ln b = \ln |b| + \mathrm{i}(\beta + 2\pi g)$ with $g \in \mathbb{Z}$,

since $b = |b|e^{\mathrm{i}(\beta+2\pi g)}$.

As **principal value** of b^z we take $e^{z\,\mathrm{Ln}\,b}$ with the principal value $\mathrm{Ln}\,b$ of $\ln b$. With this, we can for example calculate 1^{i}: $w_g = 1^{\mathrm{i}} = e^{\mathrm{i}\ln 1} = e^{\mathrm{i}2\pi\mathrm{i}g} = e^{-2\pi g} \in \mathbb{R}$:
$w_0 = 1, \quad w_1 = e^{-2\pi} = 1.87 \cdot 10^{-3}, \quad w_2 = e^{-4\pi} = 3.49 \cdot 10^{-6}$, but also $w_{-1} = e^{2\pi} = 535.49$, etc..
Surprisingly enough, even i^{i} is real: $w_g = \mathrm{i}^{\mathrm{i}} = e^{\mathrm{i}\ln \mathrm{i}} = e^{\mathrm{i}\cdot\mathrm{i}(\pi/2+2\pi g)} = e^{-(\pi/2+2\pi g)} \in \mathbb{R}$: $w_0 = e^{-\pi/2} = 0.20788$.

Exercise 8.15 Powers:

Calculate $w_g = \mathrm{i}^{\mathrm{i}}$ *for* $g = 1$ *and* $g = -1$.

Exercise 8.16 Exponentiating:

When you have a free minute, realize that $(\mathrm{i}^{\mathrm{i}})^{\mathrm{i}}$ *has more different values than* $\mathrm{i}^{(\mathrm{i}\cdot\mathrm{i})}$.

After these curious jokes, we leave behind the complex numbers. During your studies, you will often return to these things, you will learn more complex functions, for example the group of broken linear functions: $w = \frac{az+b}{cz+d}$, whose mapping preserves angles and circles. You will study and classify the different kinds of singularities as well as acquaint yourselves with the Laurent series as a generalization of the Taylor series. Particularly, in a mathematical lecture on complex analysis, you will investigate when complex functions are differentiable (= analytic = holomorphic) and learn how you can calculate difficult real integrals with help of the elegant residue theorem in the complex plane. We are sure that you will not be able to resist the charm and beauty of this mathematical theory, even though for science it is not really necessary, although very useful.

Chapter 9

VECTORS

9.1 Three-dimensional Euclidean Space

9.1.1 Three-dimensional Real Space

After we have dealt with functions of real variables and their analysis as well as with the simplest arithmetic rules of complex numbers, we turn in the last chapter of this course to the three-dimensional space in which we live and in which whole physics happens.

We all have a good visualization of three-dimensional space from our everyday experience. Typically we think of a room or a box with length, width and height in which one can place three perpendicular rulers, one measuring length, one measuring width and one height. The position for instance of the upper right front corner of your keyboard can be chosen to be the intersection point of the three rulers. Mathematicians construct **the three-dimensional real space** $\mathbb{R}^3$ according to this model logically as an "outer product" of three real straight lines $\mathbb{R}^1$, as we have used them to illustrate the real numbers: $\mathbb{R}^3 = \mathbb{R}^1_1 \otimes \mathbb{R}^1_2 \otimes \mathbb{R}^1_3$.

9.1.2 Coordinate Systems

Physicists like to be able to characterize every point of the space exactly, and they use for this purpose a coordinate system. To do so they first choose *completely arbitrarily*, but often very suitably one point of the space as **zero**, also called the **origin**. Through this point they lay once more *completely arbitrarily* three real straight lines and number them: $\mathbb{R}^1_1, \mathbb{R}^1_2$ and $\mathbb{R}^1_3$. This already defines a **coordinate system**. Nevertheless, usually they are a little more demanding and require that these three straight lines are pairwise perpendicular to one another: $\mathbb{R}^1_k \perp \mathbb{R}^1_l$ for $k, l = 1, 2, 3$, yielding what is called a **Cartesian coordinate system**. If in addition the positive halfs of the three straight lines, now called “**coordinate axes**”, are arranged or numbered such that the rotation of the positive half of the 1-axis by an angle $\pi/2$ around the 3-axis into the positive half of the 2-axis is, when viewing along the direction of the positive 3-axis, a clockwise rotation (this is usually called a **right-handed screw**) then we have constructed the ideal, a (Cartesian) **right handed coordinate system**. To some of you this numbering of the axes is known as **right-hand rule** because the positive halfs of the 1-, 2- and 3-axis are arranged like thumb, forefinger and middle finger of the spread **right hand**. In the following this kind of coordina

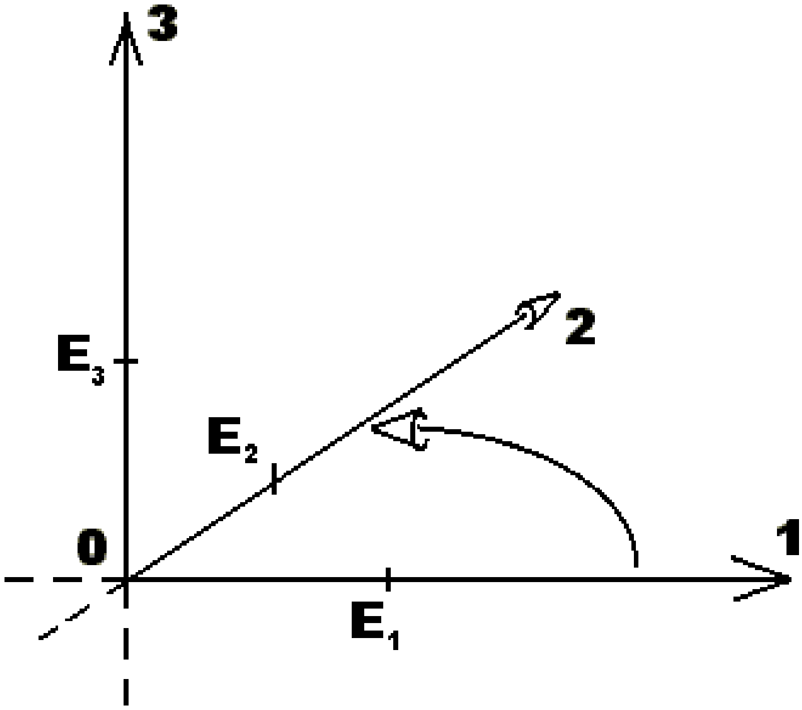


Figure 9.1: “Right-handed screw”

On the positive half of each of the three coordinate axes $\mathbb{R}^1_k$ for $k = 1, 2, 3$ lies, in each case once again *completely arbitrarily* chosen, the accompanying unit point E_k whose distance from zero fixes the length unit much as for a ruler. Thus every point $P \in \mathbb{R}^3$ of the three-dimensional space can be labeled unambiguously by a **coordinate triple** of real numbers $P = (p_1, p_2, p_3)$. The number p_k is in each case the height above the **coordinate plane** spanned up by the other two real straight lines $\mathbb{R}^1_l$ and $\mathbb{R}^1_m$, measured in the unit chosen by E_k.

9.1.3 Euclidean Space

Because we want to do measurements we need a measure of the distance between any two points $P = (p_1, p_2, p_3)$ and $Q = (q_1, q_2, q_3)$. Therefore, we introduce in the three-

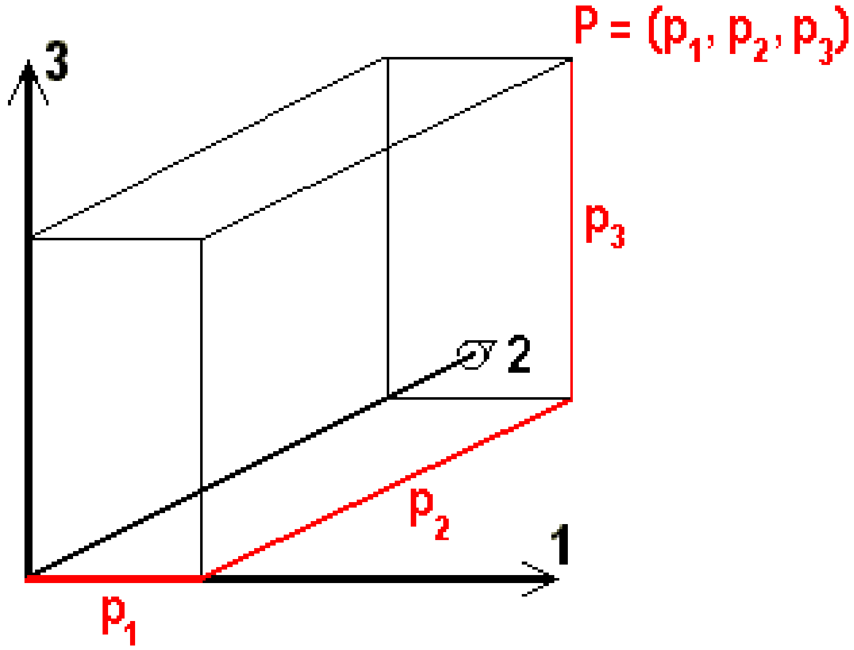


Figure 9.2: Point with its coordinate triple

dimensional space a distance measure, namely according to **Pythagoras** the root of the squares of the coordinate differences:

Distance: $|PQ| = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + (q_3 - p_3)^2}$

In particular, the distance of the point P from zero O is $|PO| = \sqrt{p_1^2 + p_2^2 + p_3^2}$.

Exercise 9.1 Distances between points:

Calculate the distance between the points $P = (2, 2, 4)$ *and* $Q = (1, -2, 0)$ *as well as the distance of each point from zero.*

Mathematicians call a space with this distance definition **Euclidean**. Then the distance between two different points $P \neq Q$ is always positive $|PQ| = |QP| > 0$, and the distance of a point from itself is zero: $|PP| = 0$. As usual the **triangle inequality** holds which says that in every triangle the sum of two side lengths is larger than the length of the third side:

Triangle inequality: $|PQ| \leq |PR| + |RQ|$

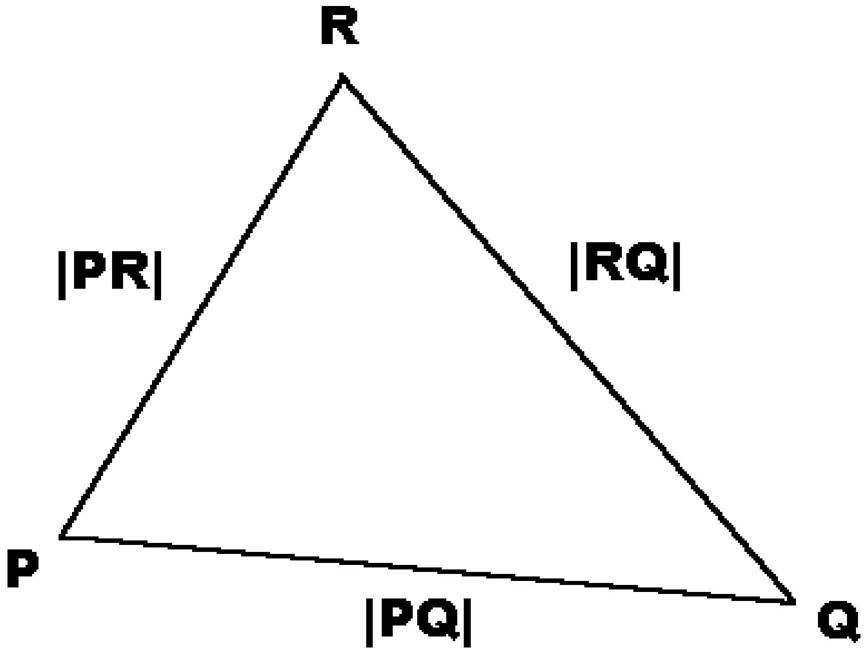


Figure 9.3: Illustration of the triangle inequality

For two different distances $|PQ| \neq |PR|$ between points P and Q or R, respectively, one naturally can decide which is the larger one, because real numbers are involved which are ordered on a straight line. The points of the space $\mathbb{R}^3$, however, can not be ordered, just as the complex numbers cannot.

9.1.4 Transformations of the Coordinate System

Hopefully you have noticed how *arbitrarily* we have proceeded with the choice of the coordinate system. Because an intelligently chosen coordinate system can be extraordinarily helpful for the daily work of physicists, the freedom in the choice of coordinates, however, and the independence of the results of physical measurements of this choice is of outstanding importance Therefore we want to discuss what would have happened had we made another choice:

In particular, **four kinds of transformations of the coordinate system** are of practical interest. We select in each case a simple but typical example:

1. **TRANSLATIONS** (SHIFTS), e.g. by a distance of 1 cm in the 3-direction:

First the freedom in the *choice of the origin* attracts our attention: How would the coordinates (p_1, p_2, p_3) of the point P look, if we had chosen instead of the point O another point, e.g., $\widehat{O} = E_3$, as zero which is shifted by the distance $|E_3O| = 1$ cm in the positive 3-direction?

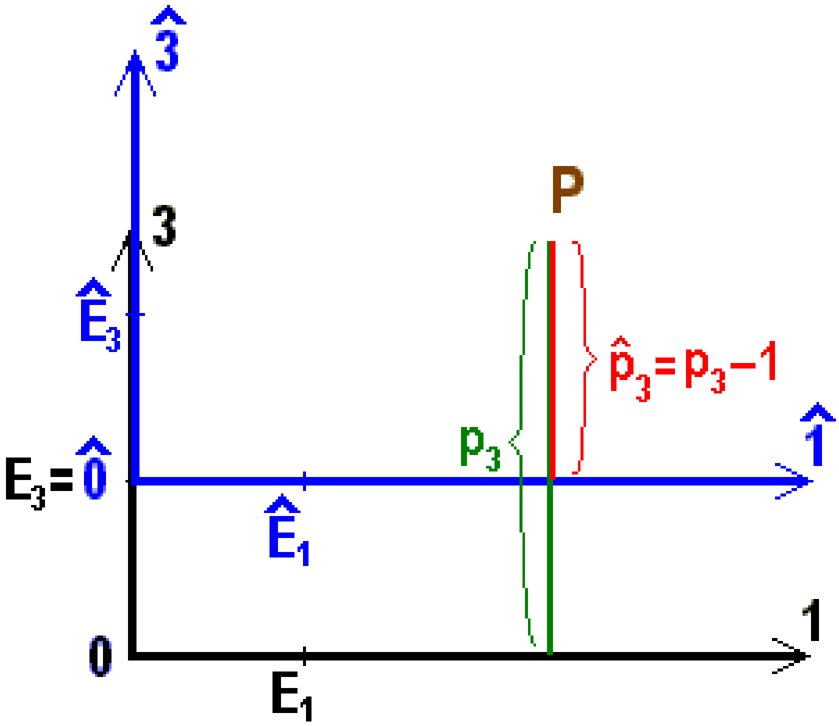


Figure 9.4: Translation by the distance 1 cm in the positive 3-direction

From the figure we directly read off that for the numbers $\hat{p}_k$ holds: $\hat{p}_3 = p_3 - 1$, while $\hat{p}_1 = p_1$ and $\hat{p}_2 = p_2$ remain unchanged, thus all together:

$$\boxed{(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (p_1, p_2, p_3 - 1).}$$

Insert: Equals sign: *Considering more carefully one finds that the well-known equals sign is used often with* ***different meanings****:*

If we write $1/2 = 2/4$ *we mean "****numerically equal****", i.e. the numerical values are identical after calculation. In this sense the equals sign is used in the equation* $(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (p_1, p_2, p_3 - 1)$*: the numerical value of the first coordinate* $\hat{p}_1$ *of the point P in the shifted coordinate system* $\widehat{S} : \hat{p}_1 = p_1$ *is equal to the number* p_1 *of the 1-coordinate in the old system. Accordingly, the numerical value of the 3-coordinate in the shifted system* $\hat{p}_3 = p_3 - 1$ *is smaller by 1 than the corresponding number in the old coordinate system.*

If, however, we write $P = (p_1, p_2, p_3)$ we mean: the point P **"is in the coordinate system S represented by"** *the three given coordinate numbers. In this equation it needs to be specified in which system the coordinates have been measured. Usually one takes the point of view, this is signified by the symbols for the coordinates which remind us of the choice of our coordinate axes. Care is also needed if we want to represent the point in the shifted system. On no account can we simply write $P = (\hat{p}_1, \hat{p}_2, \hat{p}_3)$, because from this you could conclude $(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (p_1, p_2, p_3)$ which is definitively wrong as we have seen.*

There are obviously ***three ways out*** *of this difficulty:*

(a) *Either we define a new sign for "is represented in the system ...by ...", while we add a symbol for the coordinate system to the equals sign: possibly "$\hat{=}$" with the meaning "is represented in the system $\widehat{S}$ by ...", e.g., $P\hat{=}(\hat{p}_1, \hat{p}_2, \hat{p}_3)$. But this is complicated and difficult to realize with computer fonts.*
(b) *Or we use the hat which labels the shifted coordinate system to also label the indices $p_{\hat{k}}$, e.g. $p_{\hat{k}}$, to remind us that the coordinates of the old point P are now being given with respect to the new $\hat{1}$-, $\hat{2}$- or $\hat{3}$-axis of the system $\widehat{S}$. This too is very difficult to realize on the computer, and moreover uncommon.*
(c) *Therefore, we choose here the third possibility: We place the hat on the coordinates, thus $\hat{p}_k$ as done with alternative (a),* ***avoiding****, however, the statement: "it is represented in the system $\widehat{S}$ by ..." in equations.*

Now that we have clarified this problem, there is no more reason for insecurity or misunderstanding.

It is easy to generalize this result on translations to arbitrary distances and to other directions, so that we can skip this here.

Exercise 9.2 Point coordinates:

What are the coordinates of the point P in a coordinate system whose origin lies in the point $\widehat{O} = (1, 2, -3)$?

Instead of this we now turn to other particularly important coordinate transformations with which **the origin** remains **unchanged**: first to

2. **<u>ROTATIONS</u>** (TURNINGS), e.g. by the angle φ around the 3-direction:

We consider in addition to our original coordinate system S a new one $\widehat{S}$, which with unchanged origin $\widehat{O} = O$ was rotated, e.g., by an angle φ around the 3-axis seen clockwise in the positive 3-direction:
From the figure we see immediately that $\hat{p}_1 = p_1 \cos\varphi + p_2 \sin\varphi$ and $\hat{p}_2 = p_2 \cos\varphi - p_1 \sin\varphi$ while $\hat{p}_3 = p_3$, thus:

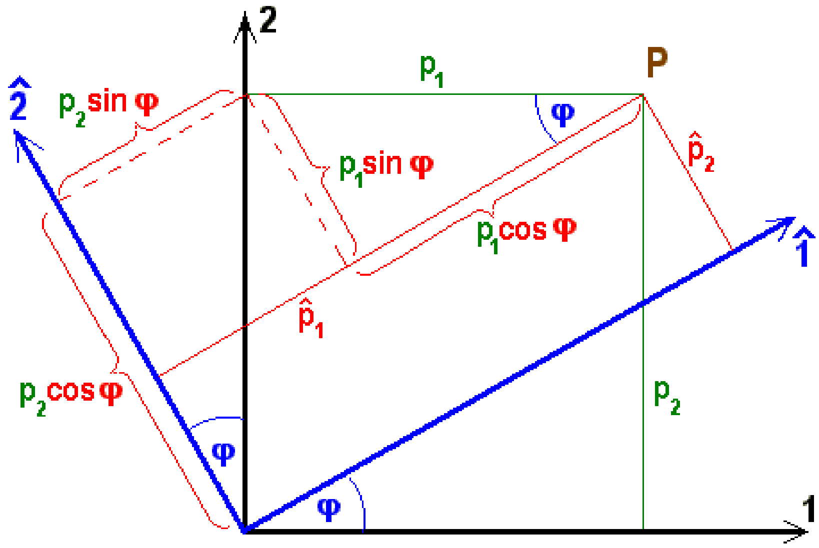


Figure 9.5: Rotation of the coordinate system by an angle φ around the 3-direction

$$\boxed{(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (p_1 \cos\varphi + p_2 \sin\varphi,\ p_2 \cos\varphi - p_1 \sin\varphi,\ p_3)}$$

e.g., for: $\varphi = \pi/2$: $\quad (\hat{p}_1, \hat{p}_2, \hat{p}_3) = (p_2, -p_1, p_3)$.

Exercise 9.3 Rotated coordinate systems:

Calculate the coordinates of the point P in a coordinate system $\widehat{S}$ which was rotated around the 3-direction relative to S by the angles $\varphi = \pi$, $\quad \varphi = \pi/4$ or $\varphi = \pi/6$.

Further very interesting transformations which leave the origin invariant are the

3. **REFLECTIONS** (MIRRORINGS), e.g. through the origin (parity transformation).

   It is sufficient to consider just one reflection, because it is possible to construct all other reflections from this and suitable rotations. We choose the point reflection through the origin which is illustrated in the following figure and is known by physicists by the name **parity transformation**:
   We see immediately from the figure that all coordinates go over in their negative:

$$\boxed{(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (-p_1, -p_2, -p_3)}$$

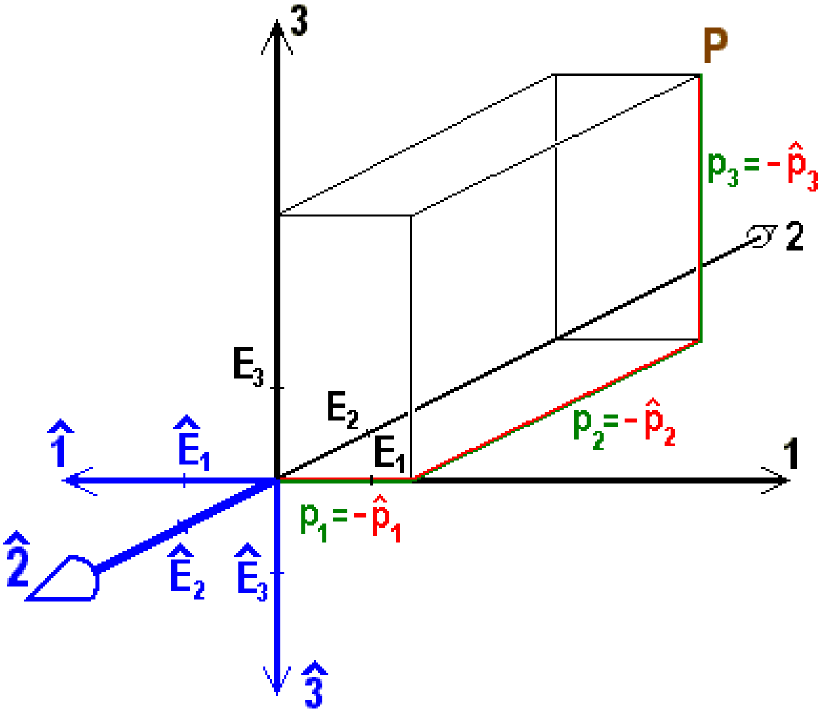


Figure 9.6: Reflection of the coordinate system through the origin

Exercise 9.4 Reflections composed of parity and rotations:
Show how one can obtain the reflection in the 1-2-plane with $\hat{\hat{p}}_1 = p_1, \quad \hat{\hat{p}}_2 = p_2$ *and* $\hat{\hat{p}}_3 = -p_3$ *from the parity transformation and a rotation.*

All reflections and especially the parity transformation have a remarkable property which we easily recognize from the figure above: If we turn namely the positive half of the $\widehat{1}$-axis by the angle $\pi/2$ into the positive half of the $\widehat{2}$-axis, this is, seen in the direction of the positive $\widehat{3}$-axis **no more a right screw**, but a left screw (anti-clockwise). Meaning: after a reflection a left coordinate system has become from our right coordinate system. For people who have agreed on the use of right coordinate systems that is no pleasant situation, but we must learn to live with it and to find means and ways to detect also a hidden reflection always immediately when we want to stay with right coordinate systems.

As the final example of transformations of the coordinate system we examine:

4. <u>**DILATATIONS**</u> (STRETCHINGS): especially of all three axes by a common factor, e.g., 10:

   Such a thing arises in practice when we want to measure lengths, instead of in centimeters cm, in decimeters dm. Under such a **scale change** the origin naturally remains invariant, and also the coordinate axes remain unchanged. Only the measure points E_k are shifted along the axes so that $\widehat{E}_k = 10E_k$, thus all distances from the origin are numerically increased according to $|\widehat{E}_k O| = 10$:

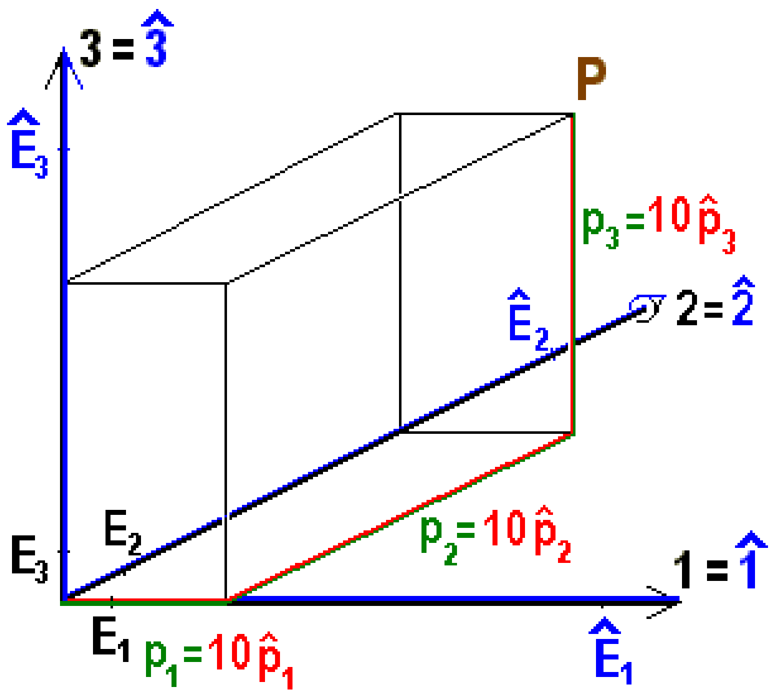


Figure 9.7: Scale change of the coordinate system by a factor 10

If you measure the keyboard of your PC instead of in cm in larger units, e.g. in dm, you will receive of course smaller numbers, namely:

$$(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (\tfrac{1}{10}p_1, \tfrac{1}{10}p_2, \tfrac{1}{10}p_3)$$

In summary we can see that the coordinates of one and same point P are substantially different in different coordinate systems so that we must always pay attention in the following to the coordinate systems involved when we want to describe physical states and processes.

Up to now, however, we have dealt only with the points of the three-dimensional Euclidean space and can describe only a static "still life" of masses, charges, etc.. However, physics becomes much more interesting when motion enters the game.

9.2 Vectors as Displacements

9.2.1 Displacements

In the following we want to investigate what happens when we **shift** a mass point or a charge by a **certain straight-line distance in a certain direction**, e.g. from the starting point $P = (p_1, p_2, p_3)$ to the final point $Q = (q_1, q_2, q_3)$ in a fixed arbitrarily chosen coordinate system. From a given starting point P with the three coordinates p_1, p_2 and p_3 (**three** numbers!) a displacement by a given distance or length a (**fourth** number!) necessarily ends somewhere on the surface of a sphere with radius a around the starting point. In addition the direction of the displacement is fixed by **two further** numbers (!), e.g., the geographic longitude and latitude on the surface of the sphere or by two other angles θ and φ. Thus all together we need **six** real numbers to fully represent a displacement in three-dimensional space.

Such **displacements**, their causes and results are in many problems the central physical entities, for instance displacement per unit time as velocity, or changes of velocity with time as accelerations. By Newton's Second Law, forces are proportional to accelerations, and therefore directly linked to displacements, consequently also forces per unit charge as electric field strengths, etc...

When we look more exactly at typical physical examples, e.g. the velocity of a car on a straight section of a motorway, we find out that firstly the car consists of a lot of points which have all identical velocity, and that secondly in most cases physics is not really interested in the special starting values of all these points. The **really important information is the displacement common to all points of a body** independent of the special starting or final points. When we take into account these physical requirements economically, we arrive at the concept of vectors:

9.2.2 Vectors

We call a displacement a **vector** $\vec{a}$ (or sometimes also a **tensor of first order**), if we suppress the special starting and final position of a shifted object, if we are only interested in the "**displacement in itself**". In other words, only the amount of the distance of the displacement and the direction are interesting, and it does not matter where in the space the displacement takes place.

Because of the homogeneity of three-dimensional space, which is without exception presumed in physics, this concept is advantageous for the formulation of **universal valid physical laws**. It means mathematically that, as for the introduction of the rational numbers where we have equated, e.g. $1/2 = 2/4 = 3/6 = ...$, we divide displacements into equivalence classes, and identify all displacements with the same distance amount and direction. For illustration, we then can elect any representative of the class if necessary, for instance the so-called **position vector**, by applying the displacement to the origin.

After we have introduced a Cartesian coordinate system for the description of points by their coordinates in three-dimensional Euclidean space, the question rises how can we characterize the vectors in this system of three pairwise perpendicular coordinate axes. To do this we choose arbitrarily as representative of our vector $\vec{a}$ (marked by a small arrow above the letter) a starting point $P = (p_1, p_2, p_3)$, shift this by the distance of the length a in the prescribed direction and reach thus the final point as in the following figure:

$$\textbf{vector: } \vec{a} = \overrightarrow{PQ} := \begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 \\ q_3 - p_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

(Except the small arrows used here above the Latin letter or above both points of the representative, other labels for vectors are underlined small letters, letters in boldface type or German letters.) In contrast to a certain special displacement of a certain point which has to be labeled by six real numbers, a vector is characterized only by **three real numbers**. In contrast to the three real "coordinates of a point" one calls these three real numbers

$$\textbf{vector components: } \quad a_k = q_k - p_k.$$

Note that we also put them, like the point coordinates, in round brackets. To distinguish both we write the components, however, usually (like above) **as a column** one **below** another instead of one after the other. If you want to have the vector components written like the point coordinates one after the other, you should add an upper index "T", as an abbreviation for "**transposed**", i.e.:

$$\textbf{transposed vector: } a^T = (a_1, a_2, a_3).$$

As we can see from the figure, the three components of a vector are the lengths of the three projections of the representative $\overrightarrow{PQ}$ on the coordinate axes or also the coordinates of the final point $A = (a_1, a_2, a_3)$ which we reach by displacement if we have chosen the origin as starting point. In this case we are using the position vector as representative: $\overrightarrow{OA} = (a_1, a_2, a_3)$. By the use of these special representatives it becomes directly clear that there is a reversibly unambiguous relation between the totality of the points of the

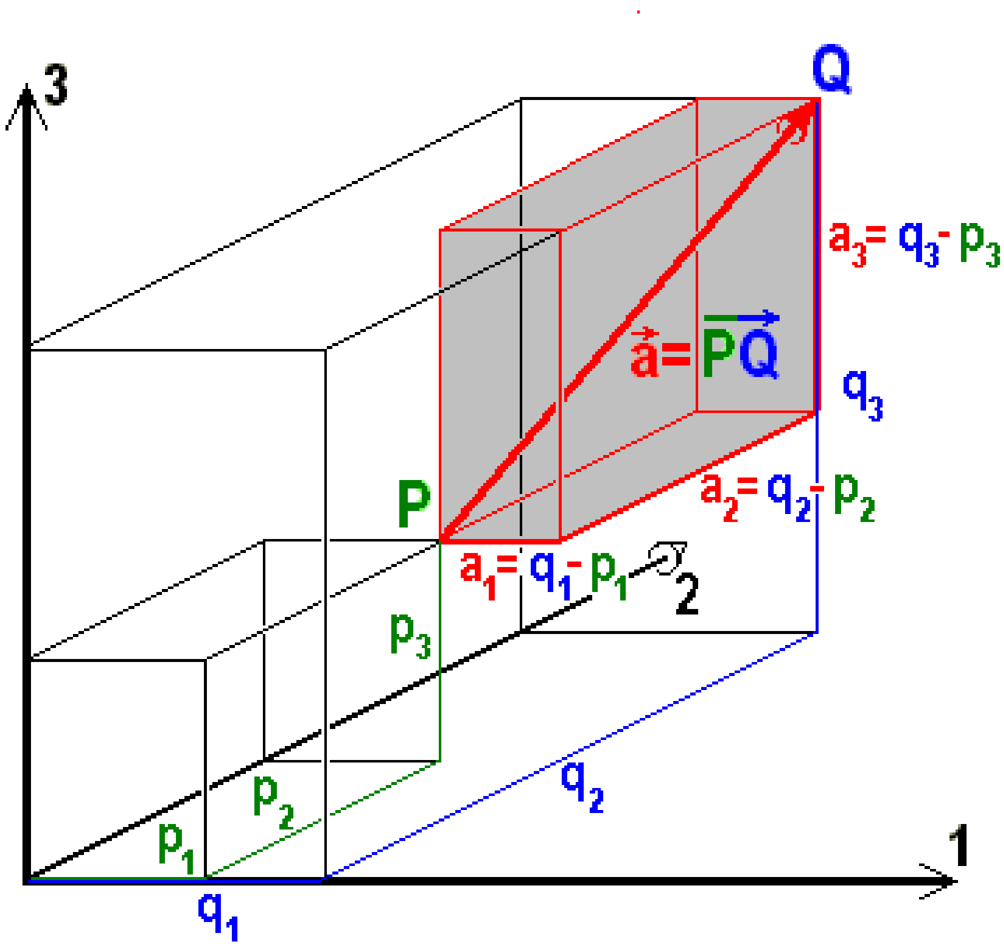


Figure 9.8: Vector components

$\mathbb{R}^3$ and the set of all vectors of the so-called **vector space**. Mathematicians call this an **isomorphism**.

Because a vector is unambiguously characterized by its three components, i.e. by the distance amount and the two direction angles, **a vector equation is equivalent to three equations for the single components**. For this reason vector notation can be a very efficient manner of writing:

vector equation: $\vec{a} = \vec{b} \iff a_k = b_k$ for $k = 1, 2, 3.$

The amount of the shift distance, i.e. **the length of a vector**, is determined from its components according to Pythagoras, just as was the distance of two points from coordinate differences:

length:

$$a := |\vec{a}| = |PQ| = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + (q_3 - p_3)^2}$$
$$= |OA| = \sqrt{a_1^2 + a_2^2 + a_3^2}.$$

Exercise 9.5 Lengths of vectors:

Determine the lengths of the following vectors: $\begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix}$ *and* $\begin{pmatrix} -1 \\ 4 \\ 1 \end{pmatrix}$.

To make completely clear the typical differences (which exist in spite of the isomorphism) between the components of a vector $\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ and the coordinates of the final point $A = (a_1, a_2, a_3)$ of the representing position vector (to which the different manners of writing should remind us over and over again) we investigate once more what happens to the vector components under transformations of the coordinate system:

9.2.3 Transformations of the Coordinate Systems

We want to investigate how the components a_k of the fixed physically-given vector $\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, possibly represented by $\overrightarrow{PQ}$, are changed if the coordinate system is subjected to the four selected special transformations of Section 9.1.4 .

We start with the:

1. **<u>TRANSLATIONS</u>** (SHIFTS): e.g. by 1 in the 3-direction:

   At first we are once again concerned with the arbitrariness of the choice of the origin: How would the components of our vector look if we had instead of the point O, selected as zero another point, e.g. $\widehat{O} = E_3$ which is shifted by the distance $|E_3O| = 1$ cm in the positive 3-direction? With the help of our results from 9.1.4.1 we get:

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \end{pmatrix} = \begin{pmatrix} \hat{q}_1 - \hat{p}_1 \\ \hat{q}_2 - \hat{p}_2 \\ \hat{q}_3 - \hat{p}_3 \end{pmatrix} = \begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 \\ q_3 - p_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}.$$

Since the translation terms of the starting and end points of the representative cancel in the difference, we obtain (as expected because of the free mobility in the vector definition) the **translational invariance of vectors**, i.e. the arbitrary choice of the origin of our coordinate system has **no** consequences for the vector components.

From this also the **length** of the vectors is a **translation invariant quantity**:

$$\hat{a} = \sqrt{\hat{a}_1^2 + \hat{a}_2^2 + \hat{a}_3^2} = \sqrt{a_1^2 + a_2^2 + a_3^2} = a.$$

However, not all vector-like quantities in physics are translation invariant and not in every physical problem, e.g., the forces acting on a rigid body away from the centre of mass, or also the field strength of an inhomogeneous electric field. Physicists speak then of **bound vectors**. In such cases, before the application of the vector algebra which we will develop in the next sections, we have to consider in each case separately to what extent the achieved results can be applied.

As the second example we examine

2. **<u>ROTATIONS</u>** (TURNINGS), e.g. by an angle φ around the 3-direction:

   Keeping the origin $\widehat{O} = O$ fixed we consider once more in addition to our old coordinate system S as in Figure 9.5 a new one: $\widehat{S}$ which was rotated e.g. by an angle φ around the 3-axis seen clockwise in the positive 3-direction and we get (for instance with the representative $\vec{a} = \overrightarrow{OA}$):

$$\hat{a}_1 = a_1 \cos\varphi + a_2 \sin\varphi, \quad \hat{a}_2 = a_2 \cos\varphi - a_1 \sin\varphi, \quad \hat{a}_3 = a_3.$$

   For the rule according to which the new coordinates can be calculated from the old ones, mathematics offers a formulation which is known to most of you from school: the **matrix formulation**:

   In order to get it, we write the three transformation equations one below the other and complete the display with zeros in the following manner:

$$\begin{aligned} \hat{a}_1 &= a_1 \cos\varphi + a_2 \sin\varphi + a_3 0 \\ \hat{a}_2 &= -a_1 \sin\varphi + a_2 \cos\varphi + a_3 0 \\ \hat{a}_3 &= a_1 0 + a_2 0 + a_3 1. \end{aligned}$$

   The factors necessary to get the new components $\hat{a}_k$ from the old ones a_l are summarized by the following (3x3)-matrix $\mathbf{D}^{(3)}(\varphi)$:

$$\textbf{rotation matrix : } \mathbf{D}^{(3)}(\varphi) := \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

with the matrix elements $D^{(3)}_{zs}(\varphi)$ for $z, s = 1, 2, 3$, where the index $\boldsymbol{z}$ marks the (horizontal) lines (German: **Z**eilen) and $\boldsymbol{s}$ the (vertical) columns (German: **S**palten). For example,

$$\begin{aligned}
D^{(3)}_{11}(\varphi) &= D^{(3)}_{22}(\varphi) = \cos\varphi \quad \text{and} \\
D^{(3)}_{12}(\varphi) &= -D^{(3)}_{21}(\varphi) = \sin\varphi, \quad \text{since the 1-2-plane is rotated,} \\
D^{(3)}_{33}(\varphi) &= 1 \quad \text{to signalize that the 3-axis remains unchanged, and} \\
D^{(3)}_{13}(\varphi) &= D^{(3)}_{31}(\varphi) = D^{(3)}_{23}(\varphi) = D^{(3)}_{32}(\varphi) = 0.
\end{aligned}$$

If instead of the single matrix elements $D^{(3)}_{zs}(\varphi)$ we mean the entire matrix with its nine elements, we use a bold capital letter $\mathbf{D}^{(3)}(\varphi)$. The three equations for the calculation of the new coordinates from the old ones is obtained in this new formulation through the following prescription of a **generalized multiplication** rule for $z = 1, 2, 3$:

$$\hat{a}_z = D^{(3)}_{z1}(\varphi)a_1 + D^{(3)}_{z2}(\varphi)a_2 + D^{(3)}_{z3}(\varphi)a_3 = \sum_{s=1}^{3} D^{(3)}_{zs}(\varphi)a_s =: D^{(3)}_{zs}(\varphi)a_s.$$

In the last term the convenient **Einstein summation convention** was used. This allows us to omit the sum symbol whenever **two identical indices** (here the two indices **s**) appear and in this way to signal the summation even without using the explicit sum symbol.

According to this prescription we get the column vector of the components in the rotated coordinate system by "multiplying" the column vector of the components in the old system from the left by the rotation matrix:

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \mathbf{D}^{(3)}(\varphi) \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

To do this we imagine most simply the column vector transposed and pushed line by line over the rotation matrix, multiply the terms lying on top of each other and add up the three products.

Exercise 9.6 Special vectors in the rotated coordinate system

What are the components of the vectors $\vec{a} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $\vec{b} = \begin{pmatrix} 0 \\ -2 \\ 3 \end{pmatrix}$ *and* $\vec{c} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$ *in a coordinate system* $\widehat{S}$ *which was rotated with respect to the original system* S *by* $\pi, \pi/2$ *or* $\pi/4$ *around the 3-direction?*

Exercise 9.7 Change of the vector components by special rotations of the coordinate system:

How are the components a_k of a vector $\vec{a}$ changed, if we rotate the coordinate system by the angle π or $\pi/2$ around the 3-axis?

In the matrix formulation the transformation equations can be most easily memorized: The number 1 occupies the position $D^{(3)}_{33}(\varphi)$ because the 3-axis as the rotation axis remains unchanged during the rotation and the 1-2 plane is rotated by φ. Also the extension to rotations around the two other axes can be imagined easily: e.g. for a rotation $D^{(1)}(\varphi)$ by φ around the 1-axis the matrix element must certainly be $D^{(1)}_{11}(\varphi) = 1$ and the 2-3-plane is rotated:

$$\mathbf{D}^{(1)}(\varphi) := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix}$$

meaning that $\hat{a}_1 = a_1, \hat{a}_2 = a_2\cos\varphi + a_3\sin\varphi$ and $\hat{a}_3 = a_3\cos\varphi - a_2\sin\varphi$.

Exercise 9.8 Rotations around the 2-axis:

Realize through a sketch like our Figure 9.5 above that the coordinates of a point A and consequently also the components of a vector $\vec{a}$ transform according to the following rotation matrix in case of a rotation around the 2-axis.

$$\mathbf{D}^{(2)}(\varphi) := \begin{pmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{pmatrix}$$

meaning that $\hat{a}_3 = a_3\cos\varphi + a_1\sin\varphi$ and $\hat{a}_1 = a_1\cos\varphi - a_3\sin\varphi$.

From this exercise you can find the transformation formulae for the three rotations from each other without much calculation simply by replacing the indices **cyclically** (i.e. in a circle), meaning 1 by 2, 2 by 3 and 3 by 1:

Exercise 9.9 Special rotation matrices:

Determine the following rotation matrices: $\mathbf{D}^{(1)}(\pi/2), \mathbf{D}^{(1)}(\pi/6)$ and $\mathbf{D}^{(2)}(\pi/4)$.

The transformation formula for the components of a vector with rotations of the coordinate system is an important characteristic feature of vectors, and sometimes vectors are simply defined as quantities whose three components transform with rotations of the coordinate system in the given manner. Indeed, when a physicist wants to find out whether a quantity having three components is a vector, he measures

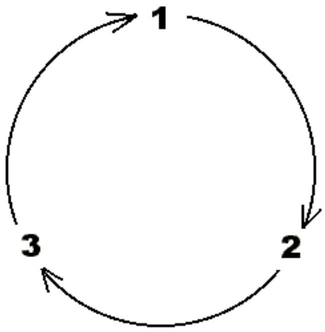


Bild 9.1: cyclic replacement

its components in two coordinate systems rotated with respect to each other and investigates whether the measured results can be connected by the corresponding rotation matrix.

We also examine the rotation behaviour of the **length of a vector**:

$$\hat{a} = \sqrt{\hat{a}_1^2 + \hat{a}_2^2 + \hat{a}_3^2} = \sqrt{a_1^2 \cos^2 \varphi + a_2^2 \sin^2 \varphi + a_1^2 \sin^2 \varphi + a_2^2 \cos^2 \varphi + a_3^2} = a$$

and we find that it is **rotation invariant** as expected.

Insert: <u>M A T R I C E S</u>:

The rotation matrices are only one example of quantities with two indices which are called matrices by mathematicians. It is possible to define calculation rules generally for $(z \times s)$-matrices, i.e. schemes with z lines and s columns and to examine their structures. We want to restrict our considerations to quadratic $(n \times n)$-matrices and even more specifically to **(3 × 3)-matrices with real elements.**

We denote the matrices by underlined capital letters, e.g. $\mathbf{A}$. Their elements A_{zs} carry two indices: the left one $\boldsymbol{z}$ denotes the (horizontal) line and the right one $\boldsymbol{s}$ the (vertical) column of the matrix:

$$\textbf{matrix:} \quad \mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$$

Some kinds of matrices have special names because of their importance:
In particular, **diagonal matrices** are of special importance, having only the three elements A_{11}, A_{22} and A_{33} along the so-called **main diagonal** (:from the left on top downward to the right) different from 0. The second diagonal (:from the right up downward to the left) is, in comparison, much less important.

$$\textbf{diagonal matrix:}\quad \mathbf{A} = \begin{pmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & A_{33} \end{pmatrix}$$

The matrices of rotations by a multiple of π are examples of diagonal matrices: $\mathbf{D}^{(1)}(\pi), \mathbf{D}^{(2)}(\pi)$ and $\mathbf{D}^{(3)}(\pi)$.

Half the way to the diagonal structure the **triangle form** is worth mentioning, which has only zeros either above or below the main diagonal:

$$\textbf{triangle matrix:}\quad \mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{pmatrix}$$

Also **matrices in box form** are especially convenient for many purposes. In these matrices non-zero elements are only in "boxes" around the main diagonal. Our rotation matrices $\mathbf{D}^{(1)}(\varphi)$ and $\mathbf{D}^{(3)}(\varphi)$ are of this kind.

$$\textbf{matrix in box form:}\quad \mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & A_{33} \end{pmatrix}$$

A simple operation that can be carried out with every matrix is **transposition**: This means the reflection of all matrix elements through the main diagonal, or in other words, the exchange of lines and columns: $A^T_{zs} = A_{sz}$

$$\textbf{transposed matrix:}\quad \mathbf{A}^T = \begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix}$$

There exist matrices for which transposition does not change anything: They are called **symmetric**.

$$\textbf{symmetric matrix:} \quad \mathbf{A} = \mathbf{A}^T = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{12} & A_{22} & A_{23} \\ A_{13} & A_{23} & A_{33} \end{pmatrix}$$

These symmetric matrices occur very often in physics and have the advantage that they can be brought to diagonal form by certain simple transformations.

As you see immediately, each symmetric matrix has only six independent elements.

If the reflection through the main diagonal leads to a minus sign, the matrix is called **antisymmetric**:

$$\textbf{antisymmetric matrix:} \quad \mathbf{A} = \begin{pmatrix} 0 & A_{12} & A_{13} \\ -A_{12} & 0 & A_{23} \\ -A_{13} & -A_{23} & 0 \end{pmatrix}$$

Of course the diagonal elements have to vanish in this case. Apparently an antisymmetric (3×3)-matrix has only three independent elements. That is the deeper reason for the existence of a vector product in three dimensions, as we will soon see in more detail.

Finally we mention a special quantity of every quadratic matrix: **The sum of the elements along the main diagonal** is called the **trace** (:in German "Spur") of the matrix:

$$\textbf{trace:} \quad \mathrm{Sp}\,\mathbf{A} = \mathrm{tr}\,\mathbf{A} = \sum_{k=1}^{3} A_{kk} \equiv A_{kk}$$

You can easily imagine that a termwise addition can be defined for the set of real (3×3)-matrices and that these form an **Abelian group** of addition with Associative Law, unique zero-matrix, exactly one negative for every matrix and Commutative Law, since the corresponding properties of the real numbers can simply be transferred to this new situation. Also the termwise **multiplication with a numerical factor** is possible and leads to the usual Distributive Laws.

Much more important for physics is however the **multiplication of two** (3×3)**-matrices** which corresponds in the case of transformation matrices to two transformations of the coordinate system carried out one after the other:

The following multiplicative instruction holds:

matrix multiplication: $\mathbf{C} = \mathbf{B}\,\mathbf{A} \iff C_{zs} := \sum_{k=1}^{3} B_{zk}A_{ks} \equiv B_{zk}A_{ks}$

In the last part above the summation symbol is omitted according to the **Einstein summation convention**, since the two identical indices signalize the summation well enough.

To calculate the product matrix element C_{zs} in the z-th line and the s-th column you may imagine the s-th (vertical) column A_{ms} of the factor matrix $\mathbf{A}$ on the right side put horizontally upon the z-th line B_{zm} of the left factor matrix $\mathbf{B}$, elements on top of each other multiplied and the three products added: e.g. $C_{12} = \sum_{k=1}^{3} B_{zk}A_{ks} \equiv B_{zk}A_{ks} = B_{11}A_{12} + B_{12}A_{22} + B_{13}A_{32}$, thus altogether:

$$\begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23}\ ) \\ A_{31} & A_{32} & A_{33} \end{pmatrix} = \begin{pmatrix} B_{1k}A_{k1} & B_{1k}A_{k2} & B_{1k}A_{k3} \\ B_{2k}A_{k1} & B_{2k}A_{k2} & B_{2k}A_{k3} \\ B_{3k}A_{k1} & B_{3k}A_{k2} & B_{3k}A_{k3} \end{pmatrix}$$

Exercise 9.10 Matrix multiplication:

Multiply the following transformation matrices:
a) $\mathbf{D}^{(1)}(\theta)\mathbf{D}^{(3)}(\varphi)$ *and compare with* $\mathbf{D}^{(3)}(\varphi)\mathbf{D}^{(1)}(\theta)$,
b) especially $\mathbf{D}^{(1)}(\pi)\mathbf{D}^{(3)}(\pi/2)$ *to be compared with* $\mathbf{D}^{(3)}(\pi/2)\mathbf{D}^{(1)}(\pi)$,
c) $\mathbf{D}^{(3)}(\theta)\mathbf{D}^{(3)}(\varphi)$ *and compare with* $\mathbf{D}^{(3)}(\varphi)\mathbf{D}^{(3)}(\theta)$,
d) $\mathbf{D}^{(1)}(\pi)\mathbf{D}^{(3)}(\pi)$ *and compare with* $\mathbf{D}^{(3)}(\pi)\mathbf{D}^{(1)}(\pi)$.

The most important discovery to be made by working through the Exercise 9.10 is the fact that generally **no commutative law** holds for rotations, and consequently not for the representing matrices. You can easily check this visually with every match box as is illustrated in the following Figure:

The examples from Exercise 9.10 have already shown to you that in some **exceptional cases** the **commutative law** nevertheless holds: all rotations around one and the same axis are for instance commutable. Also all diagonal matrices are commutable with each other. This is the reason for their popularity. If $\mathbf{A}\,\mathbf{B} \neq \mathbf{B}\,\mathbf{A}$, the so-called commutation relation $[\mathbf{A}, \mathbf{B}] := \mathbf{A}\,\mathbf{B} - \mathbf{B}\,\mathbf{A}$ promises to be an interesting quantity. This will acquire great significance in quantum mechanics later on.

Apart from commutability, matrix multiplication behaves as expected: There holds an

Associative Law: $\mathbf{C}(\mathbf{B}\,\mathbf{A}) = \mathbf{C}\,\mathbf{B}\,\mathbf{A} = (\mathbf{C}\,\mathbf{B})\mathbf{A}$.

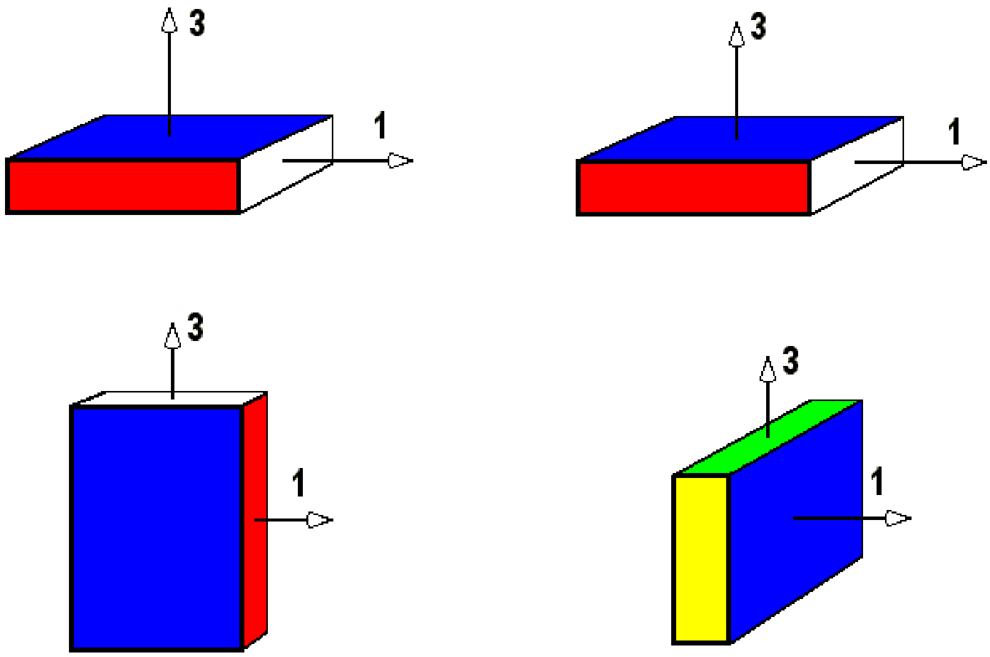


Figure 9.10: Match box, first rotated by 90° around the 3-axis and then by 90° around the 1-axis, compared with a box which is first rotated around the 1-axis and afterwards around the 3-axis.

Exercise 9.11 Associative Law for matrix multiplication

Verify the Associative Law for the Euler rotation: $\mathbf{D}^E(\psi, \theta, \varphi) := \mathbf{D}^{(3)}(\psi)\mathbf{D}^{(1)}(\theta)\mathbf{D}^{(3)}(\varphi)$ *which leads us from the space fixed coordinate system to the body fixed system of a rotating gyroscope.*

A uniquely determined

$$\boxed{\textbf{unit matrix:} \quad \mathbf{1} := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ with } \mathbf{A}\,\mathbf{1} = \mathbf{A} = \mathbf{1}\,\mathbf{A} \quad \forall \mathbf{A},}$$

exists independently whether you multiply from the right or the left.

Only with the

$$\boxed{\textbf{inverse matrix:} \quad \mathbf{A}^{-1} \text{ with } \mathbf{A}\,\mathbf{A}^{-1} = \mathbf{1}}$$

do we encounter a certain complication analogous to the condition "$\neq 0$" for division by a real number. A uniquely determined inverse matrix exists only for the so-called **non-singular** matrices. These are matrices whose *determinant* does not vanish: $|\mathbf{A}| \neq 0$. The determinants, the most important characteristics of matrices, will be treated in an extra insert in the following subsection.

For our transformation matrices, however, this constraint is unimportant. For these matrices the inverse is simply the transposed one $\mathbf{A}^{-1} = \mathbf{A}^T$ which exists in every case as we have seen: Mathematicians call these matrices orthogonal and we will inspect these carefully later on:

$$\textbf{orthogonal matrix:} \quad \mathbf{A}^{-1} = \mathbf{A}^T \text{ or } \mathbf{A}\,\mathbf{A}^T = \mathbf{1}$$

As our next transformation of the coordinate system we treat the:

3. <u>**REFLECTIONS**</u> (MIRRORINGS), e.g. through the origin (the parity transformation).

   We consider once more only the parity transformation, i.e. the reflection through the origin which transfers all coordinates and consequently all components into their negatives: Also this transformation which trivially leaves the origin invariant $\widehat{O} = O$ may be described by a matrix $\mathbf{P}$, namely by the negative of the unit matrix which we denote by $\mathbf{1}$: $\mathbf{P} = -\mathbf{1}$

$$\textbf{parity:} \quad \mathbf{P} := \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

   Thus we obtain for the components of a vector in the reflected system:

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \mathbf{P} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} -a_1 \\ -a_2 \\ -a_3 \end{pmatrix}$$

   All vectors whose components $\hat{a}_k = -a_k$ change sign through a reflection in the origin are called **polar vectors**. Again not all vectors important in physics have this property. We will soon come to physical vectors, e.g. the angular momentum, which are parity invariant. We will call these vectors **axial vectors**.

   Nevertheless, for all kinds of vectors the **length** is **parity invariant**, because in every case holds

$$\hat{a} = \sqrt{\hat{a}_1^2 + \hat{a}_2^2 + \hat{a}_3^2} = \sqrt{a_1^2 + a_2^2 + a_3^2} = a.$$

Insert: D E T E R M I N A N T S:

The most important characteristic of a matrix is its determinant. The following notations are used:

$$\textbf{determinant:} \quad \det \mathbf{A} \equiv |\mathbf{A}| = \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix}$$

Leibniz gave the following

definition of the determinant:

$$\det \mathbf{A} := \sum_{(P1,P2,P3)} (-1)^{\sigma(P1,P2,P3)} A_{1\,P1} A_{2\,P2} A_{3\,P3} =$$
$$= A_{11}A_{22}A_{33} + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32}$$
$$- A_{11}A_{23}A_{32} - A_{12}A_{21}A_{33} - A_{13}A_{22}A_{31}.$$

This yields a **real number**, namely the sum resp. the difference of terms, each of which is a product of three matrix elements. The line indices (on the left hand) are always $z = 123$ for all terms, while the column indices (on the right hand) go through all permutations Pz of these three numbers: $(P1, P2, P3) = 123$, 231, 312; 132, 213, 321. The sign of each term is fixed by the number of transpositions (:interchanges of each two indices) which are needed to get the concerned configuration from the configuration 123. The first three of the configurations above can be obtained by an even number of transpositions. They get a plus sign, the remaining three, obtained by an odd number of interchanges, were subtracted: e.g. into 132 (odd), but 231 into 312 (even),.... In the case of (3×3)-matrices we get six summands for which the even permutations can also be found through **cyclic permutation**.

Besides these generally valid definitions there exist several different methods for the calculation of the determinant of a matrix. Here we will get to know two of them: first we consider the **Sarrus rule** valid in particular for (3×3)-matrices:

To get it we write the first and second column once more on the right beside to the determinant of interest:

$$\textbf{to the Sarrus rule:} \quad \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} \begin{matrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{matrix}$$

Within this scheme we multiply the elements in the main diagonal $A_{11}A_{22}A_{33}$ with each other. To this product we add the product of the three elements standing next to the right side in direction of the main diagonal $+A_{12}A_{23}A_{31}$ and $+A_{13}A_{21}A_{32}$. From this result we subtract the product of the elements in the secondary diagonal $-A_{11}A_{23}A_{32}$ and once more two times the products of the three matrix elements next to the right side in direction of the secondary diagonal $-A_{12}A_{21}A_{33}$ and $-A_{13}A_{22}A_{31}$. This procedure yields the desired determinant.

Very often a further method leads to the answer more quickly, the so-called **development with respect to the first line**: Since this is an iterative method, we first note that the determinant of a (2×2)-matrix consists of the product of the two diagonal elements $A_{11}A_{22}$ diminished by the product of the two elements in the secondary diagonal $-A_{12}A_{21}$. Exactly this (2×2)-determinant is left over after removing both the third line and the third column from our desired (3×3)-determinant. It is called **adjoint** and characterized through the indices of the removed rows:

$$\textbf{adjoint:} \quad \text{adj}_{33}(\mathbf{A}) = A_{11}A_{22} - A_{12}A_{21}.$$

With help of these adjoints the desired (3×3) -determinant can be written in the following way:

$$\begin{aligned}|\mathbf{A}| &= (-1)^{1+1}A_{11}\,\text{adj}_{11}(\mathbf{A}) + (-1)^{1+2}A_{12}\,\text{adj}_{12}(\mathbf{A}) + (-1)^{1+3}A_{13}\,\text{adj}_{13}(\mathbf{A}) \\ &= A_{11}(A_{22}A_{33} - A_{23}A_{32}) - A_{12}(A_{21}A_{33} - A_{23}A_{31}) + A_{13}(A_{21}A_{32} - A_{22}A_{31}).\end{aligned}$$

We can see immediately that the restriction to the development with respect to the **first line** means no limitation, since the determinant has many symmetry properties. With help of those we can easily get developments with respect to other lines or columns. Here we use the term

row as the common term for **line or column**.

To reach a concise notation, we sometimes summarize matrix elements arranged one on top of the other in the form of so-called **column vectors** $\mathbf{A}_k$ e.g.

$$|\mathbf{A}| := |\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3| \text{ with column vectors: } \mathbf{A}_k := \begin{pmatrix} A_{1k} \\ A_{2k} \\ A_{3k} \end{pmatrix}$$

Determinants have a lot of interesting characteristic symmetry properties which make their calculation and handling exceptionally easy.

Symmetry properties of the determinants:

A determinant is **invariant with respect to transposition**, i.e. with respect to reflection through the main diagonal:

$$\boxed{|\mathbf{A}^T| = |\mathbf{A}|.}$$

A determinant stays unchanged when to one of its lines, a **linear combination of the other lines is added**, e.g.:

$$\boxed{|\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3| = |\mathbf{A}_1 + \lambda\mathbf{A}_2, \mathbf{A}_2, \mathbf{A}_3|.}$$

A determinant changes sign with every **permutation of two rows**: e.g.

$$\boxed{|\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3| = -|\mathbf{A}_2, \mathbf{A}_1, \mathbf{A}_3|}$$

Determinants are **homogeneous with respect to their rows**: for a real number $\lambda \in \mathbb{R}$

$$\boxed{|\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3| = \frac{1}{\lambda}|\lambda\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3|.}$$

A determinant vanishes if the **row vectors are coplanar** (: linearly dependent) or one of the row vectors is the zero vector:

$$\boxed{|\mathbf{A}_1, \mathbf{A}_2, \mathbf{0}| = 0.}$$

A determinant with an **odd number of dimensions** stays unchanged through **cyclic permutation of the rows**:

$$\boxed{|\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3| = |\mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_1|.}$$

A determinant with an **odd number of dimensions** vanishes if the matrix is **antisymmetric $\mathbf{A}^T = -\mathbf{A}$** :

$$\boxed{|\mathbf{A}^T| = -|\mathbf{A}| = 0.}$$

Determinants are **additive** if the summands differ only in one row: e.g.

$$|\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3| + |\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_4| = |\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 + \mathbf{A}_4|.$$

The determinant of a **product of two matrices** is equal to the product of the determinants of the two factor matrices:

$$|\mathbf{A}\,\mathbf{B}| = |\mathbf{A}||\mathbf{B}|$$

Exercise 9.12 Calculate the following determinants:

$$a)\begin{vmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{vmatrix} \qquad b)\begin{vmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 2 & 1 & 3 \end{vmatrix} \qquad c)\begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{vmatrix}$$

$$d)\begin{vmatrix} 1 & 0 & 2 \\ 3 & 2 & 1 \\ 2 & 2 & 1 \end{vmatrix} \qquad e)\begin{vmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 0 \end{vmatrix} \qquad f)\begin{vmatrix} \frac{17}{7} & 4 & \frac{31}{14} \\ \frac{3}{7} & \frac{3}{2} & \frac{3}{14} \\ 2 & \frac{5}{2} & 2 \end{vmatrix}$$

$$g)\begin{vmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{vmatrix} \qquad h)\begin{vmatrix} A_{11} & 0 & 0 \\ A_{21} & A_{22} & A_{23} \\ A_{31} & 0 & A_{33} \end{vmatrix} \qquad i)\begin{vmatrix} A_{11} & 0 & 0 \\ 0 & A_{22} & 0 \\ 0 & 0 & A_{33} \end{vmatrix}$$

$$j)\begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & 0 \\ A_{31} & 0 & 0 \end{vmatrix} \qquad k)\begin{vmatrix} 0 & 0 & A_{13} \\ 0 & A_{22} & 0 \\ A_{31} & 0 & 0 \end{vmatrix} \qquad l)\begin{vmatrix} A_{11} & A_{12} & b_1 \\ A_{21} & A_{22} & b_2 \\ 0 & 0 & b_3 \end{vmatrix}$$

$$m)\begin{vmatrix} 0 & a_3 & -a_2 \\ -a_3 & 0 & a_1 \\ a_2 & -a_1 & 0 \end{vmatrix} \qquad n)\begin{vmatrix} A_{11}-\lambda & A_{12} & A_{13} \\ A_{21} & A_{22}-\lambda & A_{23} \\ A_{31} & A_{32} & A_{33}-\lambda \end{vmatrix}$$

Exercise 9.13 Determinants of rotation matrices:

Calculate the determinants of $\mathbf{D}^{(1)}(\varphi)$, $\mathbf{D}^{(2)}(\varphi)$ *and* $\mathbf{D}^{(3)}(\varphi)$.

Exercise 9.14 Determinants of reflections:

Calculate the determinants of the parity matrix $\mathbf{P}$, *of the matrix* $\mathbf{D}^{(3)}(\pi)$ *and of the product* $\mathbf{P}\,\mathbf{D}^{(3)}(\pi)$.

Finally we turn to the:

4. **DILATATIONS** (STRETCHINGS): especially of all axes by a common factor, say the factor 10:

   We examine again e.g. as an example the scale change from centimeter cm to decimeter dm, while the coordinate axes remain unchanged and only the unit points E_k on the axes are shifted, so that the distances from the origin $|\widehat{E_k O}| = 10$ are increased. In this case the values of the vector components decrease. The corresponding transformation matrix of the dilatation is: $\mathbf{S} = \frac{1}{10}\,\mathbf{1}$:

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{10} & 0 \\ 0 & 0 & \frac{1}{10} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \frac{1}{10}\,\mathbf{1} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{10}a_1 \\ \frac{1}{10}a_2 \\ \frac{1}{10}a_3 \end{pmatrix}$$

   Under a scale change, of course, no vector stays invariant and also the length is reduced by the factor 1/10:

$$\hat{a} = \sqrt{\hat{a}_1^2 + \hat{a}_2^2 + \hat{a}_3^2} = \sqrt{\frac{a_1^2}{100} + \frac{a_2^2}{100} + \frac{a_3^2}{100}} = \frac{a}{10}.$$

With these investigations the most difficult part of our program to understand vectors has been accomplished. Now we can proceed to study how we can calculate with vectors, always having in mind that we deal with displacements having freely eligible starting points.

It still remains to stress, that of course there exist also physical quantities for which only one single measurement, value and unit, is involved, as for example mass, charge, temperature, current strength, etc.. These quantities are called **scalars** (or sometimes also **tensors of order zero**) in contrast to the vectors (which occasionally are called also tensors of first order) and still more complicated physical quantities, as for example the momentum of inertia.

9.3 Addition of Vectors

9.3.1 Vector Sum

While for the points of three-dimensional Euclidean space there was no reason to think of any arithmetic operations, it makes sense from the physical point of view to ask for arithmetic operations involving vectors. For example several **displacements** can be implemented **one after the other**: After we shifted a mass point e.g. from the point P in accordance with the vector $\vec{a}$ to the point Q, we can shift it further afterwards from Q according to the shift prescription of the vector $\vec{b}$ to the point R. We would have evidently reached the same end position, if we had pushed it immediately in one course from P to R in accordance with the vector $\vec{c} = \vec{a} + \vec{b}$, which we call the vector sum of $\vec{a}$ and $\vec{b}$:

vector sum: $\overrightarrow{PQ} + \overrightarrow{QR} = \vec{a} + \vec{b} = \vec{c} = \overrightarrow{PR} \iff a_k + b_k = c_k$ for $k = 1, 2, 3.$

The geometrical addition of the vectors takes place **componentwise**, meaning the algebraic addition for each of the three components separately. This is where the designation "addition" for vectors comes from.

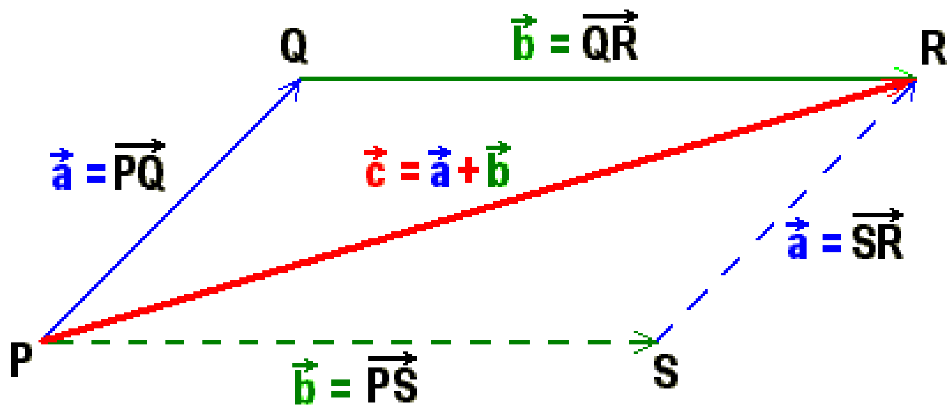


Figure 9.11: Vector addition

9.3.2 Commutative Law

If we consider the free movability of the vectors, it immediately follows from Figure 9.11 that with the auxiliary point S the

Commutative Law of addition: $\overrightarrow{PQ} + \overrightarrow{QR} = \vec{a} + \vec{b} = \vec{b} + \vec{a} = \overrightarrow{PS} + \overrightarrow{SR}$

This also results algebraically from the Commutative Law of addition of each of the components as real numbers.

This commutability of the summands leads us to a second geometrical regulation to formulate the vector product of two vectors $\vec{a}$ and $\vec{b}$: We select two representatives for the vectors with the **same starting** point, augment the figure to a parallelogram and obtain this way the sum $\vec{c}$ as the diagonal of the parallelogram. This construction was already found by Newton and is known as the **force parallelogram** to many of you, whereby the sum $\vec{c} = \vec{a} + \vec{b}$ represents the resulting force. Moreover this geometrical regulation has the advantage that it can also be used for non-translation-invariant "bound" vectors, if these "attack" at the same point, e.g. position vectors.

Exercise 9.15 On the force parallelogram:

1. *Three polar dogs pull on a sleigh with the same strength, but under relative angles of* $60°$. *Which force does the dog owner have to exert and in which direction, if he does not want the sleigh to drive off yet?*

2. *Form the sum of seven coplanar vectors of the length a with angle differences of* $30°$.

9.3.3 Associative Law

The addition of **three** vectors $\vec{a}, \vec{b}, \vec{c}$ satisfies the

Associative Law: $(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c})$,

because this law applies to the components as real numbers (see the next Figure):

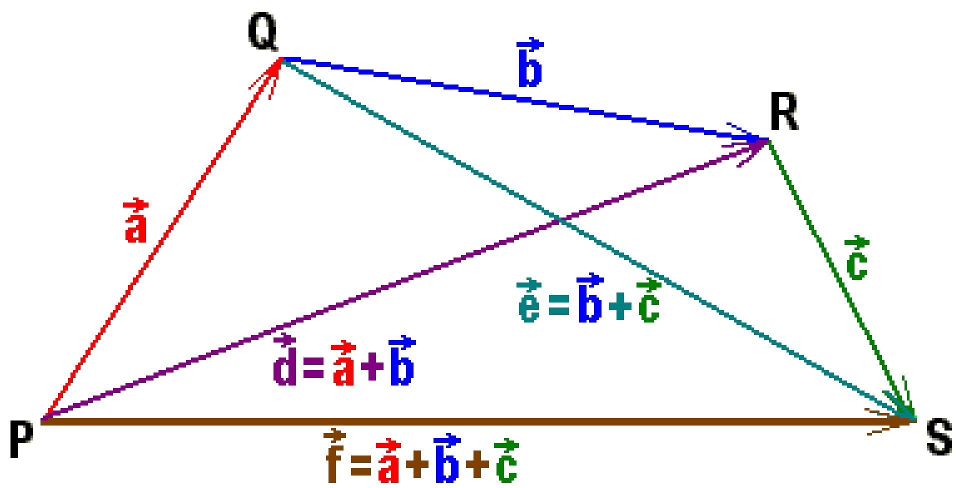


Figure 9.12: Associative Law

9.3.4 Zero-vector

It is possible that after several shifts we return to the original starting point of the first summand:

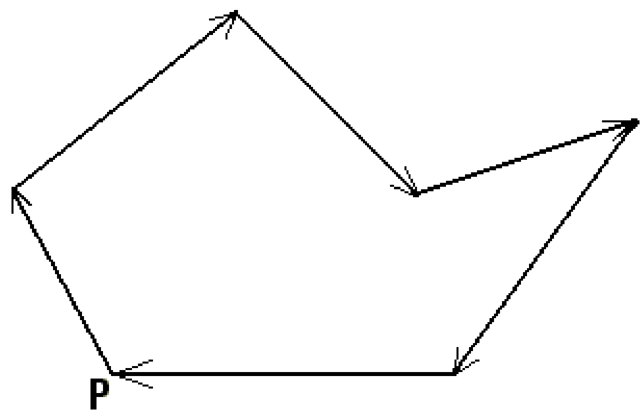


Figure 9.13: Zero-vector

From this we can conclude the existence of exactly one **zero-vector** $\vec{0}$, i.e. "no shift" with length $|\vec{0}| = 0$ and (exceptionally) an indefinite direction. Similarly as for **real numbers** the following applies:

$$\exists! \ \vec{0} \text{ with } \vec{a} + \vec{0} = \vec{a} \quad \forall \vec{a}.$$

9.3.5 Negatives and Subtraction

It is always possible to return to the starting point even after one shift. As for real numbers this means that there always exists an unambiguous reversal for each shift vector $\vec{a}$, the **negative vector**:

$$\exists! \ -\vec{a} \text{ with } \vec{a} + (-\vec{a}) = \vec{0}.$$

Simply put, the starting and final points of the representative are to be exchanged: $-\overrightarrow{PQ} = \overrightarrow{QP}$.

With these negatives of vectors a **subtraction** becomes definable also for vectors, much as for real numbers, i.e.

$$\forall \vec{a}, \vec{b} \quad \exists! \, \vec{x} \text{ with } \vec{a} + \vec{x} = \vec{b}.$$

The vector $\vec{x} = \vec{b} - \vec{a} = \vec{b} + (-\vec{a})$ with the components $x_k = b_k - a_k$ for $k = 1, 2, 3$ solves the above equation.

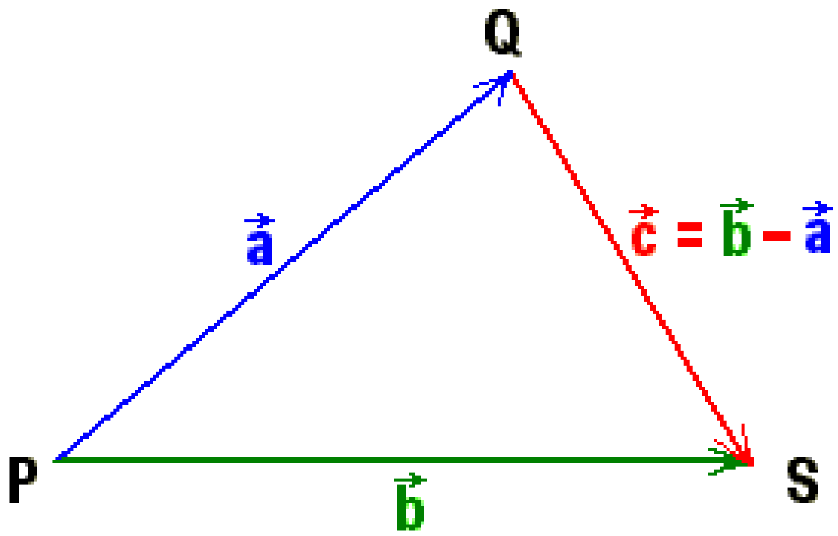


Figure 9.14: Construction of the difference vector

Exercise 9.16 Sums and differences of vectors:

Draw graphically the sum and the two possible differences of the following vectors:
a) $\vec{a} = (4, 0, 0)$, $\vec{b} = (-2, 1, 0)$; *b)* $\vec{a} = (0, -2, 0)$, $\vec{b} = (3, 0, 0)$;
c) $\vec{a} = (-3, -1, 0)$, $\vec{b} = (0, -3, 0)$; *d)* $\vec{a} = (-3, -2, 0)$, $\vec{b} = (-3, 2, 0)$;
e) $\vec{a} = (-2, -3, 0)$, $\vec{b} = (-2, -1, 0)$; *f)* $\vec{a} = (1, 3, 0)$, $\vec{b} = (4, -4, 0)$.

With the validity of the Associative Law and the existence of exactly one zero-vector and one unambiguously determined negative to each vector, the vectors form a **group of addition**, which is in fact **Abelian** because of the Commutative Law.

9.4 Multiplication with Real Numbers, Basis Vectors

9.4.1 Multiple of a Vector

If we can implement several shifts one after the other, then we naturally can also apply the same shift several times, in particular: $\vec{a} + \vec{a} = 2\vec{a}$. In this way we come to the same point, as we would have done if we had twice the distance in the same direction in one step. Similarly, this holds for any real number $\alpha \in \mathbb{R}$ as factor:

$$\textbf{multiple:} \quad \vec{b} = \alpha\vec{a} \text{ with } \alpha \in \mathbb{R} \iff b_k = \alpha a_k \text{ for } k = 1, 2, 3.$$

With $\alpha > 0$ this yields a vector with an α-fold length into the same direction as $\vec{a}$, with $\alpha < 0$ the vector points in the opposite direction. As for the length we have:

$$|\alpha\vec{a}| = \sqrt{(\alpha a_1)^2 + (\alpha a_2)^2 + (\alpha a_3)^2} = |\alpha||\vec{a}|.$$

9.4.2 Laws

Despite the imbalance of this product formation, the following characteristics are passed directly from the appropriate ones of the field of real numbers $\mathbb{R}$ to the present situation, because the vector components are real numbers. With the vectors $\vec{a}$ and $\vec{b}$ and the real numbers α and β:

the **Commutative Law**:	$\alpha\vec{a} = \vec{a}\alpha$	$\Longleftarrow$	$\alpha a_k = a_k\alpha$,	
the **Associative Law**:	$\beta(\alpha\vec{a}) = (\beta\alpha)\vec{a}$	$\Longleftarrow$	$\beta(\alpha a_k) = (\beta\alpha)a_k$	as well as
two Distributive Laws:	$(\alpha + \beta)\vec{a} = \alpha\vec{a} + \beta\vec{a}$	$\Longleftarrow$	$(\alpha + \beta)a_k = \alpha a_k + \beta a_k$	
and	$\alpha(\vec{a} + \vec{b}) = \alpha\vec{a} + \alpha\vec{b}$	$\Longleftarrow$	$\alpha(a_k + b_k) = \alpha a_k + \alpha b_k$	

The last law means for example graphically, that the diagonal of a parallelogram is stretched accordingly if the sides are stretched by a factor α.

9.4.3 Vector Space

If the elements of a set form an Abelian group of addition, and if multiplication with the elements of some field is defined, as was described above, mathematicians call this set a **vector space** or simply a **linear space**.

Indeed, we can then form **linear combinations** within the space, for instance $\vec{c} = \alpha\vec{a}+\beta\vec{b}$, whose characteristics we want to discuss briefly in the following:

9.4.4 Linear Dependence, Basis Vectors

A) Firstly we consider linear combinations of **two** different vectors: $\vec{a}_1 \neq \vec{a}_2$: Two cases can occur:

A1) If an $\alpha \in \mathbb{R}$ exists so that $\vec{a}_2 = \alpha\vec{a}_1$ applies, or differently expressed: if in $\alpha_1\vec{a}_1 + \alpha_2\vec{a}_2 = 0$ at least one of the factors $\alpha_k \neq 0$, e.g. $\alpha_2 \neq 0$, so that the equation can be solved for $\vec{a}_2$, i.e. $\vec{a}_2 = -\frac{\alpha_1}{\alpha_2}\vec{a}_1 =: \alpha\vec{a}_1$,

this means that the vector $\vec{a}_2$ can be expressed by a vector which is proportional to $\vec{a}_1$. The two vectors $\vec{a}_1$ and $\vec{a}_2$ are then called **linearly dependent**, sometimes also **collinear**.

A2) If there is no α such that $\vec{a}_2 = \alpha\vec{a}_1$ for all $\alpha \in \mathbb{R}$, thus $\vec{a}_2 \neq \alpha\vec{a}_1$ or differently expressed: if $\alpha_1\vec{a}_1 + \alpha_2\vec{a}_2 = \vec{0}$ can only be achieved when both $\alpha_1 = 0$ and $\alpha_2 = 0$,

then the two vectors $\vec{a}_1 = \overrightarrow{OA_1}$ and $\vec{a}_2 = \overrightarrow{OA_2}$ **span a plane** through the three points O, A_1 and A_2, and each point of this plane can be represented by a linear combination $\alpha_1\vec{a}_1 + \alpha_2\vec{a}_2$ with real factors α_1 and α_2.

B) Secondly we examine linear combinations of **three** different vectors: $\vec{a}_1$, $\vec{a}_2$ and $\vec{a}_3$, whereby once again two cases are possible:

B1) If two real numbers α_1 and α_2 can be found, such that $\vec{a}_3 = \alpha_1\vec{a}_1 + \alpha_2\vec{a}_2$ or differently expressed: if in $\alpha_1\vec{a}_1 + \alpha_2\vec{a}_2 + \alpha_3\vec{a}_3 = 0$ at least one $\alpha_k \neq 0$, e.g. $\alpha_3 \neq 0$, so that it can be dissolved for $\vec{a}_3$,

this means (as shown above) that $\vec{a}_3$ can be represented by a vector which lies completely in the plane defined by $\vec{a}_1$ and $\vec{a}_2$. One then calls the three vectors $\vec{a}_1$, $\vec{a}_2$ and $\vec{a}_3$ **linearly dependent**, sometimes also **coplanar**.

B2) If $\vec{a}_3 \neq \alpha_1\vec{a}_1 + \alpha_2\vec{a}_2$ or differently expressed: if $\alpha_1\vec{a}_1 + \alpha_2\vec{a}_2 + \alpha_3\vec{a}_3 = \vec{0}$ is only attainable when all three $\alpha_k = 0$ vanish,

then the three vectors $\vec{a}_1$, $\vec{a}_2$ and $\vec{a}_3$ **span the whole space** $\mathbb{R}^3$. We say then that they form a **basis** for $\mathbb{R}^3$, i.e. each three-dimensional vector is representable as a linear combination of the three basis vectors: $\forall\vec{a} = \sum_{k=1}^{3} \alpha_k\vec{a}_k$.

C) Finally, **four** vectors in $\mathbb{R}^3$ are always linearly dependent.

Exercise 9.17 Basis vectors:

a) Do the following three vectors form a basis of $\mathbb{R}^3$:
$(1, 3, -2), (2, -2, 1)$ *and* $(4, 4, -3)$*?*
b) What about the following three vectors:
$(1, 1, 0), (1, -1, 0)$ *and* $(1, 1, 1)$*?*

Particularly convenient as basis are unit vectors.

9.4.5 Unit Vectors

Unit vectors are dimensionless vectors of length of 1. Every unit vector specifies a direction in space. From any vector $\vec{a}$ we obtain the unit vector specifying to the direction by dividing by its length a, or multiplication with $1/a$:

$$\textbf{unit vector:} \quad \vec{e}_a = \frac{\vec{a}}{a} \quad \text{or} \quad \vec{a} = a\vec{e}_a.$$

Exercise 9.18 Unit vectors:

a) Determine the unit vector in direction of the vector $\vec{a} = (-1, 2, -2)$.
b) Normalize the in Lesson 9.17 assumed basis to one.

In the following we shall use the three **unit vectors** $\vec{e}_1$**,** $\vec{e}_2$ and $\vec{e}_3$ as **basis vectors** throughout, where $\vec{e}_k := \overrightarrow{OE}_k$. Sometimes the three basis vectors $\vec{e}_k$ normalized to one are called also a **trihedral**.

After we have introduced the vector components as partial shift distances along the coordinate axes (or equivalently as projections of the length of the vector on the coordinate axes) it follows directly:

$$\vec{a} = a_1\vec{e}_1 + a_2\vec{e}_2 + a_3\vec{e}_3 = \sum_{k=1}^{3} a_k\vec{e}_k \ (\equiv a_k\vec{e}_k \text{ with Einstein summation convention!}).$$

Since we selected not just any skew, but a Cartesian coordinate system at the beginning of this chapter, we know that the three unit vectors $\vec{e}_k := \overrightarrow{OE}_k$ are pairwise perpendicular one another, and thus form an **orthonormal** (i.e. orthogonal and normalized) **basis** (ONB). In order to be able to express this fact in formulae, we need a quantity connected with the angle between two vectors, which says for instance that with a right angle between them the projection of one vector onto the other disappears. This task leads us to the question of multiplication of two vectors which we will treat in the next section.

Insert: Active Viewpoint: *In our considerations concerning transformations we have always been asking only the question of, how the components of a fixed vector would look like, if we had changed the describing coordinate system or rather the basis vectors. Then the same vector $\vec{a}$ will be represented with respect to the two different* ***orthonormal bases*** *$\vec{e}_k$ rather $\hat{\vec{e}}_k$ by different components a_k and $\hat{a}_k$, respectively: $\vec{a} = a_k\vec{e}_k = \hat{a}_k\hat{\vec{e}}_k$. If in this case $\hat{\vec{e}}_k = D^{(3)}_{kl}(\varphi)\vec{e}_l$ for instance would be rotated around* *the 3-axis by an angle φ clockwise looking in positive 3-direction, the components $\hat{a}_k = D^{(3)}_{kl}(\varphi)a_l$ would rotate correspondingly, as we have seen. This situation is called the* ***passive viewpoint*** *and we will take this viewpoint consequently during the whole course. This problem is of great importance because physicists have to make sure that their laws are formulated in such a manner that they are independent of the chosen coordinate system.*

Physicists have, however, in addition another, at first sight ***fully different problem****, namely the mathematical description of mass points, vectors, etc. which really move in space (e.g. with time) for instance rotate. In this case we have to describe the original vector $\vec{a}$ and the physical vector $\hat{\vec{a}}$, which was e.g. rotated around the 3-axis by an angle ϕ clockwise looking in positive 3-direction, in one and the same coordinate system $\vec{e}_k$: $\vec{a} = a_k\vec{e}_k$ and $\hat{\vec{a}} = \hat{a}_k\vec{e}_k$. This situation is called the* ***active viewpoint****. The following figure shows you that in this case holds $\hat{a}_k = D^{(3)}_{kl}(-\varphi)a_l$.*

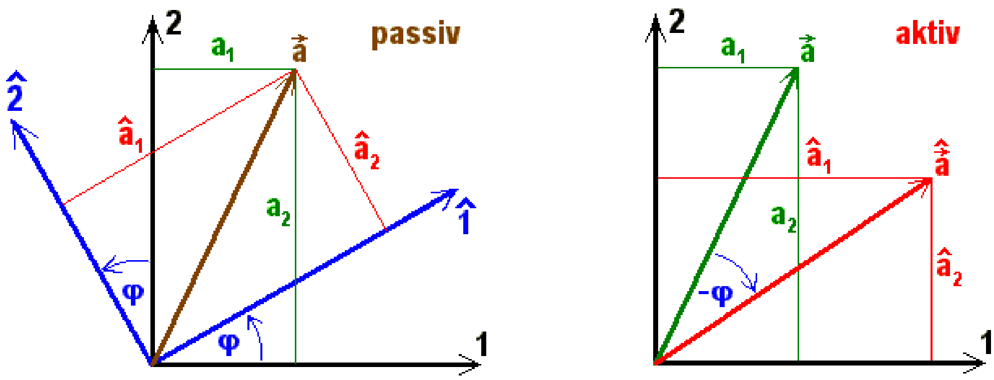


Figure 9.15: Difference between passive and active viewpoint

In both cases we have intentionally used the hat to label the new components to signalize clearly that the relation

$$\hat{a}_k = D^{(3)}_{kl}(\varphi)a_l$$

may have two fully different meanings, namely:

1. *the transformed components of a vector $\vec{a}$ as linear combination of the old ones in case of a rotation of the coordinate system by an angle φ around the 3-direction and*

2. *the components of a vector after its rotation by an angle $-\varphi$ around the 3-direction as linear combination of its components before the rotation.*

If you have clarified both situations only once, there is scarcely any danger of confusion, but rather pleasure about the fact that you are able to solve two problems at once by studying the same rotation matrices.

9.5 Scalar Product and the Kronecker Symbol

9.5.1 Motivation

In order to describe the orthogonality of our basis vectors, and also for physical reasons, we need a product of two vectors which measures the **angle** between the two factors and is connected with the **projection of one vector onto the other**. For constrained motions (for instance along rails or on an inclined plane) not the entire applied force, but only its projection (onto the rails or onto the plane) is the crucial physical quantity. Also for the computation of the work which must be done in order to displace a mass against a force (e.g. the force of gravity) not the entire force, but only its projection onto the displacement direction is the really determining quantity.

Exercise 9.19 Work:

How can we calculate the work which must be done,
a) if the mass m of a mathematical pendulum with thread length r is to be deflected by an angle φ?
b) if a mass point m is to be pushed up a distance φ along an inclined plane, with inclination angle s relative to the horizontal plane?

9.5.2 Definition

Thus we choose the following definition for a "product" between two vectors $\vec{a}$ and $\vec{b}$ and call it the "scalar product", because it gives us for every two factor vectors a real number, and as we later show that this number transforms as a scalar. Because of the dot between the two factors often used in English literature, the alternative name "dot product" is also common.

scalar product $(\vec{a} \cdot \vec{b}) := |\vec{a}| \cdot |\vec{b}| \cos \angle(\vec{a}, \vec{b}) = ab \cos \varphi.$

Besides the two trivial factors of the lengths of the two vectors, we select the **cosine** of the angle $\varphi = \angle(\vec{a}, \vec{b})$ included by both the representatives with the same starting point, because this cosine vanishes if $\varphi = 90°$ i.e. if the two vectors stand perpendicular one to the other. Clearly the expression $b \cos \varphi$ is the **projection** of the vector $\vec{b}$ on the direction of $\vec{a}$ and $a \cos \varphi$ the projection of the vector $\vec{a}$ on the direction of $\vec{b}$. These projections are multiplied in each case by the length of the vector on whose direction was projected, and are provided with a minus sign when the angle is obtuse. Thus the numerical value of the product indicates geometrically (according to the relative orientation of the vectors) the **area** of **one of the two** coloured areas shown in the following illustrations. If the angle between the two factor vectors lies between $\pi/2$ and $3\pi/2$, so that the cosine becomes negative, the area is assigned a minus sign:

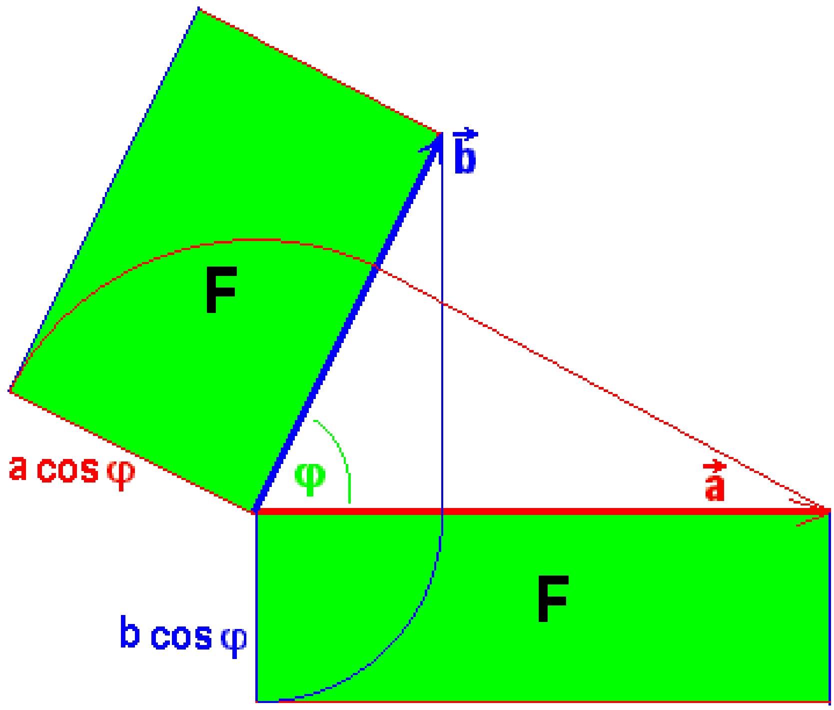


Figure 9.16: Illustration of the scalar product as a function of the angle between the two vectors: In these “luggage pictures” both the cross-section area of the “luggage body” as well as that of the more or less fitting “luggage cover” show individually the size of the product.

The following **border-line cases** are interesting:

if $\vec{a}$ and $\vec{b}$ are parallel,	it follows that $(\vec{a} \cdot \vec{b}) = ab$,
if $\vec{a}$ and $\vec{b}$ are antiparallel,	it follows that $(\vec{a} \cdot \vec{b}) = -ab$, and most importantly
if $\vec{a}$ and $\vec{b}$ are perpendicular,	it follows that $(\vec{a} \cdot \vec{b}) = 0$.

In particular is $(\vec{a} \cdot \vec{a}) = \vec{a}^2 = a^2 \geq 0$ and only the the zero-vector $\vec{0}$ has zero length

$$\boxed{(\vec{a} \cdot \vec{a}) = \vec{a}^2 = 0 \iff \vec{a} = \vec{0},}$$

because $\vec{a}$ cannot be perpendicular to itself.

Exercise 9.20 Angle in the scalar product

What does $2(\vec{a} \cdot \vec{b}) = |\vec{a}| \cdot |\vec{b}|$ mean for the angle between the two vectors?

Exercise 9.21 Cosine Theorem

a) Prove with the help of the scalar product the Cosine Theorem of plane geometry, according to which in a triangle with the side lengths a, b and c: $c^2 = a^2 + b^2 - 2ab\cos\gamma$ applies, where γ is the angle opposite to side c.

b) What follows from this theorem for $\gamma = \pi/2$?

Exercise 9.22 Schwarz inequality

Why does the Schwarz inequality hold for the absolute value of the scalar product: $|(\vec{a} \cdot \vec{b})| \leq |\vec{a}| \cdot |\vec{b}|$?

9.5.3 Commutative Law

The previous definition of the scalar product treats the two factors in a completely symmetric way. Therefore trivially the

Commutative Law: $(\vec{a} \cdot \vec{b}) = (\vec{b} \cdot \vec{a}).$

applies.

9.5.4 No Associative Law

The scalar product consisting of two absolute values and a cosine is obviously not a vector, but rather a real number. (Because of the symmetry we refrained from adding one of the two vectors. Thus our product is not an "**internal linkage**" in vector space. For this reason we prefer to avoid the sometimes used designation "inner product" used sometimes. Since the result of the multiplication is not a vector, we obviously cannot multiply it in a scalar product with a third vector. Thus, there is no **Associative Law** for the scalar product. For the same reason we cannot expect that vectors form a group of multiplication under this product.

Exercise 9.23 On the Associative Law

a) Compare the vector $(\vec{a} \cdot \vec{b})\vec{c}$ with the vector $\vec{a}(\vec{b} \cdot \vec{c})$ geometrically.

b) What is the meaning of $\vec{a}^3$?

9.5.5 Homogeneity

Multiplication of one of the vector factors with a real number $\alpha \in \mathbb{R}$, meaning multiplication of the length, is naturally possible and leads to the multiplication of the whole product. As for real numbers this property is called:

homogeneity: $(\alpha\vec{a} \cdot \vec{b}) = \alpha(\vec{a} \cdot \vec{b}) = (\vec{a} \cdot \alpha\vec{b}).$

9.5.6 Distributive Law

The vector addition law leads to the following

Distributive Law: $\Big((\vec{a} + \vec{b}) \cdot \vec{c}\Big) = (\vec{a} \cdot \vec{c}) + (\vec{b} \cdot \vec{c}).$

The proof follows by regarding the following figure from the addition of the projections on the direction of $\vec{c}$:

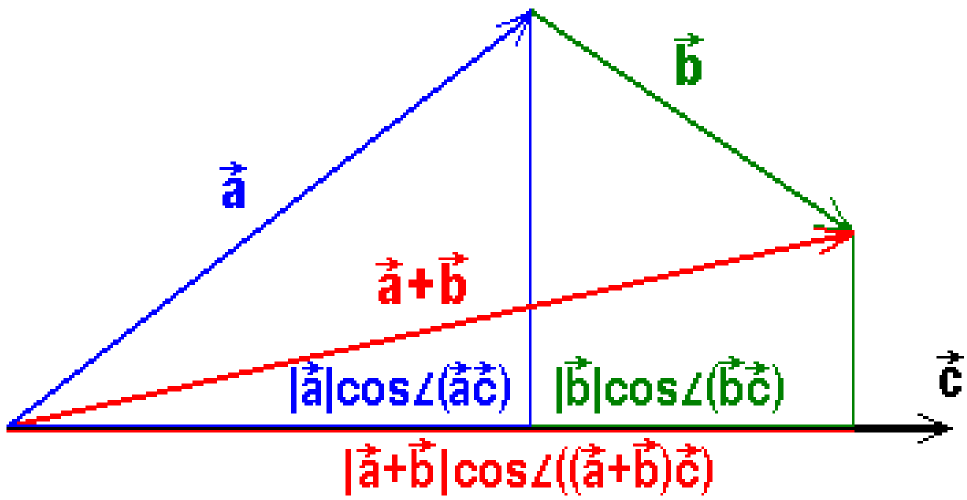


Figure 9.17: On the proof of the Distributive Law

Exercise 9.24 On the Distributive Law

Two vectors $\vec{a}$ and $\vec{b}$ span a parallelogram.

a) Calculate in addition the following scalar product $\left((\vec{a}+\vec{b})\cdot(\vec{a}-\vec{b})\right)$.

b) What is the geometrical meaning of this scalar product?

c) Determine the angle φ *between the two diagonals of the parallelogram.*

d) When are these perpendicular to each other?

9.5.7 Basis Vectors

In the definition of the scalar product, among other things we were motivated by the wish to be able to describe the **orthogonality** of the three normalized **basis vectors** $\vec{e}_k$ of the Cartesian coordinate system in a simple way. As desired we now get three equations:

$$\textbf{orthogonality:}\quad (\vec{e}_k\cdot\vec{e}_l) = |\vec{e}_k||\vec{e}_l|\cos\angle(\vec{e}_k,\vec{e}_l) = \cos\varphi_{kl} = 0 \text{ for } k \neq l = 1,2,3,$$

because $\varphi_{kl} = \pi/2$ or equivalently $\vec{e}_k \perp \vec{e}_l$ for $k \neq l$. For $k = l$ we obtain three further equations:

$$\textbf{normalization:}\quad (\vec{e}_k\cdot\vec{e}_k) = |\vec{e}_k||\vec{e}_k|\cos\angle(\vec{e}_k,\vec{e}_k) = \cos 0 = 1 \text{ for } k = 1,2,3.$$

9.5.8 Kronecker Symbol

Those nine equations contain all information about the orthogonality and normalization of the basis vectors. They can **be combined** into one single equation

$$\textbf{orthonormality:}\quad (\vec{e}_k\cdot\vec{e}_l) = \delta_{k\,l},$$

if we introduce the symbol $\delta_{k\,l}$ named after Leopold Kronecker, which is defined as follows:

$$\textbf{Kronecker symbol:}\quad \delta_{k\,l} := \begin{cases} 1 & \text{for}\quad k = l \\ 0 & \text{for}\quad k \neq l. \end{cases}$$

Like the scalar product, this number pattern is symmetrical against exchange of the two indices: $\delta_{k\,l} = \delta_{l\,k}$. In the following figure the pattern is pictorially represented in a plane:

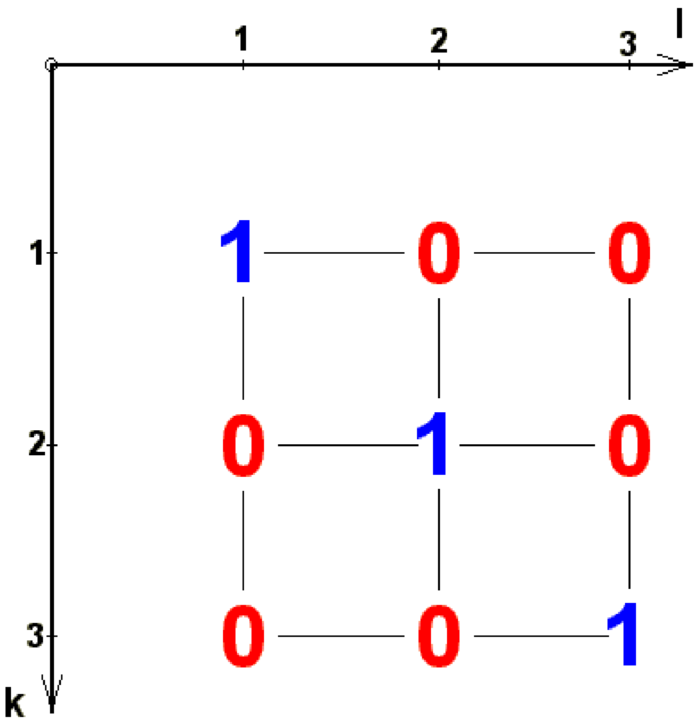


Figure 9.18: Illustration of the Kronecker symbol

The axes are arranged in such a way that we can recognize immediately the connection with the unit matrix **1**. Occasionally we need the sum of the three diagonal elements of the matrix, which is called the **trace**:

> **trace:** $\delta_{k\,k} := \sum_{k=1}^{3} \delta_{k\,k} = 3$ (with Einstein's summation convention!)

Exercise 9.25 Angle determinations

a) Determine the angle between two edges of a tetrahedron.

b) Determine the angle between two neighbouring diagonals of a cube.

9.5.9 Component Representation

Now we want to see how we can compute the scalar product if the two vectors are given in components: $\vec{a} = a_k \vec{e}_k$ and $\vec{b} = b_l \vec{e}_l$ (in both cases with summmation convention!):

$$
\begin{aligned}
(\vec{a}\cdot\vec{b}) &= (a_k\vec{e}_k\cdot b_l\vec{e}_l) && \text{whereby both over k and l are summed up}\\
&= a_k b_l(\vec{e}_k\cdot\vec{e}_l) && \text{because of the homogeneity of the scalar product}\\
&= a_k b_l\delta_{k\,l} && \text{because of the orthonormality of the basis vectors}\\
&= a_k b_k && \text{because of the Kronecker symbol only the term}\\
& && l=k \text{ of the sum over } l \text{ remains. Therefore only the sum over}\\
& && k=1,2,3 \text{ remains, thus}
\end{aligned}
$$

component representation: $(\vec{a}\cdot\vec{b}) = a_1b_1 + a_2b_2 + a_3b_3 = a_kb_k$

Exercise 9.26 Kronecker symbol

Prove the above formula in detail by explicit multiplication of the brackets $((a_1\vec{e}_1 + a_2\vec{e}_2 + a_3\vec{e}_3)\cdot(b_1\vec{e}_1 + b_2\vec{e}_2 + b_3\vec{e}_3))$ *without use of the Kronecker symbol, so that you can estimate, how much calculation work you save using the symbol.*

Exercise 9.27 Orthonormal basis

Do the three vectors, $\vec{a}_1 = \frac{1}{\sqrt{2}}(1,1,0)$, $\vec{a}_2 = \frac{1}{\sqrt{2}}(1,-1,2)$ *and* $\vec{a}_3 = \frac{1}{\sqrt{3}}(1,-1,-1)$ *form an orthonormal basis of the vector space?*

Especially for one of the three basis vectors we find:

$$(\vec{e}_k\cdot\vec{a}) = (\vec{e}_k\cdot a_l\vec{e}_l) = a_l(\vec{e}_k\cdot\vec{e}_l) = a_l\delta_{k\,l} = a_k,$$

in detail with sum signs: $(\vec{e}_k\cdot\vec{a}) = \sum_{l=1}^{3}(\vec{e}_k\cdot a_l\vec{e}_l) = \sum_{l=1}^{3} a_l(\vec{e}_k\cdot\vec{e}_l) = \sum_{l=1}^{3} a_l\delta_{k\,l} = a_k$, the k-component of the vector $\vec{a}$, because the scalar multiplication with the basis vector number k results in the projection of the vector on the k-axis. From this we can easily reconstruct the entire vector $\vec{a}$:

$$\vec{a} = \vec{e}_k a_k = \vec{e}_k(\vec{e}_k\cdot\vec{a}).$$

Insert: Completeness: *If we place two superfluous brackets "$)$" in the equation* $\vec{a} = \vec{e}_k a_k = \vec{e}_k\,(\vec{e}_k\cdot\vec{a})$, *we get:*

$$\vec{a}) = \vec{e}_k)\,(\vec{e}_k\cdot\vec{a})$$

or from this the abstract, famous symbolic

completeness relation: $\vec{e}_k)\,(\vec{e}_k = \mathbf{1}.$

After scalar multiplication from the left or right with a vector we obtain from the completeness relation the component representation of the vector $\vec{a} = \vec{e}_k a_k$ *: for instance from the right:* $\vec{e}_k)\,(\vec{e}_k \cdot \vec{a}) = \vec{e}_k a_k = \mathbf{1}\,\vec{a}) = \vec{a}$ *or from the left:* $(\vec{a} \cdot \vec{e}_k)\,(\vec{e}_k = a_k \vec{e}_k = (\vec{a}\,\mathbf{1} = \vec{a}$.

In particular for the basis vectors themselves:

$$\boxed{\vec{e}_l = \vec{e}_k \delta_{k\,l}}$$

from $\vec{e}_l = \vec{e}_l) = \mathbf{1}\,\vec{e}_l) = \vec{e}_k)\,(\vec{e}_k \cdot \vec{e}_l) = \vec{e}_k \delta_{k\,l}$.
This means that the columns or lines of the Kronecker symbol are simply the components of the basis vectors.

Exercise 9.28 Scalar product

Determine the scalar product and the lengths of the projections for the two vectors $\vec{a} = (4, -2, 4)$ *and* $\vec{b} = (-2, 3, 6)$.

Exercise 9.29 Angle with the coordinate axes

Which angles does the vector $\vec{a} = \vec{e}_1 + \sqrt{3}\vec{e}_2$ *form with the coordinate axes?*

9.5.10 Transverse Part

Apart from the decomposition in terms of components with respect to a selected coordinate system, a further decomposition of a vector $\vec{a}$ (e. g. of an oscillation vector) is needed frequently in physics, namely in a "longitudinal part" $a_{\|e}$ with respect to an arbitrarily given direction $\vec{e}$ (e.g. the direction of the propagation of a wave) and the "transverse part" $\vec{a}_{\perp e}$ with $(\vec{a}_{\perp e} \cdot \vec{e}) = 0$. From the decomposition ansatz $\vec{a} = a_{\|e}\vec{e} + \vec{a}_{\perp e}$ we get by multiplication with $\vec{e}: (\vec{a} \cdot \vec{e}) = a_{\|e}$ and from this by insertion $\vec{a} = (\vec{a} \cdot \vec{e})\vec{e} + \vec{a}_{\perp e}$ and from this by insertion

$$\boxed{\textbf{transverse part:} \quad \vec{a}_{\perp e} = \vec{a} - (\vec{a} \cdot \vec{e})\vec{e}}$$

Exercise 9.30 Transverse part:

Calculate for the vector $\vec{b} = (1, 2, -2)$ *the part which is transverse to* $\vec{a} = (3, 6, 3)$.

9.5.11 No Inverse

After we already have seen that no Associative Law holds for the scalar product, we are also not surprised that **no unambiguously determined inverse** exists: This would mean that the equation $(\vec{a} \cdot \vec{x}) = 1$ would have an unambiguously assignable vector $\vec{x}$ as a solution. You can, however, easily check that the following doubly infinite dimensional vector family fulfills the equation:

inverse vector family: $\vec{x} = \frac{\vec{e}_a}{a} + \lambda_1 \vec{e}_{\perp a} + \lambda_2 \vec{e}_{\perp a, e\perp a}$

with the two family parameters $\lambda_1, \lambda_2 \in \mathbb{R}$ and the unit vectors $\vec{e}_{\perp a}$, which is perpendicular to $\vec{a}$, and $\vec{e}_{\perp a, e\perp a}$, standing perpendicularly to $\vec{a}$ and $\vec{e}_{\perp a}$. Those, in fact, are all vectors which end in a plane lying perpendicular to $\vec{a}$, located a distance of $1/a$ from the origin. To prove this we form $(\vec{a} \cdot \vec{x}) = (\vec{a} \cdot \vec{e}_a)/a + \lambda_1(\vec{a} \cdot \vec{e}_{\perp a}) + \lambda_2(\vec{a} \cdot \vec{e}_{\perp a, e\perp a}) = a(\vec{e}_a \cdot \vec{e}_a)/a + 0 + 0 = 1$.

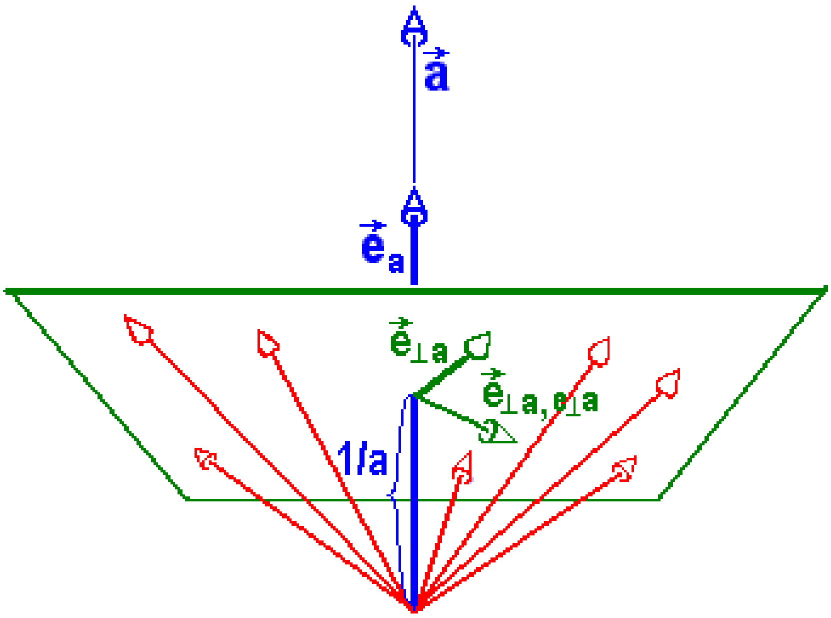


Figure 9.19: The family of inverse vectors

Accordingly **no division** by vectors is possible: The equation $(\vec{a} \cdot \vec{x}) = b$ has all vectors $\vec{x} = b\vec{e}_a/a + \lambda_1 \vec{e}_{\perp a} + \lambda_2 \vec{e}_{\perp a, e\perp a}$ as solution which have a projection of length b/a onto the direction $\vec{e}_a$ of $\vec{a}$.

Therefore it is also **impossible to cancel vectors top and bottom**, although sometimes it looks as if one could, e.g. in $(\vec{a} \cdot \vec{b})/(\vec{a} \cdot \vec{a}) = (ab/a^2) \cos \angle(\vec{a}, \vec{b}) = (b/a) \cos \angle(\vec{a}, \vec{b})$.

Exercise 9.31 Inverse vector family:

Find explicitly a family of vectors $\vec{x}$ *which solves the equation* $(\vec{a} \cdot \vec{x}) = 1$*, if* $\vec{a} = (1, 2, 2)$.

9.6 Vector Product and the Levi-Civita Symbol

9.6.1 Motivation

We now have a commutative product of two vectors. Granted, it is only a scalar one lacking associative law and a unique inverse. But it does provide a simple and compact characterization of the orthonormality of our basis vectors and, above all, a reasonable and appropriate description of all the many physical situations in which the projection of one vector onto another plays an important role.

However, we cannot settle for just this. First of all we want to at least try to find a genuine vector product which assigns to two vectors again an element of the vector space. Secondly we want a nice simple expression of the fact that our basis vectors form a right-handed coordinate system. Finally we know a whole crowd of vector-like quantities in physics which cannot be brought so simply in connection with displacements for instance angular momentum and torque - but have very much to do with rotations of extended rigid bodies around an axis. The existence of a vector product the scalar product is, by the way, a special characteristic of the three-dimensional space; for vectors in $\mathbb{R}^2$ and $\mathbb{R}^4$ such a product does not exist.

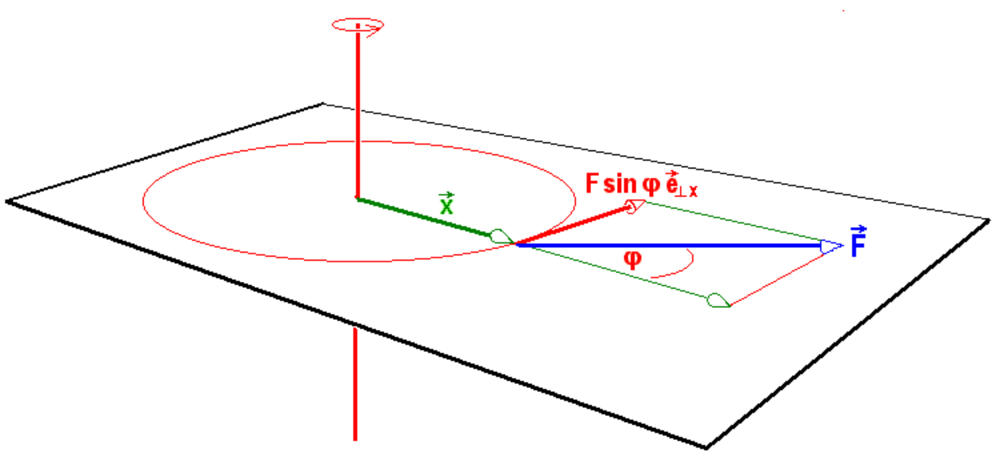


Figure 9.20: Rotating motion

As we see from the figure, no cosine determines any thing. The interacting quantity is $F \sin \varphi$, the projection of the force perpendicular to the line connecting the axis of rotation with the point of attack of the force. On the other hand apart from the vector $\vec{x}$ connecting the fulcrum to the attack point of the force, and the force vector $\vec{F}$ itself, a third direction in space is distinguished, the direction perpendicular to the plane spanned by $\vec{x}$ and $\vec{F}$ (with an orientation determined by the direction of rotation caused by the force).

9.6.2 Definition

Therefore we try the following ansatz as a genuine internal combination rule for two arbitrary vectors $\vec{a}$ and $\vec{b}$:

$$\textbf{vector product:} \quad \left[\vec{a} \times \vec{b}\right] := |\vec{a}||\vec{b}| \sin\angle(\vec{a},\vec{b})\vec{e}_{\perp a,b,\mathrm{R}}$$

Beside the **sine** of the enclosed angle and the lengths of the two vectors we have included here the **unit vector** $\vec{e}_{\perp a,b,\mathrm{R}}$, which stands perpendicular to the plane spanned by the two vectors and forms with the vectors $\vec{a}$ and $\vec{b}$ (in this order!) a *right-handed screw*. To clearly distinguish this new product from the scalar product we use a **cross** instead of the dot **and** in addition **square brackets** instead of round ones. Many people are satisfied with one of the two distinguishers: $\left[\vec{a} \times \vec{b}\right] \equiv [\vec{a}\,\vec{b}] \equiv \vec{a} \times \vec{b}$. Some call the vector product the "outer product". However, we understandably want to avoid this expression for a genuine internal combination rule in the vector space, but accept "cross product" as an alternative name.

As was the case for the scalar product we first consider various **special cases**:

For $\vec{a}$ and $\vec{b}$ **collinear**, i.e. parallel or antiparallel: $\angle(\vec{a},\vec{b}) = 0, \pi$ it follows $\left[\vec{a} \times \vec{b}\right] = 0$, and in particular:

$$[\vec{a} \times \vec{a}] := 0 \quad \forall \vec{a}.$$

For $\vec{a}$ and $\vec{b}$ **orthogonal**, i.e. $\vec{a}$ perpendicular to $\vec{b}$: $\angle(\vec{a},\vec{b}) = \pi/2$ it follows that $\left[\vec{a} \times \vec{b}\right] = |\vec{a}||\vec{b}|\vec{e}_{\perp a,b,\mathrm{R}}$,and in particular:

$$[\vec{e}_1 \times \vec{e}_2] = \vec{e}_3.$$

The length of the product vector has its maximum in this case: $\left|\left[\vec{a} \times \vec{b}\right]\right| = |\vec{a}||\vec{b}|$, i.e. the rectangle area with the lengths of the two factors as edge lengths:

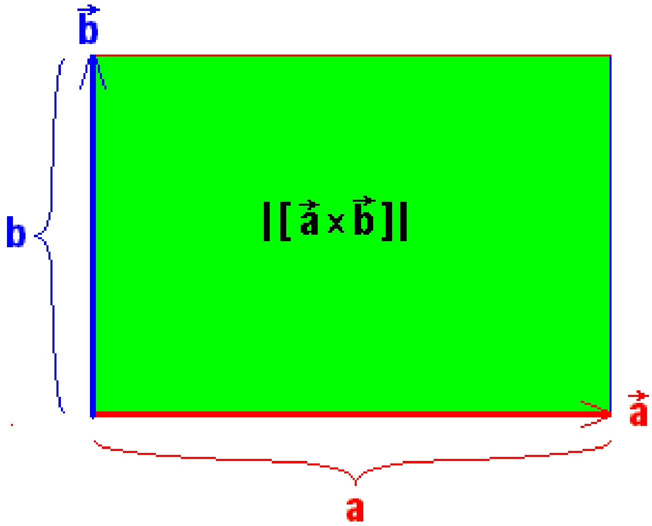


Figure 9.21: The rectangle $\left|\left[\vec{a}\times\vec{b}\right]\right| = |\vec{a}||\vec{b}|$

In the general case the **length of the product vector** is $\left|\left[\vec{a}\times\vec{b}\right]\right| = |\vec{a}||\vec{b}|\sin\angle(\vec{a},\vec{b})$, the area of the parallelogram spanned by both factors or equally one of the two rectangle areas associated with the heights of the parallelogram, shown coloured in the next picture.

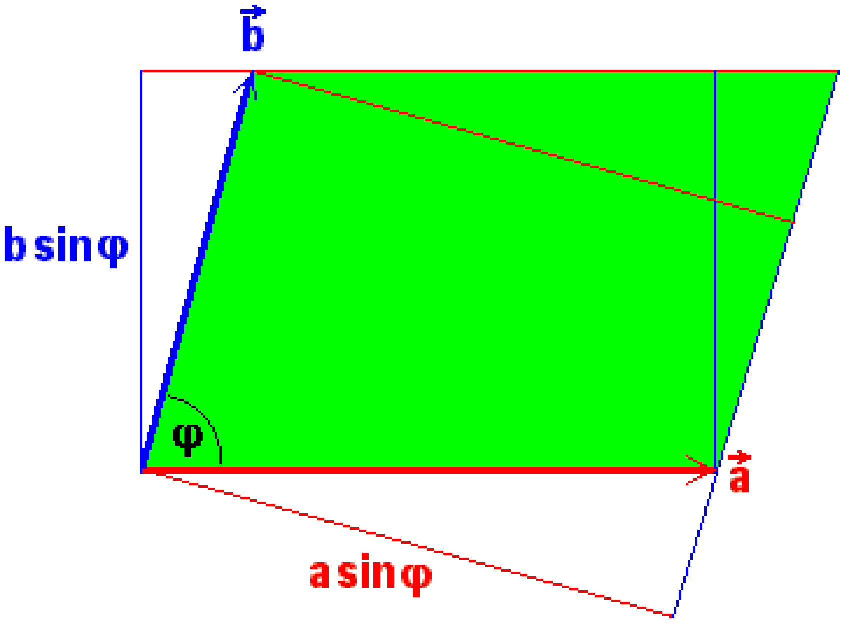


Figure 9.22: Illustration of the length of the vector product as a function of the angle between the two vectors: The areas of the parallelogram and also of each of the two rectangles with the heights of the parallelogram show the length of the vector product.

If the angle between the factor vectors exceeds π, the sine and therefore also the area becomes negative, which leads here to a reversal of the direction of the product vector.

Exercise 9.32 Physical vector products

How do you obtain:

a) the linear velocity $\vec{v}$ *during a rotating motion from the angular velocity* $\vec{\omega} = \vec{e}_\omega d\varphi/dt$ *and the position vector* $\vec{x}$*?*

b) the "area velocity" $\vec{f}$ *in the Kepler motion from the position vector* $\vec{x}$ *and the velocity* $\vec{v}$ *of the planet?*

c) the angular momentum $\vec{L}$ *from the position vector* $\vec{x}$ *and the momentum* $\vec{p}$*?*

d) the mechanical torque $\vec{D}$ *from the force* $\vec{F}$ *and the moment arm* $\vec{x}$*?*

e) the torque on an electrical dipole with dipole moment $\vec{p}$ *in a homogeneous electrical field* $\vec{E}$*?*

f) the torque on a magnetic dipole with dipole moment $\vec{m}$ *in a homogeneous magnetic field* $\vec{H}$*?*

g) the density of the electromagnetic Lorentz force $\vec{k}$ *from the velocity* $\vec{v}$ *of an electron with mass* m *and charge* e *and the magnetic induction* $\vec{B}$ *?*

h) the Poynting vector $\vec{S}$ *of the electromagnetic radiant flux from the electric field* $\vec{E}$ *and magnetic field* $\vec{H}$ *of the radiation?*

i) the magnetic field $\vec{H}$ *in the distance* $\vec{x}$ *from a linear electric current density* $\vec{j}$ *according to the Biot-Savart law?*

Exercise 9.33 Torques

Discuss the amount and the direction of the torque on a compass needle in the magnetic field of the earth, if the angle $\vartheta = \angle(\vec{m}, \vec{H})$ *between the dipole moment* $\vec{m}$ *and the field* $\vec{H}$ *: is* 0, $\pi/4$, $\pi/2$, $3\pi/4$, π, $5\pi/4$.

Exercise 9.34 Balance of the torques

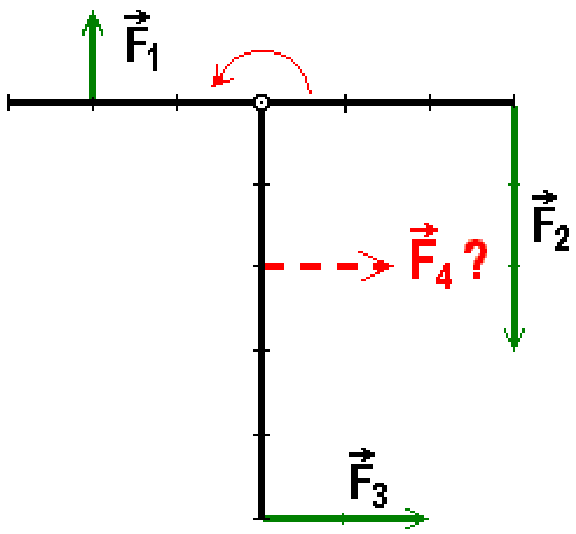


Figure 9.23: What force $\vec{F}_4$ must act at the indicated place so that the rigid T-fitting here does n o t turn around the fulcrum 0?

9.6.3 Anticommutative

The **direction** of the product vector obviously depends on the order of the factors as expressed by the **right screw rule**, which is determined by the physical interpretation of the rotation, e.g. as in the case of torque: $\vec{D} = \left[\vec{x} \times \vec{F}\right]$. Thus **no Commutative Law** holds, but the cross product is:

$$\textbf{anticommutative:} \quad \left[\vec{b} \times \vec{a}\right] = -\left[\vec{a} \times \vec{b}\right]$$

In particular we have: $[\vec{e}_2 \times \vec{e}_1] = -\vec{e}_3$ for the basis vectors.

9.6.4 Homogeneity

Just as for the scalar product, however:

$$\textbf{homogeneity:} \quad [\alpha\vec{a} \times \vec{b}] = \alpha \left[\vec{a} \times \vec{b}\right] = [\vec{a} \times \alpha\vec{b}]$$

applies to the vector product in both factors for multiplication by a real number $\alpha \in \mathbb{R}$.

9.6.5 Distributive Law

As expected there also holds a

$$\textbf{Distributive Law:} \quad [(\vec{a}_1 + \vec{a}_2) \times \vec{b}] = \left[\vec{a_1} \times \vec{b}\right] + \left[\vec{a_2} \times \vec{b}\right].$$

However, its proof is not trivial, since the transversal parts of the vectors are needed.

Insert: Distributive Law: *We call the sum* $\vec{a}_1 + \vec{a}_2 =: \vec{a}_3$, *the unit vector in the direction of* $\vec{b} : \vec{e}_b := \vec{b}/b$ *and consider the transverse parts of the two summands with respect to the direction of* $\vec{b}$ $\vec{a}_{k\perp\vec{b}} = \vec{a}_k - (\vec{a}_k\vec{e}_b)\vec{e}_b$ *for* $k = 1, 2$ *and also for* $k = 3$ *because of the Distributive Law for the scalar product. These transverse parts all lie in the plane perpendicular to* $\vec{e}_b$ *as shown in the next figure:*

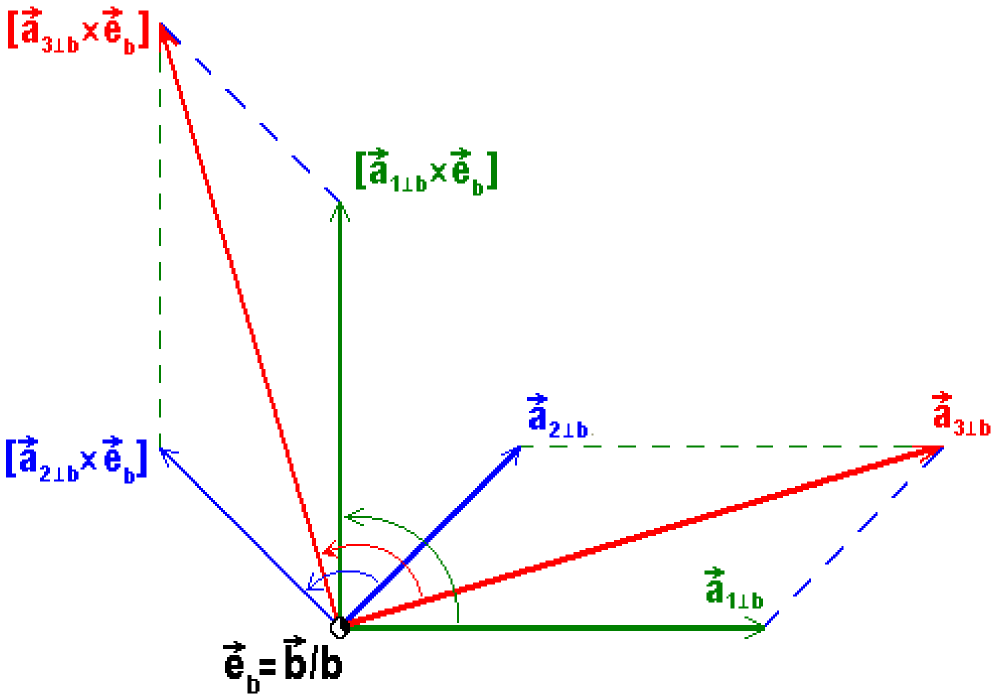


Figure 9.24: The plane perpendicular to $\vec{b}$.

The product vectors $[\vec{a}_{k\perp b} \times \vec{e}_b]$ are then vectors of length $a_{k\perp b}$ which, rotated by $90°$ counterclockwise, are perpendicular to $\vec{a}_{k\perp b}$. This means, however, that the ***whole vector addition figure was rotated by*** *$90°$, so that $[\vec{a_{3}}_{\perp b} \times \vec{e_b}] = [\vec{a_{1}}_{\perp b} \times \vec{e_b}] + [\vec{a_{2}}_{\perp b} \times \vec{e_b}]$ holds. Multiplication of this equation with b yields the desired Distributive Law.*

Exercise 9.35 Distributive law for vector products

a) Calculate: $[(\vec{a} + \vec{b}) \times (\vec{a} - \vec{b})]$.

b) How reads the Lagrange-identity: $\left[\vec{a} \times \vec{b}\right]^2 + (\vec{a} \cdot \vec{b})^2 =$?

c) What is therefore the meaning of $\sqrt{\left[\vec{a} \times \vec{b}\right]^2 + (\vec{a} \cdot \vec{b})^2} = b^2$ *for the vector* $\vec{a}$?

d) Calculate the sum of the outer normals of a tetrahedron.

9.6.6 With Transverse Parts

We can use the concept of the *transverse part* of a vector with respect to a given direction, which we defined and studied earlier, in order to shed new light on the **illustrative meaning** of the vector product: To do this, we take the **transverse part of the second factor $\vec{b}$ with respect to the direction $\vec{e}_a = \vec{a}/a$ of the first** factor: $\vec{b}_{\perp a} = \vec{b} - (\vec{b} \cdot \vec{e}_a)\vec{e}_a$ and multiply $\vec{a}$ vectorially by this:

$$\left[\vec{a} \times \vec{b}_{\perp a}\right] = \left[\vec{a} \times \vec{b}\right] - (\vec{b} \cdot \vec{e}_a)\,[\vec{a} \times \vec{e}_a] = \left[\vec{a} \times \vec{b}\right].$$

Since a corresponding argument can also be given for the other factor, this means: **In the vector product one factor can be replaced by its transverse part with respect to the other one without changing the value of the product.**

For a better understanding look once more at Figure 9.22.

9.6.7 Basis Vectors

In order to obtain a component representation for our vector product, the basis vectors first must be multiplied vectorially. As already suggested above, the vector product gives us the desired simple representation of the fact that our basis vectors form a right-handed system:

$$\textbf{right-handed system: } [\vec{e}_1 \times \vec{e}_2] = \vec{e}_3, \quad [\vec{e}_2 \times \vec{e}_3] = \vec{e}_1, \quad [\vec{e}_3 \times \vec{e}_1] = \vec{e}_2.$$

Because of the anticommutativity of the vector product three further relations with the reverse order of the factors and a minus sign can be added:

$$[\vec{e}_2 \times \vec{e}_1] = -\vec{e}_3, \quad [\vec{e}_3 \times \vec{e}_2] = -\vec{e}_1, \quad [\vec{e}_1 \times \vec{e}_3] = -\vec{e}_2.$$

Much as for the scalar product and the Kronecker symbol we summarize these six fundamental relations into one single equation by the introduction of the symbol designated after Tullio Levi-Civita:

9.6.8 Levi-Civita Symbol

We simply write:

$$\textbf{right-handed coordinate system: } [\vec{e}_k \times \vec{e}_l] = \varepsilon_{klm}\vec{e}_m \equiv \sum_{m=1}^{3} \varepsilon_{klm}\vec{e}_m$$

where the last term once more reminds us of the Einstein summation convention. Here the Levi-Civita symbol with its three indices is defined by:

$$\textbf{Levi-Civita symbol: } \varepsilon_{klm} := \begin{cases} +1, & \text{if } klm = 123, 231, 312, \\ -1, & \text{if } klm = 132, 213, 321 \text{ and} \\ 0 & \text{otherwise} \end{cases}$$

Obviously the symbol changes sign when permuting any two indices. One calls the symbol totally antisymmetric against exchanging pairs of indices:

$$\textbf{total antisymmetry: } \varepsilon_{klm} = \varepsilon_{lmk} = \varepsilon_{mkl} = -\varepsilon_{kml} = -\varepsilon_{lkm} = -\varepsilon_{mlk}$$

For all even or *cyclic* permutations of the sequence of index numbers 123 a $+1$ results, yielding exactly the three relations indicated above, characteristic of a right-handed basis. The index sequences with an odd number of permutations of two indices against the number sequence 123 or anticyclic permutations of 123 lead to -1, as in the three previously specified vector products with exchanged order of the factors. Only **six** of the 27 elements of the Levi-Civita symbol are different from 0. All **21 remaining elements are zeros**, so that you may have trouble finding the important three blue "ones" in the crowd of red zeros, arranged remarkably symmetrically from the origin in the following figure, and especially the even more important three -1 in green:

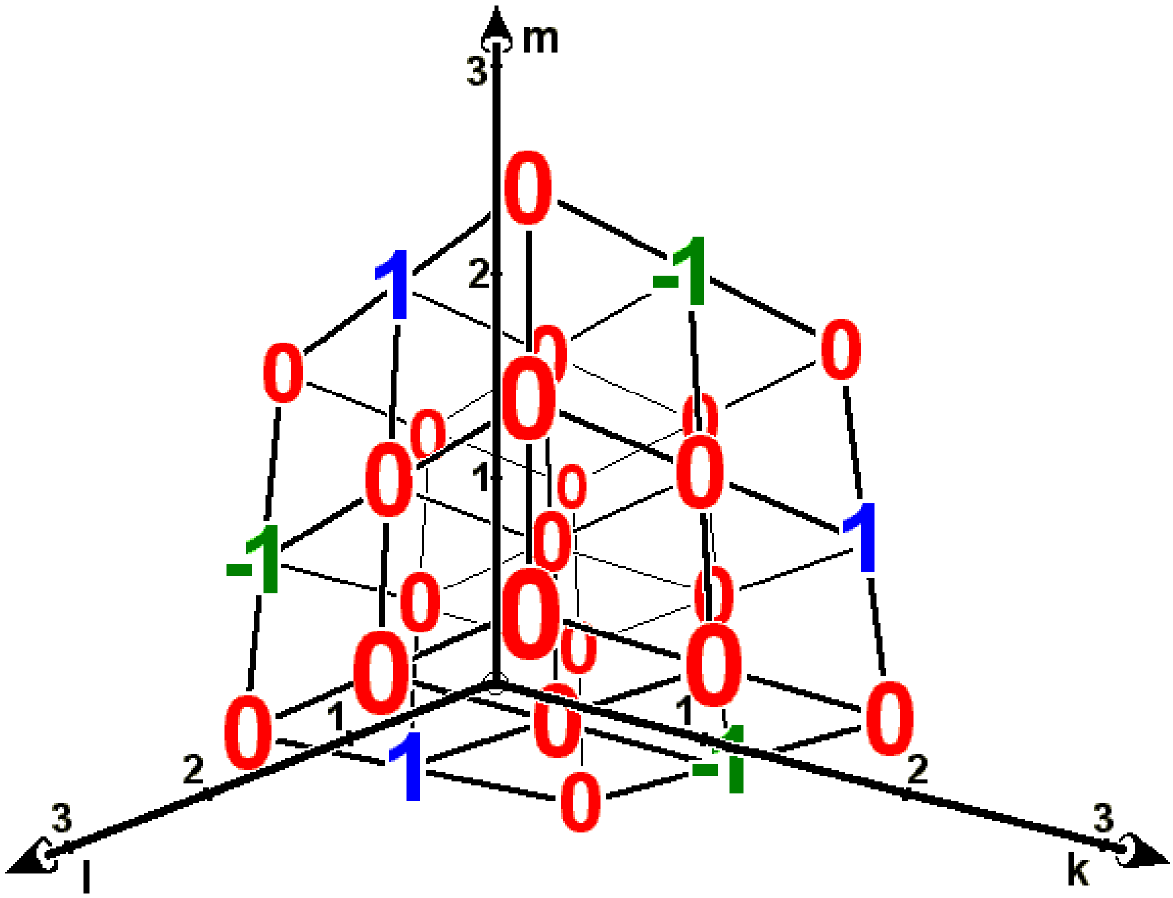


Figure 9.25: Illustration of the Levi-Civita symbol: Obviously there is no reason to fear of it! It consists almost entirely of zeros.

As you see from the figure, the Levi-Civita symbol, despite of the many zeros and its beautiful symmetry is a rather unmanageable object because of its **three** indices. Therefore we want to formulate its crucial message, i.e. the six index configurations for which it does not vanish, and its connection with the already more familiar handy *Kronecker symbols*. It is namely equal to $+1$ with $klm = 123$, -1 with 132; $+1$ with 231, -1 with 213; $+1$ with 312 and -1 with 321 :

$$\boxed{\varepsilon_{klm} = \delta_{k\,1}(\delta_{l\,2}\delta_{m\,3} - \delta_{l\,3}\delta_{m\,2}) + \delta_{k\,2}(\delta_{l\,3}\delta_{m\,1} - \delta_{l\,1}\delta_{m\,3}) + \delta_{k\,3}(\delta_{l\,1}\delta_{m\,2} - \delta_{l\,2}\delta_{m\,1}).}$$

In this form of the result we recognize (after having studied our insert on matrices) the development of the determinant of a (3×3)-matrix according to the first line or the first column:

$$\varepsilon_{klm} = +\delta_{k1}\begin{vmatrix} \delta_{l2} & \delta_{l3} \\ \delta_{m2} & \delta_{m3} \end{vmatrix} + \delta_{k2}\begin{vmatrix} \delta_{l3} & \delta_{l1} \\ \delta_{m3} & \delta_{m1} \end{vmatrix} + \delta_{k3}\begin{vmatrix} \delta_{l1} & \delta_{l2} \\ \delta_{m1} & \delta_{m2} \end{vmatrix}.$$

Thus:

$$\boxed{\varepsilon_{klm} = \begin{vmatrix} \delta_{k1} & \delta_{k2} & \delta_{k3} \\ \delta_{l1} & \delta_{l2} & \delta_{l3} \\ \delta_{m1} & \delta_{m2} & \delta_{m3} \end{vmatrix} = \begin{vmatrix} \delta_{k1} & \delta_{l1} & \delta_{m1} \\ \delta_{k2} & \delta_{l2} & \delta_{m2} \\ \delta_{k3} & \delta_{l3} & \delta_{m3} \end{vmatrix} = \begin{vmatrix} \delta_{1k} & \delta_{1l} & \delta_{1m} \\ \delta_{2k} & \delta_{2l} & \delta_{2m} \\ \delta_{3k} & \delta_{3l} & \delta_{3m} \end{vmatrix} = \ldots}$$

In the transition from the first to the second version we used that the determinant of a matrix $\mathbf{A}$ is not changed by reflection through the main diagonal: $|\mathbf{A}^T| = |\mathbf{A}|$. In the transition to the third version we used the symmetry of the Kronecker symbol against exchanging the two indices: $\delta_{k1} = \delta_{1k}$. There is obviously still an abundance of further forms of the Levi-Civita symbol as determinant, if we consider that each determinant changes its sign when exchanging two lines or columns. Using the determinant representation of the Levi-Civita symbol you should always remember that the Kronecker symbols in the determinant are nothing other than place holders, which tell you depending on the value of the indices whether there is a 1 or a 0 in this place.

Exercise 9.36 Normal vectors

Which unit vectors are perpendicular to:

a) $(\vec{e}_1 + \vec{e}_2)$ *and* $(\vec{e}_1 - \vec{e}_2)$,

b) $(\vec{e}_1 - \vec{e}_2)$ *and* $(\vec{e}_2 - \vec{e}_3)$,

c) $(\vec{e}_1 + 2\vec{e}_3)$ *and* $(\vec{e}_2 - 2\vec{e}_3)$*?*

Exercise 9.37 Tetrahedron surface

Calculate by means of vector products, the surface of a tetrahedron of edge length L by embedding it into a cube of the edge length a.

9.6.9 Component Representation

Now we are able to calculate the vector product of two vectors $\vec{a} = a_k\vec{e}_k$ and $\vec{b} = b_l\vec{e}_l$ from their components:

$$\textbf{component representation:}\ \left[\vec{a}\times\vec{b}\right] = [a_k\vec{e}_k \times b_l\vec{e}_l] = a_k b_l\,[\vec{e}_k \times \vec{e}_l] = \varepsilon_{klm} a_k b_l \vec{e}_m$$

We would like to remind you once more of the fact that the right side of this equation contains - according to the Einstein convention - three sums over $k = 1, 2, 3$, $l = 1, 2, 3$ and $m = 1, 2, 3$. Thus there are altogether 27 summands, of which 21 we do not need to be afraid of, because we know that they vanish. Only the following six terms remain non-zero, three of them with the characteristic minus signs:

$$\left[\vec{a}\times\vec{b}\right] = \vec{e}_1(a_2b_3 - a_3b_2) + \vec{e}_2(a_3b_1 - a_1b_3) + \vec{e}_3(a_1b_2 - a_2b_1)$$

Here the six summands are organized according to basis vectors, to be able to recognize more easily that e.g.: the 1-component of the product vector is constructed from components of the two factors as follows: $\left[\vec{a}\times\vec{b}\right]_1 = (a_2b_3 - a_3b_2)$. Generally for the m-th component we get:

$$\left[\vec{a}\times\vec{b}\right]_m = (\left[\vec{a}\times\vec{b}\right]\cdot\vec{e}_m) = \varepsilon_{klm}a_kb_l(\vec{e}_m\cdot\vec{e}_m) = \varepsilon_{klm}a_kb_l = \varepsilon_{mkl}a_kb_l.$$

The final way of writing is justified because of the cyclic permutability of the indices in the Levi-Civita symbol.

Our representation of the Levi-Civita symbol as a determinant of Kronecker symbols permits even more ways of writing down the component representation of the vector product. These can easily be kept in mind and may be known to some of you :

$$\left[\vec{a}\times\vec{b}\right] = a_kb_l\vec{e}_m \begin{vmatrix} \delta_{k\,1} & \delta_{k\,2} & \delta_{k\,3} \\ \delta_{l\,1} & \delta_{l\,2} & \delta_{l\,3} \\ \delta_{m\,1} & \delta_{m\,2} & \delta_{m\,3} \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \end{vmatrix} = \begin{vmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \dots$$

In order to obtain the final form in the preceding determinant, the lines were cyclically permuted. In this manner or e.g. by reflection of the determinant in its main diagonal, you can find again a whole set of representations which are all equivalent, since they lead to the same result. In practice you will naturally pay attention to existing zeros and select that form, from which you can see the desired result immediately.

In the last two forms the determinant way of writing is clearly meant symbolically and to be enjoyed with some caution, because, as you see, the basis vectors stand as elements in the determinant, and an object of this kind has actually not been defined. The meaning of this determinant is an easily remembered expression for the following frequently used development with respect to the adjuncts:

$$\left[\vec{a} \times \vec{b}\right] = \vec{e}_1 \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - \vec{e}_2 \begin{vmatrix} a_3 & a_1 \\ b_2 & b_1 \end{vmatrix} + \vec{e}_3 \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

Exercise 9.38 Vector product as determinant

Calculate the vector products of Exercise 9.36 as determinants.

9.6.10 No Inverse

With this component decomposition we have defined in convenient form for our physical use a genuine vector product of two vectors. However, again some characteristics are missing, which would be essential for the formation of a group: As is the case for the scalar product there is **no** unambiguously determined **inverse** $\vec{x}$ which solves the equation $[\vec{a} \times \vec{x}] = \vec{e}$, but again a whole family of vectors

$$\textbf{inverse vector family:} \quad \vec{x}(\lambda) = \frac{\vec{e}_{\perp a,e}}{a} + \lambda\vec{a}.$$

This inverse vector family includes all vectors with representatives which have their tips on a straight line parallel to $\vec{a}$ a distance of $1/a$ from the origin. Moreover, **no division** exists, because the vector family $\vec{x}(\lambda) = (b/a)\vec{e}_{\perp a,b} + \lambda\vec{a}$ with the family parameter $\lambda \in \mathbb{R}$ (perhaps with a dimension) solves the equation $[\vec{a} \times \vec{x}] = \vec{b}$. In order to see this, we go to a plane perpendicular to $\vec{b}$ in the next figure:

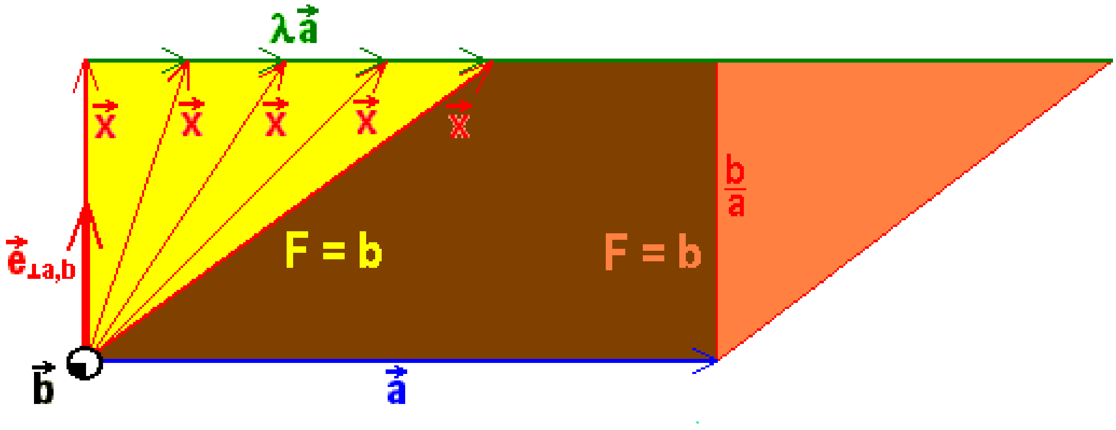


Figure 9.26: Concerning division by a vector

In this plane we draw (in the foot point of the selected representative of $\vec{a}$ perpendicular to this) the vector $(b/a)\vec{e}_{\perp a,b}$ and add to its tip the vector $\lambda\vec{a}$ parallel to $\vec{a}$ to obtain the desired vector family $\vec{x}(\lambda)$. The lengths of all product vectors $[\vec{a} \times \vec{x}]$ are the areas of the parallelograms spanned by $\vec{a}$ and $\vec{x}$ with the baseline of length a and the height b/a, which all equal b as was claimed.

9.6.11 No Associative Law

Although we have now a genuine vector product which can be multiplied again vectorially by a third vector, **no Associative Law** holds true, as we also found in the case of the scalar product. We show this in the simplest manner by a single counter-example using our basis vectors: On the one hand applies: $[[\vec{e}_1 \times \vec{e}_1] \times \vec{e}_2] = \left[\vec{0} \times \vec{e}_2\right] = \vec{0}$ and on the other hand: $[\vec{e}_1 \times [\vec{e}_1 \times \vec{e}_2]] = [\vec{e}_1 \times \vec{e}_3] = -\vec{e}_2 \neq \vec{0}$.

Thus, no Associative Law can apply. What, however, takes its place? We will deal with this question among other things in the next section.

9.7 Multiple Products

Because there are two different kinds of products for any two vectors, there are also several multiple products involving three and more vectors. **Four** of these multiple products occur in physical problems, and we want to discuss them in this section. We study first the characteristics of two kinds of products of **three** vectors and then two different products of **four** vectors. We reduce each of these more complicated products to the computation of scalar and simple vector products alone:

9.7.1 Triple Product

The simplest and most important way to multiply three vectors $\vec{a} = a_k\vec{e}_k$, $\vec{b} = b_l\vec{e}_l$ and $\vec{c} = c_m\vec{e}_m$ (in each case with summation convention!) with one another is the scalar product of a vector product with a third vector, the so-called

triple product:

$$(\vec{a}\vec{b}\vec{c}) := (\left[\vec{a}\times\vec{b}\right]\cdot\vec{c}) = a_k b_l c_m \varepsilon_{klm}$$
$$= a_k b_l c_m \begin{vmatrix} \delta_{k1} & \delta_{k2} & \delta_{k3} \\ \delta_{l1} & \delta_{l2} & \delta_{l3} \\ \delta_{m1} & \delta_{m2} & \delta_{m3} \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

Here we have first written the definitions of the two products, then we have used the determinant representation for the Levi-Civita symbol and, before performing the three summations, we have used the fact that the determinant is homogeneous in rows, as we learned in the past. According to the symmetries of determinants further formulations are possible, in particular by cyclic permutation and reflection in the main diagonal, i.e. the components of the three vectors can be organized instead of into lines also into columns of the determinant.

This variety of possibilities in the formulation of determinants gives rise to the many identities in the representations of one and the same triple product:

$$(\vec{a}\vec{b}\vec{c}) := (\left[\vec{a}\times\vec{b}\right]\cdot\vec{c}) \equiv (\vec{b}\vec{c}\vec{a}) := (\left[\vec{b}\times\vec{c}\right]\cdot\vec{a}) \equiv (\vec{c}\vec{a}\vec{b}) := ([\vec{c}\times\vec{a}]\cdot\vec{b}).$$

Note that, because of the Commutative Law of the scalar product, the two **multiplication signs** can be interchanged and therefore completely **omitted**:

$$(\vec{a}\vec{b}\vec{c}) := (\left[\vec{a}\times\vec{b}\right]\cdot\vec{c}) \equiv (\vec{c}\cdot\left[\vec{a}\times\vec{b}\right]).$$

The anticommutativity of the vector product leads to the following relations:

$$(\vec{a}\vec{b}\vec{c}) := (\left[\vec{a}\times\vec{b}\right]\cdot\vec{c}) \equiv -(\left[\vec{b}\times\vec{a}\right]\cdot\vec{c}) =: -(\vec{b}\vec{a}\vec{c})$$
$$\equiv -([\vec{a}\times\vec{c}]\cdot\vec{b}) =: -(\vec{a}\vec{c}\vec{b}) \equiv -(\left[\vec{c}\times\vec{b}\right]\cdot\vec{a}) =: -(\vec{c}\vec{b}\vec{a}).$$

The evaluation of the determinant results in a real number:

$$(\vec{a}\vec{b}\vec{c}) := (\left[\vec{a}\times\vec{b}\right]\cdot\vec{c}) = ab\sin\angle(\vec{a}\vec{b})c\cos\angle(\vec{e}_{\perp a,b,R.}\vec{c}) = Fc\cos\angle(\vec{e}_{\perp a,b,R.}\vec{c}) \in \mathbb{R}.$$

For the geometrical interpretation of this number we look at the next figure:

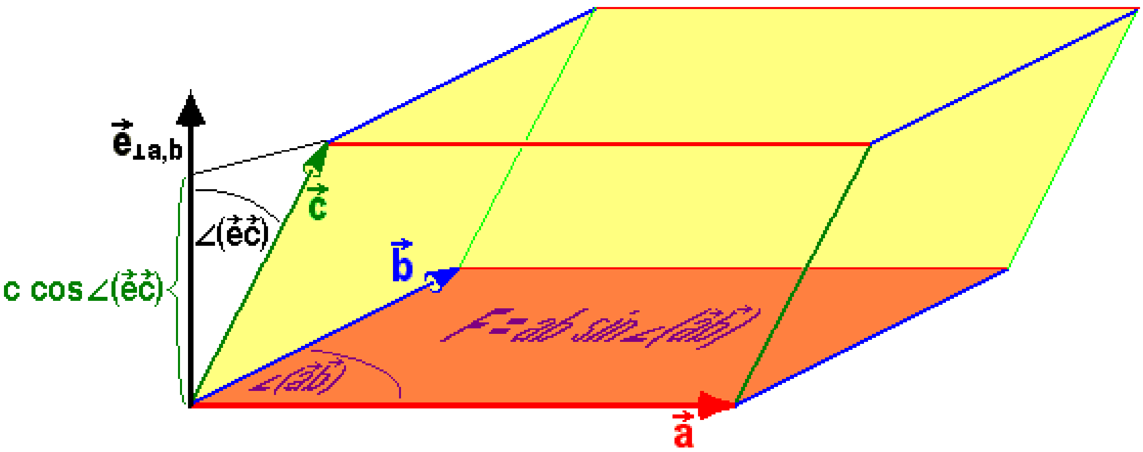


Figure 9.27: The triple product

The vector product $\left[\vec{a}\times\vec{b}\right]$ yields a vector with length equal to the parallelogram area $F = ab\sin\angle(\vec{a},\vec{b})$ and direction $\vec{e}_{\perp a,b,R}$ perpendicular to the parallelogram spanned by $\vec{a}$ and $\vec{b}$. The third vector $\vec{c}$ is now projected onto the direction of this unit vector $\vec{e}_{\perp a,b,R}$. The length of this projection $c\cos\angle(\vec{e}_{\perp a,b,R},\vec{c})$ results in the height of a **spar** (: parallelepiped) over the surface area F. The volume content of this geometrical object (however with a sign!) represents the numerical value of the triple product. After ordering of the three vectors the volume contents are to be provided with a sign. If both angles are smaller than π and the vectors in the indicated order form a right-handed screw, volume contents are to be taken positively. If you for example regard the first unit vector $\vec{a}_1 = \vec{e}_1 = (1, 0, 0)$, the unit vector $\vec{a}_2 = (\cos\varphi, \sin\varphi, 0)$ in the 1-2-plane, which forms with $\vec{a}_1$ the angle φ, and the vector $\vec{a}_3 = (\cos\psi, 0, \sin\psi)$, which in the 1-3-plane forms with $\vec{a}_1$ the angle ψ, then the triple product becomes $(\vec{a}_1\vec{a}_2\vec{a}_3) = \sin\varphi\sin\psi$, for instance $\varphi = \psi = 45°$ and $135°$, whereas $(\vec{a}_1\vec{a}_3\vec{a}_2) = -\sin\varphi\sin\psi$.

The volume content is zero if the three factor vectors are coplanar, thus linearly dependent, and in particular if two of the three factors are equal. Conversely, we can conclude from the vanishing of a determinant having three vectors as line or column vectors the linear dependence of these three vectors.

Exercise 9.39 Linear dependence

Are the vectors $(1,1,1)$, $(1,1,2)$, *and* $(1,1,3)$ *linearly independent?*

Especially for the basis vectors we get an extremely concise formulation for orthonormality, and a labeling of a right-handed coordinate system by one single equation, which on the other hand represents the Levi-Civita symbol in terms of the basis vectors:

$$\boxed{(\vec{e}_k\vec{e}_l\vec{e}_m) = \varepsilon_{klm}.}$$

Particularly, $(\vec{e}_1\vec{e}_2\vec{e}_3) = 1$,which is the volume content of the unit cube.

Exercise 9.40 Triple product

Calculate the following triple products:

a) $((\vec{a}+\vec{b})(\vec{b}+\vec{c})(\vec{c}+\vec{a}))$

b) $((1,0,0)(\sqrt{3}/2,1/2,0)(\sqrt{3}/2,0,1/2))$

c) $((\sqrt{3}/2,1/2,0)(1,0,0)(\sqrt{3}/2,0,1/2))$

d) $((1,2,3)(3,2,1)(2,1,3))$

e) $((1,2,3)(1,2,2)(3,2,1))$

Exercise 9.41 Applications of the triple product

a) Calculate the volume of the parallelepiped spanned by the following three vectors: $\vec{a}_1 = \vec{e}_1 - \vec{e}_2$, $\vec{a}_2 = \vec{e}_1 + \vec{e}_2$ *and* $\vec{a}_3 = \vec{e}_1 + \vec{e}_2 + \vec{e}_3$.

b) Calculate the volume of the triangle pyramid formed by the following three vectors: $\vec{a}_1 = \vec{e}_1 + \vec{e}_2 + \vec{e}_3$, $\vec{a}_2 = \vec{e}_1 - \vec{e}_2$ *and* $\vec{a}_3 = \vec{e}_1 + \vec{e}_2 - 2\vec{e}_3$.

c) Calculate with the help of the triple product the volume of a tetrahedron of edge length L, after you have embedded it into a cube.

d) How does the triple product of the following three vectors depend on the real number x*:* $\vec{a}_1 = (\vec{e}_1 - \vec{e}_2)/2$, $\vec{a}_2 = -\vec{e}_1 + \vec{e}_2 - \vec{e}_3$ *and* $\vec{a}_3 = 2\vec{e}_2 - x\vec{e}_3$*? Why?*

e) Find the equation of the plane which contains the three points having the following radius vectors: $\vec{a}_0 = \vec{e}_1$, $\vec{a}_3 = \vec{e}_1 + \vec{e}_2 + \vec{e}_3$ *and* $\vec{a}_2 = \vec{e}_1 - \vec{e}_2 + \vec{e}_3$.

Insert: Two Levi-Civita Symbols:: *For the calculation of further multiple products we need the following product of two Levi-Civita symbols*

$$\boxed{\varepsilon_{klm}\varepsilon_{pqn} = \begin{vmatrix} \delta_{kp} & \delta_{kq} & \delta_{kn} \\ \delta_{lp} & \delta_{lq} & \delta_{ln} \\ \delta_{mp} & \delta_{mq} & \delta_{mn} \end{vmatrix}}$$

The proof is a good exercise in the multiplication of matrices:

$\varepsilon_{klm}\varepsilon_{pqn} =$

$$= \begin{vmatrix} \delta_{k1} & \delta_{k2} & \delta_{k3} \\ \delta_{l1} & \delta_{l2} & \delta_{l3} \\ \delta_{m1} & \delta_{m2} & \delta_{m3} \end{vmatrix} \begin{vmatrix} \delta_{1p} & \delta_{1q} & \delta_{1n} \\ \delta_{2p} & \delta_{2q} & \delta_{2n} \\ \delta_{3p} & \delta_{3q} & \delta_{3n} \end{vmatrix}$$

$$= \left| \begin{pmatrix} \delta_{k1} & \delta_{k2} & \delta_{k3} \\ \delta_{l1} & \delta_{l2} & \delta_{l3} \\ \delta_{m1} & \delta_{m2} & \delta_{m3} \end{pmatrix} \begin{pmatrix} \delta_{1p} & \delta_{1q} & \delta_{1n} \\ \delta_{2p} & \delta_{2q} & \delta_{2n} \\ \delta_{3p} & \delta_{3q} & \delta_{3n} \end{pmatrix} \right|$$

$$= \begin{vmatrix} \delta_{k1}\delta_{p1} + \delta_{k2}\delta_{2p} + \delta_{k3}\delta_{3p} & \delta_{k1}\delta_{1q} + \delta_{k2}\delta_{2q} + \delta_{k3}\delta_{3q} & \delta_{k1}\delta_{1n} + \delta_{k2}\delta_{2n} + \delta_{k3}\delta_{3n} \\ \delta_{l1}\delta_{1p} + \delta_{l2}\delta_{2p} + \delta_{l3}\delta_{3p} & \delta_{l1}\delta_{1q} + \delta_{l2}\delta_{2q} + \delta_{l3}\delta_{3q} & \delta_{l1}\delta_{1n} + \delta_{l2}\delta_{2n} + \delta_{l3}\delta_{3n} \\ \delta_{m1}\delta_{1p} + \delta_{m2}\delta_{2p} + \delta_{m3}\delta_{3p} & \delta_{m1}\delta_{1q} + \delta_{m2}\delta_{2q} + \delta_{m3}\delta_{3q} & \delta_{m1}\delta_{1n} + \delta_{m2}\delta_{2n} + \delta_{m3}\delta_{3n} \end{vmatrix}$$

$$= \begin{vmatrix} \delta_{kr}\delta_{rp} & \delta_{kr}\delta_{rq} & \delta_{kr}\delta_{rn} \\ \delta_{lr}\delta_{rp} & \delta_{lr}\delta_{rq} & \delta_{lr}\delta_{rn} \\ \delta_{mr}\delta_{rp} & \delta_{mr}\delta_{rq} & \delta_{mr}\delta_{rn} \end{vmatrix}$$

$$= \begin{vmatrix} \delta_{kp} & \delta_{kq} & \delta_{kn} \\ \delta_{lp} & \delta_{lq} & \delta_{ln} \\ \delta_{mp} & \delta_{mq} & \delta_{mn} \end{vmatrix}$$

First we have replaced the two Levi-Civita symbols by two cleverly chosen determinant representations, then we have used the fact that the determinant of a product of two matrices is equal to the product of the two determinants. Afterwards we have multiplied the two matrices in the way we learned earlier. The single matrix elements show up as sums over three products of two Kronecker symbols which we bring together and write as sums over $r = 1, 2, 3$ *using the summation convention. Finally we carry out these sums, and in every case only one Kronecker symbol survives.*

Having this result before our eyes we of course realize that, because of the symmetry properties of determinants, nothing else could have resulted The final formula necessarily has to be antisymmetric within each of the two index triples klm and pqn, and symmetric with respect to the interchange of the index pairs kp, lq and mn as is our starting product.

Fortunately this general result is rather seldom encountered in calculations. Mostly the product of two Levi-Civita symbols is needed in the special case when it is ***summed over one index pair****, e.g. m:*

$$\boxed{\varepsilon_{klm}\varepsilon_{pqm} = (\delta_{kp}\delta_{lq} - \delta_{kq}\delta_{lp})}$$

Also we want to show how this important relation comes into being:

$$\begin{aligned}
\varepsilon_{klm}\varepsilon_{pqn}\delta_{m\,n} = &\quad \varepsilon_{klm}\varepsilon_{pqm} \\
= &+\delta_{m\,p}\begin{vmatrix}\delta_{k\,q} & \delta_{k\,m}\\ \delta_{l\,q} & \delta_{l\,m}\end{vmatrix} - \delta_{m\,q}\begin{vmatrix}\delta_{k\,p} & \delta_{k\,m}\\ \delta_{l\,p} & \delta_{l\,m}\end{vmatrix} + \delta_{m\,m}\begin{vmatrix}\delta_{k\,p} & \delta_{k\,q}\\ \delta_{l\,p} & \delta_{l\,q}\end{vmatrix} \\
= &\quad + \begin{vmatrix}\delta_{k\,q} & \delta_{k\,p}\\ \delta_{l\,q} & \delta_{l\,p}\end{vmatrix} - \begin{vmatrix}\delta_{k\,p} & \delta_{k\,q}\\ \delta_{l\,p} & \delta_{l\,q}\end{vmatrix} + 3\begin{vmatrix}\delta_{k\,p} & \delta_{k\,q}\\ \delta_{l\,p} & \delta_{l\,q}\end{vmatrix} \\
= &\quad (-1-1+3)\begin{vmatrix}\delta_{k\,p} & \delta_{k\,q}\\ \delta_{l\,p} & \delta_{l\,q}\end{vmatrix} \\
= &\quad (\delta_{k\,p}\delta_{l\,q} - \delta_{k\,q}\delta_{l\,p})
\end{aligned}$$

Here we have first put $n = m$ in the determinant representation of the product of the two Levi-Civita symbols, then we have developed the obtained (3×3)-determinant with respect to the last line and carried out the sums over $m = 1, 2, 3$ in the remaining (2×2)-determinants, especially we got a 3 from the trace $\delta_{m\,m}$. After interchanging the columns in the first (2×2)-determinant the result became clear.

The general structure, namely "antisymmetric in kl and pq and symmetric with respect to kp and lq and the summation index m should not appear any more", should have been guessed, but we could not have been sure that the numerical factor in front is really a 1.

Sometimes the ***twice summed product*** *of the two Levi-Civita symbols is needed which we now can get very simply:*

$$\boxed{\varepsilon_{klm}\varepsilon_{pqn}\delta_{l\,q}\delta_{m\,n} = \varepsilon_{klm}\varepsilon_{plm} = (\delta_{k\,p}\delta_{l\,l} - \delta_{k\,l}\delta_{l\,p}) = (3-1)\delta_{k\,p} = 2\delta_{k\,p},}$$

being symmetric in the index pair kp as it must be. As a joke we finally sum also over ***the third index pair****:*

$$\boxed{\varepsilon_{klm}\varepsilon_{pqn}\delta_{k\,p}\delta_{l\,q}\delta_{m\,n} = \varepsilon_{klm}\varepsilon_{klm} = 2\delta_{k\,k} = 2\cdot 3 = 3!}$$

Analogous relations hold also for totally summed Levi-Civita symbols in spaces of other dimensions, e.g. in $\mathbb{R}^4$ with the result: "number of the dimension factorial"

Exercise 9.42 Levi-Civita symbol

a) Express the Levi-Civita symbol in terms of Kronecker symbols.

b) Express the Kronecker symbol in terms of Levi-Civita symbols.

c) Express the Levi-Civita symbol in terms of unit vectors.

d) Express the Kronecker symbol in terms of unit vectors.

9.7.2 Nested Vector Product

Beside the triple product there is still another product of three vectors $\vec{a} = a_k\vec{e}_k$, $\vec{b} = b_l\vec{e}_l$ and $c = c_m\vec{e}_m$ (in each case with summation convention!): the **nested vector product**, which plays an important role in physics, e.g. in expressing the centrifugal force. Already in connection with the question about the validity of an Associative Law for the vector product we have calculated two such nested vector products as counter-examples.

For the general case we compute first the nested product with the inner vector product as **second** factor, whereby we comment on each step of the argument in detail (You should not forget to think of the summation convention!):

$$\begin{aligned}
\left[\vec{a} \times \left[\vec{b} \times \vec{c}\right]\right] &= \left[\vec{a} \times \varepsilon_{pqn} b_p c_q \vec{e}_n\right], && \text{the inner vector product was inserted,}\\
&= \varepsilon_{pqn} b_p c_q \left[a \times \vec{e}_n\right], && \text{because of the homogeneity of the vector product,}\\
&= \varepsilon_{pqn} b_p c_q \varepsilon_{lmk} a_l (\vec{e}_n)_m \vec{e}_k, && \text{also the outer vector product was inserted,}\\
&= \varepsilon_{pqn} \varepsilon_{klm} b_p c_q a_l (\vec{e}_n)_m \vec{e}_k, && \text{with } \varepsilon_{lmk} = \varepsilon_{klm} \text{ cyclically permuted,}\\
&= \varepsilon_{pqn} \varepsilon_{klm} b_p c_q a_l \delta_{n\,m} \vec{e}_k, && \text{component representation of } \vec{e}_n = \delta_{n\,m}\vec{e}_m,\\
&= \varepsilon_{pqn} \varepsilon_{kln} b_p c_q a_l \vec{e}_k, && \text{the sum over } m = 1,2,3 \text{ contributes only for } m = n,\\
&= (\delta_{k\,p}\delta_{l\,q} - \delta_{k\,q}\delta_{l\,p}) b_p c_q a_l \vec{e}_k, && \text{the product of the Levi-Civita symbols was inserted,,}\\
&= (\delta_{l\,q} b_p c_q a_l \vec{e}_p - \delta_{l\,p} b_p c_q a_l \vec{e}_q), && \text{both sums over } k \text{ were performed,}\\
&= (b_p c_q a_q \vec{e}_p - b_p c_q a_p \vec{e}_q), && \text{both sums over } l \text{ were performed,}\\
&= (\vec{b} c_q a_q - (\vec{a} \cdot \vec{b}) c_q \vec{e}_q), && \text{both sums over } p \text{ were performed,}\\
&= (\vec{a} \cdot \vec{c})\vec{b} - (\vec{a} \cdot \vec{b})\vec{c}, && \text{both sums over } q \text{ were performed.}
\end{aligned}$$

Altogether we obtain the so-called

Grassmann expansion theorem: $\left[\vec{a} \times \left[\vec{b} \times \vec{c}\right]\right] = (\vec{a} \cdot \vec{c})\vec{b} - (\vec{a} \cdot \vec{b})\vec{c},$

i.e. a vector coplanar with the factors $\vec{b}$ and $\vec{c}$ of the inner vector product.

If the Associative Law were to apply, this would be equal to $\left[\left[\vec{a} \times \vec{b}\right] \times \vec{c}\right]$. That is, however, as we have seen, not the case. Instead:

$$\left[\left[\vec{a} \times \vec{b}\right] \times \vec{c}\right] = (\vec{a} \cdot \vec{c})\vec{b} - \vec{a}(\vec{b} \cdot \vec{c}).$$

i.e. the product vector is again coplanar with the factors of the inner vector product, these are, however, now $\vec{a}$ and $\vec{b}$.

Exercise 9.43 Proof of $\left[\left[\vec{a} \times \vec{b}\right] \times \vec{c}\right] = (\vec{a} \cdot \vec{c})\vec{b} - \vec{a}(\vec{b} \cdot \vec{c})$:

Prove this relation along lines completely parallel to the way we demonstrated it above.

Exercise 9.44 Centrifugal force

How is the centrifugal force F_z connected with the angular velocity $\vec{\omega}$ during rotational motion?

Exercise 9.45 Jacobi identity

Compute the Jacobi identity: $\left[\vec{a} \times \left[\vec{b} \times \vec{c}\right]\right] + \left[\vec{b} \times [\vec{c} \times \vec{a}]\right] + \left[\vec{c} \times \left[\vec{a} \times \vec{b}\right]\right]$.

9.7.3 Scalar Product of Two Vector Products

Amongst the multiple products constructed from **four** vectors the **scalar product of two vector products** is most commonly encountered. The scalar product of two angular momenta has, for instance, this structure, and of course the square of an angular momentum.

We calculate generally for four vectors $\vec{a} = a_k\vec{e}_k$, $\vec{b} = b_l\vec{e}_l$, $\vec{c} = c_m\vec{e}_m$ and $\vec{d} = d_m\vec{e}_m$ (each with summation convention!):

$$\begin{aligned}
(\left[\vec{a} \times \vec{b}\right] \cdot \left[\vec{c} \times \vec{d}\right]) &= (\varepsilon_{klm}a_k b_l \vec{e}_m \varepsilon_{pqn} c_p d_q \vec{e}_n) \\
&= \varepsilon_{klm}\varepsilon_{pqn} a_k b_l c_p d_q (\vec{e}_m \vec{e}_n) \\
&= \varepsilon_{klm}\varepsilon_{pqn} a_k b_l c_p d_q \delta_{m\,n} \\
&= \varepsilon_{klm}\varepsilon_{pqm} a_k b_l c_p d_q \\
&= (\delta_{k\,p}\delta_{l\,q} - \delta_{k\,q}\delta_{l\,p}) \\
&= a_k b_l c_p d_q,
\end{aligned}$$

$$\text{i.e. } (\left[\vec{a} \times \vec{b}\right] \cdot \left[\vec{c} \times \vec{d}\right]) = (\vec{a} \cdot \vec{c})(\vec{b} \cdot \vec{d}) - (\vec{a} \cdot \vec{d})(\vec{b} \cdot \vec{c}).$$

Here we have inserted the component representations of the two vector products, then we took advantage of the homogeneity of the scalar product, used the orthonormality relation for the basis vectors, summed over $n = 1, 2, 3$, expressed the product of both the once-summed Levi-Civita symbols through Kronecker symbols and finally reduced the whole thing to scalar products by executing the remaining four summations.

A famous special case of this relation for $c = a$ and $d = b$ is the so-called

Lagrange identity: $\left[\vec{a} \times \vec{b}\right]^2 = a^2 b^2 - (\vec{a} \cdot \vec{b})^2$

Exercise 9.46 Momentum of inertia

In defining the momentum of inertia your lecturer uses the following equation $(\vec{e}_3 \left[\vec{x} \times [\vec{e}_3 \times \vec{x}]\right]) = [\vec{x} \times \vec{e}_3]^2$ *without any comment. Is he permitted to do so?*

9.7.4 Vector Product of Two Vector Products

We conclude our discussion of multiple products with the **vector product of two vector products**. First we decide to maintain the **second** inner vector product as long as possible untouched:

$$\begin{aligned}\Big[\big[\vec a \times \vec b\big] \times \big[\vec c \times \vec d\big]\Big] &= \Big[\varepsilon_{klm} a_k b_l \vec e_m \times \big[\vec c \times \vec d\big]\Big] \\ &= \varepsilon_{klm} a_k b_l \Big[\vec e_m \times \big[\vec c \times \vec d\big]\Big] \\ &= \varepsilon_{klm} a_k b_l \Big((\vec e_m \cdot \vec d)\vec c - (\vec e_m \cdot \vec c)\vec d\Big) \\ &= \varepsilon_{klm} a_k b_l (d_m \vec c - c_m \vec d),\end{aligned}$$

$$\textbf{i.e.} \quad \Big[\big[\vec a \times \vec b\big] \times \big[\vec c \times \vec d\big]\Big] = (\vec a\, \vec b\, \vec d)\, \vec c - (\vec a\, \vec b\, \vec c)\, \vec d, \quad \textbf{i.e. coplanar with } \vec c \textbf{ and } \vec d.$$

Here we firstly have replaced the first inner vector product by its components, then we took advantage of the homogeneity of the vector product, afterwards we have developed the nested vector product according to Grassmann, executed the projection on the components and finally reached triple products as coefficients of both the factor vectors of the second vector product, in the plane of which the result must lie.

Worried about this apparent asymmetry, we calculate the same product once more, now by maintaining the **first** inner vector product as long as possible untouched and otherwise proceeding in full analogy to above:

$$\begin{aligned}\Big[\big[\vec a \times \vec b\big] \times \big[\vec b \times \vec c\big]\Big] &= \Big[\big[\vec a \times \vec b\big] \times \varepsilon_{klm} c_k d_l \vec e_m\Big] \\ &= \varepsilon_{klm} c_k d_l \Big[\big[\vec a \times \vec b\big] \times \vec e_m\Big] \\ &= \varepsilon_{klm} c_k d_l \Big((\vec a \cdot \vec e_m)\vec b - (\vec b \cdot \vec e_m)\vec a\Big) \\ &= \varepsilon_{klm} c_k d_l (a_m \vec b - b_m \vec a),\end{aligned}$$

$$\textbf{i.e.} \quad \Big[\big[\vec a \times \vec b\big] \times \big[\vec c \times \vec d\big]\Big] = (\vec a\, \vec c\, \vec d)\, \vec b - (\vec b\, \vec c\, \vec d)\, \vec a, \quad \textbf{i.e. coplanar with } \vec a \textbf{ and } \vec b.$$

Therefore the product vector of the vector product of two vector products must necessarily lie on the mean straight line of the planes spanned by both the factor pairs of the two inner vector products.

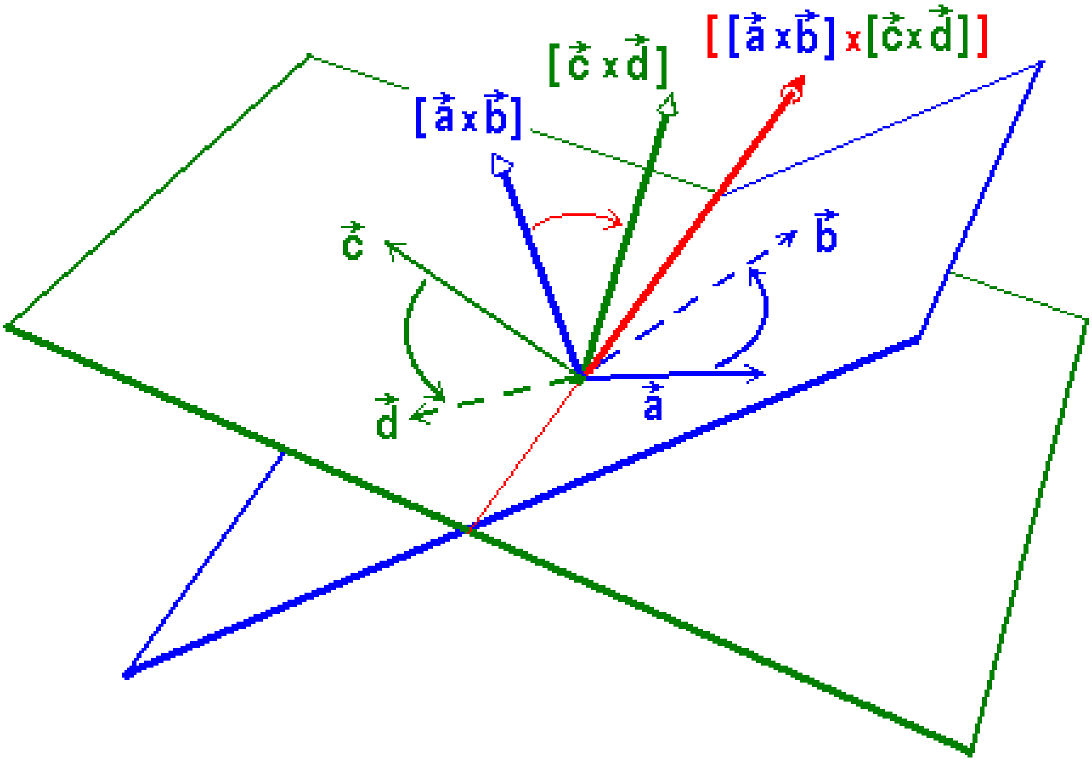


Figure 9.28 : Vector product of two vector products

To have an overview we gather together the relations for the various multiple products:

With $(\vec{a} \cdot \vec{b}) = a_k b_k$, $\left[\vec{a} \times \vec{b}\right] = \varepsilon_{klm} a_k b_l \vec{e}_m$ and $(\vec{a}\,\vec{b}\,\vec{c}) := \left(\left[\vec{a} \times \vec{b}\right] \cdot \vec{c}\right) = a_k b_l c_m \varepsilon_{klm}$ there holds:

$$\left[\vec{a} \times \left[\vec{b} \times \vec{c}\right]\right] = (\vec{a} \cdot \vec{c})\,\vec{b} - (\vec{a} \cdot \vec{b})\,\vec{c}$$

$$\left(\left[\vec{a} \times \vec{b}\right] \cdot \left[\vec{c} \times \vec{d}\right]\right) = (\vec{a} \cdot \vec{c})(\vec{b} \cdot \vec{d}) - (\vec{a} \cdot \vec{d})(\vec{b} \cdot \vec{c})$$

$$\left[\left[\vec{a} \times \vec{b}\right] \times \left[\vec{c} \times \vec{d}\right]\right] = (\vec{a}\,\vec{b}\,\vec{d})\,\vec{c} - (\vec{a}\,\vec{b}\,\vec{c})\vec{d} = (\vec{a}\,\vec{c}\,\vec{d})\,\vec{b} - (\vec{b}\,\vec{c}\,\vec{d})\,\vec{a}$$

Thus we have reduced all multiple products into scalar, vector and triple products. The only thing remaining to do is to clarify how these three kinds of products behave under changes of the coordinate system. To accomplish this we need a relation concerning the determinant of a transformation matrix, the *determinant formula* which we will obtain in the following insert.

Exercise 9.47 Triple product of vector products

Calculate the triple product $\left(\left[\vec{a}\times\vec{b}\right]\left[\vec{b}\times\vec{c}\right]\left[\vec{c}\times\vec{a}\right]\right)$.

Insert: Determinant Formula: *In all considerations concerning coordinate transformations you will again and again use the following*

> **determinant formula:** $A_{pk}A_{ql}A_{nm}\varepsilon_{klm} = |\mathbf{A}|\varepsilon_{pqn}$.

This relation looks more complicated than it really is. We want to make clear how it comes about:
In order to do this, let us consider an arbitrary (3×3)*-matrix and form:*

$$A_{pk}A_{ql}A_{nm}\varepsilon_{klm} = A_{pk}A_{ql}A_{nm}\begin{vmatrix}\delta_{k\,1} & \delta_{k\,2} & \delta_{k\,3}\\ \delta_{l\,1} & \delta_{l\,2} & \delta_{l\,3}\\ \delta_{m\,1} & \delta_{m\,2} & \delta_{m\,3}\end{vmatrix} =$$

Here we have first replaced the Levi-Civita symbol by its determinant representation with Kronecker symbols and remember the three summations over k, l *and* m. *Since determinants are homogenous with respect to their rows, we multiply the first line of the determinant with the first factor* A_{pk}, *the second one with the second factor* A_{ql} *and the third line with the third factor* A_{nm}. *Then we carry out the summations in all nine matrix elements:*

$$= \begin{vmatrix}A_{pk}\delta_{k\,1} & A_{pk}\delta_{k\,2} & A_{pk}\delta_{k\,3}\\ A_{ql}\delta_{l\,1} & A_{ql}\delta_{l\,2} & A_{ql}\delta_{l\,3}\\ A_{nm}\delta_{m\,1} & A_{mn}\delta_{m\,2} & A_{mn}\delta_{m\,3}\end{vmatrix} = \begin{vmatrix}A_{p1} & A_{p2} & A_{p3}\\ A_{q1} & A_{q2} & A_{q3}\\ A_{n1} & A_{n2} & A_{n3}\end{vmatrix} =$$

Now we undo the summation over the Kronecker symbols standing to the right just summed, by extracting the Kronecker symbols but now ***on the left side****. Then we realize that we have got the determinant of the product of two matrices:*

$$= \begin{vmatrix}\delta_{p\,k}A_{k1} & \delta_{p\,k}A_{k2} & \delta_{p\,k}A_{k3}\\ \delta_{q\,k}A_{k1} & \delta_{q\,k}A_{k2} & \delta_{q\,k}A_{k3}\\ \delta_{n\,k}A_{k1} & \delta_{n\,k}A_{k2} & \delta_{n\,k}A_{k3}\end{vmatrix} = \left|\begin{pmatrix}\delta_{p\,1} & \delta_{p\,2} & \delta_{p\,3}\\ \delta_{q\,1} & \delta_{q\,2} & \delta_{q\,3}\\ \delta_{n\,1} & \delta_{n\,2} & \delta_{n\,3}\end{pmatrix}\begin{pmatrix}A_{11} & A_{12} & A_{13}\\ A_{21} & A_{22} & A_{23}\\ A_{31} & A_{32} & A_{33}\end{pmatrix}\right| =$$

The determinant of the product of two matrices is however equal to the product of the determinants of the two factors.

$$= \begin{vmatrix}\delta_{p\,1} & \delta_{p\,2} & \delta_{p\,3}\\ \delta_{q\,1} & \delta_{q\,2} & \delta_{q\,3}\\ \delta_{n\,1} & \delta_{n\,2} & \delta_{n\,3}\end{vmatrix}|\mathbf{A}| = |\mathbf{A}|\varepsilon_{pqn}.$$

Thus we have reached the desired result which you will need very often.

Since the whole derivation could be carried out also for the transposed matrix, the determinant formula is often used also in the following form:

$$\boxed{\varepsilon_{klm} A_{kp} A_{lq} A_{mn} = |\mathbf{A}| \varepsilon_{pqn}.}$$

*It is important that our determinant formula $A_{pk} A_{ql} A_{nm} \varepsilon_{klm} = |\mathbf{A}| \varepsilon_{pqn}$ can be looked at in an entirely different way: The Levi-Civita symbol, as a quantity with three indices, can be considered also as a **tensor of third order**, and the left side of our formula as $\hat{\varepsilon}_{pqn}$, i.e. as a representation of the 27 tensor components in the transformed coordinate system: for each index one transformation matrix. Thus viewed, $\hat{\varepsilon}_{pqn} = |\mathbf{A}| \varepsilon_{pqn}$ means the rotation invariance of the tensor components, i.e. the $\pm$ ones and the zeros are in every coordinate system the same! In addition to this fact there comes from $|\mathbf{A}|$ a minus sign for reflections. Thus we are dealing with a pseudotensor. Correspondingly you will encounter the the Levi-Civita symbol, being totally antisymmetric under exchange of each two indices, under the name "**numerically rotation invariant pseudotensor of third order**".*

9.8 Transformation Properties of the Products

9.8.1 Orthonormal Right-handed Bases

After having got both products of two vectors we are able to characterize our original coordinate system S with its origin O in an elegant manner: Our basis vectors $\vec{e}_k$ with $k = 1, 2, 3$ form an OrthoNormal Right-handed Basis (:ONRB), i.e. they:

are	**1) orthonormal:**	$(\vec{e}_k \cdot \vec{e}_l) = \delta_{k\,l}$
form a	**2) right-handed system:**	$[\vec{e}_k \times \vec{e}_l] = \varepsilon_{klm} \vec{e}_m$
and are	**3) complete:**	$\vec{e}_k)\,(\vec{e}_k = \mathbf{1}$.

Besides this we consider with the same origin $\widehat{O} = O$ in addition an other coordinate system $\widehat{S}$ whose basis vectors $\hat{\vec{e}}_k$ with $k = 1, 2, 3$ can be obtained from the old basis vectors by a linear transformation matrix $\mathbf{A}$:

$$\boxed{\textbf{basis transformation:} \quad \hat{\vec{e}}_k = A_{kl} \vec{e}_l}$$

We get the elements of the (3×3)-transformation matrix $A_{kl} = (\hat{\vec{e}}_k \cdot \vec{e}_l) = \cos \angle(\hat{\vec{e}}_k, \vec{e}_l)$ through scalar multiplication with $\vec{e}_l$.

Now, physicist are interested in the following question: **What kinds of matrices are permitted, if the new basis shall form once again an ONRB?**

9.8.2 Group of the Orthogonal Matrices

In order to answer that question we first treat orthonormality. The following relation is required to hold:
$\delta_{p\,q} = (\hat{\vec{e}}_p \cdot \hat{\vec{e}}_q) = (A_{pk}\vec{e}_k \cdot A_{ql}\vec{e}_l) = A_{pk}A_{ql}(\vec{e}_k \cdot \vec{e}_l) = A_{pk}A_{ql}\delta_{k\,l} = A_{pk}A_{qk} = A_{pk}A^T_{kq}$, *i.e.*

$$\boxed{\delta_{p\,q} = A_{pk}A^T_{kq}}$$

Therefore, only matrices with the following property are permitted $\mathbf{A}\,\mathbf{A}^T = \mathbf{1}$ or $\mathbf{A}^{-1} = \mathbf{A}^T$. Mathematicians call these matrices **orthogonal**. From their nine matrix elements **only three real numbers are independent** because of the six constraint equations:
$A^2_{p1} + A^2_{p2} + A^2_{p3} = 1$ for $p = 1, 2, 3$ and
$A_{p1}A_{q1} + A_{p2}A_{q2} + A_{p3}A_{q3} = 0$ for $p \neq q = 1, 2, 3$.
With respect to matrix multiplication, the orthogonal matrices form a group called O(3) which of course can not be Abelian, since we have found the multiplication of matrices in general not to be commutative:

To verify the group property we consider first the product $C_{kl} = B_{kp}A_{pl}$ of two orthogonal matrices $\mathbf{A}$ with $A_{pn}A_{qn} = \delta_{p\,q}$ and $\mathbf{B}$ with $B_{kp}B_{lp} = \delta_{k\,l}$ and calculate:

$$\boxed{C_{kn}C_{ln} = B_{kp}A_{pn}B_{lq}A_{qn} = B_{kp}B_{lq}\delta_{p\,q} = B_{kp}B_{lp} = \delta_{k\,l}}$$

meaning the **product of two orthogonal matrices is again orthogonal**.
Furthermore the **associative law** holds true as for every matrix multiplication:

$$\boxed{\mathbf{C}(\mathbf{B}\,\mathbf{A}) = \mathbf{C}\,\mathbf{B}\,\mathbf{A} = (\mathbf{C}\,\mathbf{B})\mathbf{A}.}$$

Exactly one **unit element** exists, since the unit matrix $\mathbf{1}$ is orthogonal because from $\mathbf{1}^T = \mathbf{1}$ it follows that $\mathbf{1} \cdot \mathbf{1}^T = \mathbf{1}\,\mathbf{1} = \mathbf{1}$:

$$\boxed{\exists!\ \text{unit element } \mathbf{1} :\ \delta_{k\,m}A_{ml} = A_{kl} = A_{km}\delta_{m\,l} \text{ for all } \mathbf{A} \in O(3).}$$

For multiplication of a matrix $\mathbf{A}$ from the left or from the right with the unit matrix $\mathbf{1}$ one obtains again the old matrix.

And an unambiguously determined inverse exists for every orthogonal matrix $\mathbf{A}$, namely the transposed matrix. For this was precisely the orthogonality condition:

$$\forall \mathbf{A} \in O(3) \quad \exists! \text{ Inverse } \mathbf{A}^{-1} = \mathbf{A}^T: \quad \mathbf{A}\,\mathbf{A}^T = \mathbf{1}$$

The necessary condition for the existence of an inverse $|\mathbf{A}| \neq 0$ is fulfilled, since from $\mathbf{A}\mathbf{A}^T = \mathbf{1}$ there follows for the determinant

$$|\mathbf{A}\mathbf{A}^T| = |\mathbf{A}||\mathbf{A}^T| = |\mathbf{A}|^2 = |\mathbf{1}| = 1$$

$$|\mathbf{A}| = \pm 1 \neq 0.$$

With this, all group properties of orthogonal matrices are proven. From the determinant we see furthermore that two kinds of orthogonal matrices exist: those with determinant +1, the rotations, and those with determinant −1. The latter are just the reflections.

The defining equation for the orthogonal matrices $A_{pk}A_{qk} = \delta_{p\,q}$ opens up a fully different view on our Kronecker symbol: If we include on the left side a superfluous δ with a further summation, we obtain: $A_{pk}A_{ql}\delta_{k\,l} = \delta_{p\,q}$. If we regard the Kronecker symbol, because of its two indices, as a **second order tensor**, we find on the left side $\hat{\delta}_{pq}$, i.e. the nine elements of this tensor transformed to the new coordinate system: for each index a transformation matrix. Thus the entire equation $\hat{\delta}_{pq} = \delta_{p\,q}$ means the invariance of the matrix elements under rotations and reflections, i.e. the ones and zeros are unchanged in every coordinate system and stay at the same position: The Kronecker symbol, symmetric against interchanging the two indices, is from a higher point of view a "**numerically invariant tensor of second order**". You will frequently encounter it later on under this name.

9.8.3 Subgroup of Rotations

Distinguished by having the determinant +1 the rotations form a subgroup of the group O(3), the so-called **special orthogonal group** SO(3), since obviously (+1)(+1) = +1.

After these anticipatory discoveries, we want to investigate further how the orthogonal matrices, allowed for the transformations of coordinate systems, may be further restricted by the demand that the new basis vectors form again a **right-handed system**. That will be the case if the following equation holds true:

$$\varepsilon_{pqn}\hat{\vec{e}}_n = |\mathbf{A}|\varepsilon_{pqn}\hat{\vec{e}}_n.$$

In order to prove this let us consider

$$\varepsilon_{pqn}\hat{\vec{e}}_n = \left[\hat{\vec{e}}_p \times \hat{\vec{e}}_q\right] = A_{pk}A_{ql}\left[\vec{e}_k \times \vec{e}_l\right] = A_{pk}A_{ql}\varepsilon_{klm}\vec{e}_m = ...$$

After these calculational steps already known to us using the homogeneity of the vector product, an unusual but important step now follows: We introduce in addition to the three summations hidden in the Einstein convention over $k = 1, 2, 3$, $l = 1, 2, 3$ and $m = 1, 2, 3$ a fourth one by inserting a Kronecker symbol which seems at first sight superfluous, and sum over $r = 1, 2, 3$:

$$... = A_{pk}A_{ql}\varepsilon_{klm}\delta_{m\,r}\vec{e}_r = ...$$

The **1** introduced through the Kronecker symbol is now replaced by $\mathbf{1} = \mathbf{A}\,\mathbf{A}^T$, i.e. $\delta_{m\,r} = A_{nm}A_{nr}$ with our orthogonal transformation matrix $\mathbf{A}$:

$$... = A_{pk}A_{ql}\varepsilon_{klm}A_{nm}A_{nr}\vec{e}_r = A_{pk}A_{ql}A_{nm}\varepsilon_{klm}A_{nr}\vec{e}_r = ...$$

After exchanging the numbers ε_{klm} and A_{nm} we reach an expression which allows us to use the determinant formula $A_{pk}A_{ql}A_{nm}\varepsilon_{klm} = |\mathbf{A}|\varepsilon_{pqn}$ derived earlier:

$$... = |\mathbf{A}|\varepsilon_{pqn}A_{nr}\vec{e}_r = |\mathbf{A}|\varepsilon_{pqn}\hat{\vec{e}}_n$$

Therefore, if our new basis shall be again a right-handed system, we must permit only those transformation matrices whose determinant is $|\mathbf{A}| = +1$, i.e. only elements of the **subgroup SO(3) of the rotations**. This is, however, just what we have seen in an example at the beginning of this chapter: The parity transformation changed a right-handed coordinate system into a left-handed one, and so do all other transformations containing a reflection. The presence of coordinate reflections is reliably detected by the negative determinant.

9.8.4 Transformation of the Products

At the end of this course we want to check how our products of vectors transform under rotations and reflections of the coordinate system:

We know already that the components a_k of a vector $\vec{a}$, which emerged from a displacement as projections onto the coordinate axes $a_k = (\vec{a} \cdot \vec{e}_k)$ transform as do the basis vectors themselves:

$$\hat{a}_k = (\vec{a} \cdot \hat{\vec{e}}_k) = (\vec{a} \cdot A_{kl}\vec{e}_l) = A_{kl}(\vec{a} \cdot \vec{e}_l) = A_{kl}a_l,$$

thus

$$\boxed{\hat{a}_k = A_{kl}a_l.}$$

In particular the signs are reversed by the reflection through the origin, the parity transformation. Therefore these vectors are called **polar vectors**.
As the first product we examine the **scalar product** $c = (\vec{a} \cdot \vec{b})$ **of two polar vectors** $\vec{a}$ **and** $\vec{b}$:

$$\hat{c} := \widehat{(\vec{a} \cdot \vec{b})} = \hat{a}_k \hat{b}_k = A_{kl} a_l A_{km} b_m = A_{kl} A_{km} a_l b_m = \delta_{l\,m} a_l b_m = a_l b_l = (\vec{a} \cdot \vec{b}) =: c,$$

thus

$$\boxed{\hat{c} := \widehat{(\vec{a} \cdot \vec{b})} = c.}$$

Here we have first inserted the component representation of the scalar product in the new system with the summation over $k = 1, 2, 3$, then we have used the transformation law for the components of polar vectors, the orthogonality relation of the transformation matrices and finally after summation over $m = 1, 2, 3$ we received the component representation of the scalar product in the original coordinate system without any factor in front.

Thus we have shown that our scalar product is invariant under rotations and reflections and therefore fully deserves the name **scalar**.

Next we study the transformation properties of the **components** $v_k = \left[\vec{a} \times \vec{b}\right]_k$ **of the vector product** $\vec{v} = \left[\vec{a} \times \vec{b}\right]$ **of two polar vectors** $\vec{a}$ **and** $\vec{b}$:

$$\begin{aligned}
\hat{v}_k := \widehat{\left[\vec{a} \times \vec{b}\right]}_k &= \hat{a}_p \hat{b}_q \varepsilon_{pqk} = A_{pm} a_m A_{qn} b_n \varepsilon_{pqk} \\
&= A_{pm} A_{qn} a_m b_n \varepsilon_{pqk} = A_{pm} A_{qn} a_m b_n \varepsilon_{pqr} \delta_{r\,k} \\
&= A_{pm} A_{qn} a_m b_n \varepsilon_{pqr} A_{rl} A_{kl} = A_{pm} A_{qn} A_{rl} \varepsilon_{pqr} a_m b_n A_{kl} \\
&= |\mathbf{A}| \varepsilon_{mnl} a_m b_n A_{kl} = |\mathbf{A}| A_{kl} \left[\vec{a} \times \vec{b}\right]_l =: |\mathbf{A}| A_{kl} v_l,
\end{aligned}$$

thus

$$\boxed{\hat{v}_k := \widehat{\left[\vec{a} \times \vec{b}\right]}_k = |\mathbf{A}| A_{kl} v_l}$$

Here we have inserted the component representation of the k-th component of the vector product in the transformed coordinate system with summations over $p = 1, 2, 3$ and $q = 1, 2, 3$ then we have carried out the transformation of the vector components of both the factors and introduced a superfluous Kronecker symbol with the summation over $r = 1, 2, 3$ for the last index of the Levi-Civita symbol. This δ we have replaced through two orthogonal transformation matrices and combine the three matrices **A** with help of the determinant formula $A_{pk}A_{ql}A_{nm}\varepsilon_{klm} = |\mathbf{A}|\varepsilon_{pqn}$ into the determinant, and finally written the vector product in terms of the old vector product in the old untransformed system.

Thus, it follows that the **vector product** transforms under rotations like a displacement as a **vector**, but under reflections because the determinant $|\mathbf{A}| = -1$ introduces an extra minus sign, the vector product is **reflection invariant**. Vectors with this property are called **axial vectors**, and in fact all vectors built as vector products of two polar vectors appearing in physics are reflection invariant. As we have seen, all of these are related to rotational processes and represent in some sense - unlike the direction arrow of displacement vectors - a **turning event** which does not change when looked at in a mirror. The following figure attempts to display this.

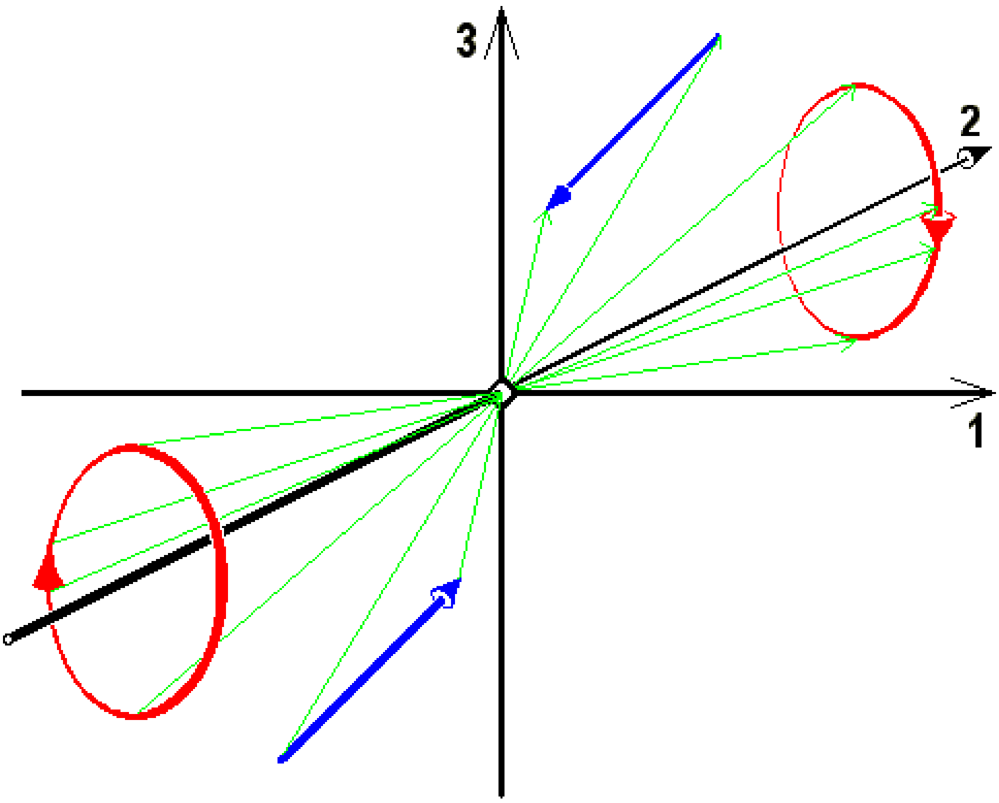


Figure 9.29: Turning circle and displacement arrow under a reflection through the origin

Exercise 9.48 Polar and axial vectors

Sort the following examples of physical vectors in two boxes according to their behavior under reflections, on the one hand the polar vectors and on the other hand the axial ones:

Position vector, momentum, angular momentum, velocity, angular velocity, force, torque, magnetic moment, electric dipole moment, magnetic field, electric current density, electric field, electric displacement, electromagnetic radiation flux density, etc.

Finally we want to have a look at the **triple product** $d = (\vec{a}\,\vec{b}\,\vec{c})$ **of three polar vectors** $\vec{a}$, $\vec{b}$ **and** $\vec{c}$ in the transformed coordinate system $\widehat{S}$:

$$\begin{aligned}\hat{d} &:= \widehat{(\vec{a}\,\vec{b}\,\vec{c})} = \hat{a}_k \hat{b}_l \hat{c}_m \varepsilon_{klm} = A_{kp}A_{lq}A_{mn}\varepsilon_{pqn}a_p b_q c_n \\ &= |\mathbf{A}|\varepsilon_{pqn}a_p b_q c_n = |\mathbf{A}|(\vec{a}\,\vec{b}\,\vec{c}) =: |\mathbf{A}|d,\end{aligned}$$

thus

$$\boxed{\hat{d} := \widehat{(\vec{a}\,\vec{b}\,\vec{c})} = |\mathbf{A}|d.}$$

We have once again transformed the polar vector components and used the determinant formula $A_{pk}A_{ql}A_{nm}\varepsilon_{klm} = |\mathbf{A}|\varepsilon_{pqn}$.
The triple product behaves as a **scalar** under rotations, but is by no means invariant under reflections, because it changes sign. Such a quantity is called a **pseudoscalar**.

Exercise 9.49 Parity violation

a) Why can we not use the triple product of the three momenta $(\vec{p}_{\mathrm{Ni}}\vec{p}_e\vec{p}_\nu)$ *to check parity symmetry for* β*-decay, e.g. in the reaction* $\mathrm{Co}^{60} \longrightarrow \mathrm{Ni}^{60} + \mathrm{e}^- + \nu$*?*

b) Which quantity has been measured instead?

Here we have treated only the simplest rules of vector algebra. You will learn much more about vectors during your studies. You will study vectors that are functions of a scalar variable, in particular the time variable, and also scalars and vectors that are functions of other vectors, usually of the position vector or the momentum, i.e. the so-called fields. You will learn to differentiate vectors, to expand them in series according to Taylor, and to integrate them in several different ways. All these vectors show the characteristic behavior under rotations of the coordinate system, and they can be distinguished as polar or axial according to their reflection behavior. Dealing with the theory of relativity you will calculate with vectors which consist of four components. In field theory you will learn to handle infinitely dimensional vectors. But the basic structures will always be the same, the structures we have gotten to know here.

Beyond scalars and vectors, in some areas of physics you will meet tensors of **second** order, e.g. the momentum of inertia, the stress tensor and the electric quadrupol moment. In four-dimensional space-time the electromagnetic fields together form a second-order tensor. The transformation theory of tensors has much in common with the transformation theory of vectors, and our considerations here have prepared you for this more advanced material. I do hope you will enjoy it!