SCRIPT PHYSICS OLYMPIAD

 THIRD EDITION

 AUTUMN 2019
Script Physics Olympiad

Third Edition

Autumn 2019
Preface

Every time I gave a lesson at a training camp I opened a book, decided what to tell and took some notes which often filled a couple of pages. Of course this preparation took me always a lot of time and sometimes I recognized during the lecture that I had forgotten a topic and had to improvize. I started asking myself if it would not be easier to write a script and every year, look up in the script what I want to tell at a lesson. Together with some members of the Physics Olympiad I started writing a script and what you now hold in your hands is the third edition of it.

As I already mentioned above the script is on the one hand helpful to give lectures. On the other hand students do not have to write everything down what the teacher tells, they simply can take notes to complete or stress out important things. Therefore this script does not replace a good book. Because in a good book there are often much more details which can not be given in a script which is written to accompany a two-hours-lecture. Additionally I highly recommend to solve problems, because then one has to think about what the abstract laws of physics are telling and how they can be used.

Now the third edition of the script is finished and it somehow reminds me of hiking: Before the beginning I was full of good ideas and expectations how it will look like. While I worked on it, I sometimes asked myself why I did all this expenditure because it was a lot of work. When it was finished I was proud to have endured all this burden and lucky to have made many interesting experiences and even learned a lot. In this sense I want to give a special thanks to the other authors, namely Cyrill, Levy, David, Quentin, Lionel, Sven and Sebastian. Also a big thanks to all the members who helped, corrected and proofread the script.

If you have any comment on the script, found a mistake or would like to help improving or developing it, please write me a mail at rafael.winkler@swisspho.ch

I wish you a entertaining and instructive reading,

Best

Rafael
Contents

1 Mathematics ......................................................... 13
  1.1 Vector algebra ................................................ 13
      1.1.1 Scalar product ....................................... 13
      1.1.2 Vector product ...................................... 16
  1.2 Differential calculus ......................................... 19
      1.2.1 Derivative of a function .............................. 19
      1.2.2 Differentiation rules ................................. 22
      1.2.3 More derivatives .................................... 25
      1.2.4 Overview about derivatives ......................... 30
      1.2.5 Higher derivatives .................................. 30
      1.2.6 Taylor approximation ............................... 31
      1.2.7 The idea of differential equations ................. 33
  1.3 Integral calculus ............................................. 35
      1.3.1 Antiderivatives ..................................... 35
      1.3.2 Integral as an area .................................. 36
      1.3.3 Fundamental theorem of calculus .................... 39
      1.3.4 More integration rules .............................. 42
      1.3.5 The idea of multidimensional integrals ............ 49
  1.4 Complex Numbers .............................................. 50
      1.4.1 Introduction ......................................... 50
      1.4.2 Representation of a complex number, Euler formula 51
      1.4.3 A first simple application ......................... 52
      1.4.4 Physical examples .................................. 53

2 Mechanics 1 ...................................................... 55
  2.1 Model .......................................................... 55
  2.2 Kinematics of Point-like Particles ......................... 56
      2.2.1 General Description ................................ 56
      2.2.2 Linear Uniform Acceleration ....................... 58
      2.2.3 Circular Motion ..................................... 61
      2.2.4 General 2 Dimensional Motion ...................... 62
## 2.3 Dynamics of Point-like Particles
- 2.3.1 Force ........................................... 63
- 2.3.2 Choice of Frame of Reference ................. 64
- 2.3.3 Newton’s Laws .................................. 64
- 2.3.4 Center of Mass .................................. 66
- 2.3.5 Equilibrium ................................. 67
- 2.3.6 Gravitational Force ......................... 68
- 2.3.7 Independence of Motion ................. 70
- 2.3.8 Volume and Surface Forces ....... 71
- 2.3.9 Friction ................................. 71

## 2.4 Momentum, Work, Energy and Power
- 2.4.1 Momentum ................................. 72
- 2.4.2 Work ........................................... 73
- 2.4.3 Energy ........................................... 74
- 2.4.4 Potential Energy ......................... 74
- 2.4.5 Kinetic Energy ................................. 77
- 2.4.6 Power ................................. 78
- 2.4.7 Rotation Energy ................................. 79
- 2.4.8 Angular Momentum ................. 79

### 3 Mechanics 2

#### 3.1 Dynamics of Circular Motion ......................... 81

#### 3.2 Frames of Reference ................................. 84
- 3.2.1 Choosing a Frame of Reference ............. 84
- 3.2.2 Accelerated Frames of Reference ............ 86

#### 3.3 Systems of Particles ................................. 88
- 3.3.1 NEWTON’s Second Law for Systems of Particles and the Center of Mass 89
- 3.3.2 Conservation of Linear Momentum .... 93
- 3.3.3 Conservation of Angular Momentum ........ 94

#### 3.4 Rigid Bodies ........................................... 98
- 3.4.1 Kinematics of Rigid Bodies ................. 98
- 3.4.2 How Forces Determine Translational Motion of a Rigid Body .... 101
- 3.4.3 How Forces Determine Rotational Motion of a Rigid Body ..... 103
- 3.4.4 Mechanical Energy of Moving Rigid Bodies .......... 107
- 3.4.5 Putting Everything Together: Rolling Rigid Bodies .......... 109
- 3.4.6 Conservation of Angular Momentum .......... 112

#### 3.5 Gravity ........................................... 112
- 3.5.1 NEWTON’s Law of Gravity ................... 112
- 3.5.2 Gravitational Fields ......................... 113
- 3.5.3 Energy and Angular Momentum in Gravitational Fields .......... 116
- 3.5.4 Two Objects Subject to Mutual Attraction ........ 118
- 3.5.5 KEPLER’s Laws of Planetary Motion* .................. 119
# 4 Thermodynamics

4.1 Important definitions .................................................. 123
4.2 The temperature scale .................................................. 124
4.3 Zeroth law of thermodynamics ......................................... 125
4.4 Thermal energy and heat capacity ..................................... 126
  4.4.1 Molar heat capacities of ideal gases ............................. 126
4.5 Ideal gas law ............................................................. 127
4.6 First law of thermodynamics ............................................ 127
4.7 Thermodynamic systems ................................................ 128
4.8 Equipartition theorem ................................................. 128
4.9 Thermodynamic processes .............................................. 129
4.10 Second law of thermodynamics ...................................... 131
4.11 Heat engines ............................................................ 132
4.12 Kinetic gas theory ...................................................... 134
4.13 Phase transitions ....................................................... 136
4.14 Real gases ............................................................... 137
4.15 Stefan-Boltzmann law ................................................. 138

# 5 Oscillations

5.1 Introduction ............................................................ 139
5.2 Harmonic Oscillations .................................................. 140
  5.2.1 Harmonic Oscillations, Spring/Mass Systems and Differential Equations 141
  5.2.2 Further Examples .................................................. 143
  5.2.3 Importance of Harmonic Oscillations .......................... 146
5.3 Beyond Harmonic Oscillations ........................................ 146
  5.3.1 Example of Forced Oscillating Systems ....................... 150

# 6 Waves

6.1 Introduction ............................................................ 151
6.2 Harmonic Waves ....................................................... 152
6.3 Waves in 3D ............................................................. 154
  6.3.1 Waves as Functions of 3D Spatial Coordinates ............... 154
  6.3.2 Waves as 3D Functions, Transversal and Longitudinal Waves, Polarization 154
6.4 Waves Propagation ..................................................... 155
  6.4.1 Sound ............................................................. 155
  6.4.2 Light ............................................................. 156
  6.4.3 Seismic Waves ..................................................... 157
  6.4.4 Transport ......................................................... 157
  6.4.5 Doppler Effect .................................................... 157
6.5 Waves Propagation at Interfaces .................................... 159
  6.5.1 Reflection .......................................................... 159
  6.5.2 Refraction .......................................................... 160
  6.5.3 Diffraction .......................................................... 162
6.6 Multi-Waves Phenomena ................................................ 162
6.6.1 Same amplitude, frequency and wavenumber .......................... 163
6.6.2 Same amplitude and frequency, opposite wavenumber .............. 163
6.6.3 Slightly different frequencies ........................................... 164
6.6.4 Fourier Analysis ............................................................ 165

7 Fluid dynamics ................................................................. 167
  7.1 Introduction .................................................................. 167
  7.1.1 What is fluid dynamics about? ...................................... 167
  7.1.2 How can we model such a fluid? ................................. 167
  7.2 Notation ....................................................................... 168
  7.3 Pressure ......................................................................... 168
    7.3.1 Compressible and incompressible fluids ................. 169
    7.3.2 Hydrostatic pressure ............................................. 169
    7.3.3 Buoyancy .................................................................. 170
  7.4 Continuity equation ........................................................ 171
  7.5 Bernoulli’s equation ....................................................... 171
    7.5.1 Derivation ............................................................ 171
  7.6 Surface tension, energy and capillary pressure ..................... 172
  7.7 Friction in fluids ............................................................ 173

8 Electro-and Magnetostatics .................................................. 175
  8.1 Electrostatics .................................................................. 175
    8.1.1 Coulomb Force ....................................................... 175
    8.1.2 Electrostatic field .................................................. 176
    8.1.3 Superposition ....................................................... 178
    8.1.4 Continuous charge distributions ....................... 178
    8.1.5 Gauss’ law ............................................................ 178
    8.1.6 Examples ............................................................. 180
  8.2 Potential and Voltage ....................................................... 183
    8.2.1 Electric potential .................................................... 183
    8.2.2 Electric potential of a point charge ..................... 184
    8.2.3 Potential of multiple charges ............................... 185
    8.2.4 Voltage ................................................................. 186
    8.2.5 Potential and conducting material .................... 186
    8.2.6 Capacity ............................................................... 187
    8.2.7 Example ............................................................... 187
  8.3 Current and magnetic field ................................................. 190
    8.3.1 Current and conservation of charge .................... 190
    8.3.2 Magnets ................................................................. 191
    8.3.3 Magnet and electric current ............................... 192
    8.3.4 Lorentz force .......................................................... 193
9 Direct current circuits

9.1 Ohm’s law ............... 195
9.2 Equivalent circuit ............... 196
  9.2.1 Wire ............... 196
  9.2.2 Series circuit ............... 196
  9.2.3 Parallel circuit ............... 197
  9.2.4 Voltage source ............... 197
9.3 Electric power ............... 197
9.4 Electric components ............... 198
9.5 Kirchhoff’s circuit law ............... 198
  9.5.1 Current law ............... 198
  9.5.2 Voltage law ............... 199
  9.5.3 Applying Kirchhoff’s law ............... 200
9.6 Examples ............... 201
  9.6.1 Maximal power from a real power source ............... 201
  9.6.2 Charging a capacitor ............... 201

10 Electrodynamics 2

10.1 Magnetism ............... 203
  10.1.1 Magnetic Field and Flux ............... 203
  10.1.2 Ampere’s Law ............... 204
  10.1.3 Magnetic Field of a Moving Point Charge ............... 205
  10.1.4 Example ............... 206
10.2 Induction ............... 210
  10.2.1 Approach and Definition ............... 210
  10.2.2 Self induction ............... 211
  10.2.3 Generator ............... 212
  10.2.4 Transformer ............... 214
10.3 Displacement current ............... 216
10.4 Maxwell’s equations and their conclusions ............... 217
  10.4.1 Maxwell’s equations ............... 217
  10.4.2 Electromagnetic wave ............... 219
10.5 Electro-magnetic field in Materials ............... 221
  10.5.1 Polarizability and dielectric constant ............... 221
  10.5.2 Electric displacement and Polarisation ............... 223
  10.5.3 Continuity equations at interfaces ............... 225
  10.5.4 Magnetic field ............... 226
  10.5.5 Maxwell’s equations in Materials ............... 228
  10.5.6 Electromagnetic waves ............... 228
10.6 Energy of the electromagnetic field ............... 230
  10.6.1 Energy density of the electric field ............... 230
  10.6.2 Energy density of the magnetic field ............... 231
  10.6.3 Poynting vector ............... 231
# 11 Alternating current (AC)

11.1 Describing alternating voltage and current ........................................ 233
  11.1.1 Fourier series .................................................. 233
  11.1.2 Usual real notation ............................................. 234
  11.1.3 Phasor ............................................................ 234
  11.1.4 Complex notation ............................................... 234

11.2 Impedance ............................................................. 236
  11.2.1 Ohmic resistor .................................................. 237
  11.2.2 Capacitor ........................................................ 237
  11.2.3 Inductor .......................................................... 238

11.3 Combinations of R,C and L ............................................ 241
  11.3.1 Kirchhoff’s laws ................................................. 241
  11.3.2 Serial and parallel circuit ..................................... 241
  11.3.3 High pass filter .................................................. 242
  11.3.4 Resonant circuit .................................................. 244

11.4 Power consideration and effective values ................................. 249
  11.4.1 Power ............................................................. 249
  11.4.2 Effective values ................................................. 250
  11.4.3 Active, reactive and apparent power ............................. 250

11.5 Three-phase electric power ............................................. 251
  11.5.1 Definition and production ....................................... 251
  11.5.2 Star and Delta circuit .......................................... 252
  11.5.3 Advantage of a three-phase system ............................. 254

# 12 Special relativity

12.1 Historical milestones ................................................ 257
  12.1.1 Aether and electromagnetic waves ................................ 257
  12.1.2 Flying electron .................................................. 258

12.2 Galileo transformations ................................................. 258
  12.2.1 Reference System ............................................... 259
  12.2.2 Inertial frame of reference ..................................... 260
  12.2.3 Galileo Transformation .......................................... 261

12.3 Lorentz transformation ................................................ 262
  12.3.1 Einstein’s Postulates ............................................. 262
  12.3.2 Synchronisation of clocks ...................................... 263
  12.3.3 Time dilation ..................................................... 263
  12.3.4 Lorentz contraction .............................................. 265
  12.3.5 Symmetry of time dilatation and Lorentz contraction .......... 267
  12.3.6 Lorentz transformation .......................................... 267

12.4 Minkowski metric ..................................................... 269
  12.4.1 Definition of Minkowski metric ................................ 270
  12.4.2 Properties of Minkowski metric ................................ 270
  12.4.3 Four vectors ..................................................... 271
  12.4.4 Note on rigorous derivation ................................... 272
## Contents

12.5 Velocities ................................................................................................................. 272  
- 12.5.1 Addition of parallel velocities ................................................................. 273  
- 12.5.2 Addition of perpendicular velocities ....................................................... 273  
- 12.5.3 Velocity four vector ..................................................................................... 274  
12.6 Dynamics .................................................................................................................. 275  
- 12.6.1 Momentum .................................................................................................... 275  
- 12.6.2 Momentum four vector .................................................................................. 276  
- 12.6.3 Energy .......................................................................................................... 277  
- 12.6.4 Acceleration and forces ................................................................................. 278  
12.7 Paradoxes .................................................................................................................. 279  
- 12.7.1 Ladder and barn ........................................................................................... 279  
- 12.7.2 Twin paradox ............................................................................................... 281  
- 12.7.3 Solution to the flying electron problem ....................................................... 281  

### 13 Quantum Mechanics 283

13.1 Experiments .............................................................................................................. 283  
- 13.1.1 Black body radiation .................................................................................... 283  
- 13.1.2 Photoelectric effect ...................................................................................... 285  
- 13.1.3 Double slit experiment .................................................................................. 288  
13.2 Laws of Quantum Mechanics .................................................................................. 290  
- 13.2.1 Wavefunction and probability ...................................................................... 290  
- 13.2.2 Measurement ............................................................................................... 291  
- 13.2.3 De Broglie hypothesis ................................................................................... 291  
- 13.2.4 Uncertainty Principle ................................................................................... 292  
- 13.2.5 Schrödinger Equation .................................................................................. 293  
13.3 Examples .................................................................................................................. 294  
- 13.3.1 Bohr model ................................................................................................... 294  
- 13.3.2 Rigorous example ......................................................................................... 296  

### 14 Introduction to Statistics 301

14.1 Location and Spread of a single Set of Data ......................................................... 301  
- 14.1.1 Bivariate Analysis ......................................................................................... 302  
14.2 Uncertainty Propagation ......................................................................................... 302  
- 14.2.1 Quantification of Uncertainty ...................................................................... 303  
- 14.2.2 Propagation of Uncertainty .......................................................................... 303  
14.3 Units ......................................................................................................................... 304  
- 14.3.1 The International System of Units (SI) ......................................................... 304  
- 14.3.2 Prefixes ......................................................................................................... 305  
- 14.3.3 Dimensional Analysis ................................................................................... 305  
14.4 Graphs ....................................................................................................................... 305  
- 14.4.1 Elements of good graphs .............................................................................. 305  
- 14.4.2 Logarithmic Plots .......................................................................................... 306
Appendices

A Further derivations
  A.1 Derivations of Statistics .................................................. 309
  A.1.1 Alternative formulations for Variance and Covariance .......... 309
  A.1.2 Derivation of the Least Squares Coefficients ..................... 310

B Tables
  B.1 List of physical constants (in SI units) .............................. 313
  B.2 List of named, SI derived units ......................................... 314
  B.3 List of material constants ............................................... 314
Chapter 1

Mathematics

This part of the script presents the most important mathematical tools. Some things will be known from school, other things maybe will be new. The goal is not to present the entire high school mathematics, but to give an overview over the concepts, which are needed in the physics part. The proofs are not really important for the physics, so one can omit them if not interested.

1.1 Vector algebra

In physics, vectors are of crucial importance, since many physical quantities are represented by vectors. The foundation of vector algebra should be known from school. Therefore we will only repeat the two most important concepts for physics. The main reference for this chapter is [1].

1.1.1 Scalar product

There is an intuitive way to add two vectors and to multiply a vector by a number. There are essentially two ways how two vectors can be multiplied. One of them is the scalar product (also known as dot product).

**Definition:** Let $\vec{a}$ and $\vec{b}$ be vectors. Then the scalar product of $\vec{a}$ and $\vec{b}$ is

$$\vec{a} \cdot \vec{b} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \cdot \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} = a_x b_x + a_y b_y + a_z b_z.$$
There is a nice connection between the scalar product and the angle $\varphi$ between the vectors $\vec{a}$ and $\vec{b}$ (see figure 1.1). We have the following alternative expression for the scalar product:

$$\vec{a} \cdot \vec{b} = a_x b_x + a_y b_y + a_z b_z = |\vec{a}| \cdot |\vec{b}| \cdot \cos(\varphi).$$

![Figure 1.1: $\vec{c} = \vec{a} - \vec{b}$](image)

**Proof:** The equality $a_x b_x + a_y b_y + a_z b_z = |\vec{a}| \cdot |\vec{b}| \cdot \cos(\varphi)$ follows from the cosine formula. We look at figure 1.1 and calculate the length of $\vec{c} = \vec{a} - \vec{b}$ in two ways. On one hand, we have:

$$|\vec{c}|^2 = |\vec{a} - \vec{b}|^2 = (a_x - b_x)^2 + (a_y - b_y)^2 + (a_z - b_z)^2$$

$$= a_x^2 + a_y^2 + a_z^2 + b_x^2 + b_y^2 + b_z^2 - 2a_x b_x - 2a_y b_y - 2a_z b_z$$

$$= |\vec{a}|^2 + |\vec{b}|^2 - 2 (a_x b_x + a_y b_y + a_z b_z).$$

On the other hand, it follows by the cosine formula, that:

$$|\vec{c}|^2 = |\vec{a}|^2 + |\vec{b}|^2 - 2 \cdot |\vec{a}| \cdot |\vec{b}| \cdot \cos(\varphi).$$

Comparing the two expressions for $|\vec{c}|^2$, we find

$$a_x b_x + a_y b_y + a_z b_z = |\vec{a}| \cdot |\vec{b}| \cdot \cos(\varphi).$$

1 The cosine formula tells us, that in a triangle with sidelengths $a$, $b$ and $c$ we have

$$c^2 = a^2 + b^2 - 2ab \cos(\varphi),$$

where $\varphi$ is the angle between $a$ and $b$. 

14
Remarks

- The scalar product is often used to calculate the angle $\varphi$ between two vectors $\vec{a}$ and $\vec{b}$. Using the two different expressions for the scalar product, we get

$$\cos(\varphi) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|} = \frac{a_x b_x + a_y b_y + a_z b_z}{|\vec{a}| \cdot |\vec{b}|}.$$ 

- The scalar product has an intuitive interpretation: One first projects the vector $\vec{a}$ on the vector $\vec{b}$ to get a vector $\vec{a}'$ of length

$$|\vec{a}'| = |\vec{a}| \cdot \cos(\varphi)$$

(see fig. 1.2). Then the scalar product of $\vec{a}$ and $\vec{b}$ is the product

$$|\vec{a}| \cdot \cos(\varphi) \cdot |\vec{b}| = |\vec{a}'| \cdot |\vec{b}|$$

of the length of $\vec{a}'$ with the length of $\vec{b}$. This means that only the part of $\vec{a}$ which is parallel to $\vec{b}$ contributes to the scalar product. The same holds, if one interchanges $\vec{a}$ and $\vec{b}$.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{projection.png}
\caption{Projection of $\vec{a}$ on $\vec{b}$}
\end{figure}
**Properties**  We summarize the most important rules for calculations involving the scalar product. Let \( \vec{a}, \vec{b}, \vec{c} \) be three vectors and \( s \) a real number. Then we have:

\[
\vec{a} \cdot \vec{a} = |\vec{a}|^2 = a_x^2 + a_y^2 + a_z^2
\]

\[
\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}
\]

\[
(s\vec{a}) \cdot \vec{b} = s \cdot (\vec{a} \cdot \vec{b})
\]

\[
\vec{a} \cdot (\vec{b} + \vec{c}) = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c}.
\]

Furthermore: If \( \vec{a}, \vec{b} \neq \vec{0} \), then \( \vec{a} \cdot \vec{b} = 0 \) if and only if \( \vec{a} \) and \( \vec{b} \) are orthogonal.

**Exercise:** Does \( (\vec{a} \cdot \vec{b}) \cdot \vec{c} = \vec{a} \cdot (\vec{b} \cdot \vec{c}) \) hold?

### 1.1.2 Vector product

The scalar product of two vectors is a number (scalar). The vector product (also known as cross product) of two vectors is a vector, concretely:

**Definition:** Let \( \vec{a} \) and \( \vec{b} \) be vectors. The vector product of \( \vec{a} \) and \( \vec{b} \) is defined as

\[
\vec{a} \times \vec{b} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \times \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} = \begin{pmatrix} a_y b_z - a_z b_y \\ a_z b_x - a_x b_z \\ a_x b_y - a_y b_x \end{pmatrix}
\]

**Properties**

1. The vector product \( \vec{a} \times \vec{b} \) is orthogonal to both \( \vec{a} \) and \( \vec{b} \), given that \( \vec{a} \) and \( \vec{b} \) aren’t parallel (see. fig. 1.3). If \( \vec{a} \) and \( \vec{b} \) are parallel then \( \vec{a} \times \vec{b} = \vec{0} \).

2. The vectors \( \vec{a}, \vec{b} \) and \( \vec{a} \times \vec{b} \) follow the right hand rule (see. fig. 1.1).

3. If \( \varphi \) is the angle between \( \vec{a} \) and \( \vec{b} \), we have \( |\vec{a} \times \vec{b}| = |\vec{a}| \cdot |\vec{b}| \cdot \sin(\varphi) \). Therefore the absolute value of the vector product is equal to the area of the parallelogram with sides \( \vec{a} \) and \( \vec{b} \) (see fig. 1.3). This means that only the part of \( \vec{a} \), which is orthogonal to \( \vec{b} \) contributes to \( \vec{a} \times \vec{b} \).
1.1. VECTOR ALGEBRA

Proof:

1. One calculates $\vec{a} \cdot (\vec{a} \times \vec{b}) = 0$ and $\vec{b} \cdot (\vec{a} \times \vec{b}) = 0$. So it follows, that $\vec{a}$ and $\vec{b}$ are orthogonal to $\vec{a} \times \vec{b}$.

2. This is not really an obvious fact. The idea is to first show the fact for $\vec{a}$ and $\vec{b}$ in the $xy$-plane and then one reduces the general case to this.

3. One has $\sin^2(\varphi) = 1 - \cos^2(\varphi)$. Therefore:

$$\left( |\vec{a}| \cdot |\vec{b}| \cdot \sin(\varphi) \right)^2 = |\vec{a}|^2 \cdot |\vec{b}|^2 \cdot (1 - \cos^2(\varphi))$$

$$= |\vec{a}|^2 \cdot |\vec{b}|^2 - \left( |\vec{a}| \cdot |\vec{b}| \cdot \cos(\varphi) \right)^2$$

$$= |\vec{a}|^2 \cdot |\vec{b}|^2 - \left( \vec{a} \cdot \vec{b} \right)^2.$$

On the other hand, one can calculate explicitly, that $|\vec{a}|^2 \cdot |\vec{b}|^2 - \left( \vec{a} \cdot \vec{b} \right)^2 = |\vec{a} \times \vec{b}|^2$. Therefore one gets $|\vec{a} \times \vec{b}| = |\vec{a}| \cdot |\vec{b}| \cdot \sin(\varphi)$. 

Figure 1.3: Vector product
More properties We summarize the most important rules for calculations involving the vector product. Let \( \vec{a}, \vec{b}, \vec{c} \) be three vectors and \( s \) a real number. Then we have:

\[
\begin{align*}
\vec{a} \times \vec{a} &= 0 \\
\vec{a} \times \vec{b} &= -\left( \vec{b} \times \vec{a} \right) \\
(s\vec{a}) \times \vec{b} &= \vec{a} \times \left( s\vec{b} \right) = s \cdot \left( \vec{a} \times \vec{b} \right) \\
\vec{a} \times (\vec{b} + \vec{c}) &= \vec{a} \times \vec{b} + \vec{a} \times \vec{c} \\
(\vec{a} + \vec{b}) \times \vec{c} &= \vec{a} \times \vec{c} + \vec{b} \times \vec{c}
\end{align*}
\]

Exercise: Does \((\vec{a} \times \vec{b}) \times \vec{c} = \vec{a} \times (\vec{b} \times \vec{c}) \) hold?
1.2 Differential calculus

Physics without calculus is impossible, since most physical laws in their general formulation use derivatives or integrals. In this chapter, we look at differential calculus. The next chapter will treat integral calculus. The main reference for this chapter is [2].

1.2.1 Derivative of a function

Given a real function \( f \), which maps a real number to another real number, we are often interested in the slope of the graph of \( f \) at some point \( x_0 \). With the slope of \( f \), we mean the slope of the tangent at the graph of \( f \) through \( f(x_0) \) (see fig. 1.5). How can we calculate this slope? For this we first think about how to calculate the slope of the secant through \( f(x_0) \) and \( f(x_0 + h) \) for some \( h > 0 \) (see. fig. 1.5): This slope is simply the difference in height divided by the difference in length, hence

\[
\frac{f(x_0 + h) - f(x)}{(x_0 + h) - x_0} = \frac{f(x_0 + h) - f(x)}{h}.
\]

If we make \( h \) smaller, the secant approaches the tangent and the slope of the secant approaches the slope of the tangent. Therefore the slope of the tangent is the limit of the
slope of the secant, when $h$ goes to 0, written as $\lim_{h \to 0}$. We call the slope of the tangent of $f$ at the point $x_0$ the derivative $f'(x_0)$ of $f$ at the point $x_0$.

**Definition:** Let $f$ be a real function and $x_0$ a real number. Then we define the derivative of $f$ at the point $x_0$ as

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}. \quad (1.1)$$

**Remarks**

- One often also writes $\frac{df}{dx}(x_0)$ for $f'(x_0)$. Then $df$ and $dx$ stand for very small ("infinitesimal") differences $f(x_0 + h) - f(x_0)$ and $(x_0 + h) - x_0$.
- In physics $f$ is often a function of the time $t$. Then, we often write

$$\dot{f}(t_0) = \frac{df}{dt}(t_0)$$

for the derivative of $f$ at the point $t_0$.
- The limit in the definition of the derivative doesn’t exist for every function. However, we will only work with functions which are "nice enough" and we will always assume the limit exists everywhere.
- One often calls the derivative $f'(x_0)$ the "instantaneous rate of change" of $f$ at the point $x_0$, because it tells how fast $f$ is changing at $x_0$. With the help of the derivative, we can approximate $f$ near $x_0$: For $\Delta x$ small, we have

$$f(x_0 + \Delta x) \approx f(x_0) + \Delta x \cdot f'(x_0), \quad (1.2)$$

since $f(x_0) + \Delta x \cdot f'(x_0)$ is the value of the tangent line at the point $x_0 + h$ (see fig. 1.6).

**Example:** We calculate the derivative of the function $f(x) = x^2$ at the point $x_0 = 1$. So we calculate

$$f'(1) = f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{h \to 0} \frac{(1 + h)^2 - 1}{h} = \lim_{h \to 0} \frac{1 + 2h + h^2 - 1}{h} = \lim_{h \to 0} \frac{2h + h^2}{h} = \lim_{h \to 0} (2 + h) = 2.$$
1.2. DIFFERENTIAL CALCULUS

Figure 1.6: Derivative as approximation

Mostly, we do not only want to know the derivative of \( f \) at some point \( x_0 \) (as in the example), but generally at an arbitrary point. We define the derivative function of \( f \) (mostly, we only say derivative of \( f \)) as the function \( f' = \frac{df}{dx} \), which maps every real number \( x \) to the derivative \( f'(x) \) of \( f \) at the point \( x \). Therefore the derivative \( f' \) of a function \( f \) is again a function.

We also say "differentiate" instead of "calculating the derivative".

**Example:** We calculate the derivative function of \( f(x) = x^2 \). We proceed just as before, but we write \( x \) instead of \( x_0 = 1 \):

\[
f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h \to 0} \frac{x^2 + 2hx + h^2 - x^2}{h} = \lim_{h \to 0} \frac{2hx + h^2}{h} = \lim_{h \to 0} (2x + h) = 2x.
\]

So the derivative of \( f(x) = x^2 \) is \( f'(x) = 2x \).

One can show analogously, that for a positive integer \( n \), the derivative of \( f(x) = x^n \) is the function \( f'(x) = n \cdot x^{n-1} \). For example for \( f(x) = x^25 \), we have \( f'(x) = 25 \cdot x^{24} \), or for \( g(x) = x \) we have \( g'(x) = 1 \) (this can be verified immediately, since the slope of \( g(x) = x \) is equal to 1 everywhere).
1.2.2 Differentiation rules

Often it is not necessary to calculate the derivative explicitly using the limit (1.1), since often the function is a combination of simpler functions, whose derivatives are already known. In this chapter we look at the corresponding differentiation rules.

**Factor rule:** Let $s$ be a real number and $g$ a function. If $f(x) = s \cdot g(x)$, then the derivative $f'(x) = s \cdot g'(x)$.

**Sum rule:** Let $g$ and $k$ be functions. If $f(x) = g(x) + k(x)$, then $f'(x) = g'(x) + k'(x)$.

**Proof:**
- Let $f(x) = s \cdot g(x)$. Then we have
  \[
  f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = \lim_{h \to 0} \frac{s \cdot g(x + h) - s \cdot g(x)}{h} = s \cdot \lim_{h \to 0} \frac{g(x + h) - g(x)}{h} = s \cdot g'(x).
  \]
- Let $f(x) = g(x) + k(x)$. Then we have
  \[
  f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = \lim_{h \to 0} \frac{g(x + h) + k(x + h) - (g(x) + k(x))}{h} = \lim_{h \to 0} \frac{g(x + h) - g(x)}{h} + \lim_{h \to 0} \frac{k(x + h) - k(x)}{h} = g'(x) + k'(x).
  \]

**Example:** We calculate the derivative of $f(x) = 3x^4 - 2x^2$. We set $g(x) = 3x^4$ und $k(x) = -2x^2$, so we have $f(x) = g(x) + k(x)$. Using the factor rule (and the rule for derivatives of powers in the last section), we get $g'(x) = 3 \cdot 4 \cdot x^3 = 12x^3$ and $k'(x) = (-2) \cdot 2 \cdot x = -4x$. Using the sum rule, we get $f'(x) = g'(x) + k'(x) = 12x^3 - 4x$. 


We are now able to differentiate sums (and also differences using the factor rule). We also want to differentiate products and quotients. The first guess \((g(x) \cdot k(x))' = g'(x) \cdot k'(x)\) can be proved wrong, for example by looking at \(g(x) = k(x) = x\). The correct rules are:

**Product rule:** Let \(g\) and \(k\) be functions. If \(f(x) = g(x) \cdot k(x)\), then we have

\[
f'(x) = g'(x) \cdot k(x) + g(x) \cdot k'(x).
\]

**Quotient rule:** Let \(g\) and \(k\) be functions. If \(f(x) = \frac{g(x)}{k(x)}\), then we have

\[
f'(x) = \frac{g'(x) \cdot k(x) - g(x) \cdot k'(x)}{(k(x))^2}.
\]

**Proof:**

- Let \(f(x) = g(x) \cdot k(x)\). Then we have
  \[
  f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{g(x+h) \cdot k(x+h) - g(x) \cdot k(x)}{h} = \lim_{h \to 0} \frac{g(x+h) \cdot k(x+h) - g(x+h) \cdot k(x) + g(x+h) \cdot k(x) - g(x) \cdot k(x)}{h} = \lim_{h \to 0} \frac{k(x) \cdot (g(x+h) - g(x)) + g(x+h) \cdot (k(x+h) - k(x))}{h} = k(x) \cdot \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} + \lim_{h \to 0} g(x+h) \cdot \lim_{h \to 0} \frac{k(x+h) - k(x)}{h} = g'(x) \cdot k(x) + g(x) \cdot k'(x).
  \]

- The quotient rule follows from the product rule: Let \(f(x) = \frac{g(x)}{k(x)}\). Then we have \(g(x) = f(x) \cdot k(x)\). Therefore by the product rule \(g'(x) = f'(x) \cdot k(x) + f(x) \cdot k'(x)\). If we solve for \(f'(x)\) and plug in \(f(x) = \frac{g(x)}{k(x)}\), we find
  \[
  f'(x) = \frac{g'(x) - f(x) \cdot k'(x)}{k(x)} = \frac{g'(x) - \frac{g(x)}{k(x)} \cdot k'(x)}{k(x)} = \frac{g'(x) \cdot k(x) - g(x) \cdot k'(x)}{(k(x))^2}.
  \]
Example:

1. We calculate the derivative of $f(x) = \frac{x-1}{x+1}$. We set $g(x) = x - 1$ and $k(x) = x + 1$. Then we have $f(x) = \frac{g(x)}{k(x)}$. Furthermore $g'(x) = 1$ and $k'(x) = 1$. Therefore by the quotient rule:

\[
 f'(x) = \frac{g'(x) \cdot k(x) - g(x) \cdot k'(x)}{(k(x))^2} = \frac{1 \cdot (x + 1) - (x - 1) \cdot 1}{(x + 1)^2} = \frac{2}{(x + 1)^2}
\]

2. For a positive integer $n$, let be $f(x) = \frac{1}{x^n} = x^{-n}$. Then we have by the quotient rule

\[
 f'(x) = \frac{(1)' \cdot x^n - 1 \cdot (x^n)'}{(x^n)^2} = \frac{0 \cdot x^n - 1 \cdot n \cdot x^{n-1}}{x^{2n}} = -n \cdot x^{-n-1} = -n \cdot \frac{1}{x^{n+1}}
\]

Especially, we have found that the rule $(x^n)' = n \cdot x^{n-1}$ also holds for negative $n$.

We are now able to differentiate all polynomial functions and all fractions of polynomial functions. However, functions often appear as compositions of two functions. For example the function $f(x) = (x^2 - 3x + 13)^{17}$, can be written as $f(x) = u(v(x))$, where $u(y) = y^{17}$ and $v(x) = x^2 - 3x + 13$. Theoretically, we could calculate $f'$ by expanding $(x^3 - 3x + 13)^{17}$ and applying the above rules, but that would not be a very satisfying solution. We now formulate the chain rule, which deals with such compositions.

Chain rule: Let $u$ and $v$ be functions and $f(x) = u(v(x))$. Then

\[
 f'(x) = u'(v(x)) \cdot v'(x).
\]

It is easy to remember the chain rule, using the notation $f' = \frac{df}{dx}$. We formally "expand the fraction" and get

\[
 f' = \frac{df}{dx} = \frac{du(v)}{dx} = \frac{du(v)}{dv} \cdot \frac{dv}{dx} = u'(v) \cdot v'
\]
Proof: Let \( f(x) = u(v(x)) \). Then we have

\[
  f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = \lim_{h \to 0} \frac{u(v(x + h)) - u(v(x))}{h} = \lim_{h \to 0} \left( \frac{u(v(x + h)) - u(v(x))}{v(x + h) - v(x)} \cdot \frac{v(x + h) - v(x)}{h} \right) = \lim_{h \to 0} \frac{u(v(x + h)) - u(v(x))}{v(x + h) - v(x)} \cdot \lim_{h \to 0} \frac{v(x + h) - v(x)}{h} = \frac{u'(v(x))}{v'(x)}.
\]

We now define \( \hat{h} = v(x + h) - v(x) \). For \( h \to 0 \), we also have \( \hat{h} \to 0 \) (because we assumed \( v \) to be "nice enough"). Therefore

\[
  \lim_{h \to 0} \frac{u(v(x + h)) - u(v(x))}{v(x + h) - v(x)} = \frac{u(v(x) + (v(x + h) - v(x))) - u(v(x))}{v(x + h) - v(x)} = \lim_{h \to 0} \frac{u(v(x)) + \hat{h} - u(v(x))}{\hat{h}} = u'(v(x)).
\]

We get \( f'(x) = u'(v(x)) \cdot v'(x) \).

Example: We calculate the derivative of \( f(x) = (x^2 - 3x + 13)^7 \). We have \( f(x) = u(v(x)) \), where \( u(y) = y^7 \) and \( v(x) = x^2 - 3x + 13 \). We differentiate \( u \) and \( v \) and get \( u'(y) = 17y^6 \) and \( v'(x) = 2x - 3 \). Therefore \( f'(x) = u'(v(x)) \cdot v'(x) = 17 \cdot (x^2 - 3x + 13)^6 \cdot (2x - 3) \).

1.2.3 More derivatives

We want to differentiate more functions. We start with the trigonometric functions sine, cosine, and tangent. We will always measure angles in radians and not in degrees. The graphs of sine and cosine are presented in fig. 1.7.

We now have a closer look at the graph of the sine function \( f(x) = \sin(x) \) (see. fig. 1.8). The slope at the point \( 0 \) seems to be approximately \( f'(0) \approx 1 \). At the point \( \frac{\pi}{2} \), the sine function obtains its maximum, therefore the derivative is \( f' \left( \frac{\pi}{2} \right) = 0 \). At the point \( \pi \), the derivative is again approximately \( f'(\pi) \approx -1 \), at the point \( \frac{3\pi}{2} \) it is again \( f' \left( \frac{3\pi}{2} \right) = 0 \), etc. We can draw the derivative function (see fig. 1.8). If we compare the figures 1.7 und 1.8, we conjecture that the derivative of the sine function is \( f'(x) = \cos(x) \).

\[ \text{Repetition: The magnitude of an angle in radians is the length of the corresponding arc of the unit circle. For example 360° correspond to } 2\pi \text{ in radians, 180° correspond to } \pi. \text{ In general: } y^\circ \text{ corresponds to } \frac{y^\circ \cdot 2\pi}{360^\circ} = \frac{y^\circ}{180^\circ} \cdot \pi \text{ in radians. } \]
One can indeed prove this, but we don’t do this here, since the proof is quite technical. Analogously, one can find the derivative of \( g(x) = \cos(x) \): It is \( g'(x) = -\sin(x) \).

For the tangent, we use

\[
\tan(x) = \frac{\sin(x)}{\cos(x)}
\]

and apply the quotient rule. Then we have (where we use that \( \sin^2(x) + \cos^2(x) = 1 \)):

\[
\tan'(x) = \frac{\sin'(x) \cdot \cos(x) - \sin(x) \cdot \cos'(x)}{(\cos(x))^2}
= \frac{\cos(x) \cdot \cos(x) - \sin(x) \cdot (-\sin(x))}{\cos^2(x)}
= \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)} = \frac{1}{\cos^2(x)}.
\]
Alternatively we can also write this differently and get:

\[ \tan'(x) = \frac{\cos^2(x) + \sin^2(x)}{\cos^2(x)} = 1 + \frac{\sin^2(x)}{\cos^2(x)} = 1 + \tan^2(x). \]

Finally, we have

\[ \tan'(x) = \frac{1}{\cos^2(x)} = 1 + \tan^2(x). \]

To summarize:

**Derivatives of trigonometric functions:**

\[
\begin{align*}
\sin'(x) &= \cos(x) \\
\cos'(x) &= -\sin(x) \\
\tan'(x) &= \frac{1}{\cos^2(x)} = 1 + \tan^2(x)
\end{align*}
\]

**Example:** We calculate the derivative of \( f(x) = \sin(x^2) + (\sin(x))^2 \) For \( g(x) = \sin(x^2) \) we get using the chain rule:

\[ g'(x) = \sin'(x^2) \cdot 2x = 2x \cdot \cos(x^2). \]

for \( k(x) = (\sin(x))^2 \) we get also using the chain rule

\[ k'(x) = 2 \cdot \sin(x) \cdot \sin'(x) = 2 \sin(x) \cdot \cos(x). \]

Using the sum rule, we finally get

\[ f'(x) = g'(x) + k'(x) = 2x \cdot \cos(x^2) + 2 \sin(x) \cdot \cos(x). \]

**Exponential function** We now want to find the derivative of the exponential function

\[ f(x) = a^x, \]

where we should have \( a > 0 \). We start with the limit (1.1). Using the power rules, we get:

\[
\begin{align*}
f'(x) &= \lim_{h \to 0} \frac{f(x + h) - f(x)}{h} = \lim_{h \to 0} \frac{a^{x+h} - a^x}{h} \\
&= \lim_{h \to 0} \frac{a^x \cdot a^h - a^x}{h} = \lim_{h \to 0} \left( a^h - 1 \right) \cdot \frac{a^x}{h} \\
&= \left( \lim_{h \to 0} \frac{a^h - 1}{h} \right) \cdot a^x \quad (1.3)
\end{align*}
\]
We see, that the limit
\[ \lim_{h \to 0} \frac{a^h - 1}{h} \] (1.4)
doesn’t depend on \( x \) anymore, so it is simply a number.

*Exercise:* What is the geometric meaning of the limit (1.4)?

We want to understand this limit better and calculate it approximately for \( a = 2, 3 \), by plugging in small numbers for \( h \) (see Table 1.1). For \( a = 2 \), the limit (1.4) is smaller than 1

<table>
<thead>
<tr>
<th></th>
<th>( h = 0.1 )</th>
<th>( h = 0.01 )</th>
<th>( h = 0.001 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{2^h - 1}{h} )</td>
<td>0.717...</td>
<td>0.695...</td>
<td>0.693...</td>
</tr>
<tr>
<td>( \frac{3^h - 1}{h} )</td>
<td>1.161...</td>
<td>1.104...</td>
<td>1.099...</td>
</tr>
</tbody>
</table>

Table 1.1: Calculation of the limit (1.4) for \( a = 2, 3 \)

and for \( a = 3 \) it is bigger 1. Therefore there exists a number \( e \), with \( 2 < e < 3 \), such that
\[ \lim_{h \to 0} \frac{e^h - 1}{h} = 1 \]

We therefore have by equation (1.3), that
\[ \frac{d}{dx} e^x = e^x. \]

The derivative of \( e^x \) is again \( e^x \). One can calculate that
\[ e = 2.7182... \]

The number \( e \) is called *Euler’s number* and plays an important role in mathematics.

**More about \( e \):** We defined \( e \) as the number, that satisfies \( \lim_{h \to 0} \frac{e^h - 1}{h} = 1 \). We want to find a better representation for \( e \). By the definition of the limit, we get for very small \( h \)
\[ \frac{e^h - 1}{h} \approx 1. \]

Therefore \( e^h \approx 1 + h \), respectively
\[ e \approx (1 + h)^{\frac{1}{h}}. \]

If we take the limit for \( h \to 0 \), the \( \approx \) gets again \( \approx \). If we set \( h = \frac{1}{n} \), we find
\[ e = \lim_{n \to \infty} \left( 1 + \frac{1}{n} \right)^n. \]

This is a widely used representation for \( e \) and is often also seen as the definition of \( e \).
We come back to the question about the derivative of $a^x$. For this, we define the natural logarithm
\[ \ln(x) = \log_e(x) \]
This means that the number $u = \ln(x)$ satisfies the equation
\[ e^u = e^{\ln(x)} = x. \]
We can use this as follows: Let again be $a > 0$ and $f(x) = a^x$. With the power rules, we have
\[ f(x) = a^x = (e^{\ln(a)})^x = e^x \ln(a). \]
With the chain rule:
\[ f'(x) = e^x \ln(a) \cdot \ln(a) = (e^x \ln(a))^x \cdot \ln(a) = a^x \cdot \ln(a). \]
The derivative of $a^x$ therefore is $\ln(a) \cdot a^x$.

An interesting result is the derivative of $\ln(x)$. It holds
\[ \frac{d}{dx} \ln(x) = \frac{1}{x}. \]

**Proof:** We use the identity $x = e^{\ln(x)}$. If we differentiate this on both sides, we get using the chain rule
\[ 1 = \frac{d}{dx} x = \frac{d}{dx} e^{\ln(x)} = e^{\ln(x)} \cdot \frac{d}{dx} \ln(x) = x \cdot \frac{d}{dx} \ln(x). \]
Thus
\[ \frac{d}{dx} \ln(x) = \frac{1}{x}. \]

**Example:**

1. We calculate the derivative of $f(x) = 2^{3x}$. Using the chain rule, we get $f'(x) = 3 \cdot \ln(2) \cdot 2^{3x}$.

2. We calculate the derivative of $f(x) = x^r$ for any real number $r$. We notice that
\[ f(x) = x^r = (e^{\ln(x)})^r = e^{r \ln(x)}. \]
We set $u(y) = e^y$ and $v(x) = \ln(x) \cdot r$. Therefore we have $f(x) = u(v(x))$. By the chain rule, we get
\[ f'(x) = u'(v(x)) \cdot v'(x) = e^{\ln(x) \cdot r} \cdot r \cdot \frac{1}{x} = r \cdot \frac{x^r}{x} = r \cdot x^{r-1}. \]
1.2.4 Overview about derivatives

We summarize in table 1.2 the most important functions and their derivatives. They appear very often and one should know them by heart.

<table>
<thead>
<tr>
<th>$f(x)$</th>
<th>$f'(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^r$</td>
<td>$r \cdot x^{r-1}$</td>
</tr>
<tr>
<td>$e^x$</td>
<td>$e^x$</td>
</tr>
<tr>
<td>$a^x$</td>
<td>$\ln(a) \cdot a^x$</td>
</tr>
<tr>
<td>$\ln(x)$</td>
<td>$\frac{1}{x}$</td>
</tr>
<tr>
<td>$\sin(x)$</td>
<td>$\cos(x)$</td>
</tr>
<tr>
<td>$\cos(x)$</td>
<td>$-\sin(x)$</td>
</tr>
<tr>
<td>$\tan(x)$</td>
<td>$\frac{1}{\cos^2(x)} = 1 + \tan^2(x)$</td>
</tr>
</tbody>
</table>

Table 1.2: Derivatives of the most important functions

1.2.5 Higher derivatives

We define the second derivative $f''(x)$ of a function $f(x)$ as the derivative of $f'(x)$. Analogously, we define the third derivative $f'''(x)$ as the derivative of the second derivative $f''(x)$, etc. In general, we define the $n$-th derivative $f^{(n)}(x)$ recursively as the derivative of the $(n-1)$-th derivative $f^{(n-1)}(x)$.

Example:

1. Let $f(x) = e^x$. Then

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

2. Let $g(x) = \sin(x)$. Then

$$g''(x) = \frac{d}{dx}(g'(x)) = \frac{d}{dx}(\cos(x)) = -\sin(x).$$

3. Let $k(x) = \frac{1}{2}x^2$. Then

$$k''(x) = \frac{d}{dx}(k'(x)) = \frac{d}{dx}(x) = 1.$$
1.2.6 Taylor approximation

We have seen at the very beginning (see equation (1.2)) that the derivative of a function $f$ at a point $x_0$ can be used to approximate the function $f$ at $x$ around $x_0$: One just calculates the value of the linear function that is tangent to $f$ at $x_0$, i.e. at $x$ around $x_0$ one approximates

$$f(x) \approx f(x_0) + (x - x_0) \cdot f'(x_0) =: T_{1,f,x_0}(x).$$

This is the best linear approximation of the function $f$ near $x_0$.

This is already a special case (and by far the most important!) of a concept called Taylor approximation. The idea is the following: Why restrict to linear approximations? How about the best quadratic approximation? For this we look for a quadratic function $T_{2,f,x_0}(x)$ such that at $x_0$ it goes through $f(x_0)$ and has the same first and second derivative as $f$ at $x_0$. It is not hard to show that $T_{2,f,x_0}(x)$ is given by

$$T_{2,f,x_0}(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2.$$

Similarly, we can find the polynomial $T_{n,f,x_0}(x)$ of degree $n$, which approximates $f$ best around $x_0$, i.e. the polynomial of degree $n$ such that for all $k \leq n$

$$T_{n,f,x_0}^{(k)}(x_0) = f^{(k)}(x_0).$$

$^3$This is essentially because a quadratic function has three "degrees of freedom", i.e. three parameters can be chosen freely.
**Definition:** This polynomial $T_{n,f,x_0}(x)$ is called Taylor polynomial of degree $n$ of $f$ at $x_0$ and given by

$$T_{n,f,x_0}(x) = f(x_0) + \sum_{k=1}^{n} \frac{1}{k!} f^{(k)}(x_0)(x - x_0)^k$$

$$= f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + \ldots + \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n$$

where $n! = n \cdot (n - 1) \cdot \ldots \cdot 2 \cdot 1$.\(^4\)

Figure 1.10 shows the Taylor polynomials up to degree 5 for the same function $f$ as above.

![Taylor polynomials](image)

Figure 1.10: Taylor polynomials of $f$ (solid line) up to degree 5.

**Example:**

1. Look at $f(x) = e^x$ and $x_0 = 0$. Since $\frac{d}{dx} e^x = e^x$, we have $f^{(n)}(x) = e^x$ for all $n$. Thus $f^{(n)}(0) = e^0 = 1$ for all $n$. Thus

$$T_{n,f,0}(x) = 1 + \sum_{k=1}^{n} \frac{1}{k!} x^k$$

is the best polynomial approximation for $e^x$ of degree $n$ around 0. Most importantly, we have

$$e^x \approx 1 + x$$

for $x$ around 0.
2. Look at \( f(x) = \sin(x) \) and \( x_0 = 0 \). Then \( f'(x) = \cos(x) \), so \( f'(0) = 1 \) and \( f''(x) = -\sin(x) \), so \( f''(0) = 0 \). Thus

\[
T_{2,f,0}(x) = f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 = x,
\]

in particular for \( x \) around 0,

\[
\sin(x) \approx x.
\]

3. Look at \( f(x) = \cos(x) \) and \( x_0 = 0 \). We have \( f(0) = \cos(0) = 1 \). Furthermore \( f'(x) = -\sin(x) \), so \( f'(0) = \sin(0) = 0 \) and \( f''(x) = -\cos(x) \), so \( f''(0) = -\cos(0) = -1 \). Thus

\[
T_{2,f,0}(x) = f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 = 1 - \frac{x^2}{2},
\]

which means that for \( x \) around 0,

\[
\cos(x) \approx 1 - \frac{x^2}{2}.
\]

One can actually state more precisely how accurate the function \( f \) is approximated by \( T_{n,f,x_0} \):

It holds that (if \( f \) satisfies certain conditions\(^5\))

\[
f(x) = T_{n,f,x_0}(x) + \mathcal{O}(|x - x_0|^{n+1}).
\]

\( \mathcal{O}(|x - x_0|^{n+1}) \) means that the approximation error is of order \( |x - x_0|^{n+1} \), i.e. there is a constant \( C > 0 \) such that when \( |x - x_0| \) is small enough,

\[
|f(x) - T_{n,f,x_0}(x)| \leq C \cdot |x - x_0|^{n+1}.
\]

### 1.2.7 The idea of differential equations

Differential equations are functional equations that also contain derivatives of the functions. They are crucial for physics since many physical laws can be formulated using differential equations.

Let us start with an example:

\[
f'(x) = f(x).
\]

This equation means that we are looking for a function \( f \) such that its derivative \( f' \) is equal to \( f \). We know already that \( f(x) = e^x \) is a solution to this differential equation. But actually, for every \( c \in \mathbb{R} \), also \( f(x) = c e^x \) is a solution. So the solution to the above differential equation is not unique. To ensure uniqueness of the solution, we also need to fix the value of \( f \) at 0. If we then present the differential equation as

\[
f'(x) = f(x), \ f(0) = 3,
\]

\(^5\)that we always assume to hold...
(a so called initial value problem) the solution is given by $f(x) = 3e^x$. One can show that this solution is actually the unique solution.

We do not develop the theory of differential equations in this script (There are books on this topic...). We just mention the general result that, if the differential equation does not behave too bad, then for any given initial value there always exists a unique solution.

Example:

1. We slightly generalize our first example. Let $c \in \mathbb{R}$ and $f_0 > 0$, and let

   $f'(x) = cf(x), \ f(0) = f_0.$

   Then the solution is given by

   $f(x) = f_0 e^{cx}.$

2. Look at the initial value problem

   $f'(x) = 1 + f(x)^2, \ f(0) = 0.$

   We already know that

   $f(x) = \tan(x)$

   is a solution.

3. We now look at a second order differential equation (i.e. also second derivatives appear). To make the solution unique, we need two initial conditions, for example on $f(0)$ and $f'(0)$. Let $\alpha > 0$ and $f_0 \in \mathbb{R}$, and let

   $f''(x) = -\alpha^2 f(x), \ f(0) = f_0, \ f'(0) = 0.$

   This is the differential equation corresponding to a harmonic oscillator. The solution to this differential equation is given by

   $f(x) = f_0 \cos(\alpha x).$

4. We generalize this equation a bit: Let $\alpha > 0$ again, and let

   $f'' + 2\alpha f' + \alpha^2 f = 0.$

   Then the solution is given by

   $f(x) = (b_1 + b_2 t)e^{-\alpha t},$

   where $b_1$ and $b_2$ are chosen according to the initial conditions. This corresponds to the case of critical damping of a harmonic oscillator.
1.3 Integral calculus

Roughly speaking, integral calculus deals with the area under the graphs of functions. We will see that differential and integral calculus are closely related. The main reference for this chapter is [2].

1.3.1 Antiderivatives

**Definition:** An antiderivative of a function \( f \) is a function \( F \), which satisfies

\[
 f(x) = F'(x) = \frac{dF}{dx}(x).
\]

Calculating an antiderivative is therefore the reverse of calculating the derivative. For example an antiderivative of \( f(x) = x^2 \) is the function \( F(x) = \frac{1}{3}x^3 \), since \( F'(x) = \frac{1}{3} \cdot 3 \cdot x^2 = x^2 = f(x) \). However, the antiderivative \( F \) isn’t unique. For example the function \( G(x) = \frac{1}{3}x^3 + 42 \) also satisfies \( G'(x) = f(x) \).

In general: If \( f \) is a function and \( F \) an antiderivative of \( f \), then for any real number \( c \), also \( F(x) + c \) is an antiderivative of \( f \). This follows because of

\[
 \frac{d}{dx}(F(x) + c) = \frac{d}{dx}F(x) + 0 = F'(x) = f(x).
\]

It even holds that for two arbitrary antiderivatives \( F \) and \( G \) of \( f \) there exists a constant \( c \) such that

\[
 G(x) = F(x) + c.
\]

(If \( F \) and \( G \) are antiderivatives of \( f \), then the derivative of the difference \( F(x) - G(x) \) is

\[
 \frac{d}{dx}(F(x) - G(x)) = \frac{d}{dx}F(x) - \frac{d}{dx}G(x) = f(x) - f(x) = 0.
\]

Therefore \( F(x) - G(x) \) is constant.)

Antiderivatives of many functions can be easily calculated. We can essentially interchange the two columns of table 1.2 and get table 1.3. In the table for each function there is only given one antiderivative. One obtains all the other antiderivatives by adding an arbitrary constant \( c \). We have a factor rule and a sum rule for antiderivatives. They are the reversal of the corresponding rules for derivatives.

**Factor rule:** Let \( f \) be a function with antiderivative \( F \) and let \( s \) be a real number. Then \( s \cdot F \) is an antiderivative of \( s \cdot f \).
### Table 1.3: Antiderivatives of the most important functions

<table>
<thead>
<tr>
<th>$f(x)$</th>
<th>$F(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^r$ für $r \neq -1$</td>
<td>$\frac{1}{r+1} \cdot x^{r+1}$</td>
</tr>
<tr>
<td>$\frac{1}{x}$</td>
<td>$\ln</td>
</tr>
<tr>
<td>$e^x$</td>
<td>$e^x$</td>
</tr>
<tr>
<td>$a^x$</td>
<td>$\frac{a^x}{\ln(a)}$</td>
</tr>
<tr>
<td>$\sin(x)$</td>
<td>$-\cos(x)$</td>
</tr>
<tr>
<td>$\cos(x)$</td>
<td>$\sin(x)$</td>
</tr>
</tbody>
</table>

**Sum rule:** Let $f$ and $g$ be functions with antiderivatives $F$ and $G$. Then $F + G$ is an antiderivative $f + g$.

### 1.3.2 Integral as an area

We define the integral of a function $f$ as "signed area" between the graph of $f$ and the $x$-axis.

**Definition:** Let $f$ be a function and let $a$ and $b$ be real numbers with $a \leq b$. Then the integral of $f$ from $a$ to $b$

$$\int_a^b f(x)dx$$

is the area between the graph of $f$ and the $x$-axis in the region between $a$ and $b$, where the area under the $x$-axis counts negative (see fig. 1.11).

So the integral is a real number.

In the example in figure 1.11, the integral is

$$\int_a^b f(x)dx = A - B + C.$$
Example: We calculate
\[ \int_{1}^{2} f(x) \, dx \]
for the function \( f(x) = x \). We look at figure 1.12 and see, that the area under the graph between 1 and 2 is a trapezoid with height \((2 - 1)\) and side lengths 1 and 2.

Therefore the area of this trapezoid is
\[
\int_{1}^{2} x \, dx = \frac{1 + 2}{2} \cdot (2 - 1) = \frac{3}{2}
\]
For \( a > b \), we define
\[
\int_a^b f(x) \, dx = -\int_b^a f(x) \, dx.
\]
In this case, the area above the \( x \)-axis counts negative and the area under the \( x \)-axis positive.

For arbitrary \( a, b \) and \( c \), we then have
\[
\int_c^a f(x) \, dx = \int_a^b f(x) \, dx + \int_b^c f(x) \, dx.
\] (1.5)

This follows because one can simply add the corresponding areas. This property is called "interval additivity".

**Approximation of integrals using Riemann sums**

Let us mention a useful way to approximate integrals, which can be generalized later. One can calculate the integral \( \int_a^b f(x) \, dx \) of a function \( f \) by approximating it using so-called Riemann sums: One divides the interval \([a, b]\) into a sequence of points \( a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b \) and calculates the area of the rectangles with width \( \Delta x_k := x_k - x_{k-1} \) and height \( f(x_k) \) for all \( k = 1, \ldots, n \) (see figure 1.13). The sum of the areas of the rectangles is
\[
\sum_{k=1}^{n} f(x_k) \cdot (x_k - x_{k-1}) = \sum_{k=1}^{n} f(x_k) \cdot \Delta x_k.
\]

![Figure 1.13: Approximation of integral with sums](image-url)
If one makes the partition \( a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b \) finer (i.e. increases \( n \)) then the sum of the areas of the rectangles will converge to the integral

\[ \int_a^b f(x) \, dx. \]

The notation of the integral is motivated by this approximation: In the limit, the \( \sum \) becomes an \( \int \) and the \( \Delta x_k \) becomes a \( dx \).

### 1.3.3 Fundamental theorem of calculus

In this section, we are going to connect antiderivatives and integrals. The idea is to let vary the upper limit of the integral. In this way, one gets a function of the upper limit of the integral. Let \( f \) be a function and \( a \) an arbitrary number. Then we define

\[ I_{f,a}(x) = \int_a^x f(t) \, dt. \]

Since \( x \) already appears in the limits of the integrals, we have to take another integration variable \( t \). The expression \( I_{f,a}(x) \) is independent of \( t \), which is just a "dummy" variable. Of course, we have for any real number \( b \) that

\[ I_{f,a}(b) = \int_a^b f(t) \, dt = \int_a^b f(x) \, dx. \]

**Example:** Let \( f(x) = x \). We calculate \( I_{f,1} \). We can proceed analogously as in the example above, we only have to replace 2 by \( x \). Then we have

\[ I_{f,1}(x) = \int_1^x f(t) \, dt = \int_1^x t \, dt = \frac{1 + x}{2} \cdot (x - 1) = \frac{x^2 - 1}{2}. \]

If we differentiate \( I_{f,1}(x) = \frac{x^2-1}{2} \) with respect to \( x \), we get

\[ I'_{f,1}(x) = \frac{d}{dx} \left( \frac{x^2-1}{2} \right) = x = f(x). \]

This is not a coincidence, but it’s exactly the statement of the fundamental theorem of calculus:

**Fundamental theorem of calculus:** Let \( f \) be a function and \( a \) a real number. Then \( I_{f,a} \) is an antiderivative of \( f \). Concretely, this means:

\[ I'_{f,a}(x) = \frac{d}{dx} (I_{f,a}(x)) = \frac{d}{dx} \left( \int_a^x f(t) \, dt \right) = f(x). \]
Proof: We want to calculate

\[ I'_f,a(x) = \lim_{h \to 0} \frac{I_{f,a}(x + h) - I_{f,a}(x)}{h}. \]

For this, we look at \( I_{f,a}(x + h) - I_{f,a}(x) \) more closely. Using the interval additivity (1.5), we have

\[ I_{f,a}(x + h) - I_{f,a}(x) = \int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt = \int_x^{x+h} f(t)\,dt. \quad (1.6) \]

We look at Figure 1.14. Let \( f_{h,\text{min}} \) be the minimal value of \( f \) between \( x \) and \( x + h \), and let \( f_{h,\text{max}} \) be the maximal value of \( f \) between \( x \) and \( x + h \). Then we surely have

\[ h \cdot f_{h,\text{min}} \leq \int_x^{x+h} f(t)\,dt \leq h \cdot f_{h,\text{max}}. \]

If we divide by \( h \), we get

\[ f_{h,\text{min}} \leq \int_x^{x+h} f(t)\,dt \leq f_{h,\text{max}}. \]

Using equation (1.6), we have

\[ f_{h,\text{min}} \leq \frac{I_{f,a}(x + h) - I_{f,a}(x)}{h} \leq f_{h,\text{max}}. \]

If we let go \( h \to 0 \), then \( f_{h,\text{min}} \to f(x) \) and \( f_{h,\text{max}} \to f(x) \), hence

\[ f(x) \leq \lim_{h \to 0} \frac{I_{f,a}(x + h) - I_{f,a}(x)}{h} \leq f(x). \]

Finally, we get

\[ I'_f,a(x) = \lim_{h \to 0} \frac{I_{f,a}(x + h) - I_{f,a}(x)}{h} = f(x) \]

and therefore \( I_{f,a} \) is an antiderivative of \( f \).
1.3. INTEGRAL CALCULUS

The fundamental theorem of calculus gives us a useful tool to calculate integrals. We want to know the value of the integral

\[ \int_a^b f(x) \, dx = I_{f,a}(b). \]

Let \( F \) be an antiderivative of \( f \). From the fundamental theorem of calculus, we know that \( I_{f,a}(x) \) also is an antiderivative of \( f(x) \). Therefore there exists a real number \( c \) such that

\[ I_{f,a}(x) = F(x) + c. \]

But then we have

\[ F(b) - F(a) = I_{f,a}(b) + c - (I_{f,a}(a) + c) = I_{f,a}(b) - I_{f,a}(a) = I_{f,a}(b), \]

since

\[ I_{f,a}(a) = \int_a^a f(x) \, dx = 0. \]

Hence we get

\[ \int_a^b f(x) \, dx = I_{f,a}(b) = F(b) - F(a). \] (1.7)

We emphasize again, that the formula (1.7) does not depend on the choice of the antiderivative.

For \( F(b) - F(a) \), we also write

\[ [F(x)]_a^b = F(b) - F(a) \]

**Example:**

1. We calculate

\[ \int_1^2 x \, dx. \]

An antiderivative of \( f(x) = x \) is given by \( F(x) = \frac{1}{2}x^2 \). So using equation (1.7),

\[ \int_1^2 x \, dx = \left[ \frac{1}{2}x^2 \right]_1^2 = \frac{1}{2} \cdot 2^2 - \frac{1}{2} \cdot 1^2 = 3 - \frac{3}{2} = \frac{3}{2}. \]

We indeed get the same result as above.

2. We calculate

\[ \int_0^1 x^2 \, dx. \]

An antiderivative of \( f(x) = x^2 \) is given by \( F(x) = \frac{1}{3}x^3 \). So we get

\[ \int_0^1 x^2 \, dx = \left[ \frac{1}{3}x^3 \right]_0^1 = \frac{1}{3} \cdot 1^3 - \frac{1}{3} \cdot 0^3 = \frac{1}{3}. \]
3. We calculate \[ \int_0^\pi \sin(x) \, dx. \]

An antiderivative of \( \sin(x) \) is given by \( -\cos(x) \). Therefore,
\[
\int_0^\pi \sin(x) \, dx = [ -\cos(x) ]_0^\pi = -\cos(\pi) - (-\cos(0)) = -(-1) + 1 = 2.
\]

The fundamental theorem of calculus motivates the following notation: Let \( f \) be a function. Then we also write
\[
\int f(x) \, dx
\]
for an arbitrary antiderivative of \( f \). Note that this notation is not really mathematically correct, since \( \int f(x) \, dx \) is not unique, but it should always be clear from the context, what is meant.

1.3.4 More integration rules

Let \( f \) be a function. If we are given an antiderivative of \( f \), it is in general easy to check that \( F \) is really an antiderivative of \( f \), since we just have to calculate the derivative \( F' \). But it can be very difficult to find an antiderivative.\(^6\) In this section we will present more methods to calculate antiderivatives in certain situations. The first one is the method of integration by parts. Integration by parts essentially is the opposite of the product rule. The second one is integration by substitution, which is the opposite of the chain rule.

**Integration by parts** We start with the product rule. Let \( u \) and \( v \) be two functions and \( f(x) = u(x) \cdot v(x) \). Then the product rule says that
\[
f'(x) = \frac{d}{dx} (u(x) \cdot v(x)) = u'(x) \cdot v(x) + u(x) \cdot v'(x).
\]

If we calculate antiderivatives on both sides, we get
\[
u(x) \cdot v(x) = \int (u'(x) \cdot v(x) + u(x) \cdot v'(x)) \, dx
\]
\[= \int u'(x) \cdot v(x) \, dx + \int u(x) \cdot v'(x) \, dx.
\]

\(^6\)There even exist functions, where it is impossible: The function \( f(x) = e^{-x^2} \) does not have an antiderivative which can be expressed by the usual functions.
This implies

\[ \int u'(x) \cdot v(x) \, dx = u(x) \cdot v(x) - \int u(x) \cdot v'(x) \, dx. \]  

(1.8)

This equation looks very abstract at first glance and will now be clarified by examples.

**Example:**

1. We calculate an antiderivative of \( e^x \cdot x \). Set \( u(x) = e^x \) and \( v(x) = x \). Then

\[ e^x \cdot x = u'(x) \cdot v(x). \]

Using equation (1.8) an antiderivative of \( e^x \cdot x \) is given by

\[ \int e^x \cdot x \, dx = \int u'(x) \cdot v(x) \, dx \]
\[ = u(x) \cdot v(x) - \int u(x) \cdot v'(x) \, dx \]
\[ = e^x \cdot x - \int e^x \cdot 1 \, dx \]
\[ = e^x \cdot x - e^x. \]

If we differentiate \( e^x \cdot x - e^x \), we see that it is indeed an antiderivative of \( e^x \cdot x \).

2. We calculate an antiderivative of \( \sin^2(x) \). We set \( u(x) = -\cos(x) \) and \( v(x) = \sin(x) \). Then \( u'(x) = \sin(x) \) and therefore

\[ \sin^2(x) = u'(x) \cdot v(x). \]

Using equation (1.8), we calculate

\[ \int \sin^2(x) \, dx = \int u'(x) \cdot v(x) \, dx \]
\[ = u(x) \cdot v(x) - \int u(x) \cdot v'(x) \, dx \]
\[ = -\cos(x) \cdot \sin(x) - \int (-\cos(x)) \cdot \cos(x) \, dx \]
\[ = -\cos(x) \cdot \sin(x) + \int \cos^2(x) \, dx. \]

We now use

\[ \cos^2(x) = 1 - \sin^2(x) \]
and get
\[
\int \sin^2(x) \, dx = -\cos(x) \cdot \sin(x) + \int \cos^2(x) \, dx \\
= -\cos(x) \cdot \sin(x) + \int (1 - \sin^2(x)) \, dx \\
= -\cos(x) \cdot \sin(x) + \int 1 \, dx - \int \sin^2(x) \, dx \\
= -\cos(x) \cdot \sin(x) + x - \int \sin^2(x) \, dx.
\]

Now we solve for \(\int \sin^2(x) \, dx\) and get
\[
\int \sin^2(x) \, dx = \frac{1}{2} \cdot (-\cos(x) \cdot \sin(x) + x) \\
= \frac{x}{2} - \frac{\cos(x) \cdot \sin(x)}{2}.
\]

Again, we can check by differentiating that this is indeed an antiderivative of \(\sin^2(x)\).

3. We calculate an antiderivative of \(\ln(x)\). This does not look like a case for integration by parts, but we can write
\[
\ln(x) = 1 \cdot \ln(x).
\]

Then we define \(u(x) = x\) and \(v(x) = \ln(x)\). Now we have (since \(v'(x) = \frac{1}{x}\))
\[
\ln(x) = u'(x) \cdot v(x).
\]

Using equation (1.8) and \(v'(x) = \frac{1}{x}\) we get
\[
\int \ln(x) \, dx = \int 1 \cdot \ln(x) \, dx = \int u'(x) \cdot v(x) \, dx \\
= u(x) \cdot v(x) - \int u(x) \cdot v'(x) \, dx \\
= x \cdot \ln(x) - \int x \cdot \frac{1}{x} \, dx \\
= x \cdot \ln(x) - \int 1 \, dx \\
= x \cdot \ln(x) - x.
\]
Integration by substitution  The second method we look at is integration by substitution. This is essentially the converse of the chain rule. We start with an example: We want to calculate an antiderivative of

\[ f(x) = 2x \cdot e^{x^2}. \]

We can write \( f \) as

\[ f(x) = k'(x) \cdot g'(k(x)), \]

where \( k(x) = x^2 \) and \( g(y) = e^y \). With the chain rule, we have

\[ f(x) = k'(x) \cdot g'(k(x)) = \frac{d}{dx} g(k(x)). \]

Therefore by definition of an antiderivative,

\[ F(x) = g(k(x)) = e^{x^2} \]

is an antiderivative of \( f \). This is actually the whole idea of integration by substitution.

Integration by substitution (version 1): Let \( u \) and \( v \) be functions and let \( U \) be an antiderivative of \( u \). Then \( U(v(x)) \) is an antiderivative of \( u(v(x)) \cdot v'(x) \). Hence

\[ \int u(v(x)) \cdot v'(x) dx = U(v(x)). \]

One checks that \( U(v(x)) \) is really an antiderivative of \( u(v(x)) \cdot v'(x) \) by differentiating using the chain rule. The simplest application of integration by substitution is the case where \( v \) is of the form

\[ v(x) = ax + b. \]

If \( u \) is a function and \( U \) is an antiderivative of \( u \), then an antiderivative of \( u(ax + b) \) is given by

\[ \int u(ax + b) \, dx = \frac{1}{a} \int u(ax + b) \cdot a \, dx \]
\[ = \frac{1}{a} U(ax + b). \]

Remark  Integration by substitution can be remembered and applied using a simple trick: We simply pretend that we can handle \( dv, \, dx \) and \( \frac{dv}{dx} \) just as normal variables. Then we can do the following formal\(^7\) calculation. Write \( v'(x) = \frac{dv}{dx} \). Then "\( v'(x)dx = dv\)", so

\[ \int u(v(x)) \cdot v'(x) dx = \int u(v) dv = U(v(x)). \]

---

\(^7\)Note that here "formal" means that we write things that are not properly defined.
Example:

1. We calculate an antiderivative of $e^{3x-2}$:

$$\int e^{3x-2} \, dx = \frac{1}{3} \cdot e^{3x-2}.$$

2. We calculate an antiderivative of $\tan(x)$. For this we write

$$\tan(x) = \frac{\sin(x)}{\cos(x)} = -\frac{\sin(x)}{\cos(x)}.$$

Set $v(x) = \cos(x)$. Then $\frac{dv}{dx} = -\sin(x)$, so we write

$$-\sin(x) \, dx = dv.$$

$$\int \tan(x) \, dx = -\int \frac{1}{v(x)} \cdot (-\sin(x)) \, dx$$

$$= -\int \frac{1}{v} \, dv = -\ln(|v|) = -\ln(|\cos(x)|).$$

Definite integrals: If one calculates definite integrals using the substitution rule, one gets the following formula:

Integration by substitution (version 2): Let $u$ and $v$ be functions and let $a$ and $b$ be real numbers. Then

$$\int_{a}^{b} u(v(x)) \cdot v'(x) \, dx = \int_{v(a)}^{v(b)} u(v) \, dv.$$

Example: We calculate

$$\int_{0}^{\pi/2} \sin(x) \cdot \cos(x) \, dx.$$

Set

$$v(x) = \sin(x).$$
Thus \( dv = \cos(x)dx \), so

\[
\int_{0}^{\frac{\pi}{2}} \sin(x) \cdot \cos(x) dx = \int_{0}^{\frac{\pi}{2}} v(x) \cdot \cos(x) dx = \int_{v(0)}^{v(\frac{\pi}{2})} v dv = \int_{\sin(0)}^{\sin(\frac{\pi}{2})} v dv = \left[ \frac{v^2}{2} \right]_{0}^{1} = \frac{1}{2}.
\]

**Two more elaborate examples** At a first glance, functions for which one can apply integration by substitution seem to have a very special form. But actually, the manipulations seen above can be applied to a wide range of functions, which we want to illustrate on two examples.

1. We want to calculate an antiderivative of

\[
\frac{1}{e^x + 1}.
\]

We set \( v(x) = e^x \). Thus

\[
dv = e^x dx.
\]

Using integration by substitution, we have

\[
\int \frac{1}{e^x + 1} dx = \int \frac{1}{(e^x)^2 + e^x} \cdot e^x dx = \int \frac{1}{v(x)^2 + v(x)} \cdot e^x dx = \int \frac{1}{v^2 + v} dv.
\]

Now we write

\[
\frac{1}{v^2 + v} = \frac{1}{v(v + 1)} = \frac{(v + 1) - v}{v(v + 1)} = \frac{v + 1}{v(v + 1)} - \frac{v}{v(v + 1)} = \frac{1}{v} - \frac{1}{v + 1}.
\]
We can simply integrate this and get that
\[
\int \frac{1}{v^2 + v} \, dv = \ln(|v|) - \ln(|v + 1|).
\]
Therefore
\[
\int \frac{1}{e^x + 1} \, dx = \int \frac{1}{v^2 + v} \, dv
\]
\[
= \ln(|e^x|) - \ln(|e^x + 1|)
\]
\[
= \ln(e^x) - \ln(e^x + 1)
\]
\[
= x - \ln(e^x + 1).
\]
If one differentiates \(x - \ln(e^x + 1)\) with respect to \(x\), one can check that it is indeed an antiderivative of \(\frac{1}{e^x + 1}\).

2. We want to find an antiderivative of
\[
\sqrt{1 - x^2}.
\]
This is a case where one can apply integration by substituion "backwards", i.e. one replaces \(x\) by some function \(x(u)\) depending on some other variable \(u\). With a bit of experience\(^8\), one sees that
\[
x(u) = \cos(u)
\]
could be a good choice for a substitution. Then we have
\[
\sqrt{1 - x^2} = \sqrt{1 - \cos^2(u)} = \sin(u).
\]
Note that \(x = \cos(u)\), so
\[
dx = -\sin(u) \, du.
\]
Using integration by substitution, we get
\[
\int \sqrt{1 - x^2} \, dx = \int \sqrt{1 - \cos^2(u)} \cdot (-\sin(u)) \, du
\]
\[
= \int \sin(u) \cdot (-\sin(u)) \, du
\]
\[
= -\int \sin^2(u) \, du
\]
\[
= -\left( \frac{u}{2} - \cos(u) \cdot \sin(u) \right)
\]
where we used the antiderivative of \(\sin^2(u)\) which we calculated in the section about integration by parts. Now, we have to write this term as a function of \(x\) again. Note that
\[
\sin(u) = \sqrt{1 - \cos^2(u)} = \sqrt{1 - x^2}.
\]
Furthermore, we write
\[ u = \arccos(\cos(u)) = \arccos(x). \]

Therefore
\[
\int \sqrt{1-x^2} \, dx = -\left( \frac{u^2}{2} - \frac{\cos(u) \cdot \sin(u)}{2} \right)
= -\left( \frac{\arccos(x)}{2} - \frac{x \cdot \sqrt{1-x^2}}{2} \right).
\]

We can differentiate this to see that it is indeed an antiderivative of \( \sqrt{1-x^2} \).

### 1.3.5 The idea of multidimensional integrals

The idea of integrals is not restricted to functions of one dimension. It can be generalized to higher dimensions. Let’s illustrate this with an example in three dimensions. Assume we have some object \( C \), at which we look as a subset of the three dimensional space \( \mathbb{R}^3 \). We would like to know the mass of \( C \), but we only know the density \( \rho \) and the volume \( V \) of \( C \). This is easy, since we can just calculate the mass \( m \) of \( C \) from this information:

\[ m = \rho \cdot V. \]

But what if the density \( \rho \) is not constant? Let us first assume that it is at least piecewise constant, i.e. we can divide \( C \) into disjoint pieces \( C_1, ..., C_n \) such that \( C_1 \cup C_2 \cup ... \cup C_n = C \) and, on each of \( C_1, ..., C_n \), the density is constant equal to \( \rho_1, ..., \rho_n \). Furthermore, let \( V_1, ..., V_n \) denote the corresponding volumes. Then the total mass of \( C \) is just the sum of all the masses of \( C_1, ..., C_n \), i.e.:

\[ m = \rho_1 \cdot V_1 + ... + \rho_n \cdot V_n = \sum_{k=1}^{n} \rho_k \cdot V_k. \]

Now let us assume that the density \( \rho \) is actually an arbitrary function of the location \( \vec{x} \) in \( C \), i.e. \( \rho = \rho(\vec{x}) \), where \( \vec{x} \in C \subset \mathbb{R}^3 \). Now we can still approximate the mass of \( C \) by dividing \( C \) into a large number of small disjoint subsets \( \Delta C_1, ..., \Delta C_n \), such that \( \Delta C_1 \cup ... \cup \Delta C_n = C \). Then for each of the \( \Delta C_k \), we choose a location \( \vec{x}_k \in \Delta C_k \). For \( k = 1, ..., n \), let \( \Delta V_k \) denote the volume of \( \Delta C_k \). Then we can approximate the mass \( m \) of \( C \) by

\[ m \approx \sum_{k=1}^{n} \rho(\vec{x}_k) \cdot \Delta V_k \]

where the approximation gets better making the partition finer. This value then converges to the actual mass \( m \) of \( C \). In analogy to the one dimensional case, we write

\[ m = \iiint_C \rho(\vec{x}) dV(\vec{x}). \]
CHAPTER 1. MATHEMATICS

As in the one dimensional case, the intuition is that in the limit the \( \sum \) gets a \( \int \int \int \) and the \( \Delta \) gets a \( d \). If we replace the density \( \rho = \rho(\vec{x}) \) by an arbitrary function \( f = f(\vec{x}) \), we have now already defined

\[
\int \int \int_C f(\vec{x})dV(\vec{x})
\]

for arbitrary\(^{10}\) functions \( f \) from \( \mathbb{R}^3 \) to \( \mathbb{R} \) and subsets \( C \subset \mathbb{R}^3 \).

At the moment, we omit the discussion of how one actually calculates those integrals. It is more important to understand the idea. This idea is not restricted to the three dimensional case. If we replace the volume \( V \) by the area \( A \), we can similarly define the integral

\[
\int \int_C f(\vec{x})dA(\vec{x})
\]

for a subset \( C \subset \mathbb{R}^2 \) and a function \( f = f(\vec{x}) \) from \( \mathbb{R}^2 \) to \( \mathbb{R} \). One can see that this integral is exactly the (signed) Volume between \( C \) and the two dimensional graph of \( f \).

We can generalize this idea even more. For example we can integrate functions on curved two dimensional surfaces in three dimensional space. In the same spirit, we can also integrate functions along one dimensional curves in three dimensional space. The idea stays always the same: One cuts the set on which one wants to integrate into small parts and then assumes the function \( f \) to be constant on those small parts. Then one just sums over all the small parts to approximate the integral.

1.4 Complex Numbers

This Chapter is (except for very small modifications) equal to Chapter 5 in [4], which was written by Lionel Philippoz.

1.4.1 Introduction

Should you encounter the following equation in a textbook

\[ x^2 = 1 \] (1.9)

and be asked to find its solutions in \( \mathbb{R} \), it would not be that difficult to conclude that there are two of them, namely 1 and \(-1\). But what happens if one slightly modifies Eq. (1.9) by changing a sign?

\[ x^2 = -1 \] (1.10)

Can you find a real solution? Actually not, since for any real number \( x \), \( x^2 \geq 0 \). In the real numbers, \( \sqrt{-1} \) is not defined. This equation thus possesses no solutions in \( \mathbb{R} \). However, one can expand the set of real numbers to the so-called set of complex numbers \( \mathbb{C} \) in which Eq. (1.10) actually has two (complex) solutions.

\(^{10}\)We implicitly assume that \( f \) is "nice enough".
1.4.2 Representation of a complex number, Euler formula

Complex numbers are not so different from real numbers, and all you actually need to know is that we define a new number $i \in \mathbb{C}$ such that $i^2 = -1$. And that’s it! You can now solve Eq. (1.10) in $\mathbb{C}$ and find its two solutions: $i$ and $-i$.

Any complex number $z \in \mathbb{C}$ can be written as the sum of two numbers:

$$z = x + iy$$

(1.11)

where $x, y \in \mathbb{R}$ and $i$ as previously defined. $x$ is also called the real part of $z$ (sometimes written as $\Re(z)$ or $Re(z)$), whereas $y$ is called the imaginary part of $z$ (written as $\Im(z)$ or $Im(z)$).

If $y = 0$, then $z$ is simply a real number, and when $x = 0$, we say that $z$ is an imaginary number. As you can see, any complex number can now be defined using two “coordinates” $x$ and $y$, which means it can actually be represented in a plane, the so-called...complex plane!

![Diagram](image.png)

Figure 1.15: $z$ can be seen as a point in the complex plane, with cartesian coordinates $(x, y)$ or polar ones $(|z|, \theta)$.

Another possibility to describe the position of a point in a plane consists in using polar coordinates, where one needs to give the distance from the origin as well as the angle between the $x$-axis (here the real axis) and the line connecting the origin and the point.

If you know your trigonometry well, you can then easily relate both coordinate systems by writing:

$$x = |z| \cos(\theta)$$
$$y = |z| \sin(\theta)$$

If you want to consider $z$ as a vector, then $x$ and $y$ are the components of that vector, $|z|$ the norm of $z$ and $\theta$ the angle between the vector and the $x$-axis.
and \( z \) can thus be written as

\[
\begin{align*}
 z &= x + iy \\
 &= |z| \cos(\theta) + i |z| \sin(\theta) \\
 &= |z| (\cos(\theta) + i \sin(\theta)) \\
 &= |z| e^{i\theta} \\
\end{align*}
\]

(1.12)

The last step is actually performed using the Euler formula:

\[
\begin{align*}
e^{i\theta} &= \cos(\theta) + i \sin(\theta) \\
\end{align*}
\]

(1.13)

which we will take as a definition of a complex exponential function in this script. This is a very important relation which allows one to switch between a trigonometric representation and an exponential function, which is much more easy to handle with. You can actually write both trigonometric functions \( \sin(\theta) \) and \( \cos(\theta) \) as a function of complex exponentials! This is done as follows:

\[
\begin{align*}
e^{i\theta} &= \cos(\theta) + i \sin(\theta) \tag{1.14} \\
e^{-i\theta} &= \cos(-\theta) + i \sin(-\theta) \\
&= \cos(\theta) - i \sin(\theta) \tag{1.15} \\
\end{align*}
\]

If you now add (resp. subtract) Eq. (1.14) and Eq. (1.15), you can eliminate the \( \sin \) part (resp. \( \cos \)) and solve for \( \cos \) (resp. \( \sin \)), which leads to

\[
\begin{align*}
\cos(\theta) &= \frac{e^{i\theta} + e^{-i\theta}}{2} \tag{1.16} \\
\sin(\theta) &= \frac{e^{i\theta} - e^{-i\theta}}{2i} \tag{1.17} \\
\end{align*}
\]

### 1.4.3 A first simple application

One useful application of the Euler formula lies in the fact that dealing with trigonometric function is not always that easy. For instance, do you know by heart how to write \( \sin(2\theta) \) or \( \cos(2\theta) \) with only functions of the angle \( \theta \), not \( 2\theta \)? If you forgot about those formulas, just switch to the complex world!
\[ e^{i(2\theta)} = \frac{\cos(2\theta) + i \cdot \sin(2\theta)}{\Re(e^{i2\theta})}\]
\[ e^{i(2\theta)} = (e^{i\theta})^2 = (\cos(\theta) + i \sin(\theta))^2 \]
\[ = \cos^2(\theta) + 2i \sin(\theta) \cos(\theta) + i^2 \sin^2(\theta) \]
\[ = \frac{\cos^2(\theta) - \sin^2(\theta)}{\Re(e^{i2\theta})} + i \cdot \frac{2 \sin(\theta) \cos(\theta)}{\Im(e^{i2\theta})} \]

where we used the fact that \( i^2 = -1 \), as you should know by now. It is now easy to conclude by equating the two expressions of the real part (and identically for the imaginary part):
\[
\begin{align*}
\sin(2\theta) &= 2 \sin(\theta) \cos(\theta) \\
\cos(2\theta) &= \cos^2(\theta) - \sin^2(\theta)
\end{align*}
\]

### 1.4.4 Physical examples

One possible application of complex numbers to physics resides in the representation of any periodic system which would require trigonometric functions to be described, such as waves for instance. If we consider a wave function \( \xi(x, t) \), it will usually be written as

\[
\xi(x, t) = \xi_0 \cos(kx - \omega t)
\]

or, using the Euler formula

\[
\xi(x, t) = \xi_0 \Re \left( e^{i(kx - \omega t)} \right)
\]

However, you will never encounter this notation with the real part (or imaginary part if you considered a \( \sin \)-function) in any physics book, where the wave is simply denoted as

\[
\xi(x, t) = \xi_0 \ e^{i(kx - \omega t)}
\]

Physicists simply assume that the real part is taken at the end of their calculations (and that way, it is shorter to write and much easier to read, don’t you agree?).

What is now the velocity \( \dot{\xi}(x, t) \equiv \frac{d\xi(x, t)}{dt} \) or the acceleration \( \ddot{\xi}(x, t) \equiv \frac{d^2\xi(x, t)}{dt^2} \)? Well, you just need to differentiate the wave function \( \xi(x, t) \), and it is much easier to do it when considering an exponential function!

---

\(12\xi(x, t) \) is the displacement of the wave, which depends on the position \( x \) and the time \( t \) and \( \xi_0 \) is the amplitude.
Chapter 2

Mechanics 1

Mechanics is one of the most important topics in physics. It describes how a body moves if the interaction of that body with other bodies is known. For example in Newtonian mechanics\(^1\) the "how bodies move" is described by the acceleration, and the interaction by the force. Newton's second law brings these two things together such that in principle we can calculate the motion. In mechanics we do not ask where the force come from, this is subject of other topics for example electromagnetism. We only describe its influence on motion.

All the following topics in physics will more or less base on mechanics, for example thermodynamics and fluid dynamics will describe the collective behavior of many particles. Oscillations and waves describe the motion of a body or many bodies where a linear force is acting on. And finally electrodynamics describes the interaction of electric and magnetic field with a body.

Since mechanics contains a lot of subjects we divide it into two parts. In this chapter we will focus on point like particles and introduce the main concepts to describe their motion. In the next chapter 3 we will then look at rigid bodies, which consist of many atoms, and generalize circular motion including the motion of planets around the sun.

2.1 Model

Any body is built-on a lot of atoms. To precisely describe the motion of that body one would need to describe the interaction of each atom with each other and also with other bodies and then compute the motion of each single atom. Obviously this calculation exceeds any computing power for systems with more than some dozens of atoms. In order to describe the motion of a body in good approximation we have to figure out a model.

- A first good approximation is to consider the body as point like and address collective properties like (total) mass and (total) charge to it. This approximation can be justified in case the internal structure of the body is small compared to its behaviour as

\(^1\)Other common descriptions are Lagrangian and Hamilton mechanics. They describe interactions by the potential energy and it is possible to generalize the formalism at the prize it gets more complicated. In the end the equations one has to solve to get the motion are the same for all description.
entire body. For example if you kick a football its travelling distance is much larger than its diameter, same holds for planets. Therefore one can in good approximation describe a football and planets as point-like particles².

- If one considers the distance between the atoms to be constant, one describes that body as rigid body. This is treated in 3.

- One can also look at the deformation of a body. The description as rigid body is then not valid any more and one has to go one step further.

2.2 Kinematics of Point-like Particles

As already mentioned above in this whole chapter 2 we will only consider point-like particles. In this section we will look how one can describe the motion of such a particle. We are not interested why the particle performs this motion. This we will discuss in the next section about Newton’s laws.

In this section, first a general description is given which is basically a little repetition of math. Afterwards we will look at the most important cases.

2.2.1 General Description

Our universe is three dimensional and the whole evolution can be parametrized by the time. Or in simpler words, the position of any body can be described by a three-dimensional vector which only depends on time and its starting point. Let’s look at this description step by step:

First of all we need to define a reference point, often denoted as \(O\). This point is the zero point in the coordinate system and the coordinates of all bodies refer to this point, see also figure 2.1. For our coordinate system we need furthermore three coordinate axis. In principle one is free to choose any point in space as zero point and also the direction of the axis. Nevertheless it is very often possible to reduce the difficulty of the problem by a good choice of reference point and direction of axis (see sections 2.2.2 and 2.2.3). In particular it is useful to take perpendicular axis since then the scalar product of two axis is always zero. Additionally one should choose the axis such that they obey the right hand rule (see also 1.1.2). This means if the \(x\)-axis points in the direction of the thumb and the \(y\)-axis in the direction of the forefinger then the \(z\)-axis should point in the direction of the middle finger. This right hand rule will get important when we look at vector products of vectors such as angular momentum or torque. The position of a point-like particle at time \(t\) is given by the position vector \(\vec{r}(t)\), namely

\[
\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}
\]

²Nevertheless there are effects that appear from their non-zero size. One then has to model it in a more complicate way.
where \( x(t) \) is the coordinate along the \( x \)-axis at the time point \( t \). This means the distance between the reference point \( O \) and the particle is in the direction of the \( x \)-axis is \( x(t) \). Note: \( x(t) \) is not simply a number, it also contains an unit such as meter or kilometer. Depending on the unit, the value of \( x(t) \) can be different but the distance is always the same. The same counts for \( y \) and \( z \).

The change of the position vector \( \vec{r}(t) \) divided by the time \( \Delta t \) needed for this change is the velocity. Assume the particle travels between the time points \( t_1 \) and \( t_2 \) from \( \vec{r}(t_1) \) to \( \vec{r}(t_2) \). The average velocity is then

\[
\vec{v}(t) = \frac{\Delta \vec{r}(t)}{\Delta t} = \frac{\vec{r}(t_2) - \vec{r}(t_1)}{t_2 - t_1} = \left( \begin{array}{c} \Delta x \\ \Delta y \\ \Delta z \end{array} \right) .
\]

To get the instantaneous velocity one looks at the limit of very small \( \Delta t \) which is the derivative. The instantaneous velocity vector \( \vec{v}(t) \) is therefore

\[
\vec{v}(t) = \frac{d\vec{r}(t)}{dt} = \left( \begin{array}{c} \frac{dx(t)}{dt} \\ \frac{dy(t)}{dt} \\ \frac{dz(t)}{dt} \end{array} \right) = \left( \begin{array}{c} v_x(t) \\ v_y(t) \\ v_z(t) \end{array} \right) .
\]

Geometrically the instantaneous velocity is the tangent at the curve \( \vec{r}(t) \) which points in the direction of motion and has length \( v = |\vec{v}(t)| = \sqrt{v_x^2 + v_y^2 + v_z^2} \), see also figure 2.1.
We can proceed and look at the change of the velocity per time. This means we take the derivative of the velocity. This quantity is the acceleration, denoted by $\vec{a}$. Formally we get

$$\vec{a}(t) = \frac{d\vec{v}(t)}{dt} = \frac{d^2\vec{r}(t)}{dt^2} = \begin{pmatrix} \frac{dv_x(t)}{dt} \\ \frac{dv_y(t)}{dt} \\ \frac{dv_z(t)}{dt} \end{pmatrix} = \begin{pmatrix} a_x(t) \\ a_y(t) \\ a_z(t) \end{pmatrix}.$$

Of course one can also look at higher derivatives but they usually have no physical importance.

2.2.2 Linear Uniform Acceleration

One of the most important non-trivial cases\footnote{Trivial cases would be no acceleration and therefore constant velocity, because then we could make a change of reference system and our object would be at rest. This is related to the definition of inertial frames, see section 2.3.2.} is the linear uniform acceleration. This means $\vec{a}$ is constant. We assume that the acceleration points in the direction of the $x$-axis and that the $y$ and $z$ coordinate do not change\footnote{Otherwise we could chose our coordinate system such that the $x$-axis points in the direction of the acceleration. Furthermore we could choose the reference point moving with the same velocity in $y$ and $z$ direction as the particle. Then the $y$ and $z$ coordinate with respect to this moving reference point would not change. The reason why we can choose a reference point that moves with constant velocity is given in section 2.3.2.}.

Obviously we don’t need to describe this motion in three dimensions and we can do the whole calculation scalar only considering the $x$-axis. For this assume\footnote{We could also choose a reference frame where at the $t_0$ the particle is at rest and at the origin. But to show how the calculations has to be done, let us consider the more general case.} that at time $t_0$, the particle is at position $x_0$ and moves with velocity $v_0$. We search for the velocity at a (later) time point $t$. Since the acceleration is the derivative of the velocity, we obtain the velocity by integrating the acceleration and choosing the integration constant such that at $t = t_0$ the velocity is $v_0$. Since the integral from $t_0$ to $t = t_0$ is zero, the integration constant\footnote{It’s not always that simple to find the integration constant. For example if the velocity $v_0$ is not given at $t_0$ but at $t_0 - 1s$.} is exactly $v_0$.

$$v(t) = v_0 + \int_{t_0}^{t} a(t') \, dt'$$

$$= v_0 + a \int_{t_0}^{t} dt' = v_0 + a(t - t_0).$$

Note that the prime of $t'$ has nothing to do with derivatives. We simply need a variable to parametrise the integral. We could also replace the $t'$ by any other sign (e.g. $x$, $\alpha$ or $\aleph$) as
it does not appear outside the integral. The position of the particle is obtained by one more integration. With the same argumentation the integration constant is \( x_0 \).

\[
x(t) = x_0 + \int_{t_0}^{t} v(t') \, dt'
= x_0 + \int_{t_0}^{t} (v_0 + a(t' - t_0)) \, dt'
= x_0 + v_0(t - t_0) + \frac{1}{2}a(t - t_0)^2.
\]

The calculation is graphically drawn in figure 2.2.

Figure 2.2: Connection between the acceleration \( a \), the velocity \( v \) and the distance \( x \) (herefore the linear uniform acceleration). Left: graph of the acceleration, which is constant (since uniform acceleration). The area under the graph (shaded) from \( t_0 \) to \( t \) corresponds to the velocity \( v - v_0 \) at time \( t \) which grows linearly. Middle: Graph of the velocity. The area corresponds to the distance \( x - x_0 \). Right: Graph of the distance which grows quadratically.

It is important to remember that the displacement (with \( a \) constant) is proportional to the square of time (for \( t_0 = 0 \)). One intuitive explanation for this square dependence is that the mean velocity is proportional to the time \( t \) and that the time of flight is proportional to \( t \). Therefore the distance is proportional to \( t^2 \), see again figure 2.2.

If the acceleration is linear (the vector \( \vec{a}(t) \) points always in the same direction) but not uniform (\( \vec{a}(t) \) not constant), we can treat the problem also in one dimension but we have to calculate the integral for the function \( \vec{a}(t) \).

Example: Let’s calculate a concrete case namely the motion of a ball when dropping it from the second floor of a house. Since in the gravitational field on earth, all bodies are accelerated with the same\(^7\) vertical acceleration\(^8\) \( g = -9.81 \text{ m/s}^{-2} \) we encounter the case of linear uniform acceleration. Take the coordinate axis to point upwards\(^9\) and the zero-point

\(^7\)Neglecting air resistance and assuming small dropping heights, such that the earth seems to be flat.
\(^8\)See also section 2.3.6.
\(^9\)This choice also influences the sign of the acceleration, i.e. if it would point downwards, it would be \( g = 9.81 \).
on the ground, see also picture 2.3. Assume the window in the second floor of the house has a height of 4m and we let the ball drop at \( t = 0 \). Then its velocity until it hits the ground is \( v = gt = -9.81t \). The velocity is negative as expected as the \( x \) coordinate of the ball gets smaller and smaller (from initially 4m towards 0). The position is then

\[
x(t) = x_0 + \int_{t_0}^{t} v(t') \, dt'
\]

\[
= x_0 + \int_{t_0}^{t} (v_0 + at') \, dt', \quad v_0 = 0
\]

\[
= 4m - \frac{1}{2} 9.81 \text{m/s}^2 t^2.
\]

To get the time the ball needs to reach the ground we have to set \( x(t) = 0 \) and solve for \( t \) leading to

\[
t = \sqrt{\frac{2 \cdot 4m}{9.81 \text{m/s}^2}} \approx 0.28 \text{s}
\]
2.2.3 Circular Motion

After we discussed the kinematic problems which can be solved in one dimension lets now look at the simplest problem which needs two dimensions. This is the circular motion. In this case the particle flies on a circle in a plane, see 2.4. Since the particle always moves on the circle we can parametrize it by an angle $\varphi(t)$.

Figure 2.4: Situation for the circular motion: A particle moves on a circle with radius $R$. Its position is described by the angle $\varphi$.

$$x = R \cos(\varphi)$$
$$y = R \sin(\varphi).$$

If the angle changes with a constant rate we call this motion uniform circular motion. The rate of change is described by the angular velocity $\omega = \frac{d\varphi}{dt}$ (being constant). This leads then to an angle $\varphi(t) = \omega t + \varphi_0$. The velocity is given by

$$v_x = \frac{dx(t)}{dt} = \frac{dR \cos(\omega t + \varphi_0)}{dt} = -R \omega \sin(\omega t + \varphi_0)$$
$$v_y = \frac{dy(t)}{dt} = \frac{dR \sin(\omega t + \varphi_0)}{dt} = R \omega \cos(\omega t + \varphi_0).$$

The velocity is obviously perpendicular to the position vector $\vec{r}$. This is related to the constriction of the motion on the circle where $|\vec{r}|$ is constant. To see this relation, assume the velocity would not be perpendicular to the position. Then there is a component of the velocity pointing in the direction of the position which means that the position vector would
get longer or smaller. The position vector would then leave the circle which contradicts to $|\vec{r}|$ being constant. The general motion in two dimensions will be discussed in the next section 2.2.4. Let’s come back to the circular motion and calculate the absolute value of $\vec{v}$:

$$|\vec{v}| = \sqrt{v_x^2 + v_y^2} = R\omega \sqrt{\cos(\omega t + \varphi_0)^2 + \sin(\omega t + \varphi_0)^2} = R\omega.$$ 

Similar to the position vector the velocity vector only changes its direction and not its absolute value. We therefore can repeat all the steps above to calculate the acceleration which is then given by

$$a_x = \frac{dv_x(t)}{dt} = -dR \sin(\omega t + \varphi) = -R\omega^2 \cos(\omega t + \varphi)$$

$$a_y = \frac{dv_y(t)}{dt} = \frac{dR \cos(\omega t + \varphi)}{dt} = -R\omega^2 \sin(\omega t + \varphi).$$

From the discussion above the acceleration was expected to be perpendicular to the velocity and since we are in two dimensions it must be parallel to the position. But position and acceleration point in the opposite direction or in mathematical language $\vec{r} \cdot \vec{a} < 0$. This means $\vec{a}$ points towards the origin $O$ of the circular motion.

### 2.2.4 General 2 Dimensional Motion

We now look at the motion in two dimension, as this occurs in many situations. Examples for motion in two dimensions are ball rolling on plane, a car driving on a street or the trajectory of a flying object with gravitational force.

Consider the following situation: A point-like particle moves on a plane. At the time $t$ the particle shall be at the position $\vec{r}(t)$ and moving with a velocity $\vec{v}(t)$ and accelerated with an acceleration $\vec{a}(t)$. The question is, what the relation between these quantities is.

Formally the relation is the one described in section 2.2.1 but there is a more intuitive approach. For this let’s first look at the relation between acceleration $\vec{a}(t)$ and velocity $\vec{v}(t)$. The acceleration describes the change of the velocity. We split the acceleration into the part parallel to the velocity, denoted by $\vec{a}_\parallel(t)$, and the part perpendicular to the velocity, denoted by $\vec{a}_\perp(t)$, see also figure 2.5. From the discussion in the section about circular motion (see 2.2.3), it is obvious that $\vec{a}_\perp(t)$ only changes the direction of $\vec{v}(t)$ but not its length. On the other hand $\vec{a}_\parallel(t)$ only changes the length of $\vec{v}(t)$ but not its direction. Similarly we can argue concerning the relation between the position $\vec{r}(t)$ and the velocity $\vec{v}(t)$.

It might seem that this intuitive approach does not lead to a lot of insight. But for example considering the motion of planets around the sun and in particular their non circular but elliptic motion can intuitively be explained with this argument.

---

10All vectors in this section shall be two dimensional vectors on the plane.
2.3 Dynamics of Point-like Particles

In the last chapter we discussed how one can describe the motion of a particle. But we did not ask what makes the particle move on that trajectory. To do this we now proceed to the description of the interaction between bodies and how the interaction is connected to the motion, i.e. what forces the particle to move the way it does. We will describe the interaction by forces\textsuperscript{11}. The relation to the motion is then described by Newton’s laws. After this rather general part we look at the most important forces and how one can categorize forces.

2.3.1 Force

Although the concept of a force is very important in this chapter and one might have an intuition from everyday life it’s not that easy to precisely define what is meant by a force in physics. It is rather the relation to other (maybe more intuitive) quantities that defines a force. For the moment let us define the force as (mechanical) resistance a body opposes. For example if someone wants to deform a body one needs to apply a force. Or if one wants to change the velocity one needs to apply a force which leads then to the acceleration (see Newton’s second law). On the other hand these two examples allow us to measure a force: For example if a body gets deformed, we know that a force is acting on it. Or if an object accelerates we know a force is causing this acceleration\textsuperscript{12}. Since the resistance as described above might be different in different directions of space it is obvious that the force in general is a vector.

\textsuperscript{11}One could also describe the interaction by the interaction energy which is done in Lagrangian and Hamiltonian mechanics.

\textsuperscript{12}This statement is not as easy as it sounds because it depends on the choice of reference frame (see 2.3.2). It might lead to fictitious forces if one does not choose an inertial frame.
2.3.2 Choice of Frame of Reference

Until now we simply took a frame of reference (usually a suitable) without thinking about its consequences to the laws of physics. To be more precise until now (meaning the kinematic part) we actually did not do physics because we only defined different quantities. This definitions are in principle math and physics comes in when we connect different definition such as force and acceleration. Depending on the choice of the frame of reference the laws of physics might look different.

An example is driving a curve with a car: In the frame of the driver the car and he/she does not move, therefore his/her velocity is always zero and hence also his/her acceleration. Nevertheless he/she feels a force in a curve without acceleration. For a person standing on the road this force is needed in order to accelerate the car such that it can take the curve (details see 3.2). So in both frames we have a force but only in one there is an acceleration. There is a very special class of reference frames in which Physics takes the simplest form. These reference frames are called inertial frames of reference. They are characterized by Newton’s first law. Most of the laws of physics are only valid in inertial frames. We will have a look at non-inertial frames in section 3.2.2. Let’s now proceed and look at Newton’s laws:

2.3.3 Newton’s Laws

The three laws that connect the description of motion (position, velocity, acceleration) and the one of interaction (force, interaction energy) are Newton’s laws (see also picture 2.6)

1. A body where no force acts on, remains in constant velocity.

2. The total force acting on a body causes an acceleration proportional to its mass $m$:

$$\vec{a} = \frac{\vec{F}}{m}.$$

3. If body A acts with a force $\vec{F}$ on body B, body B acts with a force $-\vec{F}$ on A.

Here we introduced the first time the mass $m$. This is a property of the body which tells you how inert it is or using the terminology of the definition of the force: The mass tells you how big the resistance of a body is, if a force acts on it.

The laws stated above contain a lot of information that needs to be discussed: As already mentioned in section 2.3.2, the first law gives a definition for an inertial frame. To be more precise: A reference frame is an inertial frame if and only if any body on which no (total) force acts, does not move or moves with constant velocity. Returning to the example of the car driving a curve we observe that the car is not an inertial frame: For

---

13 The relation between position and velocity and acceleration is also math because we gave the different derivatives of the position new names. The force is a completely independent concept and needs to be related to the kinematic properties.

14 Including the case where it remains in rest.

15 Be aware that defining the mass this way, the gravitational property of a mass is not included. In fact the two properties "resistance" and "two masses attract each other" are à priori two independent properties. A body could have these two properties independently so a "resistive mass" and an "attractive mass". These two properties get unified in terms of general relativity.
2.3. DYNAMICS OF POINT-LIKE PARTICLES

\[ \vec{V} = \text{const} \]

\[ \vec{F} = 0 \Rightarrow \vec{a} = 0 \]

\[ \vec{F}_{AB} = -\vec{F}_{BA} \]

Figure 2.6: Newton’s laws as pictures: Left: Newton’s first law: If no force acts, a particle moves with constant velocity. Middle: Newton’s second law: Force is equal to mass times acceleration. Right: Newton’s third law: The force acting from a particle to the other is equal to minus the force acting from the other to the one.

this we only look at the motion on the plane where the car is driving and do not take into account the vertical gravitational force. A body that is not attached at the car will then be accelerated in the frame of the car although there is no force acting on it\(^{16}\). In the frame of the road, that body will simply move straight forward as we would expect it from a body on which no force acts.

The second law is the most important one because it answers the question how a body behaves if a force is acting on it. One might think that the first law is contained in the second one. This is only partially true, because the second law is only valid in inertial frames. Therefore if we want to link the motion and the forces (as given in the second law) we first have to ensure we have an inertial frame and for this we need the first law. If multiple forces act on a body one has to sum them up (as vector) to get the total force. According the the second law, the total force is then equal to \( \vec{F}_{\text{tot}} = m\vec{a} \). Or to be more clear: If multiple forces act on a body such that they cancel each other, that body will not be accelerated. In principle you can solve all the problems simply by summing up all forces, calculate the acceleration and then the path. But in general the force depends on the position or the velocity of the body and you end up in complicated differential equations. One can often avoid this difficult way and get the result easier using conservation laws and some tricks.

The third law is in one to one correspondence to the conservation of momentum. We will look at this closer in section 2.4.1. The third law might sound contradictory to the second law. Because if a body A acts with a force \( \vec{F} \) on B and B with a force \( -\vec{F} \) on A, the total force is zero. And indeed, the total system composed of A and B will not accelerate (meaning the center of mass will not accelerate, see also 2.3.4). But as long as there is nothing that inhibits the two bodies to accelerate (for example a rod between the two bodies) the total force on each body is non zero and both will accelerate\(^ {17}\). If there is for example a rod

\(^{16}\)This might sound contradictory to the statement in section 2.3.2 were we stated that in both frames a force is acting. But there we looked at the driver and the driver is attached to the car and therefore a force is needed in order to make the curve (otherwise the driver would leave the car).

\(^{17}\)They accelerate such that the total momentum is conserved. Or equivalently the center of mass is not accelerating. This is ensured by the third law.
separating the two bodies, one has also to take into account the forces of the rod acting on each particle. In this case the total force on each body is in fact zero.

The second law of Newton can be formulated in a more general way\textsuperscript{18} as

\[
\frac{d(m\vec{v})}{dt} = \vec{F}
\]  

(2.1)

where the first term \( m\vec{v} \) is the momentum (see section 2.4.1). If the mass \( m \) is constant, we can take it out of the derivative and we get the formula above. In most cases one can use the simpler version of Newton’s second law.

### 2.3.4 Center of Mass

Consider a body consisting of \( N \) point like particles, see figure 2.7. Assume these particles interact with each other such that there is a force between each pair of particle (in case two particle do not interact, their force is zero). In addition assume there is an additional force acting on all particles. Let us enumerate all the particles and focus on one particular with the number \( i \). We can split the total force acting on this particle into the contributions of all other particles and the external force \( \vec{F}_{i,\text{ext}} \)

\[
\vec{F}_i = \sum_{j \neq i} \vec{F}_{ji} + \vec{F}_{i,\text{ext}}
\]

Formally we get

\[
\vec{F}_i = \sum_{j \neq i} \vec{F}_{ji} + \vec{F}_{i,\text{ext}}
\]

where \( \vec{F}_{ji} \) is the force\textsuperscript{19} acting from a particle \( j \neq i \) on \( i \). The total force acting on the entire body is given by the sum of all forces acting on all particles

\[
\vec{F}_{\text{tot}} = \sum_i \vec{F}_i = \sum_i \left( \sum_{j \neq i} \vec{F}_{ji} + \vec{F}_{i,\text{ext}} \right)
\]

\textsuperscript{18}Newton already stated it in that way.

\textsuperscript{19}In some literature the indices \( i \) and \( j \) are swapped. So \( \vec{F}_{ji} \) is the force from \( i \) on \( j \).
The first term (the sum over all forces $\vec{F}_{ji}$) is called internal force and the second term (sum over all external forces) is called external force. According to Newton’s third law, the force acting from particle $i$ on particle $j$ is opposite the one from $j$ on $i$. As a consequence the sum over all $\vec{F}_{ji}$ vanishes and the total force acting on the body is only

$$\vec{F}_{\text{tot}} = \sum_i \vec{F}_{i,\text{ext}}.$$ 

According to Newton’s second law, the total force acting on a particle $i$ is equal to the acceleration $\frac{d^2 r_i}{dt^2}$ time its mass $m_i$ which leads to

$$\vec{F}_{\text{tot}} = \sum_i \vec{F}_{i,\text{ext}} = \sum_i m_i \frac{d^2 r_i}{dt^2} = M \frac{d^2 \sum_i m_i r_i}{M} \tag{2.2}$$

where $M = \sum_i m_i$ is the total mass. This is a very useful statement because it means that the motion of the entire body can be described as if it were a point like particle at the position

$$\vec{R}_C = \frac{\sum_i m_i r_i}{M}.$$ 

The position $\vec{R}_C$ is called center of mass. Equation (2.2) can then be written as

$$\vec{F}_{\text{tot}} = M \frac{d^2 \vec{R}_C}{dt^2}$$

which is exactly the equation of motion of a point like particle at position $\vec{R}_C$. This is especially remarkable since we do not have to know anything about the internal forces, we can completely neglect them. Be aware that the center of mass $\vec{R}_C$ only describes the motion of the entire body and not internal motion as rotations or oscillations of its constituents. In case of a rigid body have also a look at 3.4

### 2.3.5 Equilibrium

In different topics of physics we will encounter equilibriums. In general an equilibrium is a state which does not change with time. Usually they are easier to calculate than a general state where an explicit time dependence needs to be calculated.

A body is in a mechanical equilibrium if it remains in rest or moves with constant speed. We focus on resting bodies. If a body remains in rest, its velocity and acceleration is zero. Due to Newton, the total force acting on a body must therefore also be zero. There are three possibilities how this can happen, see also figure 2.8.
Stable
The body is in a (local) minimum of energy. If it is displaced slightly, a force pushes it back towards the minimum, it returns to its position of equilibrium. See also the two examples in figure 2.8. For small displacements, the restoring force is in most cases proportional to the displacement. If slightly displaced, the Body then performs a harmonic oscillation, see also 5.2.

Unstable
In this case the body is energetically in a (local) maximum. A small displacement causes the body to be pushed away from the initial position. Examples are shown in the middle of figure 2.8. If the body is placed exactly at the maximum of the energy then the total force vanishes also.

Indifferent
A third interesting case is the indifferent equilibrium. If the body is slightly displaced, there is no force acting at all. See also the right part of figure 2.8.

2.3.6 Gravitational Force
One of the most important forces is gravity. According to Newton’s laws, a mass has the property that it opposes the change of velocity. Or in other words, a mass is inert. But masses have also an other property, they attract each other.
Consider two masses $m_1$ and $m_2$ separated by a distance $r$. Then the attractive force\(^{20}\) between the two masses is given by

\[ F = \frac{G m_1 m_2}{r^2} \]

\(^{20}\)Since the form of this formula is highly related to the electric force of two point charges, further
\[ F = \frac{G m_1 m_2}{r^2} \]

where \( G = 6.67 \cdot 10^{-11} \text{m}^3\text{kg}^{-1}\text{s}^{-2} \) is the gravitational constant. Obviously the force is proportional to each of the two masses. This has a very important implication: If the only force acting on a body is gravity, then its motion is independent of its mass. To see this assume we double the mass of a body. Then its force is also doubled. To get the acceleration of that body we have to divide the force by its mass. And there the factor two cancels out\(^{21}\). Like in electromagnetism, we can define a field strength of a gravitational field. This field strength is called gravitational acceleration\(^{22}\). The field strength is the quotient of the force divided by the mass of one body. For example if we look at the earth with mass \( M \), the gravitational acceleration at the surface of the earth is given by

\[ g = \frac{GM}{r^2} \approx 9.81 \text{m} \cdot \text{s}^{-2} \]

where for the numerical value we took the radius of the earth\(^{23}\). This means that a body with mass \( m \) is attracted by the mass \( M \) with a force \( F = mg \).

Looking how gravity looks vectorial, we have to take into account that two masses attract each other. Denoting \( \vec{r} \) as the vector pointing from mass \( M \) to mass \( m \) (see figure 2.9), the force acting on mass \( m \) is given by

\[ \vec{F}_{\text{grav}} = -\frac{GMm}{|\vec{r}|^3} \vec{e}_{\vec{r}} = -\frac{GMm}{|\vec{r}|^3} \vec{r} \]

where \( \vec{e}_{\vec{r}} = \frac{\vec{r}}{|\vec{r}|} \) is the unit vector along \( \vec{r} \). The minus sign is important as it represents the fact that two masses attract each other.

properties in case of the electric field can be found in chapter 8.1

\(^{21}\)Although in electromagnetism, the force for point-like charges looks basically the same. But since the force is proportional to the charge \( q \), only bodies with same quotient \( \frac{e}{m} \) have equal motion.

\(^{22}\)This terminology is a bit unfortunate. Because one should think of a field strength and not an acceleration. The gravitational acceleration corresponds to the electric field (strength) in electromagnetism.

\(^{23}\)The earth is obviously not a point-like particle and to use the formula for point like particles need to be justified. The proof that this is valid is the same as in case of the static electric field, we refer the reader to 8.1. 
### 2.3.7 Independence of Motion

In many cases it is useful to split the three dimensional movement in its \( x \), \( y \) and \( z \) coordinate and calculate the motion along each coordinate separately.

Example Let’s look at the case of a projectile motion where after we shoot a body (with mass \( m \)), the only force acting is the gravitational force. Shooting it at time \( t = 0 \) with an initial velocity \( v \) at an angle of \( \alpha \) (see figure 2.10), the horizontal \( \vec{v}_x \) and vertical velocity \( \vec{v}_y \) are

\[
\vec{v}_x = \cos(\alpha) \vec{e}_x \\
\vec{v}_y = \sin(\alpha) \vec{e}_y
\]

Putting the origin to the place, where we shoot the body, leads to \( \vec{r}(t = 0) = 0 \). We now can calculate the motion along the different axis independently (for \( t > 0 \)).

- As there is no initial velocity in the \( z \) direction and no force acts in the \( z \) direction, the \( z \) component will not change. So we have \( z(t) = 0 \).

- Along the \( x \) axis, there is no force acting, so no acceleration. But we have an initial velocity so we get a linearly growing \( x \) component. \( x(t) = v_x t \) where \( v_x = |\vec{v}_x| \).

- In \( y \) direction we have the gravitational force acting as well as an initial velocity. Therefore we can use the linear uniform acceleration from section 2.2.2 with \( t_0 = 0 \), \( x_0 = 0 \), \( v_0 = v_y = |\vec{v}_y| \) and \( a = -g < 0 \) where the \( g > 0 \) is due to the gravity pointing in opposite direction than the \( y \) axis. We then have the motion \( y(t) = v_y t + \frac{g}{2} t^2 \).

Due to this splitting we were able to simplify this 3 dimensional motion into the known motion in one dimension. We can even go a step further and calculate the \( y \) component as a function of the \( x \) component meaning that for each point along the \( x \) axis we know the height of the body. For this we solve \( t \) for \( x \) and get \( t = \frac{x}{v_x} \) and substitute it into the function for the \( y \) component:

\[
y(x) = v_y t(x) - \frac{g}{2} t(x)^2 = v_y \frac{x}{v_x} - \frac{g}{2} \left( \frac{x}{v_x} \right)^2 = \frac{v_y}{v_x} x - \frac{g}{2v_x^2} x^2
\]
2.3.8 Volume and Surface Forces

There are two conceptually different ways a force can act on a body. On one hand a force can act on the entire body or only on its surface. Correspondingly one calls these forces volume and surface force. Obviously these concepts do not hold for point like particles since they have no volume and no surface. In case of the volume force, the force acts on the entire body. It is not necessary that it acts on the entire body the same. Examples of volume forces are gravity, magnetic field or electric field in case of insulators.

This is different to surface forces where the force only acts on the surface of the body. Examples of these are pressure and friction. One can even split up the surface forces in normal and tangent forces. A normal force acts perpendicular on the surface. For example if a stone is placed on a horizontal table, the table opposes the gravitational force of the stone by pushing it up. This up-pushing force only acts on the surface touching the stone and nowhere else. Since gravity acts vertically, the force between the table and the stone is also vertical and therefore perpendicular to the surface. This is different in case of a tangent force which is acting parallel to the surface. For example if the stone is pushed over the table, there is friction which acts parallel to the surface. For more details, see also the next section.

2.3.9 Friction

A exact description of friction (including fluid resistance) is pretty difficult. Nevertheless there are different models where we will look at the simplest one. This model states that the friction of a (moving) body is proportional to its normal force, see also figure 2.11. The proportionality constant is called coefficient of friction \( \mu \). One can distinguish three different types of friction, static friction, dynamic friction and rolling resistance. Each of these three types has its own coefficient.

Static friction happens if two bodies are placed one onto the other but do not move. For example a stone on a table. As long as a horizontal force is smaller than the static friction, the stone does not move. This maximal static friction is given by \( F_s = \mu_s F_N \) where \( \mu_s \) is the static friction coefficient and \( F_N \) is the normal force of the stone, e.g. its gravitational force \( F_N = mg \). Be aware that this is only the maximal horizontal force one can apply before the stone starts sliding. If one applies a force \( F < F_s \), the friction is equal to \( F \) and not \( F_s \).

If the force is bigger than the static friction or if the body is already sliding, dynamic friction happens. This means the friction is \( F_d = \mu_d F_N \) where \( \mu_d \leq \mu_s \) is the dynamic friction coefficient. In our example, the stone slides with a constant speed if the applied force is equal to the dynamical friction \( F = F_d \). If the applied force is bigger, the stone accelerates with an acceleration \( a = \frac{1}{m} (F - F_d) \).

---

\(^{24}\) In fact these concepts make only sense in terms of bodies which consist of a lot of particles and we can describe them as continuum.

\(^{25}\) In case of conductors in an electric field, the charge is distributed on the surface of the body. See also 8.2.5.

\(^{26}\) Otherwise the total horizontal force would not vanish and the stone should accelerate horizontally.
Instead of sliding, a body could also roll. In this case one calls the "friction" rolling resistance. One can also assign a coefficient $\mu_r$ to this resistance. For well-formed bodies (as cylinders or spheres), the rolling resistance is much smaller than dynamic friction. Be aware that rolling resistance is highly related to static friction. Because in case of rolling resistance, at each instance one point of the body is not moving (the one touching the table). This point is the subject to static friction. For better understanding imagine driving a bike. If you brake it is not the rolling resistance that allows you to brake: As long as the wheel is still turning, your braking force is limited by the static force between the ground and the wheel. If you brake harder, you overcome the static force and you start sliding. This is worse because the dynamic friction is smaller than the static one and you brake slower (smaller acceleration). That is the reason most cars have ABS: At an emergency braking, the board computer regulates the braking such that it never blocks the wheels, they always brake with static friction and not dynamic friction.

### 2.4 Momentum, Work, Energy and Power

In principle one can always proceed as described in the previous sections. These equations of motion are very often too difficult to calculate. This does not mean we cannot say anything about the behaviour of complicated system. We can will use conserved quantities, as Momentum and Energy, which we treat in this chapter.

#### 2.4.1 Momentum

The momentum already appeared in the most general formulation of Newton’s second law (see equation (2.1)). It is defined as $\vec{p} = m\vec{v}$. As long as no force is acting, the time derivative of the momentum is zero, meaning momentum is conserved/constant. Or in other words, momentum is only changing if a force acts. The change of momentum $\Delta\vec{p}$ is then given by

$$\Delta\vec{p} = \vec{p}(t_2) - \vec{p}(t_1) = \int_{t_1}^{t_2} F(t) dt.$$
2.4.2 Work

The concept of work is pretty important in physics as it connects energy and force. Roughly speaking, work is one possibility to convert one form of energy into another. For example if someone drops something, e.g. a stone, it gets accelerated and the potential energy converts into kinetic energy (see also sections below). This transformation of energy happens due to the gravitational force acting along the path of the stone.

For a constant force $F$ always pointing in the direction of movement, the work $W$ is defined as

$$\Delta W = F \Delta s$$

where $s$ is the length of the considered path. The unit of work is newton meter (N·m) or Joule (J) or watt seconds (W·s). For example dropping a stone of 1kg from a height of 2m., the work done by gravity is $W = F s = 1kg \cdot 9.81m\cdot s^{-2} \cdot 2m = 19.62$J.

In case the force points perpendicular to the direction of motion, no work is done. Intuitively this can be seen in case of circular motion, where no energy is applied but a force acts radially and therefore perpendicular to the movement. Or in case of gravity: Moving our stone horizontally does not change its potential energy, so there is no work done by gravity\textsuperscript{27}. In general we can split up the force in the component parallel and perpendicular to the motion, see also figure 2.12. Then only the component parallel does contribute to the work. This is achieved by taking the scalar product of force and path

$$\Delta W = \vec{F} \cdot \Delta \vec{s}$$

\textbf{Figure 2.12:} A body is elevated along $\Delta \vec{s}$ (only vertical movement) by a force $\vec{F}$. Only the component $\vec{F} \parallel$ along $\vec{s}$ contributes to the work $\Delta W = \vec{F} \parallel \Delta \vec{s} = \vec{F} \cdot \Delta \vec{s}$.

In general the force might not be constant along the path and the path not a straight line. Then the calculation can get pretty tedious\textsuperscript{28}. In this general case we split up the path in

\textsuperscript{27}Although one might need to apply work in order to move the stone for example in case pushing it over a table where friction occurs. Then the work is only the friction times the path length.

\textsuperscript{28}In most cases the general calculation is not needed but it is important to understand what this general formula (intuitively) means.
small pieces $\Delta \vec{s}$ where we can assume the force $\vec{F}$ to be constant and the piece a straight line. Then we calculate the work done on this small piece and sum all this work up. The limit for the splitting into pieces of very small pieces (length going toward zero) is mathematically an integral. In this integral we integrate (=sum) the small work pieces $dW = \vec{F} \cdot d\vec{s}$ which leads then to the formula

$$\Delta W = \int \vec{s}^{\vec{s} + \Delta \vec{s}} \vec{F} \cdot d\vec{s}.$$ 

### 2.4.3 Energy

The definition of energy is the ability to perform work. This means if we want to perform work, we must apply some kind of energy which is then transferred (by the work) to another kind of energy. Energy is also a conserved quantity. Nevertheless there exist a lot of different forms of energy. The conservation of energy is stated as first law of thermodynamics\(^{29}\). In order to apply the conservation one needs to consider all kinds of energy which makes it in some cases less applicable than momentum.

### 2.4.4 Potential Energy

If a force acts between two bodies, in some cases it is possible to attribute a potential energy. Since this is in general a bit tricky, we will first have a look at an easy case and then generalize.

The easiest case is the one of a homogeneous field, e.g. the gravitational field on earth\(^{30}\). We choose the coordinate system such that the $z$-axis points vertically up (see also figure 2.13). The force on a body with mass $m$ is then $\vec{F}_{\text{grav}} = -gm\vec{e}_z$. If we lift a body vertically we have to apply a force $\vec{F}_{\text{e}} = -\vec{F}_{\text{grav}}$. Lifting it up by vertical distance $s$, we have to apply the work $W = |\vec{F}_{\text{e}}|s = \vec{F}_{\text{e}} \cdot s\vec{e}_z = -\vec{F}_{\text{grav}} \cdot s\vec{e}_z > 0$. If we take the $xy$ plane as reference, we can denote for each height $z$ an energy we have to apply. This energy is called potential energy $E_{\text{pot}}$ and is given by

$$E_{\text{pot}} = -\vec{F}_{\text{grav}} \cdot \vec{r} = |\vec{F}_{\text{grav}}|z$$

where $\vec{r}$ is the vector pointing at a certain position (see also figure 2.13). We therefore managed to attribute a gravitational energy for every point in space. Knowing the potential energy, we can also go back to the force by taking the derivative

$$\vec{F}_{\text{grav}} = -\frac{dE_{\text{pot}}}{dz} \vec{e}_z.$$  

\(^{29}\)In Newtonian mechanics this cannot be proven. But in Lagrangian mechanics this is associated to the assumption that physics is independent of time, meaning the laws of physics are true yesterday, today, tomorrow and at any other time.  

\(^{30}\)Here we only look near the surface of the earth such that earth looks like a plane.
Figure 2.13: A body in the gravitational field is lifted from ground to a coordinate $z$. The work is $W(z) = |\vec{F}_{\text{grav}}|z$. This means the work done is stored as potential energy. With the reference point $E_{\text{pot}}(z = 0)$ we get $E_{\text{pot}}(z) = W(z) = |\vec{F}_{\text{grav}}|z$

Here we used, that we already know the direction of the force. A general expression is given in the next paragraph, where we look at general potentials.

For the general case we consider two bodies denoted by their masses $M$ and $m$ which interact which each other. This interaction leads to a force $\vec{F}_{r}$ between them where $\vec{r}$ denotes the distance from $M$ to $m$, see also picture 2.14. In some cases, we can attribute a potential energy to this configuration. If this is possible, we can proceed as follows. We chose a reference point $\vec{r}_0$ where we attribute the energy $E_{\text{pot}} = 0$. The choice of this reference point is arbitrary, due to equation (2.5). Then we calculate the work $W$ that needs to be done to move $m$ from $\vec{r}_0$ to $\vec{r}$. The potential energy is then

$\text{31}$The condition to succeed is given in equation (2.4).

$\text{32}$We could also move $M$ which yields the same result. But since $\vec{r}$ points to $m$ it is more intuitive to move $m$. 

75
\[ E_{\text{pot}}(\vec{r}) = W(\vec{r}_0 \rightarrow \vec{r}) = - \int_{\vec{r}_0}^{\vec{r}} \vec{F} \cdot d\vec{s} \]  

(2.3)

where the minus sign takes into account that we have to apply the external force \( \vec{F}_e = -F \) in order to move the body \( m \).

This definition of the potential energy is only meaningful if it does not depend on the path from \( \vec{r}_0 \) to \( \vec{r} \). In particular this means we can go from point \( \vec{r}_0 \) to \( \vec{r} \) and back (by a different path) and gain no energy. This leads to the constraint the interaction between the two bodies have to fulfil in order to describe it by a potential energy: The work we have to apply for a closed path must be zero or as formula

\[ W = \oint \vec{F}_e \cdot d\vec{s} = 0 \]  

(2.4)

where the circle in the integral symbolizes that we take a closed path. Such an interaction is associated to a (so called) conservative field, further information see in the section about electromagnetism, chapter 8.2. Not all interactions fulfil this constraint, for example friction or many time-varying fields as for example a time dependent magnetic field\(^{33}\).

If we can describe an interaction with a potential energy, many calculations simplify. For example we can easily calculate the energy between two points\(^{34}\) \( \vec{r}_1 \) and \( \vec{r}_2 \) we can simply take the potential energy at \( \vec{r}_2 \) and subtract the one from \( \vec{r}_1 \)

\[ \Delta E_{\text{pot}} = - \int_{\vec{r}_1}^{\vec{r}_2} \vec{F} \cdot d\vec{s} = - \int_{\vec{r}_1}^{\vec{r}_2} \vec{F} \cdot d\vec{s} - \left( - \int_{\vec{r}_0}^{\vec{r}_1} \vec{F} \cdot d\vec{s} \right) = E_{\text{pot}}(\vec{r}_2) - E_{\text{pot}}(\vec{r}_1). \]  

(2.5)

\(^{33}\)This is used in a transformer, where the electrons "flying" around a varying magnetic field gain energy, meaning a voltage builds up.

\(^{34}\)The difference to the calculation before is that no one of these points must be the reference point \( \vec{r}_0 \).
On the other hand if the potential energy of an interaction is given, we can also go back to the force. Similar to the example above with the homogeneous gravitational field, we can apply some derivative on the potential energy. This is also clear from an analytic point of view: The potential energy is the integral from the force, and the "inverse" of the integral is the derivative. So applying an appropriate derivative on the potential energy should give back the force. Indeed we can get the force by

\[ \vec{F}(\vec{r}) = - \left( \begin{array}{c} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{array} \right) E_{\text{pot}}(\vec{r}) \]

where these curly derivative signs \( \partial \) are the usual derivatives but indicate that the potential energy not only depends on one parameter but on \( x, y, \) and \( z \). The minus sign is there because of the same reason as in the definition of the potential energy, see equation (2.3).

### 2.4.5 Kinetic Energy

Taking a bouncy ball and let it fall, it bounces up again. When flying up, it gains potential energy and in particular work is done in order to bring it up again. Hence there must be another kind of energy which allows the ball to get potential energy again. This other energy is the kinetic energy (neglecting elastic energy when it bounces). In order to calculate the kinetic energy, we make use of the conservation of energy. This means the potential energy lost must be converted into kinetic energy. When the ball falls for a time \( t \) (starting from rest), it passes a distance \( h = \frac{1}{2} gt^2 \) and loses the potential energy \( mgh \). The same energy must then be gained as kinetic energy. Therefore the kinetic energy is

\[ E_{\text{kin}} = mgh \]
\[ = \frac{1}{2} mv^2 = \frac{1}{2}pv \]

where \( v \) is the final velocity of the ball.

A 2 times bigger velocity leads to four times more kinetic energy. This is clear because to double the velocity one need double as many time and therefore a four time longer distance. As a consequence one converts four time more potential energy into kinetic energy.

---

35Since this includes more advanced math, you do not need to know this for the exams.
2.4.6 Power

An other interesting quantity is the amount of work per time, which is called power \( P \). If during a period of time \( T \) the work is constant \( W \), the power is simply \( P = \frac{\Delta W}{\Delta T} \). Therefore if one needs longer for the same work his power is smaller and vice versa. Similar to the discussion about mean velocity and instantaneous velocity in section 2.2.1, we can look at mean and instantaneous power. The instantaneous power is defined as

\[
P(t) = \frac{dW}{dt}.
\]

This instantaneous power describes how much work per second is done at a certain time. To get the mean power we choose a slightly different approach, which is more common in terms of power and work. Let's calculate the mean power between \( t_1 \) and \( t_2 \). For this we average the instantaneous power. Intuitively speaking we split the time \( T = t_2 - t_1 \) into \( N \) small pieces and consider the power being constant during each piece. Then we sum all these pieces up and divide by the total number of pieces. Letting \( N \) going towards infinity we end up by an integral

\[
\bar{P} \approx \frac{1}{N} \sum_{j=1}^{N} P(t_1 + \frac{Tj}{N}) = \sum_{j=1}^{N} \frac{P(t_1 + \frac{Tj}{N})}{\frac{N}{T}} \frac{1}{T} \to \frac{1}{T} \int_{t_1}^{t_2} P(t) \, dt = \bar{P}
\]

where we used \( \frac{T}{N} \to dt \) for \( N \) going towards infinity. The unit of Power is Watt which is one joule per second. Using this, the unit of energy is sometimes written as kW·h, 1kW·h = 3.6MJ which is the work done by the power of one Watt during one hour.

If you're not used thinking in terms of work and power, you might mix them (as many politicians and journalists do). For example a light bulb has a power of 40W, this means each second it converts 40J electric energy in light and heat. If the bulb shines one hour it has "worked" (not mechanical work) 40W · 3600s = 144000J which is of course an energy. Other example: A house need (let's say) 7500kW·h per Year. This is a power, because you have energy per year which is equivalent to work per time.
2.4.7 Rotation Energy

One other form of mechanical energy is the rotation energy. Consider a rotating body consisting of (multiple) mass point(s) $m_i$. Then the rotation energy is given by the kinetic energy of all the mass points

$$E_{\text{rot}} = \sum_i E_{\text{kin},i} = \sum_i \frac{1}{2} v_i^2 m_i = \omega^2 \sum_i \frac{1}{2} r_i^2 m_i = \frac{1}{2} \omega^2 I$$

where $\omega$ is the angular frequency and we used that the velocity $v_i = \omega r_i$ and $I = \sum_i r_i^2 m_i$ is the moment of inertia. We will have a closer look to rotating bodies in the next chapter 3.3.3.

Example:
Someone is driving with a bike with total mass $M = 80$kg with a velocity $v = 10\text{m\cdot s}^{-1}$ then the kinetic energy is $E_{\text{kin}} = \frac{1}{2} M v^2 = 4000\text{J}$. Additionally the wheels are rotating, meaning they have additionally rotation energy. Assuming (two) cylindrical wheels with mass $m = 0.6$kg and radius $R = 0.5$m, the momentum of inertia is $I = mR^2 = 0.15\text{kg\cdot m}^2$. Using $\omega = \frac{v}{R} = 20\text{s}^{-1}$ we get a rotational energy $E_{\text{rot}} = 2 \cdot \frac{1}{2} \omega^2 I = 60\text{J}$. The total energy is therefore $E_{\text{tot}} = 4060\text{J}$.

2.4.8 Angular Momentum

In terms of conservation laws one should clearly mention the angular momentum which we will also investigate more precisely in the next chapter. Nevertheless the most important formulas shall be given already now.

For a point like particle at position $\vec{r}$ and mass $m$ the angular momentum $L$ is defined as

$$L = rmv_\perp$$

where $v_\perp$ is the velocity component perpendicular to $\vec{r}$ and $r = |\vec{r}|$ is the absolute value of the position. For symmetric bodies consisting of many particles, we can use the momentum of inertia introduced in section 2.4.7 and we get

$$L = I \omega.$$
Chapter 3
Mechanics 2

In this section we have a closer look to circular motion and then proceed to many particle systems and rigid bodies.

3.1 Dynamics of Circular Motion

Uniform Circular Motion

We consider a particle of mass \( m \) moving uniformly with angular speed \( \omega \) along a circular trajectory of radius \( R \). Its acceleration vector is given according to chapter 2.2.3, NEWTON’s second law then assumes the form

\[
\vec{F} = -m\omega^2\vec{r}
\]

if the origin of the position vector \( \vec{r} \) lies in the center of the circle. From this, we can conclude:

If a particle moves uniformly along a circular trajectory of radius \( R \) at angular speed \( \omega \), the total force acting on it must be directed radially towards the center of the circle, and its magnitude amounts to

\[
m\omega^2R.
\]

(3.1)

The total force acting on a body moving uniformly along a circle is often called a centripetal force. Note that the term ‘centripetal force’ does not tell you anything about the nature of the force, but just something about its effect on motion of the object it acts on. Gravitational attraction of a planet, the tension force by which a rope pulls on a bucket filled with water, a normal force exerted by the wall of a spinning cylinder – all can keep a particle on a circular path and act as centripetal forces.

Non-Uniform Circular Motion

Olympiad problems often deal with objects moving along circular paths at varying angular velocity. In this paragraph, you will see how the laws governing uniform circular motion can be modified to handle such situations.
Consider, again, a point-like body of mass \( m \) moving along a circular trajectory of radius \( R \).

In the coordinate frame we already used in paragraph 2.2.3, the position \( \vec{r} \) of the particle at time \( t \) is given by

\[
\vec{r}(t) = (R \cos \varphi(t)) \ \hat{x} + (R \sin \varphi(t)) \ \hat{y}.
\]

To get an expression for the velocity vector \( \vec{v} \), we take the time derivative and introduce the instantaneous angular velocity you might already know as

\[
\omega(t) := \dot{\varphi}(t),
\]  

(3.2)

obtaining

\[
\vec{v}(t) = \frac{d}{dt} \vec{r} = (-R \ddot{\varphi} \sin \varphi) \ \hat{x} + (R \dot{\varphi} \cos \varphi) \ \hat{y} = R \omega (- \sin \varphi \cdot \hat{x} + \cos \varphi \cdot \hat{y}).
\]

So far, this expression does not look too different from the one valid for uniform circular motion. Indeed, the only difference is that \( \omega \) is now dependent on time and therefore \( \varphi(t) \) cannot be expressed as \( \omega t + \varphi(0) \). As for uniform circular motion, the velocity vector \( \vec{v} \) is always tangential to the circle and its now time-dependent magnitude is given by

\[
v(t) = R \omega(t).
\]  

(3.3)

Acceleration of the body is obtained by repeated differentiation:

\[
\vec{a} = \frac{d}{dt} \vec{v} = (-R \dddot{\varphi} \sin \varphi - R \ddot{\varphi}^2 \cos \varphi) \ \hat{x} + (R \dddot{\varphi} \cos \varphi - R \dot{\varphi}^2 \sin \varphi) \ \hat{y} = -R \omega^2 (\cos \varphi \cdot \hat{x} + \sin \varphi \cdot \hat{y}) + R \alpha (- \sin \varphi \cdot \hat{x} + \cos \varphi \cdot \hat{y}),
\]

where we just introduced angular acceleration as the time derivative of angular velocity:

\[
\alpha(t) := \dot{\omega}(t) = \ddot{\varphi}(t).
\]  

(3.4)

We see that the acceleration vector can be split into two components, one along the time-dependent unit vector

\[
\ddot{\varphi} = - \sin \varphi \cdot \hat{x} + \cos \varphi \cdot \hat{y}
\]  

(3.5)
and the other along the time-dependent unit vector

\[ \hat{r} = \cos \varphi \cdot \hat{x} + \sin \varphi \cdot \hat{y}, \quad (3.6) \]

thus

\[ \vec{a} = R\alpha \hat{\varphi} - R\omega^2 \hat{r}. \]

You are welcome to verify that \( \hat{r} \) is always directed radially along the position vector, that \( \hat{\varphi} \) is always at a right angle to \( \hat{r} \) pointing in the direction of increasing polar angle \( \varphi \) and that both are actually unit vectors. In other words: The two vectors \( \hat{r} \) and \( \hat{\varphi} \) form a set of orthogonal unit vectors the body carries around as it moves along the circle. The corresponding components of the acceleration vector along these unit vectors are the tangential acceleration

\[ \vec{a}_t = R\alpha \hat{\varphi} \]

and the generalized centripetal acceleration

\[ \vec{a}_c = -R\omega^2 \hat{r}, \]

such that the total acceleration vector is written as their sum:

\[ \vec{a} = \vec{a}_t + \vec{a}_c. \]

In contrast to the case of uniform circular motion, the acceleration vector now has a tangential component in addition to the radial one. Recall how, in the uniform case, centripetal acceleration was only responsible for changing the direction of the velocity vector, thereby keeping the particle on the circle, but had no effect on the magnitude of velocity. In the case of non-uniform motion, however, an additional tangential component is responsible for the change in angular velocity.

The expression for centripetal acceleration is just a time-dependent version of the uniform case and is responsible for changing the direction of the velocity vector, whereas tangential acceleration changes its magnitude.

Again, a summary will be useful:

Consider a particle moving along a circular trajectory of radius \( R \). At any time, the velocity vector is given by

\[ \vec{v} = R\omega \hat{\varphi} \quad (3.7) \]

The acceleration vector can be split up into two components:

\[ \vec{a} = \vec{a}_t + \vec{a}_c \quad (3.8) \]

where

\[ \vec{a}_t = R\alpha \hat{\varphi} \quad (3.9) \]

is the tangential acceleration and

\[ \vec{a}_c = -R\omega^2 \hat{r} \quad (3.10) \]

the centripetal acceleration.
According to the second law of motion, the equation of motion of a particle of mass $m$ moving along a circular trajectory of radius $R$ is generally given by

$$\vec{F} = mR\alpha \hat{\phi} - mR\omega^2 \hat{r},$$

(3.11)

where $\vec{F}$ is the total force acting on the particle. As with the acceleration vector, we can decompose $\vec{F}$ into a radial component along $\hat{r}$ and a tangential component along $\hat{\phi}$. Each of these components will be responsible for the corresponding acceleration.

Circular motion also has a very interesting property that can be useful for problem solving:

If, in a given coordinate system, the position and velocity vector of a particle are at any time orthogonal to each other, then the particle’s trajectory lies in a circle around the origin of the coordinate frame. Furthermore if the acceleration vector is also perpendicular to the velocity vector at any time, then this circular motion is even uniform.

To see why this holds, consider how the square of the position vector of the particle, which is the square of its distance from the center of the coordinate system, changes over time:

$$\frac{d}{dt} |\vec{r}|^2 = \frac{d}{dt} (x^2 + y^2 + z^2) = 2 (x \dot{x} + y \dot{y} + z \dot{z}) = 2 \vec{r} \cdot \vec{v} = 0$$

where we made use of the chain rule of differentiation and of the fact that the scalar product $\vec{r} \cdot \vec{v}$ vanishes at any time since the two vectors are assumed to be perpendicular to each other. This means that the distance $|\vec{r}|$ of the particle from the origin of the coordinate frame is constant over time as its time derivative vanishes. An analogous calculation shows that with the condition $\vec{a} \cdot \vec{v} = 0$ at any time, the magnitude $|\vec{v}|$ of the velocity vector is also a constant, that is, circular motion is uniform.

### 3.2 Frames of Reference

#### 3.2.1 Choosing a Frame of Reference

You might have realized, or just know from school, that kinematic quantities are not *absolute* in that they depend on the point of view from which they are measured. Take, for example, a person walking the deck of a cruise ship. The velocity of that person with respect to someone leaning on the railing might be, say, $4\text{ km h}^{-1}$. With respect to someone walking towards them, to a seagull flying over the ship, or to an island in the sea, the velocity of that person might be very different. In physical terms, we say that velocities and all other kinematic quantities are *relative*. We say that they depend on the frame of reference (or reference frame) in which they are measured.

Even though the two terms are very similar, a reference frame is *not the same as a coordinate frame*. You can think of the reference frame as of a certain point of view, whereas a coordinate frame is more of a mathematical tool associated with a reference frame.\(^1\)

\(^1\)Again, the true story is way more complicated. And again… Don’t worry about that.
3.2. FRAMES OF REFERENCE

Take, for example, a situation often encountered in physics competitions: a rigid body, such as a sphere, rolling down an inclined plane. As you can read in section 3.4.5, such problems are best described in a reference frame where the inclined plane is at rest, such that the only object moving in this frame is the rolling body. For the choice of a coordinate system, we have many options. On a first thought, it seems reasonable to take an immobile coordinate frame with the $xy$-plane parallel to the ground and the $z$-axis pointing upwards. On the other hand, it turns out that choosing a moving coordinate frame with $xy$-plane parallel to the incline such that its origin coincides with the center of the rolling body and the $x$-axis points along the direction of rolling is very useful. Remember, the reference frame is still at rest with respect to the inclined plane. It is just the set of coordinates you will use to describe various points inside the rolling body that moves around. You’ll see more about this in section 3.4.

Let’s figure out how kinematic quantities can be converted from one reference frame to another. Consider a point-like particle moving in a reference frame we call $A$ and a coordinate system associated with it. We shall write $\vec{r}_A$ for the position of the particle with respect to $A$. Let $B$ be a second reference frame with associated coordinate system, such that at any time, the origin of this system is located at position $\vec{r}_B^{(A)}$ with respect to the origin of the coordinate system associated with $A$. Conversely, the origin of $A$’s coordinates lies at position $\vec{r}_A^{(B)}$ with respect to $B$’s coordinates. It is not too difficult to see that

$$\vec{r}_A^{(B)} = -\vec{r}_B^{(A)}$$

must hold. The particle will have two different position vectors with respect to the two systems, let’s call them $\vec{r}^{(A)}$ and $\vec{r}^{(B)}$, respectively. From a simple geometrical consideration, it follows that

$$\vec{r}^{(B)} = \vec{r}^{(A)} + \vec{r}_A^{(B)}$$
$$= \vec{r}^{(A)} - \vec{r}_B^{(A)}.$$  \hspace{1cm} (3.13)

Such a relationship is called a coordinate transformation between the two frames of reference $A$ and $B$. By taking its time derivative, we obtain a transformation rule for the velocity of the particle:

$$\vec{v}^{(B)} = \vec{v}^{(A)} + \vec{v}_A^{(B)}$$
$$= \vec{v}^{(A)} - \vec{v}_B^{(A)}.$$  \hspace{1cm} (3.14)
Similarly, for accelerations,

$$\vec{a}^{(B)} = \vec{a}^{(A)} + \vec{a}_A^{(B)} = \vec{a}^{(A)} - \vec{a}_B^{(A)}.$$  \hfill (3.15)

Make sure you develop an intuitive approach to such transformations by thinking about some examples. What about transformation of forces? This question is somewhat more difficult. In our framework, we regard forces as fundamental interactions between bodies, such as gravity or electrostatic attraction. Therefore it seems natural to assume that a force vector does not depend on the frame of reference. That is, if $\vec{F}^{(A)}$ denotes a force measured in frame $A$ and $\vec{F}^{(B)}$ the same force measured in frame $B$, then

$$\vec{F}^{(A)} = \vec{F}^{(B)}$$  \hfill (3.16)

must hold. Let’s say that this force acts on a body of mass $m$, which we also assume to be a quantity independent of the frame of reference. NEWTON’s second law would state that $\vec{F}^{(A)} = m\vec{a}^{(A)}$ and $\vec{F}^{(B)} = m\vec{a}^{(B)}$, which just implies $\vec{a}^{(A)} = \vec{a}^{(B)}$. We see from (3.15) that this can only hold if the relative acceleration $\vec{a}_A^{(B)}$ vanishes, that is, if the two systems move with constant velocity with respect to each other. The case where the relative acceleration does not vanish, however, poses us a bigger problem: either NEWTON’s second law is not valid or forces must somehow transform between frames. Either way, we need to rethink our concepts.

### 3.2.2 Accelerated Frames of Reference

Imagine, for example, to be sitting on a bus at night, getting home from an exhausting training session at the Physics Olympiad, as the cutest kitten suddenly appears on the street a couple of meters ahead of the bus. The driver, who is very alert despite late hours, almost immediately pushes down the brake pedal. You feel how your body is somehow dragged forward and you almost fall off your seat. It feels as if some magic force acting on your body had appeared at the instant of braking. Anyway, thanks to the fast reflexes of the driver, a fatal accident could be prevented. Some minutes later, as your blood adrenalin levels get back to normal, you start thinking. By what was the force exerted that made you fall forward? Most probably, it was not the bus, as you could still feel a drag even after your back had lifted off the seat. Was it the air? Guess not, how is that supposed to work? On the other hand, think about how the mother of the kitten might have seen it from the roadside. For her, as much as she might have been scared, the whole situation was nothing but a manifestation of NEWTON’s first law of motion. The bus would decelerate, and you would just keep moving at your initial speed along the direction of travel, no magic force...
involved. Still, from your point of view, the magic force must have existed, as otherwise Newton’s first law would have been violated in your reference frame. Indeed, as we saw at the end of the previous section, it turns out that Newton’s laws of motion are not valid in all frames of reference. Examples like the situation with the bus and the kitten make us realize that there might be two different kinds of frames of reference: the ones where Newton’s laws are valid, which we call inertial frames, and those where they aren’t. From the considerations at the end of the previous section, we see that non-inertial frames are nothing but just accelerated frames, or, more precisely, they are frames accelerated with respect to any inertial frame. The frame of reference of the kitten’s mother is an inertial frame as also one associated with the road or the earth would be, approximately. Your frame of reference on the bus is accelerated with respect to these inertial frames, which is why Newton’s first law loses its validity and you feel an acceleration in forward direction even without any force acting.

Fortunately there is a trick enabling us to rescue Newton’s second law in accelerated frames. Let $S$ be an inertial frame and $S'$ a frame moving with constant acceleration $\vec{a}_{S}$ with respect to $S$. Consider a particle moving with acceleration $\vec{a}$ in $S$ and $\vec{a}'$ in $S'$. From (3.15), we know that $\vec{a}' = \vec{a} - \vec{a}_{S}^{(S)}$. If a force $\vec{F}$ acts on this particle, which is the same force in $S'$ as in $S$, we know from Newton’s second law in $S$ that $\vec{F} = m\vec{a}$, where $m$ is the mass of the particle. Therefore,

$$m\vec{a}' = m\vec{a} - m\vec{a}_{S}^{(S)} = \vec{F} - m\vec{a}_{S}^{(S)}.$$

In $S'$, the second law is satisfied except for an extra term $-m\vec{a}_{S}^{(S)}$ having the dimension of a force. We might as well just cheat and deliberately call this term a force too. This force would only exist in $S'$ and not in $S$. That’s why we refer to it as a fictitious force.

Let’s apply this concept to our example with the bus and the kitten. $S$ shall be the frame of the road, $S'$ the frame of the bus, and $\vec{a}_{\text{brake}}$ the acceleration of the bus in $S$ as it brakes, pointing opposite to the direction of travel. The fictitious force acting on your body is then given by $-m\vec{a}_{\text{brake}}$, where $m$ is your mass, and points towards the front of the bus, and your equation of motion in $S'$ becomes $m\vec{a}'^{(S')} = -m\vec{a}_{\text{brake}}$ for your acceleration $\vec{a}^{(S')}$. Finally, the fictitious force enables us to explain from the point of view of $S'$ the forward drag you could feel in during the braking manoeuvre in a fashion more consistent with Newton’s laws.

Let $S'$ be a frame moving with acceleration $\vec{a}_{\text{rel}}$ with respect to the inertial frame $S$. The validity of Newton’s laws of motion in $S'$ is preserved if you introduce a fictitious force given by

$$\vec{F}_{f} = -m\vec{a}_{\text{rel}} \quad (3.17)$$

acting on every body in $S'$.

As you know, the earth is not really an inert object. With respect to the sun, it moves along an elliptical orbit spinning about itself, meaning that any frame associated with the earth is actually an accelerated frame. The solar system rotates about the center of the milky way, and the milky way itself takes part in the expansion of the universe. The reason
why we can often treat the earth as an inertial frame is that the fictitious forces involved are in general extremely small compared to the other forces considered in the problem. I’m sure you realize that going through some examples to quantify this statement would be a nice exercise. Things get somewhat different when space or time scales involved are much larger though. For problems dealing with planetary motion as in section 3.5, the accelerated motion of the earth has to be taken into account.

One more remark about rotational motion. Consider a particle of mass \( m \) rotating uniformly with angular velocity \( \omega \) along a circular trajectory of radius \( r \) in some inertial frame. For the particle to move this way, the total force \( F \) acting on it must account as a centripetal force for centripetal acceleration, that is, \( F = m\omega^2/r \). Look at what happens if you describe the situation in a frame rotating with the particle. This frame is accelerated by \( \omega^2/r \) with respect to any inertial frame, that’s why one has to take into account a fictitious force of magnitude \( m\omega^2/r \) directed opposite to this acceleration. This fictitious force points in outward radial direction, that’s why we call it the centrifugal force acting on the particle. In the rotating frame, the total force acting on the particle is just \( F - m\omega^2/r = 0 \), which is consistent with the fact that the particle is at rest in this system.

### 3.3 Systems of Particles

Don’t worry, you won’t spend your whole physics Olympiad career working on problems dealing with point-like masses and particles only. I bet you are eager to discover how to deal with questions concerning more interesting objects such as extrasolar planets, hula hoops, dumbbells, revolving doors and spinning tops. In general, such objects can be described as systems consisting of a given number \( n \) of point masses \( m_i \), \( i = 1, \ldots, n \), where \( n \) may be just equal to 2 or a number so large that writing \( \infty \) instead would save a lot of space. With knowledge from the previous chapter about Newton’s laws of motion, one should basically be able to formulate equations of motion for all point masses out of which such a system is composed. However, there are two major difficulties:

- Generally, a system is composed of many particles, that is, \( n \) is very large. It’s o.k. as long as you are dealing, say, with a system composed of the sun and the earth only, each of which can often be treated as a single point mass (we shall see why very soon), i.e. \( n = 2 \) and you have to handle two equations of motion only. But as soon as you start working on something like a wooden sphere rolling down an inclined plane or a figure skater spinning on the rink, you might realize that solving an almost infinite number of equations of motion can give you quite a headache. Therefore, problem number one is the often insanely huge number of point masses involved.

- Secondly, let’s take a closer look at the equations of motion we would formulate for such a system. On each particle of mass \( m_i \), a force \( \vec{F}_i \) shall act according to

\[
m_i\vec{a}_i = \vec{F}_i, \quad i = 1, \ldots, n
\]

Let’s take as an example a billiard sphere rolling freely on a table. The forces \( \vec{F}_i \) are of course not the same for all particles inside the sphere. While every single one
experiences gravitational acceleration $m_i \vec{g}$ from the earth, only the particles in direct contact with the table will experience friction. That shouldn’t be a great problem, as you know how to calculate a friction force as soon as you know the normal force and the friction coefficient. There is more to it though. A rolling billiard sphere remains a billiard sphere thanks to internal forces holding together all of its particles. There is no chance how you would be able to determine these internal forces without exact knowledge of the position and velocity of every single particle in space. That is, problem number two is lack of knowledge about internal forces of the system.

As we will see, we can overcome these difficulties thanks to a set of laws that enable us to treat systems of point masses as if they were concentrated in one single particle.

### 3.3.1 Newton’s Second Law for Systems of Particles and the Center of Mass

For a start, we will look at the effect of forces on motion of a system consisting of two point masses $m_1$ and $m_2$ only. We shall later see how our results can be generalized to any arbitrary number $n$ of particles, but let’s keep things simple first.

Let $\vec{F}_1$ and $\vec{F}_2$ denote the total forces acting $m_1$ and $m_2$, respectively. According to Newton’s second law, if we assume the masses of the particles to be constant over time,

$$
\vec{F}_1 = m_1 \frac{d^2 \vec{r}_1}{dt^2}, \\
\vec{F}_2 = m_2 \frac{d^2 \vec{r}_2}{dt^2}.
$$

(3.18)

So far, nothing new. In case we know how to compute the forces $\vec{F}_1$ and $\vec{F}_2$, it is not too difficult to set up the equations of motion. You might have realized that the number of equations involved is $3 \times 2 = 6$, one for each of the 3 coordinates of the 2 particles. O.k., this isn’t too much, but handling six equations simultaneously can take you quite some time – that’s why we might want to look for some simplification by digging a bit deeper into our formulas.

One idea is to consider the particle system and its surroundings as two different entities undergoing some kind of interaction. Take the forces $\vec{F}_1$ and $\vec{F}_2$, for instance. In general, the force acting on the $i$-th particle can be decomposed into one total internal force $\vec{F}_{i,\text{int}}$ arising from the rest of the system itself, that is, exerted by all other particles of the system on the $i$-th particle, and one total external force $\vec{F}_{i,\text{ext}}$ arising from the surroundings of the system. The internal force can be gravity, electrostatic attraction, a superposition thereof, a force two particles experience only during a very short time interval as they collide, and many others. The external force can be, for example, gravitational attraction from the earth, a Lorentz force exerted by a magnetic field, or air friction. For now, we do not care too much about the nature of these forces, but merely about the decomposition

$$
\vec{F}_i = \vec{F}_{i,\text{ext}} + \vec{F}_{i,\text{int}}.
$$

(3.19)
With this in mind, we now look at what happens if we sum the two equations of motion (3.18) and rearrange stuff using our decomposition:

\[
\vec{F}_1 + \vec{F}_2 = (\vec{F}_{1, \text{ext}} + \vec{F}_{1, \text{int}}) + (\vec{F}_{2, \text{ext}} + \vec{F}_{2, \text{int}}) = (\vec{F}_{1, \text{ext}} + \vec{F}_{2, \text{ext}}) + (\vec{F}_{1, \text{int}} + \vec{F}_{2, \text{int}})
\]

for the left-hand side and

\[
m_1 \frac{d^2}{dt^2} \vec{r}_1 + m_2 \frac{d^2}{dt^2} \vec{r}_2 = \frac{d^2}{dt^2} (m_1 \vec{r}_1 + m_2 \vec{r}_2) = (m_1 + m_2) \frac{d^2}{dt^2} \frac{m_1 \vec{r}_1 + m_2 \vec{r}_2}{m_1 + m_2}
\]

as a very fancy-looking version of the right-hand side. Together,

\[
(\vec{F}_{1, \text{ext}} + \vec{F}_{2, \text{ext}}) + (\vec{F}_{1, \text{int}} + \vec{F}_{2, \text{int}}) = (m_1 + m_2) \frac{d^2}{dt^2} \frac{m_1 \vec{r}_1 + m_2 \vec{r}_2}{m_1 + m_2}.
\]

Now stare at the internal forces \( \vec{F}_{1, \text{int}} \) in this last equation for some time. In our case of just two particles, we might as well write \( \vec{F}_{2 \rightarrow 1} \) instead of \( \vec{F}_{1, \text{int}} \) and something analogous for the second particle. From NEWTON’s third law, we know that \( \vec{F}_{1 \rightarrow 2} + \vec{F}_{2 \rightarrow 1} = 0 \), that is, the sum of all internal forces of the system vanishes. The right-hand side also has something to offer: define a point in space given by

\[
\vec{r}_{CM} := \frac{m_1 \vec{r}_1 + m_2 \vec{r}_2}{m_1 + m_2}
\]

and introduce the total mass \( M = m_1 + m_2 \) of the system and the total external force \( \vec{F}_{\text{ext, tot}} = \vec{F}_{1, \text{ext}} + \vec{F}_{2, \text{ext}} \) acting on it. Our equation can then be rewritten in a rather appealing fashion:

\[
\vec{F}_{\text{ext, tot}} = M \frac{d^2}{dt^2} \vec{r}_{CM}.
\]

As you can see, by just manipulating the equations of motion of the two particles and by introducing a nice notation, we were able to retrieve a simple equation containing a lot of information about motion of the system. The beauty of the equation lies in the following statement:

Motion of a system of two point-like masses \( m_1 \) and \( m_2 \) can be described by NEWTON’s second law applied on a particle of mass \( M = m_1 + m_2 \) at position of the center of mass \( \vec{r}_{CM} \) of the two particles:

\[
\vec{F}_{\text{ext, tot}} = M \frac{d^2}{dt^2} \vec{r}_{CM},
\]

where \( \vec{F}_{\text{ext, tot}} \) denotes the sum of all external forces acting on the two particles.
In other words, we were able to reduce the system of two particles to motion of one single virtual particle of mass $M$ at position $\mathbf{r}_{CM}$. We know from section 2.3.3 about Newton’s laws how to handle motion of single point masses. You will see in the examples how this reduction comes in very handy. But first, let’s take a closer look at our equation.

The vector defined by (3.20) denotes a point in space called the center of mass of the two particles. You can think of the center of mass as a point in space at any time associated with the two particles in motion. Its position with respect to either particle varies as the particles move.

In order to develop an intuitive approach to the concept of center of mass, you might want to work on the following questions:

- For a one-dimensional system, that is, two particles $m_1$ and $m_2$ moving along a straight line, where does the center of mass lie with respect to the line?
- What is its distance to either particle?
- What if $m_1 > m_2$? What if $m_1 \gg m_2$? What if $m_1 = m_2$?
- If you haven’t already done so, introduce a coordinate axis along the line connecting the two particles with origin at the position of $m_1$. What does the expression for the coordinate of the center of mass with respect to this axis look like?

It is not too difficult to perform similar calculations for an arbitrary number $n$ of particles of masses $m_i$ at positions $\mathbf{r}_i$. It turns out that one can define the center of mass of the system as

$$\mathbf{r}_{CM} = \frac{\sum_i m_i \mathbf{r}_i}{\sum_i m_i}$$

(3.22)

and that the equation of motion of the system assumes the following form:

**Newton’s Second Law for a System of Particles**

$$\mathbf{F}_{ext, tot} = M \frac{d^2}{dt^2} \mathbf{r}_{CM},$$

(3.23)

where $\mathbf{F}_{ext, tot}$ is the sum of all external forces acting on the system and $M = \sum_i m_i$ its total mass. Alternatively,

$$\mathbf{F}_{ext, tot} = \frac{d}{dt} \mathbf{p}_{CM},$$

(3.24)

where

$$\mathbf{p}_{CM} = \sum_i \mathbf{p}_i = M \mathbf{v}_{CM}$$

(3.25)
is at the same time the total momentum of the system and the momentum of its center of mass.

Here again, you might want to spend some more time thinking about the concept of center of mass for a system consisting of more than just two particles.

Take, for instance, \( n \) equal point masses \( m \) placed on the edges of a regular polygon.

- Where does the center of mass of the polygon lie?
- What happens to its position if the mass of one of the particles increases or decreases, the others kept fixed?

Equation (3.23) has an interesting consequence:

If the sum of all external forces acting on a system of particles vanishes, the speed of their center of mass remains constant.

Consider, for instance, a radioactive nucleus decaying in flight into three particles going off in separate directions. If the nucleus is far away from any other object that could interact with it, the speed of the center of mass of the three decay products must be the same as the speed of the nucleus before decay.

Consider a system consisting of various extended objects we call \( A_1, A_2, \ldots \), such as blocks of wood, soccer balls or human beings, of masses \( m_1, m_2, \ldots \). Each of these single objects is itself a system consisting of many particles, let’s say \( A_1 \) consists of point masses \( m_{A_1,i} \) located at \( \vec{r}_{A_1,i} \) and so on for \( A_2 \) etc. The center of mass of \( A_1 \) lies at \( \vec{r}_{CM,1} = (\sum_i m_{A_1,i} \vec{r}_{A_1,i})/m_1 \), similarly for \( A_2 \) etc. The center of mass of the whole system is then obtained as

\[
\vec{r}_{CM} = \frac{\sum_i m_{A_1,i} \vec{r}_{A_1,i} + \sum_i m_{A_2,i} \vec{r}_{A_2,i} + \cdots}{\sum_i m_{A_1,i} + \sum_i m_{A_2,i} + \cdots} = \frac{m_1 \vec{r}_{CM,1} + m_2 \vec{r}_{CM,2} + \cdots}{m_1 + m_2 + \cdots},
\]

which just means the following:

The center of mass of a system of extended objects \( A_1, A_2, \ldots \) of masses \( m_1, m_2, \ldots \) whose centers of mass are located at \( \vec{r}_{CM,1}, \vec{r}_{CM,2}, \ldots \), respectively, lies at

\[
\vec{r}_{CM} = \frac{m_1 \vec{r}_{CM,1} + m_2 \vec{r}_{CM,2} + \cdots}{m_1 + m_2 + \cdots}, \tag{3.26}
\]

meaning that each of the objects contributes to the position of the center of mass of the whole system as if it was a point-like particle concentrated in its own center of mass.
3.3.2 Conservation of Linear Momentum

We already know from the last chapter that the center of mass of a system of particles moves at constant speed if the sum of all external forces acting on it vanishes. That is,

\[ M \vec{v}_{CM} = \sum_i m_i \vec{v}_i = \sum_i \vec{p}_i \]

is constant. What looks very trivial is in fact one of the most fundamental laws in mechanics, the Law of Conservation of Linear Momentum.

If the sum of all external forces acting on a system of particles vanishes, the total momentum of the system is constant over time. In particular, the total momentum of an isolated system not undergoing any interaction with its surroundings is constant over time.

I’m pretty sure you already know this law from your high school physics and how you can apply it to problems. In particular, you might know that conservation of linear momentum plays an important role for problems involving collision, often together with the law of conservation of energy.

Just as a reminder: elastic collisions are the ones where kinetic energy, and thus total mechanical energy, is conserved. An inelastic collision occurs as soon as kinetic energy is dissipated by the impact. Whenever particles stick together in a collision, you know that the collision process must be inelastic: Suppose a particle of momentum \( m_1 \vec{v}_1 \) collides with a particle \( m_2 \) at rest to build a larger particle of momentum \( (m_1 + m_2) \vec{v}' \). Conservation of momentum states that \( m_1 \vec{v}_1 = (m_1 + m_2) \vec{v}' \), from which

\[ \vec{v}' = \frac{m_1}{m_1 + m_2} \vec{v}_1. \]

The corresponding kinetic energies are \( E_{\text{kin, start}} = m_1 v_1^2 / 2 \) before and \( E_{\text{kin, end}} = (m_1 + m_2) v'^2 / 2 \) after the collision, that is,

\[
\frac{E_{\text{kin, end}}}{E_{\text{kin, start}}} = \frac{(m_1 + m_2) v'^2 / 2}{m_1 v_1^2 / 2} = \frac{(m_1 + m_2) \cdot \left( \frac{m_1}{m_1 + m_2} \right)^2 v_1^2}{m_1 v_1^2} = \frac{m_1}{m_1 + m_2} < 1,
\]

meaning that kinetic energy is always dissipated during a collision process after which particles stick together.
3.3.3 Conservation of Angular Momentum

Very often, you’ll have to deal with problems taking place in a plane, that is, where all vectorial quantities such as positions, momenta and forces as well as any of their time derivatives can be described as two-dimensional. Moreover, such problems often involve what we call central forces, that is, forces always directed to a fixed point of the plane. For instance, these problems might involve a planet subject to gravitational attraction by a star or an electrically charged particle subject to COULOMB interaction with a large central charge. Both gravity and the electrostatic COULOMB interaction are central forces. Let \( \vec{r} \) be the position of a particle and \( \vec{F} \) such a central force acting on it in a two-dimensional situation. If we take the origin of a coordinate system at the center towards which the force \( \vec{F} \) is directed, then \( \vec{F} \) and \( \vec{r} \) must be parallel at any time. In terms of their components in \( xy \)-coordinates in this plane,

\[
r_x : r_y = F_x : F_y,
\]
as long as both \( r_y \) and \( F_y \) are nonvanishing. This condition can be rewritten as

\[
r_x F_y - r_y F_x = 0,
\]
and you can verify that this still holds if one of \( r_y \) or \( F_y \) is zero. With NEWTON’s second law, we can write

\[
0 = r_x F_y - r_y F_x = r_x p_y - r_y p_x
\]

\[
= v_x p_y + r_x p_y - v_y p_x - r_y p_x
\]

\[
= \frac{d}{dt} (r_x p_y - r_y p_x),
\]
meaning that the quantity given by \( r_x p_y - r_y p_x \) remains constant over time. You are totally right if you think that the expression \( r_x p_y - r_y p_x \) looks very familiar: it is nothing but the \( z \)-component of the cross product \( \vec{r} \times \vec{p} \). You might verify that the \( x \)- and \( y \)-component of this cross product are both 0 at any time, since the position and momentum vector of the particle are always assumed to lie in the \( xy \)-plane. Therefore, we can say that the vectorial quantity

\[
\vec{L} = \vec{r} \times \vec{p}
\]

remains constant over time as long as the particle is subject to central forces only. This quantity is called the angular momentum of the particle with respect to the origin of the coordinate system.

We might want to generalize this result to particles not necessarily constrained to move in a plane. Consider a particle moving with momentum \( \vec{p} = m \vec{v} \) and position vector \( \vec{r} \) subject to a total force \( \vec{F} \). If we take the time derivative of its angular momentum vector \( \vec{L} = \vec{r} \times \vec{p} \),
we get
\[
\frac{d}{dt} \vec{L} = \frac{d}{dt} (\vec{r} \times \vec{p}) \\
= \left( \frac{d}{dt} \vec{r} \right) \times \vec{p} + \vec{r} \times \left( \frac{d}{dt} \vec{p} \right) \\
= \vec{v} \times \vec{p} + \vec{r} \times \vec{F} \\
= \vec{r} \times \vec{F},
\]
where we used the second law of motion \(d\vec{p}/dt = \vec{F}\) and the product rule of differentiation for vector products\(^2\). Note that the velocity \(\vec{v}\) and momentum vector \(\vec{p}\) are parallel at any time, their cross product vanishes. The quantity \(\vec{r} \times \vec{F}\) is called the **torque** \(\vec{\tau}\) by which \(\vec{F}\) acts on the particle with respect to the origin. You probably know about the torque and its meaning from school, that’s why we won’t discuss it too extensively here. With this definition, we now have a very elegant law:

\[
\frac{d}{dt} \vec{L} = \vec{\tau}. \tag{3.28}
\]

The quantity \(\vec{L}\) can in some way be understood as an ‘amount of rotational motion’, similar to linear momentum as a quantity describing the ‘amount of linear motion’ – that’s why \(\vec{L}\) is called **angular** momentum. On the other side, the torque is nothing but a measure for the extent to which a force is able to change angular motion of a rotating object. It vanishes if the force is radial, that is when \(\vec{r}\) and \(\vec{F}\) are parallel, and it is maximal for a force at right angles to the position vector.

The relationship of (3.28) somehow reminds us of NEWTON’s second law, where the force acting on a particle and its linear momentum were linked. Because of this analogy, we call (3.28) **NEWTON’s second law for angular motion**.

Let’s apply (3.28) to a system of masses \(m_i\) by adding up the equations for every single particle:

\[
\frac{d}{dt} \sum_i \vec{r}_i \times \vec{p}_i = \sum_i \vec{r}_i \times \vec{F}_i,
\]
where \(\vec{F}_i\) denotes the total force acting on the \(i\)-th particle at position \(\vec{r}_i\). We now use the same trick as in section 3.3.1 and divide all forces acting on the particles of our system into an internal component \(\vec{F}_{i,\text{int}}\) and an external component \(\vec{F}_{i,\text{ext}}\). The calculation that follows is quite elaborate; you can find it below. Anyway, it turns out that all internal forces cancel out of the picture the same way they already did for linear motion in section 3.3.1.

---

\(^2\)Given any two vectorial quantities \(\vec{A}\) and \(\vec{B}\) dependent on time, the time derivative of their vector product satisfies the rule

\[
\frac{d}{dt} (\vec{A} \times \vec{B}) = \left( \frac{d}{dt} \vec{A} \right) \times \vec{B} + \vec{A} \times \left( \frac{d}{dt} \vec{B} \right),
\]

which you can verify by writing out the components of \(\vec{A} \times \vec{B}\) and taking their time derivatives.
The rate of change of the total angular momentum $\vec{L}$ of a system of particles is entirely determined by the total torque $\vec{\tau}_{\text{ext, tot}}$ acting on them due to forces external to the system:

$$\frac{d}{dt}\vec{L} = \vec{\tau}_{\text{ext, tot}} \tag{3.29}$$

Again, this leads us to one of the most central conservation laws of mechanics:

**Law of Conservation of Angular Momentum**

If the total torque due to external forces acting on a system vanishes, its angular momentum remains constant over time.

In particular:

- The total angular momentum of a system of particles subject to radial forces only is constant over time.
- The total angular momentum of an isolated system of particles is constant over time.

This example illustrates an important principle:

Conservation of angular momentum is a useful tool for problems where systems of point masses undergo some process where they rotate in space.

**Derivation of Newton’s Second Law for Angular Motion**

Consider a system, not necessarily a rigid body for now, of $n$ particles $m_i$. We denote their positions by $\vec{r}_i$, their velocities by $\vec{v}_i$ and their linear momenta by $\vec{p}_i = m_i\vec{v}_i$. The total angular momentum of this system is given by

$$\vec{L} = \sum_i \vec{r}_i \times \vec{p}_i.$$ 

We take its time derivative and get

$$\frac{d}{dt}\vec{L} = \sum_i \left( \frac{d}{dt}\vec{r}_i \times \vec{p}_i + \vec{r}_i \times \frac{d}{dt}\vec{p}_i \right)$$

$$= \sum_i \vec{v}_i \times \vec{p}_i + \sum_i \vec{r}_i \times \frac{d}{dt}\vec{p}_i$$

by the product rule. Now for all $i$ we have that

$$\vec{v}_i \times \vec{p}_i = \vec{v}_i \times (m_i\vec{v}_i) = m_i \vec{v}_i \times \vec{v}_i = 0$$

96
and the first sum above vanishes. For the second sum we use NEWTON’s second law for each single particle
\[ \frac{d}{dt} \vec{p}_i = F_i, \]
where \( F_i \) is the total force acting on the \( i \)-th particle. If we divide this force again into an external and an internal part, \( \vec{F}_{i,\text{ext}} \) and \( \vec{F}_{i,\text{int}} \), respectively, we get:
\[ \frac{d}{dt} \vec{L} = \sum_{i=1}^{n} \vec{r}_i \times \vec{F}_{i,\text{ext}} + \sum_{i=1}^{n} \vec{r}_i \times \vec{F}_{i,\text{int}} \]
and we easily identify the first sum as the external, the second as the internal torque. To get rid of the internal torque, we make the following observation: The forces between the particles are subject to NEWTON’s third law. We can split up each \( \vec{F}_{i,\text{int}} \) into the components arising from all particles but the \( i \)-th itself:
\[ \vec{F}_{i,\text{int}} = \sum_{j=1, j \neq i}^{n} \vec{F}_{j \rightarrow i}, \]
where \( \vec{F}_{j \rightarrow i} \) is the force by which particle \( j \) acts on particle \( i \). NEWTON’s third law states that
\[ \vec{F}_{j \rightarrow i} = -\vec{F}_{i \rightarrow j}. \]
Note that every term of the kind
\[ \vec{r}_i \times \vec{F}_{j \rightarrow i} \]
inside the double summation
\[ \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \vec{F}_{j \rightarrow i} \]
has a unique ‘counterpart’
\[ \vec{r}_j \times \vec{F}_{i \rightarrow j} \]
within the sum. We can group the sum into such pairs of terms and get
\[ \vec{r}_i \times \vec{F}_{j \rightarrow i} + \vec{r}_j \times \vec{F}_{i \rightarrow j} = (\vec{r}_i - \vec{r}_j) \times \vec{F}_{j \rightarrow i} \]
for each pair of distinct particles \( i \) and \( j \), from the third law. In our purely classical problems, the forces will be of gravitational or electromagnetic nature only. As such, they are known to act along the vector \( \vec{r}_i - \vec{r}_j \) joining two particles \( i \) and \( j \), meaning that the cross products
\[ (\vec{r}_i - \vec{r}_j) \times \vec{F}_{j \rightarrow i} \]
all vanish! The whole internal torque of the system therefore vanishes and we are left with the final result
\[ \frac{d}{dt} \vec{L} = \sum_{i=1}^{n} \vec{r}_i \times \vec{F}_{i,\text{ext}} = \vec{\tau}_{\text{ext}}, \]
which was to be shown.
3.4 Rigid Bodies

An example will show you why this section is even necessary. At a first glance, this looks like your standard problem involving NEWTON’s second law in which you have to first determine and then solve the equations of motion for an object. You might realize more or less rapidly that application of NEWTON’s second law for point masses doesn’t lead you anywhere: the rod is definitely not a point-like particle but an extended system of particles. Maybe one of the tools described in section 3.3 can help. Take (3.28), for instance. In order to apply this equation, we should calculate the total angular momentum of the rod by summing the angular momenta of all particles building it. This seems to be quite complicated.

The rod is a system of point masses of a very special kind in that its particles won’t change their positions relative to each other as the rod moves. Such objects are modelled in mechanics as what we call rigid bodies. For our purposes, only two properties of rigid bodies are important:

A rigid body is

- a system consisting of many particles
- such that the distance between any two of the particles remains constant over time.

In this section, we will see how the second condition gives rise to laws governing the motion of rigid bodies following from the laws already stated in the last section about system of particles.

In general, rigid bodies occurring in Olympiad problems will be more or less regular with a more or less symmetric shape, such as spheres, rods, hollow cylinders, prisms, or even various combinations thereof. Such objects are relatively easy to deal with from a mathematical point of view. In particular, competition problems will require you to answer questions concerning rotational motion of rigid bodies – that’s why they usually display some kind of rotational symmetry around some axis.

3.4.1 Kinematics of Rigid Bodies

First, we need to describe motion of rigid bodies in a useful way by means of kinematic quantities. Unlike point masses, rigid bodies do not only change their position in space by ‘mere displacement’, but they can also rotate. Indeed, general motion of a rigid body satisfies the following rule:

Consider a rigid body moving in space. One can always choose a point \( P \) associated with this body (and not necessarily a part of it), that is, the relative position of \( P \) with respect to any part of the rigid body does not change as the body moves, such that motion of the rigid body can be described as a superposition of
• translation of $P$ in space and

• rotation of the rigid body about an axis passing through $P$. This axis might change direction in space as the rigid body moves.

This division into translation and rotation is of fundamental importance, as different sets of equations govern these two components of motion.

Choice of the point $P$ is of course not unique. You will see in the examples throughout this section that often $P$ is best chosen as the center of mass of the rigid body, sometimes however as a point fixed in space during motion of the rigid body such that, in the corresponding set of coordinates, motion will be described as a pure rotation without any translational component. There is no better or worse choice of $P$, since in the end all choices must lead to physically equivalent results. You will find, however, that in certain situations some choices will come with less complicated algebra and lead you to the solution of a problem faster than other choices.

It is not hard to figure that rotational motion of a rigid body can, at any instant, be described as a change of orientation with respect to a certain axis in space. Imagine a rigid body rotating around an axis whose position in space does not change with time. If you look onto the body from either up or down along the axis, you can see how every particle of the body moves on a circular trajectory around the axis in a plane perpendicular to it. Let $\vec{v}_i$ be the speed of the $i$-th particle and $\vec{r}_{\perp i}$ a position vector pointing perpendicularly from the axis of rotation to the position of the particle, therefore lying in the plane where the $i$-th particle moves. The angular speed of rotation in the plane perpendicular to the axis is the same for all particles of the rigid body, as all inter-particle distances do not change over time. If we call this angular speed $\omega$, then the relationship

\[ v_i = r_{\perp i}\omega \]  

for the absolute value $v_i$ of the speed of the $i$-th particle holds. As long as the axis of rotation remains invariant in time, description of rotational motion of a rigid body is basically not much different from circular motion of a single particle in a plane. However, the rigid body is a three-dimensional object and forces acting on it causing changes in its motion will generally not lie in the same plane as the velocities of its point-like components. Therefore, it sounds reasonable to use a vectorial quantity for description of rotational motion of the rigid body. One could start by introducing a unit vector $\hat{n}$ pointing in the direction of the axis of rotation. This enables us to give (3.30) a vectorial character, since

\[ \vec{v}_i = \hat{n} \times \omega \vec{r}_{\perp i} \]

holds. If we introduce the angular velocity vector

\[ \vec{\omega} := \omega \hat{n}, \]  

(3.31)
this expression becomes very compact:

\[ \vec{v}_i = \vec{\omega} \times \vec{r}_{i\perp}. \]  

(3.32)

Notice how the angular velocity vector \( \vec{\omega} \) contains a lot of information about rotational motion of the body. Not only it shows in direction of the axis of rotation, but it also tells you how fast the body rotates and in which sense: from (3.32), you can deduce that, if you look through the axis of rotation in direction of \( \omega \), the body rotates clockwise. In other words:

If the long fingers of the right hand are curled in the sense of rotation of the body, then \( \vec{\omega} \) points in the direction shown by the stretched right thumb along the axis of rotation.

Something still doesn’t look too elegant in (3.32): that subscript \( \perp \) for position vectors. Suppose we work in such a coordinate frame that the axis of rotation passes through its origin. Every position vector \( \vec{r}_i \) relative to the origin can then be split into \( \vec{r}_i = \vec{r}_{i\perp} + \vec{r}_{i\parallel} \), where \( \vec{r}_{i\parallel} \) is a component of \( \vec{r}_i \) lying on the axis of rotation and \( \vec{r}_{i\perp} \) defined as before. With this notation, we see that

\[ \vec{v}_i = \vec{\omega} \times (\vec{r}_i - \vec{r}_{i\parallel}) = \vec{\omega} \times \vec{r}_i - \vec{\omega} \times \vec{r}_{i\parallel} \]

\[ = \vec{\omega} \times \vec{r}_i, \]

since \( \vec{r}_{i\parallel} \) is parallel to \( \vec{\omega} \) so that \( \vec{r}_{i\parallel} \times \omega = 0 \). In other words, if the origin of our coordinate system lies on the rotation axis,

\[ \vec{v}_i = \vec{\omega} \times \vec{r}_i \]

holds for all particles \( i \).

In physics, one often encounters the concept of degree of freedom, by which one usually means the number of parameters of a system that may vary independently. In mechanics, you can think of the degrees of freedom of an object as the number of ‘settings’ one has to adjust to give a unique description of its position in space. For instance, a given point particle in space has three degrees of freedom, as each of its three coordinates can be varied independently and all of them are necessary to locate it. A general system of \( n \) particles, accordingly, has \( 3n \) degrees of freedom, as long as all particles move independently of each other. A rigid body, no matter how many particles it is made of, has six degrees of freedom. Three coordinates are needed to describe the position of its center of mass, or any other point \( P \) defined as previously, and three more variables, like the three components of a unit vector along a certain axis of the body, to determine its orientation in space.
3.4. RIGID BODIES

3.4.2 How Forces Determine Translational Motion of a Rigid Body

Now that we have tools for description of motion of rigid bodies, we might as well get to the laws governing it.

The first law we look at is very simple and follows directly from the previous section about systems of particles:

Given the total sum of external forces \( \vec{F}_{\text{tot, ext}} \) acting on a rigid body of mass \( m \), the acceleration \( \vec{a}_{\text{CM}} \) of its center of mass is given by

\[
m \vec{a}_{\text{CM}} = \vec{F}_{\text{tot, ext}}.
\]

(3.33)

For a body to be rigid, a great deal of internal forces must act between its particle-like components to hold it together. As we saw from the discussion of systems of particles in section 3.3, these internal forces cancel out and we can consider motion of the center of mass only.

From (3.33), you can easily derive:

If the sum of all external forces \( \vec{F}_{\text{tot, ext}} \) acting on a rigid body vanishes, the velocity of its center of mass \( \vec{v}_{\text{CM}} \) remains constant.

This is analogous to Newton’s first law for point masses.

Even with (3.33), we have a substantial problem. External forces acting on a rigid body can be many, especially if a rigid body is subject to a force field like a gravitational field. In such a case, one would have to add all forces acting on every single point-like particle building the rigid body. Take a uniform gravitational field with acceleration \( \vec{g} \), for instance. The \( i \)-th particle of mass \( m_i \) experiences a weight \( m_i \vec{g} \), and, according to (3.33):

\[
\vec{a}_{\text{CM due to gravity}} = \frac{1}{m} \sum_i m_i \vec{g} = \left( \frac{1}{m} \cdot \sum_i m_i \right) \vec{g}
\]

\[
= \vec{g}.
\]

This means:

The effect of a homogeneous gravitational field \( \vec{g} \) on translational motion of a rigid body of mass \( m \) can be described as a force \( m \vec{g} \) acting on the center of mass of the body.

In other words: concerning translational acceleration in a homogeneous gravitational field, a rigid body can be treated as a particle of same mass as the total mass of the rigid body and at position of its center of mass.

That’s why the center of mass of a rigid body is often called its center of gravity. Note that as soon as one considers an inhomogeneous field, the concept of center of gravity does not make sense any more.
The center of mass of a rigid body can in principle be determined with equation (3.22), that is,

\[ \vec{r}_{CM} = \frac{\sum_i m_i \vec{r}_i}{\sum_i m_i}, \]

provided we know what kind of particles a rigid body is made of, what their masses are and how they are distributed inside the body. You see, the task of finding the center of mass of a rigid body does not seem to be very easy. We usually deal with objects in the macroscopic world, meaning that the number \( n \) of particles involved is extremely large and the mass \( m_i \) of one single particle extremely small. For once, this immensity is an advantage: it turns out that the exact number of particles and the exact mass of every single one of them is not very relevant as long as their distribution approximates the macroscopic distribution of mass density, which we will denote by the symbol \( \varrho \), of the rigid body. This quantity can be understood as follows: a very small volume \( dV \) of the rigid body located at \( \vec{r} \) carries the mass \( \varrho \cdot dV \). In general, a rigid body will be inhomogeneous, meaning that its mass is not evenly distributed across its volume and that \( \varrho \) becomes a quantity dependent on position. That’s why we might as well write \( \varrho(\vec{r}) \). The position of the center of mass can now be expressed as

\[ \vec{r}_{CM} = \frac{\sum_i \varrho(\vec{r}_i) \vec{r}_i dV_i}{M}, \]

with the choice of very small volume elements \( dV_i \) located at \( \vec{r}_i \). In the limit of these volume elements becoming infinitesimally small, you surely know that a sum such as in the right-hand side of the previous equation approaches an integral. That is, we can write

\[ \vec{r}_{CM} = \frac{1}{M} \int_{\text{rigid body}} \varrho(\vec{r}) \vec{r} \, dV \]

How this formula can be applied to rigid bodies of different shapes is left to your math courses. For now, it is relevant that you know about some important properties of the center of mass:

- The position of the center of mass always somehow reflects the symmetry of a body. The center of mass of a homogeneous sphere, for instance, is located at its geometrical center. The same holds for a homogeneous plane square or a homogeneous one-dimensional rod.

- In particular, the center of mass of a rigid body can lie outside the body. For example, the center of mass of a homogeneous circular hoop lies at its geometrical center.

- For a system consisting of various rigid bodies, equation (3.26) applies.
3.4.3 How Forces Determine Rotational Motion of a Rigid Body

In general, (3.33) won’t be sufficient to fully determine motion of a rigid body in space: in general, there are six degrees of freedom and (3.33) involves three coordinates only. In section 3.3, we developed a quantity strongly associated with rotations of systems of particles: it was angular momentum. Of course, equation (3.29) can also be applied to a rigid body:

\[ \frac{d}{dt} \vec{L} = \vec{\tau}_{\text{ext, tot}}. \]

The angular momentum \( \vec{L} \) of a rigid body about a certain point turns out to be quite difficult to compute, as one has add up \( \vec{r}_i \times \vec{v}_i \) for all particles inside the body (even turning the sum into an integral won’t help too much here). However, one can show that for some very symmetrical situations, the angular momentum of a rigid body rotating about a fixed axis at angular velocity \( \vec{\omega} \) about any point on the axis is just given by

\[ \vec{L} = I \vec{\omega}, \tag{3.34} \]

where \( I \) is the so-called moment of inertia of the body about this axis. O.k., now step by step:

- When does (3.34) apply? In situations with a high degree of symmetry such as
  - a sphere rotating about any axis in space,
  - a cylinder rotating about its own axis of rotational symmetry,
  - a cylinder rotating about any axis parallel to its rotational axis of symmetry
  - a cylinder rotating about any axis perpendicular to its rotational axis of symmetry
  - a cube rotating about one of its edges,
  - a single point-like particle rotating about any other point in space,

  and many others. You get the principle. Plus, the axis has to be fixed in space. The equation is not valid for axes of rotation moving around.

- What is \( I \)? In general, the moment of inertia of a rigid body about a given axis is calculated as

\[ I = \sum_i m_i r_{i\perp}^2, \tag{3.35} \]

where \( r_{i\perp} \) denotes the distance of the \( i \)-th particle from the axis of rotation. In many problems, rigid bodies won’t be simple rigid systems of point masses, but be extended systems with continuous mass distributions \( \rho \), and one can calculate \( I \) as...
\[ I = \int_{\text{rigid body}} r_{\perp}^2 \, dm = \int_{\text{rigid body}} \varrho (\vec{r}) \, r_{\perp}^2 \, dV. \] (3.36)

In most cases, you won’t have to calculate moments of inertia by yourself, they will usually be given along with the task. The SI unit of the moment of inertia is kg \cdot m^2.

- The moment of inertia is an additive quantity: If a rigid body consists of various parts \( A_1, A_2 \ldots \), the total moment of inertia \( I_{\text{tot}} \) of the whole body about an axis is just the sum of the moments of inertia \( I_i \) of its parts about this axis:

\[ I_{\text{tot}} = \sum_i I_i. \] (3.37)

- With respect to what point is \( \vec{L} \) calculated? Equation (3.34) is valid for any point along the axis of rotation as the origin. (You can easily see that \( \vec{L} \) must be the same for all points along the axis of rotation from )

- From (3.34) follows that in the symmetric cases where this law applies, the angular momentum of a rigid body is parallel to the axis of rotation. Note that this won’t be the case in general for situations not displaying a high degree of symmetry.

We won’t give a derivation of (3.34), so feel free to try to figure out how it arises (it is not too difficult if you make the right assumptions about symmetry of the system). Anyway, textbooks like [5] and [6] provide you full derivations and more advanced discussions of this topic.

For you to get more acquainted with the concept of moment of inertia, here are some examples.

- Let a point-like particle of mass \( m \) move along a circle of radius \( r \) with angular velocity \( \omega \). The moment of inertia of the particle about an axis perpendicular to the plane of the circle passing through its center is equal to \( mr^2 \). Note how in this case, with origin of the coordinate frame at the center of the circle, equation (3.34) becomes \( \vec{L} = mr^2 \vec{\omega} \), and you may verify that this is indeed equal to \( \vec{r} \times \vec{p} \) for a position vector \( \vec{r} \) with respect to the center of the circle and linear momentum \( \vec{p} \) of the particle. The second
law for rotational motion then states that

\[
\frac{d}{dt} \vec{L} = \vec{r}_{\text{ext, tot}}
\]

\[\Rightarrow \quad \frac{d}{dt} (mr^2 \vec{\omega}) = \vec{r} \times \vec{F}
\]

\[\Rightarrow \quad mr^2 \ddot{\omega} = rF_t
\]

\[\Rightarrow \quad F_t = \frac{mr}{m} \dot{\omega}
\]

where \( F_t \) denotes the tangential component of a force \( \vec{F} \) acting on the particle. This is a result we already derived in section 3.1 about non-uniform circular motion.

• Consider a homogeneous ring of mass \( m \) and radius \( r \). A small mass element \( dm \) of the ring has a moment of inertia \( r^2dm \) about an axis perpendicular to the plane of the ring passing through its center. The moment of inertia of the whole ring about this axis is obtained by summing up over all infinitesimal mass elements building it: \( \int_{\text{ring}} r^2 \, dm = r^2 \int_{\text{ring}} dm = mr^2 \).

• Consider a homogeneous flat disc of mass \( m \) and radius \( R \) and an axis perpendicular to it passing through its center. To calculate the moment of inertia of the disc about this axis, we can decompose it into many small rings of radius \( r \), \( 0 < r < R \), and infinitesimal thickness \( dr \). The area of such a ring is \( 2\pi r \, dr \), its mass therefore

\[
2\pi rd\!r \cdot \frac{M}{\pi R^2} = \frac{2rdr}{R^2} \cdot M
\]

and its moment of inertia about the central axis becomes \( 2Mr \, dr \cdot r^2 = 2Mr^3 \, dr/R^2 \).

The moment of inertia of the disc can be obtained by integrating over the moments of inertia of all these small rings:

\[
\int_{r=0}^{R} \frac{2Mr^3}{R^2} \, dr = \frac{1}{2} mR^2.
\]

• Consider a homogeneous cylinder of mass \( m \), radius \( R \), length \( L \) and its axis of rotational symmetry. For the corresponding moment of inertia, think of the cylinder as consisting of many circular discs of infinitesimal thickness \( d\ell \), each of mass \( d\ell \cdot m/L \), stacked on each other. The moment of inertia of every single disc is \( mR^2d\ell/(2L) \), and the total moment of inertia of the disc becomes

\[
\int_{\ell=0}^{L} \frac{mR^2}{2L} \, d\ell = \frac{1}{2} mR^2,
\]

a result independent of the length \( L \) of the cylinder.

Computation of more complex examples is left to your math courses.

The moment of inertia satisfies a property often coming in very handy:
Parallel Axis Theorem (Steiner’s Theorem)
Let $I_{CM}$ denote the moment of inertia of a rigid body of mass $M$ with respect to an axis passing through its center of mass and $I'$ be its moment of inertia about a second axis parallel to the first one at distance $d$ from the center of mass. The following relationship holds:

$$I' = I_{CM} + Md^2.$$  \hspace{1cm} (3.38)

With (3.34), equation (3.29) takes the form

**NEWTON’s Second Law for Rotation of Rigid Bodies**

$$I \frac{d}{dt} \alpha = \vec{\tau}_{\text{ext, tot}},$$ \hspace{1cm} (3.39)

where angular acceleration $\alpha = \dot{\omega}$ is introduced. This equation will be your most fundamental tool for solution of problems involving rotating rigid bodies.

In particular, let’s look at the effect of gravity. Consider a rigid body exposed to a homogeneous gravitational field $\vec{g}$. The total torque due to the weight $m_i \vec{g}$ of each particle constituting the body with respect to some axis is

$$\vec{\tau} = \sum_i \vec{r}_i \times \vec{g} = \left( \sum_i m_i \vec{r}_i \right) \times \vec{g} = \left( \sum_i m_i \right) \vec{r}_{CM} \times \vec{g},$$

that is,

The total torque due to gravity $\vec{g}$ acting on a rigid body of mass $M$ is obtained as

$$\vec{\tau} = M \vec{r}_{CM} \times \vec{g},$$  \hspace{1cm} (3.40)

where $\vec{r}_{CM}$ is the position of the center of mass of the rigid body.

Together with the result of section 3.4.2, we have a very important result:

For both translational and rotational motion, can treat the weight of a rigid body as if it was one single force acting on its center of mass.
3.4.4 Mechanical Energy of Moving Rigid Bodies

We didn’t talk too much about conservation of energy when we looked at systems of particles. That’s because, in general, the kinetic energy of a system of particles is not very easy to calculate, as one would have to sum over all terms $m_i v_i^2 / 2$ belonging to each particle:

$$E_{\text{kin}} = \frac{1}{2} \sum_i m_i v_i^2.$$  \hfill (3.41)

The situation turns out to look quite simple for rigid bodies though:

The kinetic energy of a rigid body of mass $M$ in motion can be split up into two components, one for translational motion of its center of mass and one for rotational motion about an axis passing through the center of mass:

$$E_{\text{kin}} = \frac{1}{2} M v_{CM}^2 + E_{\text{kin, rot}},$$  \hfill (3.42)

where $v_{CM}$ denotes the velocity of the center of mass. This result is derived below. In situations displaying a high degree of symmetry, the kinetic energy of rotation about the axis passing through the center of mass can be calculated as

$$E_{\text{kin, rot}} = \frac{1}{2} I \omega^2,$$  \hfill (3.43)

where $\omega$ is the angular velocity of this rotational motion and $I$ the moment of inertia of the rigid body about this axis of rotation.

Let’s say motion of a rigid body can be described as pure rotation of its center of mass about an axis distant $d$ from it at angular velocity $\omega$. At the same time, one can always describe such motion as the superposition of translation of the center of mass at speed $v_{CM} = \omega d$ about this axis and rotation of the body about another parallel axis passing through the center of mass at angular velocity $\omega$. According to (3.42), one can write

$$E_{\text{kin}} = \frac{1}{2} M (\omega d)^2 + \frac{1}{2} I \omega^2$$
$$= \frac{1}{2} \left( I + M d^2 \right) \omega^2$$
$$= \frac{1}{2} I' \omega^2,$$

where $I$ is the moment of inertia of the body about the axis passing through its center of mass and $I'$ about the other axis with $I' = I + M d^2$ according to the parallel axis theorem.
If motion of a rigid body can entirely be described as rotation of its center of mass about an axis at angular velocity $\omega$, its kinetic energy is then given by

$$E_{\text{kin}} = \frac{1}{2} I \omega^2,$$

where $I$ denotes the body’s moment of inertia about this axis of rotation.

We already saw how the weight of a rigid body can be treated for both translational and rotational motion. We turn now to potential energy of a rigid body in a homogeneous gravitational field. As usual, we can obtain the potential energy of a rigid body by adding up the potential energies of all masses $m_i$ it is made of:

$$E_{\text{pot}} = \sum_i m_i g h_i = g \cdot \sum_i m_i h_i = g \cdot \left( \sum_i m_i \right) h_{\text{CM}},$$

where $h_i$ is the height of the $i$-th particle and $h_{\text{CM}}$ the height of the center of mass above the zero of potential energy. That is,

Potential energy of a rigid body of mass $M$ in a homogeneous gravitational field $\bar{g}$ is given by

$$E_{\text{pot}} = M g h_{\text{CM}}$$

with the height $h_{\text{CM}}$ of the center of mass above the zero level of potential energy.

**Derivation of the Kinetic Energy Expression for Moving Rigid Bodies**

The total kinetic energy $E_{\text{kin}}$ of a rigid body moving as in 3.4.4 can be expressed as

$$E_{\text{kin}} = \sum_i \frac{1}{2} m_i v_i^2$$

where the $i$-th particle is of mass $m_i$ and moving at speed $\bar{v}_i$. If we split up motion into translation of the center of mass of the rigid body and rotation about an axis passing through the center of mass, we can write

$$\bar{v}_i = \bar{v}_{\text{CM}} + \bar{v}_i^{(\text{CM})},$$
where $\vec{v}_{CM}$ is the velocity of the center of mass of the rigid body and $\vec{v}_i^{(CM)}$ the velocity of the $i$-th particle with respect to the center of mass. Therefore,

$$E_{\text{kin}} = \frac{1}{2} \sum_i m_i \left( \vec{v}_{CM} + \vec{v}_i^{(CM)} \right)^2$$

$$= \frac{1}{2} \sum_i m_i \left( v_{CM}^2 + v_i^{(CM)2} + 2 \vec{v}_{CM} \cdot \vec{v}_i^{(CM)} \right)$$

$$= \frac{1}{2} \left( \sum_i m_i \right) v_{CM}^2 + \frac{1}{2} \sum_i m_i v_i^{(CM)2} + \vec{v}_{CM} \cdot \left( \sum_i m_i \vec{v}_i^{(CM)} \right)$$

$$= \frac{1}{2} m v_{CM}^2 + \frac{1}{2} \sum_i m_i v_i^{(CM)2}$$

### 3.4.5 Putting Everything Together: Rolling Rigid Bodies

One particular kind of motion of rigid bodies is rolling. We consider rigid bodies with a somehow circular cross section moving on a plane surface. Let’s say that rolling of a rigid body occurs at angular velocity $\omega$ about its geometrical axis of rotational symmetry, which is perpendicular to its direction of translational motion. You can convince yourself by drawing a sketch that with respect to an axis parallel to the first one but passing through the point of contact with the surface, every point of the rigid body also rotates at angular velocity $\omega$.

In particular, the geometrical center of the body moves at speed $\omega r$ with respect to the axis through the point of contact if $r$ is the radius of the body. The ever changing, instantaneous point of contact with the underlying surface is always at rest with respect to the surface. You will also agree that the condition for rolling without slipping to occur is that $r \omega$ is equal to translational velocity $v$ of the geometrical center of the body:

A rigid body of cross-sectional radius $r$ rolls without slipping on a surface if and only if the translational velocity $v$ of its geometrical center and its angular velocity $\omega$ of rotation about its geometrical center satisfy

$$v = \omega r. \quad (3.46)$$

Rolling motion of a rigid body can be described as the superposition of translational motion of its center of mass and rotational motion of the body about the center of mass. The center of mass is assumed to lie on the axis of rotational symmetry of the body. We saw in sections 3.4.2 and 3.4.3 how to set up the equations of motion for translational and rotational motion, respectively. Translational motion does not pose much of a problem, as application of equation (3.33) is quite straightforward. Rotational motion is a bit trickier though: equation (3.39) is only valid if the axis of rotation of the rigid body is fixed in space, but the axis of rotation of a rolling body moves along with it. Fortunately, it turns out that an equation analogous to (3.39) is always valid in the reference frame where the center of mass is at rest, independently of whether it is accelerated or not! That is,
NEWTON’s Second Law for Rolling Bodies

\[ \vec{\tau}^{(\text{CM})}_{\text{ext, tot}} = I^{(\text{CM})} \alpha, \]  

(3.47)

where \( I^{(\text{CM})} \) is the moment of inertia of the body about its axis of rotational symmetry. Angular acceleration \( \alpha \) is the same in a reference frame moving with the center of mass of the body as in any inertial frame. A derivation of this equation is shown below.

Note how we introduced a static friction force, as discussed in section 2.3.9. One can think of rolling of a rigid body down an incline as follows: in absence of static friction, the body would just slide down the incline. With static friction acting, the point of contact of the body with the incline is prevented from sliding. Therefore the rigid body will start rotating under influence of its weight exerting a torque due to the inclination of the center of mass with respect to the point of contact. If the slope of the incline is not too large, static friction can prevent the body from sliding all along it, resulting in rolling motion.

**Derivation of the Equations of Motion for Rolling Rigid Bodies***

Equation (3.47) is a consequence of a much more general principle:

Total angular momentum \( \vec{L}_{\text{tot}} \) of a system of particles with respect to a certain point of reference can be expressed as a sum of the total angular momentum of the system with respect of the center of mass \( \vec{L}^{(\text{CM})}_{\text{tot}} \) plus the angular momentum of the center of mass \( \vec{L}_{\text{CM}} \) with respect to that point:

\[ \vec{L}_{\text{tot}} = \vec{L}^{(\text{CM})}_{\text{tot}} + \vec{L}_{\text{CM}}. \]  

(3.48)

To see why this hold, consider the transformations \( \vec{r}_i = \vec{r}^{(\text{CM})}_i + \vec{r}_{\text{CM}} \) and \( \vec{v}_i = \vec{v}^{(\text{CM})}_i + \vec{v}_{\text{CM}} \) for the position and velocity vector, respectively, of the \( i \)-th particle between our point of reference and the center of mass system. Then

\[
\vec{L}_{\text{tot}} = \sum_i m_i \left( \vec{r}^{(\text{CM})}_i + \vec{r}_{\text{CM}} \right) \times \left( \vec{v}^{(\text{CM})}_i + \vec{v}_{\text{CM}} \right)
\]

\[
= \left( \sum_i m_i \vec{r}^{(\text{CM})}_i \right) \times \vec{v}_{\text{CM}} + \sum_i m_i \vec{r}^{(\text{CM})}_i \times \vec{v}^{(\text{CM})}_i
\]

\[
+ \vec{r}_{\text{CM}} \times \sum_i m_i \vec{v}^{(\text{CM})}_i + \left( \sum_i m_i \right) \vec{r}_{\text{CM}} \times \vec{v}_{\text{CM}}
\]

\[
= 0 + \sum_i m_i \vec{r}^{(\text{CM})}_i \times \vec{v}^{(\text{CM})}_i + 0 + \left( \sum_i m_i \right) \vec{r}_{\text{CM}} \times \vec{v}_{\text{CM}}
\]

\[
= \vec{L}^{(\text{CM})}_{\text{tot}} + \vec{L}_{\text{CM}},
\]
where

\[ \sum_i m_i \vec{r}_i^{(CM)} = 0 \]

since the center of mass of the system lies at the origin, or 0, of the center of mass frame and

\[ \sum_i m_i \vec{v}_i^{(CM)} = \sum_i m_i (\vec{v}_i - \vec{v}_{CM}) = \sum_i m_i \vec{v}_i - \left( \sum_i m_i \right) \vec{v}_{CM} = 0. \]

The total external torque in the center of mass frame can be expressed as

\[ \vec{\tau}_{\text{ext, tot}} = \sum \vec{r}_i \times \vec{F}_{i, \text{ext}} = \sum \left( \vec{r}_{CM} + \vec{r}_i^{(CM)} \right) \times \vec{F}_{i, \text{ext}} = \vec{r}_{CM} \times \left( \sum \vec{F}_{i, \text{ext}} \right) + \sum \vec{r}_i^{(CM)} \times \vec{F}_{i, \text{ext}} = \vec{r}_{CM} \times \vec{F}_{\text{ext, tot}} + \vec{\tau}_{\text{ext}}^{(CM)}. \]

The time derivative of the total momentum with respect to the reference point is, with (3.48),

\[ \frac{d}{dt} \vec{L}_{\text{tot}} = \frac{d}{dt} \vec{L}_{CM} + \frac{d}{dt} \vec{L}_{(CM)}^{(CM)} = \frac{d}{dt} (\vec{r}_{CM} \times \vec{p}_{CM}) + \frac{d}{dt} \vec{L}_{(CM)}^{(CM)} = \left( \frac{d}{dt} \vec{r}_{CM} \right) \times \vec{p}_{CM} + \vec{r}_{CM} \times \left( \frac{d}{dt} \vec{p}_{CM} \right) + \frac{d}{dt} \vec{L}_{(CM)}^{(CM)} = 0 + \vec{r}_{CM} \times \vec{F}_{\text{ext, tot}} + \frac{d}{dt} \vec{L}_{(CM)} \]

With \( \vec{\tau}_{\text{ext, tot}} = \frac{d}{dt} \vec{L}_{\text{tot}} \), it follows that

\[ \vec{\tau}_{\text{ext}}^{(CM)} = \frac{d}{dt} \vec{L}_{(CM)}^{(CM)}, \quad (3.49) \]

which is just a more general form of (3.47).
3.4.6 Conservation of Angular Momentum

From (3.33) follows that angular momentum of a rigid body is constant if the total external torque acting on it vanishes. Of course, this is just the law of conservation of angular momentum for a system of particles. For rigid bodies, conservation of angular momentum has some very interesting consequences. Suppose a rigid body rotates at angular speed $\omega$ around an axis with a moment of inertia $I$ with respect to this axis. Suppose that this body is not entirely rigid, such that due to some internal forces its moment of inertia changes from $I$ to $I'$. Since only internal forces are considered, angular momentum is conserved and the angular speed of the body must change as its moment of inertia changes:

$$I\omega = I'\omega' \iff \omega' = \frac{I}{I'}\omega.$$ 

Therefore angular velocity increases if the moment of inertia decreases and vice versa. This gives an explanation for phenomena like figure skaters modulating their spins by holding their extremities tight to increase spinning velocity or holding them out to slow down.

3.5 Gravity

3.5.1 Newton’s Law of Gravity

I guess there’s no need to explain you what gravitational attraction is, you’ve surely heard about it in school. Let’s get straight to the point by introducing a notation for gravitational forces that takes their vectorial nature into account:

**Newton’s Law of Gravity**

A point-like mass $m_1$ located at $\vec{r}_1$ exerts the gravitational force

$$\vec{F}_{1\rightarrow 2} = G \frac{m_1m_2}{|\vec{r}_1 - \vec{r}_2|^3} (\vec{r}_1 - \vec{r}_2) \quad (3.50)$$

on another point-like mass $m_2$ located at $\vec{r}_2$, where

$$G = 6.67 \cdot 10^{-11} \text{ m}^3 \text{ s}^{-2} \text{ kg}^{-1} \quad (3.51)$$

is the universal gravitational constant.

In school you might have encountered something more like

$$F = G \frac{m_1m_2}{r^2} \quad (3.52)$$

for the magnitude $F$ of gravitational attraction between the point-like masses $m_1$ and $m_2$ at distance $r$ apart. This formula looks indeed simpler than (3.50), but it contains no
information about the direction of the force. From (3.50), you see that the force acting from \( m_1 \) onto \( m_2 \) is an attractive force, as \( \vec{r}_1 - \vec{r}_2 \) shows from \( m_2 \) towards \( m_1 \). We leave it to you to show that formula (3.52) is easily derived from (3.50).

Look at how the third law of motion is contained in (3.50): by swapping the subscripts 1 and 2, you see that

\[
\vec{F}_{2\rightarrow 1} = -\vec{F}_{1\rightarrow 2},
\]

that is, the two bodies 1 and 2 act on each other by means of opposite gravitational forces of equal magnitude.

### 3.5.2 Gravitational Fields

Consider a large mass distribution, like a planet or a star, exerting gravitational forces on other smaller objects in its surroundings, like meteoroids, comets, satellites, or human beings. These objects are assumed to be so small that they do not affect motion of the larger mass distribution.

To calculate the gravitational force by which an extended mass distribution acts on a small mass \( m \), we can think of the large mass distribution as being composed by a large amount of point masses \( m_i \), every of which will exert its ‘own’ gravitational attraction \( \vec{F}_i \) on \( m \) according to

\[
\vec{F}_i = G \frac{m_i m}{|\vec{r}_i - \vec{r}|^3} (\vec{r}_i - \vec{r}),
\]

where \( \vec{r}_i \) denotes the position of the \( i \)-th point mass \( m_i \) and \( \vec{r} \) the position of \( m \). To calculate the total gravitational attraction on \( m \), we make use of the extremely important

**Principle of Superposition**

Consider three point masses \( m_A, m_B \) and \( m \). If \( m_A \) and \( m \) were alone, let \( \vec{F}_A \) be the gravitational force \( m_A \) would exert on \( m \). Analogously, let \( \vec{F}_B \) be the gravitational force by which \( m_B \) would act on \( m \) if \( m_B \) and \( m \) were alone. The gravitational force acting on \( m \) if both \( m_A \) and \( m_B \) are around is then given by

\[
\vec{F}_A + \vec{F}_B.
\]

In other words: The gravitational force by which a system of point masses acts on another point mass is given as the vector sum of all gravitational forces by which each of the point masses of the system alone would act on this other point mass.

Applied to our situation, this simply means that we can calculate the total gravitational attraction acting on \( m \) as

\[
\sum_{\text{all masses } m_i} \vec{F}_i = \sum_i G \frac{m_i m}{|\vec{r}_i - \vec{r}|^3} (\vec{r}_i - \vec{r}) = m \cdot \left( G \sum_i \frac{m_i}{|\vec{r}_i - \vec{r}|^3} (\vec{r}_i - \vec{r}) \right).
\]
The very last term in brackets is a quantity describing how the distribution of the masses \( m_i \) in space exerts gravitational forces on a mass located at \( \vec{r} \). We will call this term the gravitational field \( \vec{g}(\vec{r}) \) produced by this distribution of masses at \( \vec{r} \), such that we can write

\[
\vec{F} = m \vec{g}(\vec{r}).
\] (3.53)

with

\[
\vec{g}(\vec{r}) = G \sum_i \frac{m_i}{|\vec{r}_i - \vec{r}|^3} (\vec{r}_i - \vec{r}).
\] (3.54)

So far for systems consisting of point masses. Calculation of gravitational fields produced by extended mass distributions turns out to be somewhat more difficult, the same way as determining the center of mass of a rigid body by means of finite sums used to be in section 3.4. Exactly as we did there, we can treat continuous mass distributions as continuous by introducing their mass density \( \varrho(\vec{r}) \) and turning the sum into an integral:

The gravitational field produced in point \( \vec{r} \) by a continuous mass distribution \( \varrho \) confined to the portion of space \( \Omega \) is given by

\[
\vec{g}(\vec{r}) = G \int_{\Omega} \frac{\varrho(\vec{r}')}{|\vec{r}' - \vec{r}|^3} (\vec{r}' - \vec{r}) \, dV'.
\] (3.55)

Again, your math courses will show you such integrals can be computed in various situations. You might realize that dealing with the integrand of (3.55) can be a rather tedious affair. Fortunately, a very elegant law exists that makes calculation of gravitational fields in particularly symmetric situations quite simple:

**Gauss’ Law for Gravitational Fields**

Let \( \Omega \) be a volume in space and \( \partial \Omega \) its outer surface. For gravitational field \( \vec{g} \), the following identity is valid:

\[
\oint_{\partial \Omega} \vec{g} \cdot dA = -4\pi Gm_{\Omega},
\] (3.56)

where \( m_{\Omega} \) denotes the total mass located inside \( \Omega \).

You probably already know a very similar law by Gauss for electric fields and charge distributions. Indeed, both the gravitational and the electrostatic versions of Gauss’ law are nothing but a direct consequence of the inverse-square-law nature of Newton’s law of gravity and Coulomb’s law of electrostatic interaction, respectively. Again, your math
courses will tell you what the integral in the left-hand side of (3.56) means and how to
calculate it.
We won’t discuss Gauss’ law in great detail as you can learn more about it in your electro-
dynamics courses. However, we will show you one very important application.
Many Olympiad problems involving gravity deal with planets or stars, which are in most
cases treated as spheres of homogeneous mass density. We will use Gauss’ Law to calculate
the field induced by such a homogeneous sphere of total mass $M$ and radius $R$ in any point
of space. This situation shows perfect spherical symmetry in space, that is, one could rotate
everything about the center of the sphere by any angle and you wouldn’t be able to tell the
difference. Therefore, magnitude of the gravitational field must be the same at any point
located at the same distance $r$ from the center of the planet. Furthermore, still due to
spherical symmetry, the vector $\vec{g}$ must point in radial direction anywhere in space: spherical
symmetry would be broken if $\vec{g}$ had at some point in space a non-vanishing component not
pointing in radial direction. Choose a coordinate system with origin at the center of the
sphere. The gravitational field at any point in space $\vec{r}$ can be described as

$$\vec{g}(\vec{r}) = g(r) \hat{r}$$

with the projection $g(\vec{r})$ of the gravitational field on $\hat{r}$. First, we consider a point located
inside the sphere, that is, $r < R$. Now apply (3.56) on a sphere $V(r)$ of radius $r$ concentric
to the planet. The surface integral on the left-hand side is obtained by subdivision of the
surface $\partial V(r)$ of the sphere into many infinitesimal area elements $dA$ with normal vectors
$\hat{n}$ and taking the sum over the scalar products $\vec{g} \cdot dA \hat{n}$ at any such area element, that is,
something like

$$\int_{\partial V(r)} dA \ \vec{g}(\text{at this area element}) \cdot \hat{n} (\text{at this area element})$$

$$= \int_{\partial V(r)} dA \ (g(r)\hat{r}) \cdot \hat{r}$$

$$= g(r) \int_{\partial V(r)} dA$$

$$= g(r) \cdot \text{surface area of } \partial V(r)$$

$$= g(r) \cdot 4\pi r^2,$$

where we used $\hat{n} = \hat{r}$ for all surface elements on the sphere. Our result,

$$\int_{\partial V(r)} \vec{g} \cdot d\vec{A} = 4\pi r^2 \cdot g(r),$$

is, according to (3.56), equal to $-4\pi G$ times the total mass $m_{V(r)} = M \cdot (r/R)^3$ contained
within $V(r)$, that is,

$$4\pi r^2 \cdot g(r) = -4\pi G \cdot \frac{M}{R^3}r^3$$

$$\Leftrightarrow \ g(r) = -\frac{GM}{R^3}r.$$
Inside the sphere, the magnitude of gravitational field increases linearly with the distance from its center.

Now apply the same strategy for any point outside the sphere, that is, \( r > R \). The mass \( m_{V(r)} \) contained inside a spherical surface of radius \( r \) around the center is now just the total mass \( M \) of our spherical body. GAUSS’ law now yields

\[
4\pi r^2 g(r) = -4\pi GM \\
\Leftrightarrow g(r) = -\frac{GM}{r^2},
\]

which is the same as the field produced by a point mass \( M \) located at the origin. This is a very important result:

The gravitational field inside a homogeneous sphere of mass \( M \) and radius \( R \) increases linearly with the distance from the center of the sphere:

\[
\vec{g}(\vec{r}) = -\frac{GM}{R^3} r^3 \hat{r}, \quad 0 < r < R
\]

Outside the sphere, the gravitational field behaves as if the sphere was a point mass \( M \) located at its own center:

\[
\vec{g}(\vec{r}) = -\frac{GM}{r^2} \hat{r}, \quad r > R.
\]

### 3.5.3 Energy and Angular Momentum in Gravitational Fields

In school, you might already have encountered the expression

\[
E_{\text{pot}} = -G \frac{Mm}{r}
\]

for potential energy of a small mass \( m \) subject to gravitational field of the larger mass \( M \) at distance \( r \) from it. This expression arises from the definition of potential energy of the particle as the amount of work an external force would have to perform to move the mass inside the field from infinity to a point at distance \( r \) from the center of the field.

Consider a point \( \vec{r} \) at distance \( r \) from the center of the field and another arbitrary point \( \vec{r}' \) at distance \( r' \) from the center. The amount of work done by an external force when displacing the mass from \( \vec{r}' \) to \( \vec{r} \) along a path \( \gamma \) connecting \( \vec{r}' \) to \( \vec{r} \) is the line integral

\[
W_\gamma = -\int_\gamma m\vec{g} \cdot d\vec{s}
\]

Decompose the path \( \gamma \) into very small segments \( d\vec{s} \) in such a way that their direction is either radial or perpendicular to the radial direction. The force vector \( m\vec{g} \) is always directed
radially, therefore only terms with radial $d\vec{s}$ will contribute to the integral, since all other terms vanish with $\vec{g} \cdot d\vec{s} = 0$. Therefore, the value of $W_\gamma$ is independent of the exact shape of $\gamma$ as long as it connects $\vec{r}'$ to $\vec{r}$. Furthermore, the exact position of both $\vec{r}'$ and $\vec{r}$ is not relevant for the value of $W_\gamma$, they could be replaced by any points located at distance $r'$ and $r$, respectively, from the center of $M$. For calculation of $W_\gamma$, we can therefore replace $\vec{r}'$ and $\vec{r}$ by the two points $(r',0,0)$ and $(r,0,0)$, respectively:

$$W_\gamma = \int_{x=r'}^{r} \left( \frac{GMm}{x^2} \right) \cdot (dx \, \hat{x})$$

$$= GMm \int_{x=r'}^{r} \frac{dx}{x^2}$$

$$= GMm \left( \frac{1}{r'} - \frac{1}{r} \right),$$

that is,

Suppose a large mass $M$ generates a gravitational field at the origin of a coordinate system. The work an external force has to do to bring a small mass $m$ from a point at distance $r'$ from the origin to another point distant $r$ from the origin is equal to

$$GMm \left( \frac{1}{r'} - \frac{1}{r} \right).$$

(3.60)

Now back to the concept of potential energy. In a homogeneous gravitational field, one can define potential energy of an object as the amount of work an external force has to perform to bring the object from a determined reference height to the height of the object’s current position. For central gravitational fields, we can apply an analogous definition. It seems reasonable to choose infinity as a distance of reference: if we let $r' \to \infty$, the corresponding term of (3.60) vanishes. That is, we are left with equation (3.59).

Problems involving gravity can often be solved by considering conservation of mechanical energy of an object moving in a gravitational field:

Total mechanical energy

$$E = \frac{1}{2} mv^2 - G \frac{mM}{r}$$

(3.61)

of an object of mass $m$ moving in a gravitational field produced by a large mass $M$ at the origin is constant over time.

Gravitational fields as from point masses or homogeneous spheres are radial. In particular, angular momentum of particles is preserved as they move in such fields.

117
3.5.4 Two Objects Subject to Mutual Attraction

So far, we have only considered small masses subject to an external gravitational field produced by a mass distribution so large not to be affected by the small mass. We turn now to the discussion of systems consisting of two masses \( m_1, m_2 \) of similar orders of magnitudes, such that motion of each is influenced by gravitational attraction from the respective other mass. In such a system, if not subject to any other external forces, the following conservation laws hold:

- Position of the center of mass \( \frac{(m_1 \vec{r}_1 + m_2 \vec{r}_2)}{(m_1 + m_2)} \) is constant over time.
- Total energy
  \[
  E = E_{\text{kin}, 1} + E_{\text{kin}, 2} + E_{\text{pot}} \\
  = \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 - G \frac{m_1 m_2}{r_{12}}
  \]
  is a conserved quantity, where \( r_{12} \) denotes the distance between the two particles.
- Total angular momentum
  \[
  \vec{L} = m_1 \vec{r}_1 \times \vec{v}_1 + m_2 \vec{r}_2 \times \vec{v}_2
  \]
  is a conserved quantity.

Note how the expression for total energy consists of three parts: one term for kinetic energy of each particle, \( E_{\text{kin}, 1} \) and \( E_{\text{kin}, 2} \), and one for potential energy

\[
E_{\text{pot}} = -G \frac{m_1 m_2}{r_{12}}. \tag{3.62}
\]

This expression for potential energy of a system of two point masses looks very similar to the expression derived in the previous section for a single particle moving in a central gravitational field. The two situations are slightly different though. In the last section, we would deal with a gravitational field and potential energy of the particle was defined as the amount of work done an external force when dragging the particle from infinity to its current position in the field. Here we are dealing with two point masses and using the concept of field as in the last section doesn’t make too much sense as either mass influences motion of the other one with its gravitational attraction. We can, however, define the potential energy of the system as the amount of work one would have to perform to assemble it in its current spatial configuration. This assembling might take place in two steps as follows: take \( m_1 \) from infinity and put it to \( \vec{r}_1 \) first, then take \( m_2 \) from infinity and place it into \( \vec{r}_2 \). The first step does not require any work, as \( m_1 \) is all alone in its way to \( \vec{r}_1 \) and no force acts on it. For
the second step, we need to move $m_2$ against gravitational attraction exerted by $m_1$, and the work required is exactly the same as if we thought of $m_2$ being moved in a gravitational field produced by $m_1$ and the amount of work required is therefore just the term on the right-hand side of (3.62). Note that while $m_2$ is put into $\vec{r}_2$, some force must prevent $m_1$ from slipping off its position at $\vec{r}_1$ due to attraction by the approaching $m_2$. This force does not contribute any mechanical work, since $m_1$ is not displaced under its influence.

3.5.5 **KEPLER’s Laws of Planetary Motion***

Here as well, you most probably already know these laws from your high school physics courses. This section will give you short clarifications about the concepts involved, as they might provide a deeper understanding of the principles governing gravity and NEWTON’s laws of motion.

**KEPLER’s Laws of Planetary Motion**

1. The orbit of a planet around the sun is a plane ellipse, the sun lies at one of its two foci.

2. Consider the radius vector pointing from the sun to the planet as it moves along its orbit. During equal amounts of time, the radius vector sweeps out segments of the ellipse of equal area.

3. The third power of the semi-major axis of an elliptical planet orbit is proportional to the square of its orbital period around the sun.

**The First Law**

To understand the first law, we need to clarify some geometry. An ellipse can be defined in different but equivalent ways:

A: As the figure you get by stretching a circle by any factor with respect to a straight line.

B: As the locus of all points in the plane such that the sum of their distance to two given, fixed points, called the foci of the ellipse, is a constant.

C: As the locus of all points in the plane whose distance from a given point, called a focus of the ellipse, is proportional to their distance from a given straight line, called the directrix, with a constant of proportionality $< 1$.

D: As the figure you get by intersecting a cone with a plane under such an angle that the figure is closed.
You might find it an interesting exercise in geometry to prove the equivalence of all these definitions. Anyway, for Olympiad problems, the following two parametrizations of an ellipse in the $xy$-plane are useful:

\[
\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1 \tag{3.63}
\]

in Cartesian coordinates and

\[
r(\vartheta) = \frac{p}{1 + \varepsilon \cos \vartheta} \tag{3.64}
\]

in polar coordinates. You might verify that both equations describe the same curve in the plane under certain conditions for $a$, $b$, $p$, $\varepsilon$.

Equation (3.63) is probably the most useful to understand the concept of major and minor axis. These are defined as the largest and smallest diameter, respectively, of the ellipse. It is easy to show that their values are given by $2a$ and $2b$, respectively, if the ellipse is described as in (3.63). Furthermore, they coincide with the $x$- and $y$-axis. Note how an ellipse is also supposed to have two foci, such that the sum of the distances of any point on the ellipse to the two foci is a constant. You might verify that the foci of the ellipse as introduced in definition B are situated at $(-s, 0)$, $(s, 0)$ with $s = \sqrt{a^2 - b^2}$ if $a > b$ or at $(0, -s)$ and $(0, s)$ with $s = \sqrt{b^2 - a^2}$ if $b > a$. In particular, this means the foci are always situated on the major axis. Conversely, by starting with definition B with two foci lying on either the $x$- or $y$-axis, you might retrieve equation (3.63) for description of the ellipse in the $xy$-plane.

In (3.64), the quantities $p$ and $\varepsilon$ are called semi-latus rectum and eccentricity of the ellipse, respectively. You can verify that the curve described by this equation satisfies the condition of definition B with one focus at the origin and the other located at $(-2\varepsilon p / (1 - \varepsilon^2), 0)$ on the $x$-axis. Definition C is satisfied for the focus at the origin with a directrix described by the equation $x = p/\varepsilon$, the constant of proportionality between the distance from the points to the origin (which corresponds to $r$) and the distance from the points to the directrix (which corresponds to the points’ $x$-coordinates) is equal to $\varepsilon$. The ellipse intersects the $y$-axis at two points distant $p$ from the origin.

I bet you can’t wait to see how an equation like (3.63) or (3.64) can arise from the equations of motion of a body subject to universal gravitation. The calculations involved are somewhat elaborate and may not be of great help for solving physics competition problems.

At this point, it is just important for you to understand what an ellipse is, what its foci are, and to keep in mind that closed planetary orbits are ellipses with the sun at one focus. In addition, it sure wouldn’t hurt to feel comfortable around the different geometrical definitions of an ellipse, so feel free to practice your skills in analytic geometry and play around.
3.5. GRAVITY

The Second Law

Kepler’s second law is a direct consequence of conservation of angular momentum. Put the whole orbit into a polar coordinate system with the sun at the origin. The position of the particle can be written as

\[ \vec{r} = (r \cos \varphi) \hat{x} + (r \sin \varphi) \hat{y}, \]

at any time. The velocity vector is obtained by taking the time derivative

\[ \vec{v} = \dot{\varphi} \left( (\dot{r} \cos \varphi - r \sin \varphi) \hat{x} + (\dot{r} \sin \varphi + r \cos \varphi) \hat{y} \right). \]

The angular momentum vector is then found to be

\[ \vec{L} = m \vec{r} \times \vec{v} = r^2 \dot{\varphi} \hat{z} \]

if \( m \) is the mass of the planet.

During a small amount of time \( t \), let the position vector of the planet sweep out a small area \( dA \) of the ellipse, thereby covering an angle \( d\varphi = \omega \, dt \) with \( \omega = \dot{\varphi} \). The small area can be approximated as a triangle, such that it amounts to \( \frac{1}{2} \cdot r \cdot r \, d\varphi = \omega r^2 dt/2 \). Due to conservation of angular momentum, we know that \( \omega r^2 \) is a conserved quantity. This just means that the area \( dA \) swept out during \( dt \) is a constant and that it does not depend on the position of the planet along its orbit. We see that this is also true for any finite time interval of a given length \( \Delta t \) between any times \( t_1 \) and \( t_2 = t_1 + \Delta t \), as the area swept out by the position vector is just calculated as

\[ \int_{\text{time interval } [t_1, t_2]} dA = \int_{t_1}^{t_2} \frac{1}{2} \omega r^2(t) \, dt = \frac{L}{2m} \int_{t_1}^{t_2} dt = \frac{L}{2m} \cdot \Delta t, \]

which does not depend on the exact position of the time interval as long as its length is \( \Delta t \) – just the statement of the second law.

The Third Law

It is quite easy to verify the third law for a circular orbit. Suppose the planet of mass \( m \) moves on a circular orbit of radius \( R \) around the sun of mass \( M \) with a time period \( T \). Then gravitational attraction of the sun on the planet must account for centripetal acceleration, and, by the second law of motion:

\[ m \left( \frac{2\pi}{T} \right)^2 R = G \frac{mM}{R^2}, \]

which is equivalent to

\[ \frac{R^3}{T^2} = \frac{GM}{4\pi^2}, \]

in other words, circular orbits satisfy Kepler’s third law. Calculations for general elliptic orbits are somewhat more complex.
Chapter 4

Thermodynamics

Thermodynamics is a phenomenological treatment of classical macroscopic systems and their properties. The systems looked at in thermodynamics involve many degrees of freedom ($\sim 10^{23}$). Therefore it is impossible to solve the equations of motion exactly.

In this chapter, we will first look at some important definitions, before looking at ideal gases and heat engines. In the end, we will also have a quick look at real gases and the Stefan-Boltzmann law.

4.1 Important definitions

Mole: base unit for the amount of substance. One mole of a substance contains exactly $N_A$ particles. (See figure 4.1.)

Avogadro constant $N_A$: the number of molecules in one mole. $N_A \approx 6.022 \times 10^{23} \text{ mol}^{-1}$.

Boltzmann constant $k_b$: an important proportionality constant in statistical physics to relate energy and temperature. $k_b \approx 1.380 \times 10^{-23} \text{ JK}^{-1}$

Gas constant $R$: proportionality constant to relate energy to a mole of particles at a stated temperature. Molar version of $k_b$: $R = k_b \cdot N_A$

Volume $V$: space in which the system is confined in.

Temperature $T$: measured quantity of a thermometer.

Pressure $p$: force applied per unit area.

Number of molecules $N$: number of molecules confined in the considered volume.
**Amount of substance** $n$:

molar version of $N$. (See figure 4.1.)

**Work** $W$:

energy transferred from or to the system by expansion/compression.

**Heat** $Q$:

transferred energy between systems due to temperature differences.

**Internal energy** $U$:

also sometimes called just energy, all the energy inside the system.

**Ideal gas**:

a collection of neutral atoms or molecules where the interaction between individual particles can be neglected.

![Diagram of mass, volume, number of particles, and amount of substance](image)

**4.2 The temperature scale**

Since there is no easy way to define temperature, we just state that the temperature is a physical quantity that tells us in which direction the heat flows. Heat always flows in the direction of the body with the lower temperature. A way to measure this temperature is by using a thermometer, which uses thermal properties of materials to measure the temperature. Examples are the thermal expansion of mercury, an electrical resistance that changes with temperature or the volume of a gas at constant pressure.

The temperature is a scalar (a number) measure which is constant in an isolated system in thermodynamic equilibrium. There are 3 main scales for temperature (Kelvin, Celsius and
4.3 ZEROTH LAW OF THERMODYNAMICS

Fahrenheit), but in physics, only Kelvin and, in rare cases, Celsius are used. In figure 4.2, a comparison between the scales can be seen.

The Kelvin scale starts at absolute zero, which is equal to $-273.15^\circ\text{C}$. At $0\text{K}$ all thermal motion freezes. There is no temperature that is $< 0\text{K}$. The Kelvin scale also uses the same temperature step as the Celsius scale and therefore we can easily switch between the two:

$$T[\circ\text{C}] + 273.15 = T[\text{K}]$$

\hspace{1cm} (4.1)

Note that most equations in physics only work properly if you calculate everything in the Kelvin temperature scale and with no other.

Figure 4.2: Comparison of different temperature scales related to the energy scale via the Boltzmann constant $k_b$ (e.g. for $T=300\text{ K}$ $\Rightarrow k_bT = 4.1 \times 10^{-21} \text{ J}$) with the corresponding energies in Zeptojoule ($1\text{ zJ} = 1 \times 10^{-21} \text{ J}$).[11]

4.3 Zeroth law of thermodynamics

The zeroth law of thermodynamics states that if we bring two bodies ($A$ & $B$) with different temperatures ($T_A$ & $T_B$) into contact, the two bodies will reach a thermal equilibrium ($T_{eq}$) after a certain time. Since the temperature is a measure of thermal movement of atoms/molecules (see chapter 4.12), the equilibrium temperature $T_{eq}$ is between $T_A$ and $T_B$:

$$T_A < T_{eq} < T_B \text{ or } T_A > T_{eq} > T_B,$$

\hspace{1cm} (4.2)

depending on which body had the higher temperature in the beginning.
4.4 Thermal energy and heat capacity

When two bodies with different temperatures touch each other, the heat of one body flows to the other until they have the same temperature. This thermal energy \( Q \) is often (especially in chemistry) measured in calories (cal):

\[
1 \text{ cal} = 4.1868 \text{ J},
\]

where one calorie is the heating energy one needs to heat up 1 g water at normal pressure \((p = 1 \text{ bar})\) from 287.65 K to 288.65 K. Today the SI-unit Joule should be used, but in many textbooks one can still find calories.

Different materials also take different amounts of energy \( \Delta Q \) to increase the temperature by a certain extent \( \Delta T \). Therefore we define the heat capacity \( C \) which is a measure of how much thermal energy the body needs to heat up.

\[
C = \frac{\Delta Q}{\Delta T} \iff \Delta Q = C \cdot \Delta T \tag{4.4}
\]

Usually the heat capacity is normalized to the mass of the body (specific heat capacity \( C_m \)) or the amount of substance (molar heat capacity \( C_M \)), i.e.

\[
C_m = \frac{1}{m} \frac{\Delta Q}{\Delta T} \tag{4.5}
\]

and

\[
C_M = \frac{1}{n} \frac{\Delta Q}{\Delta T} \tag{4.6}
\]

4.4.1 Molar heat capacities of ideal gases

In general there are two different types of heat capacities (for gases). The heat capacity of a substance depend on how one measures it: If you measure it while leaving the pressure constant \((C_p)\) the specific heat is higher compared to when you leave the volume constant \((C_V)\). This means that one needs different amounts of energy to heat the same amount of particles, depending on how one heats them up. This is because additional work is done when expanding the system while keeping the pressure constant. This can be seen from figures 4.3 and 4.4. The difference between the two molar heat capacities is exactly the gas constant

\[
R = C_p - C_V \tag{4.7}
\]

For a derivation look at equation (4.15) and divide both sides by \( n \Delta T \).
4.5 Ideal gas law

The most important equation in classical thermodynamics is the ideal gas law

\[ pV = nRT, \tag{4.8} \]

which can also be written as

\[ pV = Nk_bT. \tag{4.9} \]

It is a combination of the empirical Boyle’s law \((p \propto \frac{1}{V})\), Charle’s law \((V \propto T)\) and Avogadro’s law \((V \propto n)\) and is only valid for ideal gases. Such an ideal gas cannot be liquefied like a real gas. For real gases there is among others the Van-der-Waals gas equation to describe not only the behaviour in the gaseous phase, but also the liquid phase as well as the phase transition between them.

4.6 First law of thermodynamics

Temperature is connected to the kinetic energy of the particles in the gas (see chapter 4.12). This internal energy \(U(p, V, T)\) is a function of the parameters of the system: pressure \(p\), volume \(V\) and temperature \(T\). These are state variables like described in 4.7. A change in internal energy only depends on the states at the beginning and the end of the process.
\( \Delta U = U_E - U_A \). This also means that it doesn’t matter how one got from one state to the other.

The first law of thermodynamics states that the energy of a system does change if thermal energy \( Q \) is added or when mechanical work \( W \) is performed on the system:

\[
dU = \delta Q + \delta W
\]  

(4.10)

This means that for an isolated systems, in which we do not add thermal energy or do work on, the energy is constant. Note that here we introduced the convention that work done onto the system, or thermal energy added to the system is positive and the work the system does on it’s surroundings is negative.

### 4.7 Thermodynamic systems

A thermodynamic system is a collection of particles in thermal equilibrium and is characterized by so called state variables like volume \( V \), temperature \( T \), pressure \( p \), entropy \( S \), number of particles \( N \), amount of substance \( n \) and many more. In contrast heat \( Q \) and work \( W \) are not state variables, but process functions, since they do not describe an (equilibrium) state.

The different types of thermodynamic systems are characterized as follows.

**Isolated systems**

An isolated system has no exchange of matter or energy with its surroundings. It is completely isolated and will stay in the thermodynamic equilibrium.

**Closed systems**

Closed systems have, like the isolated system, no exchange of matter with the surroundings. But they can exchange energy, for example with a thermal contact or by performing work on each other.

**Open systems**

Open systems may exchange energy as well as matter with their surroundings.

### 4.8 Equipartition theorem

The equipartition theorem states that the mean energy of each molecule with \( f \) degrees of freedom (number of independent motions that are allowed, e.g. moving in \( x \), \( y \) and \( z \) equals to 3 degrees of freedom) is given by

\[
\text{mean energy per molecule} = \frac{f}{2} k_b T.
\]  

(4.11)
Thus the energy of an ideal gas with n moles of molecules is given by

\[ U = nN_A \frac{f}{2} k_b T = n \frac{f}{2} RT. \]  \hspace{1cm} (4.12)

Using \( \Delta Q = nC_V \Delta T \), we also find that \( C_V = \frac{f}{2} R \) and therefore \( C_p = \frac{f^2}{2} R \). Due to quantum mechanical effects, degrees of freedom can be "frozen" out at low temperatures. This means they cannot be excited and therefore don’t contribute to the inner energy. This can be seen in figure 4.5.

![Figure 4.5: Specific heat of H₂ (schematically). The y-axis corresponds to \( \frac{f}{2} = \frac{C_V}{R} \). The vibration gives two degrees of freedom, since one can store energy in potential or kinetic energy.][11]

### 4.9 Thermodynamic processes

In this chapter we look at different processes that change the system by doing/extracting work onto it or by heating/cooling it. These processes are mostly characterized by the variables they leave constant. The different processes in a \( p-V \) diagram can be seen in 4.6.

#### Isobaric processes

In an isobaric process the pressure is held constant. The work done on an expanding gas from the outside is given by

\[ W^\vdash = \int_{V_a}^{V_b} \delta W^\vdash = - \int_{V_a}^{V_b} p dV, \]  \hspace{1cm} (4.13)

where \( V_a \) is the volume at the beginning and \( V_b \) at the end of the process. By using that the pressure is constant over the whole expansion, we can take the pressure out of the integral and get the total work

\[ W^\vdash = -p \int_{V_a}^{V_b} dV = -p(V_b - V_a) = -nR(T_b - T_a). \]  \hspace{1cm} (4.14)
By using the equipartition theorem, we can also find the change in internal energy $\Delta U = nC_V \Delta T$ and therefore, according to the first law of thermodynamics and the ideal gas law, the heat exchanged is

$$Q^\uparrow = \Delta U - W^\uparrow = nC_V \Delta T + nR \Delta T = nC_p \Delta T.$$

(4.15)

**Isothermal processes**

For isothermal processes the temperature is held constant ($T = \text{const.}$) and therefore, according to the equipartition theorem, the internal energy is constant ($U = U(T) = \text{const.}$). Using the ideal gas law we also find that the product $pV = \text{const}$ as well. Since we know that the internal energy remains constant, the first law of thermodynamics tells us that the incoming heat is entirely converted to work:

$$Q^\uparrow = \int \delta Q^\uparrow = \int -\delta W^\uparrow = -W^\uparrow = W^\downarrow$$

(4.16)

Using the ideal gas law we can then calculate the work done by the system as

$$W^\downarrow = \int_{V_a}^{V_b} pdV = nRT \int_{V_a}^{V_b} \frac{dV}{V} = nRT \ln \left( \frac{V_b}{V_a} \right).$$

(4.17)

**Isochoric processes**

In isochoric processes the volume doesn’t change ($V = \text{const.}$). As a consequence no work is being done and therefore the first law of thermodynamics tells us

$$dU = \delta Q^\uparrow.$$

(4.18)
Therefore the change in energy is simply given by

\[ \Delta Q = \int_{T_n}^{T_h} nC_V dT = nC_V \Delta T. \] (4.19)

### Adiabatic processes

If during a process the system does not exchange heat with the surroundings, one speaks of adiabatic processes. This is for example the case when the system is thermally isolated or when the process is so fast that the heat exchange with the surroundings is negligible.

We can write down the change in work and energy for the process

\[ \delta W = -pdV = -\frac{nRT}{V}dV \] (4.20)

\[ dU = nC_V dT. \] (4.21)

Since we know that there is no heat exchanged \( \Delta Q = 0 \), we find that for any adiabatic process \( nC_V dT + \frac{nRT}{V}dV = 0 \). By dividing through \( nT \) and integrating we get

\[ \begin{align*}
\int \frac{C_V}{T} dT &= -\int \frac{R}{V}dV \\
C_V \ln(T) &= -R \ln(V) + \text{const.}
\end{align*} \] (4.22)

\[ \Rightarrow T = V^{-\frac{R}{C_V}} + \text{const.}, \] (4.23)

which is exactly

\[ TV^{\kappa-1} = \text{const.}. \] (4.25)

or equivalently

\[ pV^\kappa = \text{const.} \] (4.26)

\[ T^{\kappa}p^{1-\kappa} = \text{const}'. \] (4.27)

Where \( \kappa \) is the ratio \( \kappa = \frac{C_p}{C_V} \) which implies \( \frac{R}{C_V} = \kappa - 1 \) (using equation (4.7)).

### 4.10 Second law of thermodynamics

The first law of thermodynamics was about the exchange of energy with the surroundings. The second law of thermodynamics is about the distribution of the molecules within the volume of the system. The entropy of a system is a measure for the disorder in the system. There are many formulations, but the most known one is

\[ S = k_b \ln \mathcal{W}, \] (4.28)

where \( S \) is the entropy and \( \mathcal{W} \) the probability for the system to be in a given state. (To be more precise it is the probability to be in a given macrostate defined by state variables. For
The second law of thermodynamics states that the entropy of a system can only increase
\[ dS \geq \frac{\delta Q}{T}. \] (4.29)

There are also other formulations of the second law, like the one from Rudolf Clausius:

"Heat can never pass from a colder to a warmer body without some other change, connected therewith, occurring at the same time." [15]

or Lord Kelvin:

"It is impossible, by means of inanimate material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects." [15]

The second law also splits processes up into those which are reversible and conserve the entropy, and those which are irreversible and do not conserve entropy. For reversible processes, the system can be returned to its initial state. For reversible processes we therefore have

\[ dS = \frac{\delta Q}{T}. \] (4.30)

There is also the third law of thermodynamics which fixes the entropy at absolute zero \( S(0\, \text{K}) = 0 \). For more information have a look at [14].

### 4.11 Heat engines

Heat engines do work by transferring heat between two reservoirs at different temperatures. One can imagine it a bit like a water mill, where the water from the higher level produces mechanical work by running to the lower level.

Heat engines can be characterized by the associated cycle, which is the closed curve on a p-V diagram. There are many different heat machines, like conventional car engines, but we will have a look at the theoretically most efficient process, the Carnot cycle. The Carnot process, which is depicted in figure 4.7, features two isothermal and two adiabatic processes. To see how efficient the Carnot cycle is we have to write down the equation for every step, shown in table 4.1.
Figure 4.7: The Carnot cycle with its two isotherms and two adiabats. The surface that is
enclosed by the Carnot process equals the total work done. [14]

Table 4.1: The Carnot process.

<table>
<thead>
<tr>
<th>step</th>
<th>process</th>
<th>$T$</th>
<th>$W$</th>
<th>$Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:</td>
<td>a→b adiabatic</td>
<td>$T_2 \rightarrow T_1$</td>
<td>$\delta W_1^a = nC_V(T_1 - T_2)$</td>
<td>0</td>
</tr>
<tr>
<td>2:</td>
<td>b→c isothermal</td>
<td>$T_1$</td>
<td>$\delta W_2^b = nRT_1 \ln \frac{V_c}{V_b}$</td>
<td>$\delta Q_2^b = nRT_1 \ln \frac{V_c}{V_b}$</td>
</tr>
<tr>
<td>3:</td>
<td>c→d adiabatic</td>
<td>$T_1 \rightarrow T_2$</td>
<td>$\delta W_3^c = nC_V(T_1 - T_2)$</td>
<td>0</td>
</tr>
<tr>
<td>4:</td>
<td>d→a isothermal</td>
<td>$T_2$</td>
<td>$\delta W_4^d = nRT_2 \ln \frac{V_d}{V_a}$</td>
<td>$\delta Q_4^d = nRT_2 \ln \frac{V_d}{V_a}$</td>
</tr>
</tbody>
</table>

Furthermore we know that $TV^\kappa^{-1} = \text{const.}$ during adiabatic processes

\[ T_2 V_a^\kappa^{-1} = T_1 V_b^\kappa^{-1} \quad (4.31) \]
\[ T_2 V_d^\kappa^{-1} = T_1 V_c^\kappa^{-1} \quad (4.32) \]

and if we divide the two equations, we get the condition

\[ V_a \cdot V_c = V_b \cdot V_d. \quad (4.33) \]
With all of that we can define the Carnot efficiency

\[ \eta_C = \frac{\text{work done}}{\text{supplied heat}} = \frac{\delta W_1 - \delta W_2 + \delta W_3 - \delta W_4}{\delta Q_2} = \frac{T_1 - T_2}{T_1}. \] (4.34)

The Carnot cycle defines a thermodynamic cyclic process to produce work with the highest achievable efficiency. All reversible heat engines between two heat reservoirs are equally efficient as a Carnot engine operating between the same reservoirs.[15] Therefore all real processes are less efficient

\[ \frac{\text{work done}}{\text{supplied heat}} = \eta_{\text{real}} \leq \eta_C = \frac{T_{\text{hot}} - T_{\text{cold}}}{T_{\text{hot}}}. \] (4.35)

It is also nice to see that the entropy for a fully reversible cycle does not increase:

\[ \Delta S = \int dS = \int \frac{\delta Q}{T} = 0 \] (4.36)

This is known as the Clausius equality for reversible processes.[16]

### 4.12 Kinetic gas theory

In this chapter we look at how microscopic motion explains the macroscopic properties of a system. Solving the problem exactly is impossible due to the huge number of particles within a macroscopic volume, but with some approximations we can model the real system very accurately. The most important approximations are:

- The gas consists of very many, small particles and we approximate them as points, so we can neglect the volume they take.

- All particles are identical.

- The particles move with a constant, random velocity.

A full list of the approximations can be found on Wikipedia[17].

We look at a cube of side length \( L \) in which the particles are confined, like the one in figure 4.8. When a particle hits the wall, the change in momentum in the direction (e.g. \( x \)) is given by

\[ \Delta P = P_{\text{before},x} - P_{\text{after},x} = P_{\text{before},x} - (-P_{\text{before},x}) = 2P_{\text{before},x} = 2mv_x, \] (4.37)

where \( P \) is the momentum. Further the particle hits any given wall periodically in time with

\[ \Delta t = \frac{2L}{v_x}. \] (4.38)
between collisions, which gives an average force onto the wall of one particle of

\[ F_{\text{particle}} = \frac{\Delta P}{\Delta t} = \frac{mv_x^2}{L}. \] (4.39)

This results in a total force onto the wall of

\[ F = N \cdot \bar{F}_{\text{particle}} = N\frac{mv_x^2}{L}, \] (4.40)

where the bar over the force and the velocity indicates mean values over all particles. Since we do not have a bias in any direction, the average squared speed in any direction should be the same

\[ \bar{v}^2 = \bar{v}_x^2 + \bar{v}_y^2 + \bar{v}_z^2 = 3\bar{v}_x^2. \] (4.41)

This leads to a pressure of

\[ p = \frac{F}{L^2} = \frac{Nm\bar{v}^2}{3V}, \] (4.42)

with a volume \( V = L^3 \).

Comparing with the ideal gas law

\[ pV = \frac{Nm\bar{v}^2}{3} = Nk_bT, \] (4.43)
and knowing that $k_b$, $m$ and $N$ are constant, we see that the velocity squared is directly connected to the temperature

$$k_b T = \frac{mv^2}{3}. \quad (4.44)$$

By using this we can write down the average kinetic energy per particle

$$\frac{1}{2}mv^2 = \frac{3}{2}k_b T, \quad (4.45)$$

which we recognize as the equipartition theorem with particles having 3 degrees of freedom (they are able to move in x,y and z).

### 4.13 Phase transitions

Ideal gases do not show any phase transitions (e.g. condensation), but real gases like described in chapter 4.14 do. For every point in the $p$-$T$ diagram, there is exactly one phase that minimizes the energy. The system always tries to minimize this energy and therefore phase transitions happen when the system goes from one region to another. During phase transitions the two phases coexist.

When a phase transition happens usually depends on the pressure as well as on the temperature, as can be seen in figure 4.9. In the figure one can see the phase diagram of water, with the freezing point at 0°C and the boiling point at 100°C for normal atmospheric pressure. But with changing pressure, the necessary temperature to boil/freeze water gets shifted and one can even get water that is under 0°C cold.

When a system transitions from one phase to the other, heat is either released to the surrounding (e.g. condensation) or taken from the surroundings (e.g. evaporation). This heat is called the latent heat $L$ and is usually found as specific latent heat $L_m$ which is normalized by the mass. Therefore one can calculate the heat $Q$ needed to evaporate a material only by knowing its mass and the specific latent heat

$$Q = mL_m. \quad (4.46)$$

This is of course only valid if the system is already at the right temperature.
Figure 4.9: Phase diagram of H$_2$O. On the very left is the solid phase, in the middle the liquid phase and on the right, extending all the way to the left at low pressure, is the gaseous phase. [19]

4.14 Real gases

In this script we only looked at ideal gases, but just as an outlook we shall briefly discuss real gases.

As a first approximation we can introduce two empirical parameters to the ideal gas law, bringing us to the van der Waals gas:

$$\left( p + \frac{a n^2}{V^2} \right) (V - nb) = nRT.$$ \hspace{1cm} (4.47)

The two terms are:

1) There is not only the outer pressure, but the interaction of the molecules is also taken into account by the factor $a$.

2) The volume that the gas can move in is reduced by the volume the molecules occupy. This is taken care of by the factor $b$.

The probably biggest advantage of using the van der Waals equation is that with this model one can also take phase transitions into account, which is not included in the ideal gas law.
4.15 Stefan-Boltzmann law

Every body radiates power in the form of electro-magnetic waves because of its temperature. For example molten metals send out light in the visible range. The power radiated per area by a black body (a body that absorbs all incoming light) is given by the Stefan-Boltzmann law

\[ P = \sigma T^4. \]  \hspace{1cm} (4.48)

\( \sigma \) is the Stefan-Boltzmann constant and is given by

\[ \sigma = \frac{2\pi^5 k^4}{15 c^2 h^3} \approx 5.670 \times 10^{-8} \text{ Wm}^{-2}\text{K}^{-4}, \]  \hspace{1cm} (4.49)

where \( c \) is the speed of light and \( h \) is the Planck constant.
Chapter 5

Oscillations

5.1 Introduction

Important sources for this chapter include [20] and [21].
The term oscillation generally refers to the (periodic) variation of some physical quantity with time around a point of equilibrium.
There is no strong consensus among physicists for a proper definition of oscillations. Some prefer using broader definitions, to the point of including phenomena that do not "oscillate" in the common sense, while some others prefer a narrower definition, with the risk of excluding movements showing the characteristic "back-and-forth" variation. Damping is the common example of a phenomenon of the "grey zone", more on this later.
A way of physically defining oscillatory systems (oscillators) is to require that the varying quantity has an equilibrium value, and that the system naturally tends to go back to that value if taken out of equilibrium by an external perturbation.
From that definition, we already see that oscillators are very common in physics, virtually in any domains; oscillations, in particular, happen each time the considered quantity (which can be almost any mathematical quantity: a position, the strength of a field, the voltage of an alternating current circuit, etc.) is bound to a potential with a local minimum.
In the following, we will take a more precise look at simple oscillations, notably harmonic oscillations and their extensions - damping and resonance.
As a preamble, let us first define periodicity. This term refers to the idea of having some time-dependent quantity that takes the same values over and over, in a regular pattern. For example, considering the time-dependent variable $x(t)$, we call it periodic if we can find some $T$ such that

$$x(t) = x(t + T)$$

for all times $t$.
As can be seen, if we find such a $T$, then $2T$ also satisfies the criterion, as well as any multiple of $T$. Thus, we define the period of a periodic system as the (positive) smallest such $T$.  

139
It is clear from this definition that the period $T$ has the units of an interval of time, thus is
given in s (seconds) in the SI, and measures the time it takes an oscillator to accomplish an
oscillation cycle.

We also define the frequency $f$ (also commonly referred to as $\nu$): it is the inverse of the
period, i.e. $1/T$, and it measures how many oscillation cycles an oscillator accomplishes in
a time unit. Its SI unit is $s^{-1}$, also called Hz (Hertz).

## 5.2 Harmonic Oscillations

Harmonic oscillations are considered the most simple type of oscillation. Despite this (or
more probably, because of this), it also provides a convenient model for most types of (more
complicated) physical oscillations. As is often the case, mathematical and physical simplicity
come together.

Mathematically, a quantity $x(t)$ following a harmonic oscillation has the following form:

$$x(t) = A \sin(\omega t + \varphi), \quad (5.1)$$

that is, it simply follows a sine function, up to some scaling and shifting factors:

- $A$ is called *amplitude*, and has the same units as $x$. It corresponds to the maximal
  value of $x$.

- $\omega$ is called *angular frequency* (sometimes also *pulsation*) and corresponds to $2\pi f$. Its
  units are that of an angle over a time, i.e. radians per second in the SI. Mathematically,
  it is a scaling factor that allows transforming the parameter $t$ into an object of the
  right dimension for the sine function to absorb, i.e. $\omega t$ should have units of an angle.

- $\varphi$ is called *phase* and has units of an angle (for the same reason as just explained for
  $\omega t$), usually given in radians.

**Note:**

- some textbooks refer to $2A$ as the amplitude. Be careful!

- another definition can be created using a cosine instead of a sine. Those definitions
  are equivalent and only differ by a shift of $\pi/2$ in the phase.

It is easy to check that harmonic oscillations are periodic, it simply follows from the math-
ematical periodicity of the sine function:

$$x(t + T) = A \sin(\omega(t + T) + \varphi)$$
$$= A \sin(\omega t + \omega T + \varphi)$$
$$= A \sin(\omega t + 2\pi f T + \varphi)$$
$$= A \sin(\omega t + 2\pi T/T + \varphi)$$
$$= A \sin(\omega t + \varphi + 2\pi)$$
$$= A \sin(\omega t + \varphi)$$
$$= x(t).$$
5.2.1 Harmonic Oscillations, Spring/Mass Systems and Differential Equations

At this point, we want to look at a first concrete example of harmonic oscillator and try to relate the physical equations of such a system with the mathematical form of harmonic oscillations described above.

The simplest example of harmonic oscillator is perhaps the spring/mass system, where an extremity of the (massless) spring is fixed and the other is attached to a free mass \( m \). We will only consider here the 1D case, where the mass can only move along one dimension. We do not take any gravitational force into account.

We will later see that by accounting for further interactions (friction, driving by an external movement, etc.) we will be able to generalize this simple example and therefore study various interesting oscillatory cases.

![Spring-mass system](image.png)

Figure 5.1: Spring-mass system. Adapted from [22].
Newton’s equation for the mass is as follows:

$$\sum_i F_i = ma .$$

The only force acting on $m$ is the restoring force of the spring (Hooke’s law):

$$F = -kx ,$$

where $k$ is the spring constant (defining its stiffness) in N·m$^{-1}$ (equivalent to kg·s$^{-2}$). The origin of $x$ is taken at the point of equilibrium of the spring. Thus we have, by reordering and using that the acceleration is the second time derivative of the position ($a = \ddot{x}$):

$$\ddot{x} + \frac{k}{m}x = 0 .$$

(5.2)

Such an equation is called a homogeneous second-order linear constant coefficient ordinary differential equation:

- **differential**, because it contains both a function ($x(t)$) and some of its derivatives ($\ddot{x}(t)$), and the goal is to determine $x(t)$;
- **ordinary**, because the function $x$ depends on only one variable $t$;
- **constant coefficient**, because the function and its derivatives only show up with constant coefficients in the equation;
- **linear**, because the function and its derivatives only show up with degree 1 in the equation (no square or such);
- **second-order**, because the highest derivative of $x$ present is the second one;
- **homogeneous**, because there is no "free" constant or function of $t$ in the equation (all terms contain an $x$ or one of its derivatives at least).

Due to Newton’s law, most of mechanics revolve around finding solutions to second-order differential equations.

Our hypothesis above was that spring/mass systems are harmonic oscillators, i.e. we believe that such systems perform harmonic oscillations (with the variable quantity being the position of the mass).

To verify this, we have to check whether 5.1 and 5.2 are compatible, i.e. whether 5.1 is a solution of 5.2.
Let’s insert 5.1 into 5.2:

\[ x(t) = A \sin (\omega t + \varphi) \]
\[ \dot{x}(t) = \omega A \cos (\omega t + \varphi) \]
\[ \ddot{x}(t) = -\omega^2 A \sin (\omega t + \varphi) \]

\[ \ddot{x} + \frac{k}{m} x = -\omega^2 A \sin (\omega t + \varphi) + \frac{k}{m} A \sin (\omega t + \varphi) \]
\[ = \left( -\omega^2 + \frac{k}{m} \right) A \sin (\omega t + \varphi) \]
\[ \overset{!}{=} 0 \, . \]

For the last equality to hold for all \( t \), we need to have

\[ -\omega^2 + \frac{k}{m} = 0 \, , \]

i.e.

\[ \omega = \sqrt{\frac{k}{m}} \, . \]

Thus, a spring/mass system with the above defined characteristics does behave as a harmonic oscillator with angular frequency \( \sqrt{k/m} \).

We can also conclude that any system with a single force \(-Kx\) for some constant \( K \) acting on it is harmonic. Such a system has a potential \( V(x) = \frac{1}{2} Kx^2 \), which also suffices to characterize it.

### 5.2.2 Further Examples

Beyond the spring/mass, other simple systems represent harmonic oscillators:

- **Torsion Pendulum**

  Hooke’s law also holds for the angular position \( \varphi \) of an object of moment of inertia \( J \) attached to a torsion spring of angular spring constant \( \kappa \) (with the torque \( M \)):

  \[ M = -\kappa \varphi \, , \]

  thus with Newton’s law for rotations:
\[
\sum_i M_i = J\alpha \\
\ddot{\varphi} + \frac{\kappa}{J} \varphi = 0.
\]

And we again find a solution of the form \(\varphi(t) = \Phi \sin \left(\sqrt{\frac{\kappa}{J}} t + \theta\right)\).

- Simple Pendulum

A simple pendulum is composed of a (massless) rod of length \(l\) fixed at one of its extremities to a pivot, and attached at the other end to a moving mass.

In the common case, one restricts oneself to the 2D. Using the angle \(\varphi\) of the rod with the vertical, one can then parametrize the system with one single variable - thus the system has one single degree of freedom.

Two forces act on the mass, the gravity \(\vec{F}_P\) and the tension \(\vec{F}_T\).

Seen from the pivot point, the tension produces no torque (because of its collinearity with the rod, \(M_T = 0\)), while the weight of the mass creates

\[
M_P = -mgl \sin(\varphi)
\]
Thus we get, using Newton’s law in rotationary form (we assume the mass to be point-like, so \( J = ml^2 \) by Steiner’s theorem):

\[
\ddot{\varphi} + \frac{g}{l} \sin(\varphi) = 0 .
\]  

As can be seen, we again find a homogeneous second-order constant coefficient ordinary differential equation, but this time it is not linear, due to the sine! Therefore the simple pendulum is not harmonic; as can be shown, 5.1 is not a solution of 5.3.

However, in the case of small oscillations, the angle \( \varphi \) remains itself small and we can linearize the equation. This is because of the Taylor expansion of sine:

\[
\sin(x) = x - \frac{x^3}{6} + \frac{x^5}{120} + \mathcal{O}(x^7) ,
\]

which allows to approximate \( \sin(x) \approx x \) for small \( x \).

Thus, for \( \varphi \ll 1 \) (with \( \varphi \) in radians), we have

\[
\ddot{\varphi} + \frac{g}{l} \varphi \approx 0
\]

and we can say that the pendulum is approximatively harmonic with angular frequency \( \sqrt{g/l} \). Note that the condition \( \varphi \ll 1 \) is necessary for the approximation to hold!
5.2.3 Importance of Harmonic Oscillations

The fact that non-harmonic oscillations can be approximated in the vicinity of equilibrium points by harmonic oscillations is general, and is the reason why harmonic oscillations are so common and important.

More concretely, if we have a system with potential $V(x)$ we can Taylor-expand (say around 0):

$$V(x) = V(0) + V'(0)x + \frac{V''(0)}{2}x^2 + \mathcal{O}(x^3) .$$

If the potential has a local minimum at 0, we have $V'(0) = 0$ and $V''(0) > 0$, thus for small $x$ we can neglect the higher orders and get

$$V(x) \approx V(0) + \frac{V''(0)}{2}x^2 ,$$

which is the potential of a harmonic oscillator ($V(0)$ is arbitrary and we can redefine $V$ in order to get rid of it). In terms of force we get

$$F = -\frac{\partial V}{\partial x} = -V''(0)x$$

which is obvious linear in $x$ ($V''(0)$ is the value of the second derivative at 0 and therefore simply a number).

5.3 Beyond Harmonic Oscillations

Here we will consider our mass-spring system, but accounting for further phenomena. In fact, harmonic oscillations represent an idealized case. In reality, we often have two more ingredients in oscillatory systems:

- **Damping**
  Damping occurs everytime we have friction in the system, opposing the motion.

- **Forcing**
  Forcing (also called driving force) appears when an external element interacts with the oscillating system. Depending on this interaction, this can typically greatly increase or reduce the oscillation amplitude.

In this script we will consider the following simple situation only: linear damping (with a force proportional to the velocity, $-bv$) and sinusoidal forcing (with an external excitation $B\sin(\Omega t)$).
Newton’s law gives for such a system:

\[-kx - bv + B \sin(\Omega t) = ma,\]

which leads to the (in general inhomogeneous due to the forcing) differential equation:

\[\ddot{x} + 2\zeta \omega \dot{x} + \omega^2 x = \frac{B}{m} \sin(\Omega t),\]

where \(\zeta = \frac{b}{2\sqrt{mk}}\) is called the damping ratio.

The full mathematical treatment of this differential equation is beyond the scope of this script (it is partly covered in the analogous case of oscillatory circuits in the AC part), so we only provide the solutions without resolution steps below. We distinguish the following cases:

- **\( B = 0 \)**

  This is the free case, i.e. there is no forcing. The behavior of the system depends on the damping ratio:

  - \( \zeta = 0 \)
    
    There is no damping here, thus this is the common harmonic oscillator, as discussed above.

    As soon as \( \zeta \) gets bigger than 0, the oscillator looses energy in the damping (friction) and is not periodic anymore, but rather tends to go back to its equilibrium point and stop there.

  - \( 0 \leq \zeta < 1 \)
    
    This case is called underdamped and corresponds to solutions of the form

    \[x(t) = Ce^{-\zeta\omega t} \sin(\tilde{\omega}t + \tilde{\varphi}),\]

    with \( C \) and \( \tilde{\varphi} \) arbitrary constants depending on the initial conditions. As can be seen, the movement represents a sinusoidal modulated (squashed) by an exponential decay. It is not periodic, but one can define \( \tilde{\omega} \) as a so-called pseudoperiod for the oscillation. We have

    \[\tilde{\omega} = \omega \sqrt{1 - \zeta^2} .\]
\[ x(t) = (C_1 t + C_2) e^{-\zeta \omega t}, \]

where \( C_1 \) and \( C_2 \) are again arbitrary constants depending on the initial conditions.

The system here approaches its equilibrium point in the fastest possible way, which makes it technically useful in real-world applications, such as springs for automatically closing doors (that should close fast but without oscillating) or shock absorbers on vehicles.

- \( \zeta > 1 \)

Here we have the overdamped case, with the system going back to its point of equilibrium without oscillating, but slower that in the critical case described above.

The solutions have the form

\[ x(t) = C_1 e^{-\left(\zeta - \sqrt{\zeta^2 - 1}\right) \omega t} + C_2 e^{-\left(\zeta + \sqrt{\zeta^2 - 1}\right) \omega t}, \]

with \( C_1 \) and \( C_2 \) defined in an equivalent way as above.

\[ x(t)/x(0) \]

\( \omega t \)

\[ \zeta = 0.0 \quad \zeta = 0.5 \quad \zeta = 1.0 \quad \zeta = 1.5 \]

Figure 5.4: Linear damping with various \( \zeta \) values. Adapted from [27].

- \( B \neq 0 \)

In the presence of forcing, we will have to separate to cases:

- We will first have a transitory solution depending on initial conditions.
- Then, we will have a steady-state solution depending only on the forcing.
The general solution is a sum of both transitory and steady-state solutions; here we will consider the steady-state solution only. It has the form

$$x(t) = \frac{B}{m \sqrt{4\omega^2\Omega^2\zeta^2 + (\omega^2 - \Omega^2)^2}} \sin(\Omega t + \Phi)$$

with some phase $\Phi$ that we won’t discuss in this script. The interesting part is the square root $Z = \sqrt{4\omega^2\Omega^2\zeta^2 + (\omega^2 - \Omega^2)^2}$.

For $\zeta < \frac{1}{\sqrt{2}}$, we will have a minimum of $Z$ for a forcing frequency $\Omega_r = \omega \sqrt{1 - 2\zeta^2}$, thus the amplitude of the oscillation for that case of forcing will be strongly increased. In the undamped case ($\zeta = 0$), the amplitude in that case will even diverge to infinity, meaning that for that forcing frequency $\Omega_r$ (called resonance frequency) the forcing will continuously add energy to the system.

Figure 5.5: Amplitude resonance with sinusoidal forcing. The maxima line corresponds to $\Omega_r$. Adapted from [26].
5.3.1 Example of Forced Oscillating Systems

One of the most common examples of forcing is the swing: one has to furnish energy at the right time (i.e. with the right frequency) for the swing to increase its oscillation amplitude. Conversely, any sufficiently little damped mechanical system has a resonance frequency that can be used to bring it to strongly oscillating, even with a relatively small amplitude driving interaction.

During the engineering of any mechanical product, resonance has to be taken into account in order to prevent unwanted oscillations that could even make it break. Among others, vehicles parts should not be brought to resonance by the vibrations induced by the engine. Buildings are another type of mechanical systems where resonance has to be well controlled. Modern bridges and skyscrapers often contain mechanical parts designed to absorb oscillations.

One of the most well-known examples of so-called resonance disaster is the destruction of the Tacoma Narrows Bridge in 1940 due to wind passing through its structure, bringing it to twisting and eventually leading to its collapse.

Figure 5.6: The Tacoma Narrows Bridge collapsing [24]. See also [25].
Chapter 6

Waves

6.1 Introduction

In the previous chapter, we spoke about oscillations. Now we can ask ourselves the question "what happens if we couple many oscillators together into a network, so that oscillations on one of them also influence its neighbors?"

What happens is that the perturbation will little by little, neighbor after neighbor, "travel" across the oscillators network. Such a propagating perturbation is called progressive wave. Most waves are progressive waves, but there are some counterexamples such as standing waves.

An exact definition of waves is difficult to construct, but one can anyways say that waves are perturbations (variations of some quantity, typically in an oscillating way) in space and time. This is the main difference with oscillations, which occur in more or less abstract quantities with time: waves add the idea of space.

Mathematically, waves are often represented as functions of time and space \( u(\vec{r}, t) \) \((u(x, t)\) in the 1D case). Sometimes the oscillations composing the wave also have a determined direction in space, thus the wave in this case is represented by a vector function \( \vec{u}(\vec{r}, t) \).

We will not go into much details in this script, but the temporal evolution of those functions is fixed by the wave equation, which in 1D has the form

\[
\frac{\partial^2 u(t, x)}{\partial t^2} = v^2 \frac{\partial^2 u(t, x)}{\partial x^2}.
\]

(6.1)

This is a partial differential equation (as opposed to an ordinary DE), meaning that it contains partial derivatives \(u\) is function of multiple variables and gets differentiated - such as in \( \frac{\partial u}{\partial t}\) - with respect to one single of those variables at a time, by letting other variables constant). This is a direct consequence of our letting space play a role in the description of waves, in opposition to the ordinary differential equations we saw for oscillations, where only time plays a role.

The \(v\) in the wave equation can already let us think about the velocity of a wave, which is in general an important property. Sometimes \(c\) is used instead of \(v\).

Waves can be seen as generalizations of oscillations, so in the following we will begin by taking a look at general properties of waves - which may recall us about oscillations. Then
we will talk about their propagation and the media associated. Finally the interference phenomenon (two or more waves interacting with each other) will be discussed.

6.2 Harmonic Waves

We will provide a more in-depth description of waves using the example of harmonic waves. Those are waves that can be described using a sinusoidal function:

\[ u(x,t) = A \sin(\omega t - kx + \varphi) \, . \]  

(6.2)

As can be seen, this mathematical form is quite similar to the one we used for harmonic oscillations, with the difference of the argument of the sine gaining a spatial term. Sometimes, the reversed convention is used, \( u(x,t) = A \sin(kx - \omega t + \varphi') \). Both conventions are equivalent, up to a change in the phase from \( \varphi \) to \( \varphi' \). Here we use a positive time and a negative space, as it more directly relates to the case of oscillations.

As was the case before, \( u \) as a function of \( t \) (provided that \( x \) is held constant) has a sinusoidal form. This means that, at any given point, the "quantity the wave moves in" oscillates with time. But it is also to be noted that the spatial coordinate \( x \) has an equivalent place in the formula as the temporal coordinate \( t \). Thus the other way round also holds: at any given time, the "quantity the wave moves in" oscillates with space. It is therefore very important, when looking at a plot describing the amplitude of a harmonic wave, to check whether it drawn against time or space. In both cases the figure will appear sinusoidal.

When talking about oscillations, we defined \( \omega \) as the angular frequency, a measure of the variation "rapidity" of the quantity with time. In the formula above, \( x \) has a similar role as \( t \) and it also has some factor \( k \) associated, called wavenumber. And again similarly to the time, where we defined the period \( T = 2\pi/\omega \) as the time interval between (say) two peaks of the oscillation (or wave, at a given point), we can define the wavelength \( \lambda = 2\pi/k \), which is nothing but the space interval (i.e. the distance) between (say) two peaks of the wave (at a given, fixed time).

From this, we can derive a very important equation by inserting 6.2 into 6.1:
6.2. HARMONIC WAVES

\[ \frac{\partial u}{\partial t} = \omega A \cos(\omega t - kx + \varphi) \]
\[ \frac{\partial^2 u}{\partial t^2} = -\omega^2 A \sin(\omega t - kx + \varphi) \]
\[ \frac{\partial u}{\partial x} = -kA \cos(\omega t - kx + \varphi) \]
\[ \frac{\partial^2 u}{\partial x^2} = -k^2 A \sin(\omega t - kx + \varphi) \]
\[ -\omega^2 A \sin(\omega t - kx + \varphi) = \frac{\partial^2 u}{\partial t^2} \]
\[ = \frac{v^2 \partial^2 u}{\partial x^2} \]
\[ = v^2 (-k^2) A \sin(\omega t - kx + \varphi) \]

\[ \Rightarrow \omega^2 = v^2 k^2 \]
\[ \Rightarrow v = \omega/k = \lambda/T = \lambda f , \]

where we assume \( v, \omega \) and \( k \) to be positive (which is mostly a matter of convention).
This relates the temporal and spatial characteristics of waves through a parameter, the velocity of the wave (which is not very surprising, considering that velocities express in fact the division of a space interval by a time interval).

6.3 Waves in 3D

6.3.1 Waves as Functions of 3D Spatial Coordinates

When going from waves in one dimension to three, the first element to look at is the fact that now the function $u(x,t)$ is modified into $u(\vec{r},t)$.

In order to adapt our formula for harmonic waves, we define $\vec{k}$, the wavevector and write:

$$u(\vec{r},t) = A \sin(\omega t - \vec{k} \cdot \vec{r} + \varphi) . \quad (6.3)$$

While $\vec{k}$ still indicates the "rapidity" of variation of the wave with space (through its norm $k = |\vec{k}|$), it now also has a direction, which actually is the direction of propagation of the wave itself.

The formula 6.3 creates so-called planar waves: at a given time, all points in planes perpendicular to the wavevector share the same oscillation state.

6.3.2 Waves as 3D Functions, Transversal and Longitudinal Waves, Polarization

Depending on the "quantity the wave moves in", not only the oscillations defining the wave can depend on all the three spatial coordinates, but the oscillations themselves can have a well-defined direction in space, which is mathematically indicated by a vectorial function $\vec{u}(\vec{r},t)$. In the harmonic case, in general we simply modify 6.3 by using a vectorial amplitude $\vec{A}$ to take this oscillation direction into account.

As we saw before, waves in 3D have a particular direction determined by $\vec{k}$, the propagation direction. For waves whose oscillations themselves have a well-defined direction, we have two possible cases:

- The oscillation direction can be perpendicular to the propagation direction, in which case the wave is named transversal.

- Or it can be parallel, and we have a longitudinal wave.

In the first case, even for a fixed propagation direction the oscillation direction is not unique, but can be freely chosen in the plane perpendicular $\vec{k}$. We thus define the polarization, which is the oscillation direction. Two transversal waves going in the same direction can nevertheless have different polarization states.
6.4 Waves Propagation

In the introduction to the present chapter, we said that waves are perturbations travelling through a "network" of coupled oscillators. Later on, we used the expression "quantity the wave moves in" in several sentences. It is now time to try answering the question "what does a wave move in?", i.e. find what is the propagation medium of waves.

This in turn depends on the type of wave, and in particular on the nature of the oscillators. In the following, we will introduce a couple of examples of wave forms as concrete cases for the different properties that were discussed in the previous sections. This will allow us to also get an insight into the different propagation media.

6.4.1 Sound

Sound is made of so-called mechanical waves, i.e. waves whose oscillators are positions of matter particles - be it in a solid, liquid or gas. In the case of sound, those waves are longitudinal, so matter particles get pushed away in the direction of propagation by previous ones, where they hit next ones, thus propagating the wave further and letting them bounce back (in average) to their original position.

Like any mechanical waves, sound cannot propagate in the absence of physical medium (like in vacuum). Even if an enormous asteroid closely passes by your spacecraft in outer space,
The frequency of sound is directly related to the perceived *height* - a higher frequency leading to a higher *pitch*. It also leads to a shorter wavelength, as the sound travels at approximately the same velocity in a given material (round $343 \text{ m/s}^{-1}$ in air) under given conditions, independently of its frequency.

### 6.4.2 Light

While light also composes itself of waves - and while we also can sense it (in some cases), one could not imagine anything more different from sound than light!

The first fundamental difference between them is that light is not a material wave, i.e. it is not made of matter particles brought to oscillating. Moreover, it does not need any matter to propagate (the question on the nature of a hypothetic physical medium - "luminiferous aether" - for light propagation actually participated in the development of relativity).

While light does not need any material to propagate, the presence of a material can have an impact on its propagation, such as changing its velocity, direction, etc. We will talk about these general effects later on.

We said that waves travel on "networks of oscillators" so, even for light there has to be something supporting it. This "something" is the *electromagnetic field*, and thus light is actually an electromagnetic wave. Concretely, this means that light beams are perturbations across space that bring electric and magnetic fields to locally oscillate.

In most cases (and most importantly in vacuum), light is a transversal wave. Thus, in any point on the light path, both $\vec{E}$ and $\vec{B}$ are perpendicular to $\vec{k}$ ($\vec{E}$ and $\vec{B}$ are actually also perpendicular to each other).

As a transversal wave, light is subject to polarization, given by the direction of the electric field. The existence of polarization filters, which can block light depending of its polarization orientation, is crucial for many technologies in both research and engineering; but it is also at the origin of the most common 3D cinema system, where two images are projected on the screen simultaneously, with perpendicular polarizations. Eyeglasses similarly have two lenses with corresponding polarization filters, so that each lens exactly lets one of the images pass through while blocking the other one.

Wavelength and frequency of electromagnetic waves cover a huge range of magnitude orders, which also leads to very different types of interaction with matter.
6.4. WAVES PROPAGATION

6.4.3 Seismic Waves

Like sound, seismic waves - waves created by geological events in the inner of the Earth, such as earthquakes - are material waves. Unlike them, though, they are not necessarily longitudinal. In fact, single events often create a bunch of seismic waves of different types - both longitudinal and transversal, but also waves travelling through the Earth or across its surface, etc. All those different types come in general together with different wavelengths, frequencies and velocities, which also depend on the propagation material. In turn, seismology can take advantage of all these properties to better understand the Earth’s inner structure as well as to study geological risks.

6.4.4 Transport

While they may transport energy, it is important to note that waves do not transport matter (not taking into account the small local displacements in material waves).

This may seem counterintuitive, for example in waves on a water surface: the waves peaks move across the surface, but the water particles themselves only move up and down. Thus an object floating on the water - such as a duck - is itself brought to (almost) only oscillate vertically.

6.4.5 Doppler Effect

Doppler Effect is a phenomenon occurring when the emitter (an object creating a wave) and/or the receiver (an hypothetic object observing wave oscillations) of a wave move with respect to each other and, in the case of material wave, with respect to the propagation medium.

In this script we will restrict ourselves to the 1D case for sound waves.
Let’s consider an emitter travelling with velocity \( v_{em} \) and creating a sound wave of frequency \( f_{em} \) and velocity \( v \). Let’s also assume that a receiver travelling with velocity \( v_{re} \) is at a distance \( d \) from the emitter at a given time (which we will set to zero together with the emitter’s initial position for the sake of simplicity).

Note that we do not make any assumption on the signs of \( v_{em} \) and \( v_{re} \) but, as we are considering the 1D case, each object is moving either directly towards or away from the other one. We will simply define \( v_{em} \) and \( v_{re} \) as positive when they are in the same direction as the vector going from the emitter to the receiver, as negative otherwise.

Moreover, both \( v_{em} \) and \( v_{re} \) are relative to the medium.

![Emitter and receiver](image)

Figure 6.4: Emitter and receiver. Due to the Doppler effect, the frequency of the sound will be different for the emitter, the receiver and a neutral observer at rest with respect to the medium.

Let \( t_1 \) be the time taken by the sound emitted at the initial instant to arrive to the receiver. We have:

\[
v t_1 = d + v_{re} t_1,
\]

Thus \( t_1 = d/(v - v_{re}) \). Similarly, let \( t_2 \) be the time taken by the sound emitted after a period of the sound (1/\( f_{em} \)) to arrive to the receiver, so:

\[
v t_2 = d - v_{em} \frac{1}{f_{em}} + v_{re} \frac{1}{f_{em}} + v_{re} t_2,
\]

so we have \( t_2 = \left( d + (v_{re} - v_{em}) \frac{1}{f_{em}} \right)/(v - v_{re}) \).

From \( t_1 \) and \( t_2 \) we can calculate the frequency of the sound \( f_{re} \) as observed by the receiver, as a function of \( f_{em} \) (\( t_2 - t_1 \) representing the difference in periods between the sound received and emitted):
\[ f_{re} = \frac{1}{t_2 - t_1 + \frac{1}{f_{em}}} = \frac{1}{\frac{v - v_{re}}{v - v_{em}} + \frac{v - v_{em}}{v - v_{re}}} = \frac{v - v_{re}}{v - v_{em}} f_{em}. \]

Note: the signs in the fractions depend on the way we defined the velocities. They vary among textbooks and one should be careful when applying the formula.

The Doppler effect typically leads to a higher pitch for approaching source and lower when it recedes from the observer. An everyday example is the siren of emergency vehicles: one can well hear the frequency drop when the vehicle passes by.

It can also be used in technical applications, by sending a sound signal and analysing the frequency of the signal reflected back: radars and blood flow imaging in cardiology are typical uses.

6.5 Waves Propagation at Interfaces

Under this title we understand phenomena where, in contrast to what we saw before with (implicitly assumed) homogeneous, isotropic and infinite propagation spaces, there is some element in the path, preventing the wave from freely propagating further.

An important principle for understanding those phenomena is Fermat’s principle: it states that light (and it can be generalized on most waves) always "chooses" the locally fastest way between two points, i.e. the way that it takes the shortest time for the wave to travel. We will see later on why this "locally" is important, when studying reflections.

6.5.1 Reflection

Reflection happens when a wave bounces back at an interface between its propagation medium - be it material or not - and some material element.

Supposing that the element’s interface with the medium is sufficiently smooth, we can define the incidence angle \( \varphi_i \) and reflection angle \( \varphi_r \) between the surface normal and the respective path elements.

Using Fermat’s principle, it can be shown that this is exactly the case when both angles are equal, \( \varphi_i = \varphi_r \). We also easily see the local aspect of the principle: of course this path is not the globally shortest one (a straight line between the extremities of the path would be shorter), but of the continuum of paths going from one point to the other while hitting the interface, it is the shortest one.

A straightforward way to see how Fermat’s principle implies equal angles is to imagine that the mirror does not exist, but that instead one of the extremities of the path is mirrored, i.e. on the other side of the mirror’s plane. Then it is clear that to minimize the distance, one has to consider the straight line from that mirrored extremity to the other, unmodified one. And it is also directly visible that the equality of angles should be respected in that case.
Figure 6.5: Reflection on a plane surface of a wave travelling from \( A \) to \( B \). By constructing the image \( \tilde{A} \) of \( A \) by symmetry w.r.t. the plane, one sees that \( \phi_i = \phi'_i \) as well as \( \phi_r = \phi'_i \) (Fermat’s principle requiring \( \tilde{A}B \) be a straight line). Hence \( \phi_i = \phi_r \).

### 6.5.2 Refraction

*Refraction* occurs when a wave arrives at an interface between two propagation media, which typically have different properties, such that the wave has a different propagation velocity in each of them.

In order to take these velocities into account, one defines - in case of light - the *refractive index* \( n \) of a medium as the fraction of the vacuum velocity \( c \) by the velocity in the medium \( v \): \( n = c/v \). The refractive index is slightly dependent on the light wavelength, which is one of the main reasons for *chromatic aberration* in refractive optical parts. We will nevertheless forget this detail in this script.

As can be shown using Fermat’s principle, the angles of incidence \( \phi_1 \) and of refraction \( \phi_2 \) at the interface between media of refractive indices \( n_1 \) and \( n_2 \) are related by Snell-Descartes law:

\[
 n_1 \sin(\phi_1) = n_2 \sin(\phi_2) .
\]
Figure 6.6: Top: Refraction at a medium interface of a wave travelling from $A$ (in medium with refraction index $n_1$) to $C$ (index $n_2$). Here, $n_2 < n_1$ is implied. Middle: limit case with $\varphi'_2 = \frac{\pi}{2}$. $\varphi'_1$ is thus called the critical angle. Bottom: For $\varphi''_1 > \varphi'_1$, no refraction is possible anymore, only reflection. This phenomenon is therefore called total internal reflection.
6.5.3 Diffraction

Diffraction is what happens when part of the wavefront (the most advanced part of the perturbation, with all points in the same oscillation state) hits an object: secondary waves are formed at the hitting points, which make the wave "turn around" the object, following its curvature a bit. For the diffraction to have an important effect, the wavelength of the wave has to be of roughly the same scale as the object in the way. This is typically the reason why we can hear a sound without actually seeing its source, because the objects in the way are at our scale - which is roughly also the sound’s scale - but are much too big for visible light to get diffracted enough to come to us.

![Figure 6.7: Diffraction of a plane wave passing through a slit. Notice how the diffracted wave is approximately circular and thus also reaches zones (even if attenuated) that are "hidden" sideways of the slit, along the barrier. [28]](image)

6.6 Multi-Waves Phenomena

Until now, we only considered single waves going through space. However, as waves are simply perturbations, nothing forbids two waves from being at the same place at the same time. Thus we have to consider the so-called wave superposition, or wave interference. The main idea here is the superposition principle, which states that, given two waves \( u(\vec{r}, t) \) and \( v(\vec{r}, t) \) of same type (i.e. sharing the same oscillators), the perturbation resulting from their cumulated effect is simply their algebraic sum \( u(\vec{r}, t) + v(\vec{r}, t) \). The same goes with vectorial waves as well.

The resulting perturbation can have different forms depending on what the original waves look like, but the most important property is that the result is again a wave. In the following,
we will study the most common and interesting interference phenomena for 1D harmonic waves.
In general, we will therefore look at

\[ u(x, t) + v(x, t) = A_1 \sin(\omega_1 t - k_1 x + \varphi_1) + A_2 \sin(\omega_2 t - k_2 x + \varphi_2) \]

and try to understand what is going on depending of the relation between \( A_1 \) and \( A_2 \), \( \omega_1 \) and \( \omega_2 \), \( k_1 \) and \( k_2 \), and \( \varphi_1 \) and \( \varphi_2 \).

### 6.6.1 Same amplitude, frequency and wavenumber

The simplest case is when \( A_1 = A_2 = A \), \( \omega_1 = \omega_2 = \omega \) and \( k_1 = k_2 = k \), i.e. the waves are very similar but can simply be shifted by some phase. For the sake of simplicity, we will set \( \varphi_1 = 0 \) and \( \varphi_2 = \varphi \).

We can use the sine addition formula to rewrite the interference:

\[ u(x, t) + v(x, t) = A \sin(\omega t - k x) + A \sin(\omega t - k x + \varphi) = 2A \sin \left( \omega t - k x + \frac{\varphi}{2} \right) \cos \left( \frac{\varphi}{2} \right). \]

As can be seen, the resulting wave is again harmonic, with the same frequency and wavenumber. Depending on \( \varphi \), its amplitude can vary:

- For \( \varphi = 0, \pm 2\pi, \pm 4\pi \), etc., the amplitude is maximal and equal to \( 2A \). This is the perfect addition of two identical waves.
- For \( \varphi = \pm \pi, \pm 3\pi \), etc., the amplitude is null, i.e. the waves are exactly out of phase and thus completely annihilate.

### 6.6.2 Same amplitude and frequency, opposite wavenumber

This situation corresponds to two identical waves propagating in opposite directions, so \( A_1 = A_2 = A \), \( \omega_1 = \omega_2 = \omega \), \( k_1 = -k_2 = k \). We can set \( \varphi_1 = \varphi_2 = 0 \) without loss of generality.

Again using the addition of sines:

\[ u(x, t) + v(x, t) = A \sin(\omega t - k x) + A \sin(\omega t + k x) = 2A \sin(\omega t) \cos(k x). \]

Very interestingly, by summing the waves we get a separation of \( \omega t \) and \( k x \), which are therefore uncoupled. This means that the oscillations of the resulting wave are static - some places will constantly have zero amplitude due to the \( \cos(k x) \) ("nodes") while others will
have maximal amplitude ("anti-nodes"). Similarly, at some times the whole wave will have zero amplitude everywhere due to the $\sin(\omega t)$, etc.

Such a wave, that does not seem to travel along its medium, is called a stationary wave or standing wave. It is quite common, as it forms everytime some wave gets reflected back and interferes with itself.

Note that the distance from node to node and from antinode to antinode is $\frac{\lambda}{2}$ (for space) and $\frac{T}{2}$ (for time).

### 6.6.3 Slightly different frequencies

Another interesting case is $A_1 = A_2 = A$, $k_1 = k_2 = k$ but $\omega_1 \neq \omega_2$. We can set both phases to zero without loss of generality.

Using once more the addition of sines:

$$u(x, t) + v(x, t) = A \sin(\omega_1 t - kx) + A \sin(\omega_2 t - kx)$$

$$= 2A \sin \left( \frac{\omega_1 + \omega_2}{2} t - kx \right) \cos \left( \frac{\omega_1 - \omega_2}{2} t \right).$$

What we have here is a conventional harmonic wave of angular frequency $\frac{\omega_1 + \omega_2}{2}$ - i.e. oscillating more rapidly than the original waves - modulated by an oscillation (not a wave - note the absence of $x$) of angular frequency $\frac{\omega_1 - \omega_2}{2}$. In the case where $\omega_1 \approx \omega_2$, this modulation will be slow with respect to the oscillation of the wave itself.
Figure 6.9: The resulting stationary wave \( w(x, t) = u(x, t) + v(x, t) \) represented in both space and time. In both plots, nodes and antinodes remain at the same \( x \), resp. \( t \) coordinate, hence "stationary". For our choice of wave equation and \( \varphi \)'s, spatial antinodes are at \( n\lambda/2 \), while temporal nodes are at \( nT/2 \), \( n \in \mathbb{Z} \). Note that \( w(x, 0) = 0 \) for all \( x \)'s.

The result is the so-called beat phenomenon, with an almost harmonic wave periodically increasing and decreasing in intensity. \( \omega_1 \approx \omega_2 \) is needed for the phenomenon to be visible/audible.

### 6.6.4 Fourier Analysis

We just saw that the sum of two harmonic waves is again a periodic wave, and in some cases is even harmonic itself. We can generalize this idea and prove that any superposition of harmonic waves is periodic.

Conversely, an important result says that any periodic oscillation can be decomposed in a sum of (potentially infinitely many) harmonic oscillations. Even more interestingly, it exists a well-defined mathematical operation, called Fourier-transform, that does exactly that. More precisely, given a periodical signal, it allows to calculate its spectrum, i.e. the distribution of amplitudes for each frequency of fundamental harmonic oscillation.

Fourier Analysis can be thought of as the counterpart of Taylor expansion in terms of periodical signals instead of polynoms, and is a crucial instrument in the modern digital world, among other uses. For example, analogic audio signals can be converted (digitalized) into a sequence of numbers representing a finite approximation of the signal’s spectrum, thus allowing an efficient digital treatement and storage.
Figure 6.10: The modulated beat wave is the result of the multiplication of the dotted wave by a temporal oscillation (dashed) and an amplitude factor, $2A$. Note that both the beat and the dotted lines represent waves, and as such also move in space. In contrast, the dashed line represents a pure temporal oscillation that only "modulates" (constrains the amplitude of) the beat wave. Note also that the "envelope" of the beat wave is a periodic, non-sinusoidal oscillation with a period half of the modulating cosine.
Chapter 7

Fluid dynamics

In this chapter we discuss the behaviour of fluids macroscopically. Macroscopically means that we do not look at the individual particles forming that fluid but the fluid as continuous system of particles. This will then lead to important concepts such as hydrostatic pressure or buoyancy. Furthermore we will encounter very fundamental equations such as the continuity equation or Bernoulli’s equation.

7.1 Introduction

We will first explain the basic assumptions here and then present some important results in the next sections.

7.1.1 What is fluid dynamics about?

Fluid dynamics describes the dynamic properties of fluids. Dynamic properties (or systems) are those who can change in time, as opposed to static ones\(^1\). The other part of the name is fluid. A fluid is generally something which can flow\(^2\) (i.e. it is a gas or a liquid). However the distinction between a fluid and a solid is not that easy in general, as something may look solid if we only look at it for a short time but it flows on larger timescales (e.g. a glacier or the pitch drop experiment - google it).

7.1.2 How can we model such a fluid?

The first formal concept used is that of a trajectory. Intuitively this is the path of some small particle put into the fluid, e.g. a leaf in a stream. For each point in space we can find exactly one trajectory passing through this point (at any given point in time). The second - more important - concept is that of a velocity field, which is just a function giving us the velocity of a fluid at each point in space and time (see figure 7.1). If we have such a velocity field we can define flow lines (analogously to field lines in electrodynamics). This lines are

\(^1\)Normally static systems are seen as an (mostly) easier special case of the corresponding dynamic systems. I.e. we will also look at static properties in this chapter.

\(^2\)Technically speaking, we could say that in contrary to a solid, in a fluid two initially neighbouring particles can move arbitrarily far away from each other.
such that they are tangent to the vectors of the vector fields at each point. In the cases we look at here, they are identical to the trajectories\(^3\).

![Image of velocity field and flowline](image.png)

**Figure 7.1**: A velocity field and a flowline (dashed).

### 7.2 Notation

Before we start a short summary of notations used throughout this chapter:

- **Velocity** \( \vec{v} \): the velocity of the fluid at some point (in space and time).
- **Speed** \( v = |\vec{v}| \): the absolute value of the velocity.
- **Pressure** \( p \): force applied per unit area.
- **Density** \( \rho \): the density of a fluid (in mass per volume).

### 7.3 Pressure

Pressure in fluid dynamics is the same concept as in thermodynamics. It is a force per area which acts on any surface a fluid touches. The force is always perpendicular to the surface on which it acts.

**Note**: In contrast to thermodynamics, the pressure here can also depend on the position in the fluid (see below).

\(^3\)In general they are not, namely if the velocity field depends explicitly on time.
7.3.1 Compressible and incompressible fluids

One big difference between water and air is the compressibility. We say air is compressible, that means we can change the density of it e.g. by applying pressure. We cannot do this with water\(^4\). To be able to describe this we would need some equation relating pressure and density. One example would be the ideal gas law (exercise: why?). But we will not calculate compressible flows.

7.3.2 Hydrostatic pressure

Suppose we have a long vertical pipe with radius \(r\), closed at the bottom with a lid (see fig. 7.2). If we now fill it with water up to some height \(h\), the total volume of water above the lid is \(V = \pi r^2 h\). The force of all the water pushing on the lid is thus given by \(F = V \rho_{\text{water}} g\), where \(g = 9.81 \text{ m/s}^{-2}\) is the gravitational acceleration. The pressure on the lid is now given by the force divided by the area, which leads to \(p = \rho gh\). The astonishing thing is that the area of the lid cancels. This means that the pressure at some depth \(d\) (measured from the water surface) is independent of the form of the tube. It also means that if we have connected tubes the pressure at any height \(h\) should be the same in every tube\(^5\).

\(\text{Figure 7.2: A pipe filled with water.}\)

\(^4\)Of course we can, but the effects are much smaller than in air, and we neglect them here.

\(^5\)Taking tubes with open tops, we can deduce that they have the same water level.
7.3.3 Buoyancy

Buoyancy is a phenomenon which is due to the depth-dependence of the hydrostatic pressure. This also means that we need a force like gravity acting on our fluid to have buoyancy. To find a formula for this effect, consider a small cube of side length $a$ fully submersed in water (fig. 7.3). Assume the top and bottom of the cube are perpendicular to the gravitational field. Now we look at the forces due to the pressure. By symmetry, the two forces for the sides (front, back) and (left, right) cancel. Assume the top is at a depth $d$ then the force on the top is $F_{\text{top}} = pa^2 = \rho g a^2$. The bottom is at depth $d + a$ and the force is $F_{\text{bottom}} = \rho g (d + a) a^2$. So the total (upward) buoyancy force on the cube is given by $F_{\text{up}} = F_{\text{top}} - F_{\text{bottom}} = \rho g a^2 = \rho g V_d$. Where $V_d$ is the Volume of the water displaced by the cube$^6$. If the cube is only partially submersed, we just set $F_{\text{top}} = 0$ and arrive at the same equation. The same formula holds also for other bodies (i.e. boats), intuitively just think of them as being built from small cubes, then also the same formula holds$^7$. A body can now float in water if $F_{\text{up}} > F_g = mg$. We can translate this into an equation of of the densities, namely a homogeneous body floats if $\rho_{\text{body}} < \rho_{\text{water}}$.

![Figure 7.3: A cube submersed in water.](image)

$^6$As long as the cube is fully submersed, this is the volume of the water.

$^7$You could also do surface integrals over the whole body, which is much more tedious to do and leads to the same results.
7.4 Continuity equation

The continuity equation is a consequence from the conservation of mass. In a first step assume we have water flowing through a tube. We take the tube to have different cross-section areas, say $A$ at point $a$ and $B$ at point $b$. The total mass passing point $a$ in a time interval $\Delta t$ is given by $v_a A \rho \Delta t$. And analogously at $b$: $v_b B \rho \Delta t$. Time cancels on both sides and we are left with

$$v_a A \rho a = v_b B \rho b$$

Now we look at a different system. Lets say we have some sort of bottle (of constant volume) and fill it with air. We can again look at the total mass flowing into the bottle in some $\Delta t$, which is given by $\Delta M = v A \rho \text{incoming} \Delta t$. As the air cannot escape, the mass inside the bottle increases and so does the density: $\Delta M = \Delta \rho V$. We set this equal to the incoming mass (conservation of mass) and get $\Delta \rho V = v A \rho \text{incoming} \Delta t$. We can simplify this by introducing the time derivative of $\rho$: $\frac{d\rho}{dt} = \frac{v}{V} \rho \text{incoming}$. We need to be a bit careful here, as the density of the incoming fluid $\rho \text{incoming}$ is independent of the change of density inside the bottle $\frac{d\rho}{dt}$. Combining the two parts, i.e. looking at a tube with changing density between two points $a$ and $b$ we find:

$$v_a A \rho_a - v_b B \rho_b = V_{ab} \frac{d\rho}{dt},$$

where $V_{ab}$ is the volume between the cross-sections $A$ and $B$.

7.5 Bernoulli’s equation

Let us just start with the full equation and then explain what it says:

$$\frac{v^2}{2} \rho + \rho gh + p = \text{const.}$$

(7.2)

This equation relates the velocity, the gravitational potential and the pressure. Note however that it is only valid for incompressible fluids and that $\text{const.}$ needs to be taken along one flow line (but in most problems here this constant will be the same for all flow lines).

7.5.1 Derivation

Assume we have a tube with a piston at each end, filled with water (see figure 7.4). Denote the two ends and all quantities there with $a$ and $b$ respectively. Suppose piston $i$ has an area of $A_i$ and is at height $h_i$ above ground. Let the water have pressure $p_i$ at the piston. If we now displace piston $a$ by a small distance $d_a$, piston $b$ will move $d_b = \frac{A_b}{A_a} d_a = \frac{V}{A_a}$. The work needed to displace piston $a$ is $W_a = d_a A_a p_a = V p_a$ and for piston $b$: $W_b = -d_b A_b p_b = -V p_b$, where the minus sign comes from the fact that the force coming from pressure now points into the other direction (compared to piston $a$). So the total work we put into the system is given by $W_{\text{tot}} = W_a + W_b = V (d_a - d_b)$. If we look at the energy of the fluid, we find two effects: We displace a volume $V$ of water from $h_a$ to $h_b$, this gives a change in

---

8We assume $v$ to be constant over the whole cross-section.

9$\rho$ would also cancel here, but as we also want to look at gases, in general $\rho$ is not the same at $a$ and $b$. 

potential energy of $\Delta E_{\text{pot}} = \rho V g (h_b - h_a)$. The second effect is the increase in kinetic energy $\Delta E_{\text{kin}} = \frac{\rho V}{2} (v_b^2 - v_a^2)$. If we now combine everything $W_{\text{tot}} = \Delta E_{\text{pot}} + \Delta E_{\text{kin}}$ divide by $V$ and rearrange the terms we get Bernoulli’s equation.

![Figure 7.4: Derivation of Bernoulli’s equation [46].](image)

### 7.6 Surface tension, energy and capillary pressure

Surface tension comes from the fact that molecules at the surface of a liquid have no molecules 'above' them and are attracted by those below. So there is a force acting on them which needs to be compensated (e.g. by pressure) to have a static surface. Assume that we increase the surface of a liquid by a small area $\Delta A$ (e.g. by bulging it out just a little bit). In general we need to do some work $\Delta E$ to achieve this. The surface tension is now given by $\sigma = \frac{\Delta E}{\Delta A}$. Capillary pressure is the pressure forcing water up a thin tube with radius $a$. This is also a surface effect, as the water molecules need less energy when they are at the surface to the wall compared to somewhere in the liquid (i.e. they stick to the wall). This pressure is given by:

$$p_c = \frac{2\gamma \cos(\Theta)}{a}$$  \hspace{1cm} (7.3)

where $\gamma$ is the surface tension relative to the wall and $\Theta$ is the contact angle. To get the height the liquid rises to, equate this pressure with the hydrostatic pressure and solve for $h$. 

172
7.7 Friction in fluids

As for now we silently assumed that our flows do not have friction. As friction is in general very complicated, we will only state two approximate results which hold for objects moving through a fluid (e.g. a submarine, a ball, a car). There are two types of flows, laminar and turbulent, which have different formulas for friction. In the laminar case we get a force proportional to the velocity. An exact formula (only valid for a sphere with radius $R$)

$$F_r = 6\pi \eta R v$$

(7.4)

where $v$ is the velocity and $\eta$ is the dynamic viscosity. The turbulent friction is given by:

$$F_r = c_W A \rho \frac{v^2}{2}$$

(7.5)

Where $c_W$ is the drag coefficient (a constant depending on the material and form of the object) and $A$ is the area of the object perpendicular to the velocity (i.e. the area you see if you look in the direction of flow of the fluid).

---

10A farmer wants to improve the milk production of his cows. He asks a biologist, an economist and a physicist to help him. All three of them come to his farm and observe everything. After a day the economist gives the advice to fire all cows and outsource the farm to China to improve the production by 2%. The farmer doesn’t like this suggestion and waits for the other two. After a week the biologist presents the idea of replacing the cows by gene manipulated algae to produce 10% more milk. The farmer decides to wait for the physicist. After several weeks, the physicist appears with tons of paperwork and claims to have found an idea to improve the production by over 60%. The farmer is really interested and asks the physicist to explain his idea in more detail. The latter starts: "Assume cows to be spherical and in a vacuum (see fig. 7.5)."

![Figure 7.5: A cow in a vacuum.](image-url)
Chapter 8

Electro-and Magnetostatics

Electric attraction and repulsion is a fundamental property of charged particles. It is described by the electric field. When charged particles move they generate an additional field, the magnetic field\(^1\). The theory about electrodynamics describes these fields and how they interact between each other or with charges and also how they evolve in time. Since the general case of time dependent systems is pretty complicated, we will focus on static setups. This means that the charges, wires or whatever we are looking at do not move and have always the same position. The main sources for this chapter are [30] and [31].

8.1 Electrostatics

This chapter examines the interaction between point charges and how this interaction can be described. Knowing the physics of point charges one can easily derive the physics for charge distributions.

8.1.1 Coulomb Force

The basic experimental observation of electrodynamics is that there exist "things" that can attract or repel each other in a way that is not due to gravity. These "things" we call charges and the attracting or repelling force electric force or Coulomb force. One observes that there exist two types of charges, we call them positive and negative charge. A charge which has a very small spreading compared to the distance to other charges we call a point charge. An ideal point charge is just a point in space with a charge.

\(^1\)In relativity, the electric and magnetic field are not two independent fields, they have a strong relation to each other. That is why it is often called the electromagnetic field. The relativistic treatment is out of scope for this course.
If we take two point charges \( q_1 \) and \( q_2 \) which are separated by a constant distance \( r \) we measure a force \( \vec{F} \) acting on \( q_2 \). This force is related to the charges according to

\[
\vec{F}_{12} = k \frac{q_1 q_2}{r^2} \vec{e}_r
\]

where \( \vec{e}_r = \frac{\vec{r}}{|\vec{r}|} \) is the vector that points from \( q_1 \) to \( q_2 \) and has length 1. Such a vector with length 1 we call unit vector. \( k \) is a constant which depends how we define the basic unit of the electrodynamic theory. There exist different systems of units and as a consequence \( k \) is different in each of those unit systems. We use the SI system, where the current and the time are defined and therefore also the charge (see 8.3.1). The unit of charge is Coulomb C (see exact definition in section 10.1.4). One electron has a charge of \( 1.602 \cdot 10^{-19} \) C. We then get \( k = \frac{1}{4\pi\epsilon_0} \) where \( \epsilon_0 = 8.85 \cdot 10^{-12} \) Fm\(^{-1} \). The reason why \( k \) contains a factor \( 4\pi \) is explained in chapter 8.1.5.

The important properties we learn from this formula are:

- The coulomb force always points in the direction of the connecting line of the two charges. The force acting on \( q_1 \) has the same amount but opposite direction to the force acting on \( q_2 \) which agrees to Newtons actio equal reactio.
- If both charges are positive or negative the two charges repel each other, if the charges are different they attract each other.
- The force has a \( 1/r^2 \) dependence as we know it from gravity (see also 8.1.5).
- Opposite the gravitation force, there is also repulsion possible. As a consequence it is possible to shield a charge from the influence of other charges.

### 8.1.2 Electrostatic field

In the 18\(^{th} \) and 19\(^{th} \) century, when the theory about electro-magnetism was developed, there was a big discussion how the force described by equation (8.1) can act over distances. One point of view was that the force acts directly and that the two charges interact immediately. This view contradicts with some statements from relativity which state that information can maximal propagate with speed of light. Farady successfully described the Coulomb force by introducing the electric field: The space has an additional property, the electric field, which is influenced by the presence of charges and the charges interact with the field. If we apply this field theory to the example above with the two point charges we get that a charge, for example \( q_1 \) influences the field in such a way that the interaction between \( q_2 \) and the field at the position of \( q_2 \) results in the Coulomb force. The simplest way to describe this interaction is defining the electric field \( \vec{E} \) as quotient of Force \( \vec{F} \) and charge \( q_2 \)

\[
\vec{E}_{q_1} = \vec{F}_{12} \frac{q_1}{q_2} = \frac{q_1}{4\pi\epsilon_0 r^2} \vec{e}_r
\]
To get the electric field \( \vec{E}(\vec{r}) \) at the position \( \vec{r} \) of a given setup, one calculates the force \( \vec{F} \) that would act on a very small charge \( q_0 \) at the position \( \vec{r} \) and divides the force through the charge

\[
\vec{E}(\vec{r}) = \frac{\vec{F}(\vec{r})}{q_0}
\]

The reason why \( q_0 \) has to be very small is because otherwise it might influence the other charges and therefore the electric field. This gets important if one looks at time dependent fields which we do not treat here.

Fields are often drawn by field lines, see figure 8.1. Putting a very small charge in a field line picture the force on the small charge points always in the direction of the field line. Therefore in a field line picture the field is always tangential to the field lines and the strength of the field is proportional to the density of the field lines. Electric field lines have the following qualitative properties: They start and end at the sources, in the electric case they start at the positive charge or in infinity and end at the negative charges or at infinity. Additionally field lines want to be as short as possible but they repel each other. This attraction in length and repulsion of each other defines a stable state for the field lines which they will take. Of course this is more a qualitative reason for how the field looks like but often one gets a first intuition for a problem.

The advantage of describing the interaction of charges by a field is that it is possible to describe the changing of the field with a finite speed. Therefore two charges interact only with the speed of light and not immediately.

![Field lines of two charges](image)
8.1.3 Superposition

The Coulomb force allows us also to examine the field of multiple charges because every pair of charges interacts according to equation (8.1). Therefore the total force on a charge \( q \) is the sum of all forces between \( q \) and the other charges \( q_1, \ldots, q_N \). The definition of the electric field is still the same as defined in equation (8.2) and we get

\[
\vec{E} = \frac{\vec{F}}{q} = \sum_{j=1}^{N} \frac{q_j q}{4\pi\varepsilon_0 r_j^2} \vec{e}_{r_j} = \sum_{j=1}^{N} \frac{q_j}{4\pi\varepsilon_0 r_j^2} \vec{e}_{r_j}
\]  

(8.3)

where \( \vec{e}_{r_j} \) is the unit vector pointing from \( q_j \) to \( q \).

8.1.4 Continuous charge distributions

When there are a lot of point charges involved (1C = 6.2 \cdot 10^{18} \text{ electrons}) it is useful to use the charge density \( \rho = \frac{q}{V} \) instead of describing the electric field of every point charge. The charge density contains the information how much charge \( q \) a volume \( V \) contains. Therefore the charge in the volume is given by \( q = \rho V \) if \( \rho \) is constant all over \( V \). It can also happen, that the density depends on the position \( \vec{y} \), we then write \( \rho(\vec{y}) \). To compute the electric field at the position \( \vec{x} \) we treat the charge that is in a very small volume \( dV(\vec{y}) \) as point charge located at the position \( \vec{y} \). We then integrate the charge over all these \( dV(\vec{y}) \) which are located inside the total volume \( V \) (so \( \vec{y} \) is inside \( V \)). This leads to the formula

\[
\vec{E}(\vec{x}) = \iiint_{V} \frac{\rho(\vec{y})}{4\pi\varepsilon_0 |\vec{x} - \vec{y}|^2} \frac{\vec{x} - \vec{y}}{|\vec{x} - \vec{y}|} dV(\vec{y})
\]

(8.4)

where the \( \rho(\vec{y}) \) symbolises that one has to take the charge density at the location where the \( dV(\vec{y}) \) is. It is not important to be able to calculate the electric field for a difficult charge distribution. It is more important to understand the concept (see application 8.1.6). Often the problems have some nice symmetries which makes it easier to solve.

There might also be two dimensional surface charge distributions or one dimensional charge distributions which can be treated analogously to the three dimensional case discussed above (see 8.1.6).

8.1.5 Gauss’ law

The \( 1/r^2 \) dependence in the Coulomb law (8.1) has an important consequence. When we look at a point charge \( Q \) around which we place an imaginary sphere, we observe that the number of field lines that pass through the sphere does not depend on the radius of the sphere because field lines start and end at charges or in infinity. Let’s formulate this more mathematically. The electric flux \( \Psi \) for a homogeneous electric field \( \vec{E} \) is defined as \( \Psi = \vec{E} \cdot \vec{A} \) where \( \vec{A} \) is the surface vector for a plane surface. The surface vector points perpendicular to the surface. The length of the vector is equal to the area of the surface. If the sign of \( \Psi \) is positive, the electric field goes through the surface in the same sense as the surface
vector is pointing, if it’s negative in the opposite sense. Visually spoken the flux indicates
how many field lines pass through the surface. If we now have a curved surface, as it is the

case at a sphere, we have to split the surface $A$ of the sphere into small pieces $dS$ and we
get a small amount of the flux by $d\Psi = \vec{E} \cdot d\vec{S}$ (see figure 8.2).

![Curved surface with a small surface vector $d\vec{S}$ and the electric field $\vec{E}$ going through the surface](image)

Since the surface $A$ of the sphere is always perpendicular to the radius vector, $\vec{E}$ and $d\vec{S}$
are parallel and the flux is simply $d\Psi = \pm EdS$, where $E$ and $dS$ are the absolute values of
$\vec{E}$ and $d\vec{S}$ (the $\pm$ indicates that one has to take into account weather the field lines go into
the sphere which is the case if $Q < 0$ or if the field lines go out of the sphere for $Q > 0$).

As we stated the whole flux through the sphere should be independent of the radius and
indeed, the calculation also states this:

$$\Psi = \iint_A d\Psi = \iint_A \vec{E}d\vec{S} = \iint_A EdS$$

$$= E \iiint_A dS = \frac{Q}{4\pi \varepsilon_0 r^2} 4\pi r^2 = \frac{Q}{\varepsilon_0}$$  (8.5)

The two integral symbols indicate that the integral is over a two dimensional surface and
the circle in the integral symbolizes that the area is closed (it is the surface of a volume).
The step from the second last to the last line is right since the electric field is everywhere on
the surface parallel to $d\vec{S}$ and has the same strength $E = \frac{Q}{4\pi \varepsilon_0 r^2}$ (because on the sphere the
distance to the charge is everywhere the same), therefore it does not depend on $dS$ and we
can take it out of the integral. The last integral is simply the integral of the small surface
areas $dS$ all over $A$ and therefore simply $A = 4\pi r^2$ itself.
Following the idea of field lines it is also clear that in any volume $V$ that has no charge in it the number of field lines that enter the volume is the same as the number of field lines that leave the volume. In mathematical notation this means that the total flux is zero:

$$\Psi = \iiint_{\partial V} \vec{E} \, d\vec{S} = 0 \quad (8.6)$$

where $\partial V$ is the surface of the volume. If we now combine it with the result from equation (8.5) we get that independent of the volume the flux only depends on the amount of charge that is inside the volume:

$$\Psi = \iiint_{\partial V} \vec{E} \, d\vec{S} = \frac{Q_{\text{in}}}{\epsilon_0} = \frac{1}{\epsilon_0} \iiint_V \rho(y) \, dV(y) \quad (8.7)$$

This formula is called Gauss law and its statement is that the flux through a closed surface only depends on the electric charge inside that volume and not on the form of the surface. One can conclude that the field lines are produced by the charges. It is pretty useful for example to calculate the electric field in symmetric problems (see 8.1.6) or to find out how charge is distributed.

### 8.1.6 Examples

We want now use the laws above to calculate the electric field of some configurations.

**infinite plain**

Let’s assume there is a thin (height = 0) plate with a surface charge density $\sigma$ (charge per area) at the x-y plane (see figure 8.3). We again use Gauss law 8.1.5 by using a small cylinder which has above and under the x-y plane a surface parallel to the x-y plane with area $S$ each.

Since the plane is infinite and the charge density is uniform, the electric field has no component that points parallel to the x-y plane so the field only points in z-direction. Therefore the whole flux through the surface of the volume goes through the surface above and under the plane and using Gauss law one gets

$$\frac{Q_{\text{in}}}{\epsilon_0} = \iint_{2S} \vec{E} \, d\vec{S} \quad (8.8)$$

$$\frac{\sigma A}{\epsilon_0} = E \cdot 2A \quad (8.9)$$

Therefore the electric field is $\vec{E} = \pm \frac{\sigma}{2\epsilon_0} \hat{z}$. The $\pm$ depends weather one calculates the electric field above or under the plate.
infinite long wire

Let’s assume there is an infinite long, straight wire on the x-axis which has a constant charge density (per length) \( \lambda \). The goal is to calculate the electric field at every point. We’ll use two different approaches:

At the first approach we use Gauss law 8.1.5 and some symmetries to calculate the electric field. To use Gauss law we need some volume and since the problem is rotations symmetric around the \( z \)-axis we choose a cylinder with length \( l \) and radius \( r \) around the wire (see figure 8.4). Since the wire is infinitely long there is no midpoint of the wire and as a consequence the electric field must point radial. As a consequence the flux through the left and right circle \( S \) (see figure 8.4) is zero. Furthermore the electric field has the same strength everywhere on the side of the cylinder and is pointing outside, parallel to the surface vector. Therefore the scalar product is the same as the multiplication of the absolute values of the surface vector \( dS \) and the electric field \( E \). Now using Gauss law one gets

\[
\frac{\lambda l}{\epsilon_0} = \frac{Q_{\text{in}}}{\epsilon_0} = \iiint_{\text{surface of cylinder}} \vec{E} \cdot d\vec{S} = \int_A E dS = E \int_A dS = E 2\pi rl
\]

where \( A \) is the area of the side of the cylinder with \( A = 2\pi rl \).
The electric field is therefore

\[
\vec{E} = \frac{\lambda}{\varepsilon_0 2\pi r} \vec{e}_r = \frac{\lambda}{\varepsilon_0 2\pi (x^2 + y^2)} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}
\]

Figure 8.4: Cylinder around the wire. The wire is along the z-axis.

The second approach uses the superposition principle for continuous distribution (see 8.1.4) and is more complicated. Since the whole problem is rotation symmetric around the x-axis and also translation invariant along the z-axis the electric field only depends on the distance \( r \) from the wire and to simplify the calculation we look at the electric field at the point \( x = z = 0 \) and \( y = r \). The density in this problem is per unit length therefore the \( dV \) is replaced by a \( ds \) in formula (8.4). So the integral looks like

\[
\vec{E} = \int_{z-axis} \frac{\lambda}{4\pi \varepsilon_0 \sqrt{x^2 + y^2 + z^2}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} ds
\]

\[
= \int_{-\infty}^{\infty} \frac{\lambda}{4\pi \varepsilon_0 \sqrt{r^2 + z^2}} \begin{pmatrix} 0 \\ r \\ z \end{pmatrix} dz
\]

So we can look at the \( z \) and \( y \) component of the \( \vec{E} \) field separately. For the \( z \) Component we get 0 because \( \frac{z}{\sqrt{x^2 + y^2}} \) is an odd function (it is point symmetric to the origin) and the integral from an odd function over a symmetric interval is always zero. Physically this can be interpreted the following way: when we compute the contribution to the electric field from a point on the wire at \( z_0 \) we find a point \(-z_0\) which contributes the same amount to the \( z \)-direction of the electric field but in opposite direction. So the contributions from \( z_0 \)
and \(-z_0\) to the z-component cancel out. Now let’s calculate the y-component of the electric field which we have to do by solving the integral. Since \(\lambda\) is constant over the whole wire we can take it out of the integral

\[
E_y = \frac{\lambda}{4\pi\epsilon_0} \int_{-\infty}^{\infty} \frac{r}{\sqrt{r^2 + z^2}} dz
\]

We now apply the substitution \(z = r \sinh(u)\) and \(dz = r \cosh(u) du\) and we get

\[
E_y = \frac{\lambda}{4\pi\epsilon_0} \int_{-\infty}^{\infty} \frac{r^2 \cosh u}{(r^2 (\sinh(u)^2 + 1))^{3/2}} du
\]
\[
= \frac{\lambda}{4\pi\epsilon_0} \int_{-\infty}^{\infty} \frac{1}{r \cosh(u)^2} du
\]
\[
= \frac{\lambda}{4\pi\epsilon_0} \left[ \frac{1}{r} \tanh(u) \right]_{-\infty}^{\infty}
\]
\[
= \frac{\lambda}{4\pi\epsilon_0 r} 2 = \frac{\lambda}{2\pi\epsilon_0 r}.
\]

This is the same result as we got at the first approach.

### 8.2 Potential and Voltage

An electric field produces a force on a charged particle. If one displaces the particle work has to be done. This chapter looks at this work and the energetic properties of the electric field.

#### 8.2.1 Electric potential

The force \(\vec{F}\) on a charged particle \(q\) in an electric field \(\vec{E}\) is given by \(\vec{F} = q\vec{E}\). If one wants to move the particle from one position \(P_1\) to an other position \(P_2\) one has to overcome the force \(\vec{F}\) therefore one has to apply the force \(F_{ext} = -\vec{F} = -q\vec{E}\). Moving \(q\) along a path \(S\) one has to effort the work \(W\)

\[
W = \int_{P_1}^{P_2} F_{ext} \cdot d\vec{s} = - \int_{P_1}^{P_2} \vec{F} \cdot d\vec{s} = -q \int_{P_1}^{P_2} \vec{E} \cdot d\vec{s}
\]

The \(-\) sign causes the work to be positive if one drags a positive charge \(q > 0\) against the electric field, therefore one has to apply work (actively). If one pulls a charge \(q > 0\) in direction of the electric field one gets energy. Therefore if the work from \(P_1\) to \(P_2\) is positive the charge has at \(P_2\) a higher potential energy than at the position \(P_1\). If the energy at the
reference point \( P_1 \) is chosen to be \( E_0 \) then the energy at the point \( P_2 \) the energy is exactly \( E_p = W + E_0 \). Since only the energy difference between the two points can be used, \( E_0 \) can be set to zero without changing the behaviour of the system. As a consequence the energy at \( P_2 \) is proportional to the charge and we define a new property, the electrostatic potential \( \varphi(P_2) \) (often only called electric potential) which is defined as

\[
\varphi(P_2) = \frac{E_p}{q} = -\int_{P_1}^{P_2} \vec{E} \cdot d\vec{s}
\] (8.10)

This formula is only true in the electrostatic case, for the dynamic case one can also define an electric potential but one has to pay attention to more things. The electrostatic potential describes the influence of the electric field to the energy of a charged particle \( q \).

### 8.2.2 Electric potential of a point charge

If we understand the electric potential of a point charge we can easily generalize it to multiple point charges or even to a general charge distribution.

Let \( Q \) be a point charge at the origin of the 3-dimensional coordinate system and \( q \) an other point charge which we move. We want to examine the potential energy of \( q \) depending on its position. Lets first move \( q \) from a point \( P_1 \) to a point \( P_2 \) where both points have the same distance \( r \) from the origin. As path \( S \) we chose a path where we have always the same distance \( r \) from \( Q \) (which is at the origin). Therefore we never move radial and therefore always perpendicular to the electric field since the electric field points radial (this means in the same direction as the connecting line of \( Q \) and \( q \), see chapter 8.1.1). To calculate the potential at \( P_2 \) we integrate according to equation (8.10) and since the electric field and the path are always perpendicular to each other the scalar product is zero and as a consequence the whole integral. Therefore the energy of \( q \) only depends of the distance to \( Q \), which is reasonable since the whole problem is spheric symmetric.

Let’s now examine the dependence of \( r \) on \( \varphi \). The point \( P_1 \) shall have the distance \( r_1 \) and \( P_2 \) the distance \( r_2 \) from the origin. Then the potential difference is

\[
\varphi_2 - \varphi_1 = -\int_{r_1}^{r_2} \vec{E} \cdot d\vec{s} = -\int_{r_1}^{r_2} \frac{Q}{4\pi\varepsilon_0 r^2} dr = -\frac{Q}{4\pi\varepsilon_0} \left[ -\frac{1}{r} \right]_{r_1}^{r_2} = \frac{Q}{4\pi\varepsilon_0} \left( \frac{1}{r_2} - \frac{1}{r_1} \right)
\] (8.11)

Since the reference energy can be chosen freely, it is also possible to choose the reference point \( r_1 \) freely and the simplest way is to choose \( r_1 = \infty \) and we get.

\[
\varphi(R) = -\int_{\infty}^{R} \frac{Q}{4\pi\varepsilon_0 r^2} dr = \frac{Q}{4\pi\varepsilon_0} \left( \frac{1}{R} \right)
\] (8.12)

If we want to calculate the potential difference \( \Delta \varphi \) between two distances \( r_1 \) and \( r_2 \) we can use equation (8.12)
\[ \Delta \varphi = - \int_{r_1}^{r_2} \frac{Q}{4\pi \epsilon_0 r^2} dr \]
\[ = - \int_{r_1}^{\infty} \frac{Q}{4\pi \epsilon_0 r^2} dr - \int_{\infty}^{r_2} \frac{Q}{4\pi \epsilon_0 r^2} dr \]
\[ = \varphi(r_2) - \varphi(r_1) \]

which makes it easy to calculate potential differences of point charges.

### 8.2.3 Potential of multiple charges

We want to examine the potential energy \( E_p \) of a point charge \( q \) at position \( \vec{r} \) in a system of \( n \) charges \( q_1...q_n \) at the positions \( \vec{r}_1...\vec{r}_n \), where infinity has potential energy zero. Since the force at any point is the sum of the forces between \( q \) and one of the other charges the total potential energy is also the sum of the energy between \( q \) and each other charge:

\[
E_p(\vec{r}) = - \int_{\infty}^{\vec{r}} \sum_{j=1}^{n} \frac{qq_j}{4\pi \epsilon_0 |\vec{r} - \vec{r}_j|^3} (\vec{r} - \vec{r}_j) d\vec{s}
\]
\[= \sum_{j=1}^{n} \int_{\infty}^{\vec{r}} \frac{qq_j}{4\pi \epsilon_0 |\vec{r} - \vec{r}_j|^3} (\vec{r} - \vec{r}_j) d\vec{s} \]
\[= \sum_{j=1}^{n} \int_{\infty}^{r_j'} \frac{qq_j}{4\pi \epsilon_0 r^2} dr = \sum_{j=1}^{n} \frac{qq_j}{4\pi \epsilon_0 r_j'} \]

where \( r_j' = |\vec{r} - \vec{r}_j| \) is the distance between \( q \) and \( q_j \). The step from the second last to the last line is possible because inside the sum we treat the interactions separately and we can therefore apply equation (8.12). Therefore the potential of the \( n \) charges is given by

\[
\varphi(\vec{r}) = \sum_{j=1}^{n} \frac{q_j}{4\pi \epsilon_0 r_j'}
\]
We also find the potential of a continuous charge distribution \( \rho \) in a volume \( V \) by applying the same argument as when we looked at the electric field in chapter 8.1.4 and we find the potential at the position \( \vec{r} \) by

\[
\varphi(\vec{r}) = \int_V \frac{\rho(\vec{y})}{4\pi\varepsilon_0|\vec{r} - \vec{y}|} dV(\vec{y})
\]

where \( \vec{y} \) is the position of the infinite small charge \( dq = \rho(\vec{y})dV(\vec{y}) \) inside the volume.

### 8.2.4 Voltage

A potential difference is called voltage. The voltage between two points indicates how much energy per charge a charge gains if it moves from one point to the other. The unit of the voltage is volt \( V \).

If one coulomb is moved between two points with one volt then this coulomb charge gains one joule energy. Therefore one volt is defined as \( 1V = 1J/1C \). Once the volt is defined one can also define a new energy unit: The energy one electron gains passing one volt is called one electronvolt eV.

The quantity volt plays an important role in electric circuits (see chapter 9).

The fact that the potential is only a function of the place allows us to make an important statement about the electric field. If we calculate the energy \( W \) we have to apply to move a charge \( q \) from a starting point \( \vec{r}_0 \) over a closed path \( \gamma \) with end point equal to the starting point \( \vec{r}_0 \) we see that the applied energy is zero \( W = 0 \) since the potential difference between \( \vec{r}_0 \) and \( \vec{r}_0 \) is zero. Since \( W \) is proportional to the charge \( q \) the integral of the electric field along the path must be zero:

\[
\oint_{\gamma} \vec{E} \cdot d\vec{s} = 0
\]  

(8.13)

This property makes the electric field an irrotational field (opposite to the magnetic field, see chapter 10.1.2). The statement that the electric field is an irrotational field is only true for the static electric field. Therefore the electrostatic field is an irrotational source field which means that equation (8.13) holds and that the electric field begins and ends at sources which we called charge.

### 8.2.5 Potential and conducting material

If a body made of conducting material is placed in an external electric field the charges in that material will move according to the electric field. At the end there will be a stable state with some interesting properties:

1. The total electric field (sum of external field and field of moved charges in the conducting material) points perpendicular to the surface at every point on the surface. Because if there would also be a parallel component the charge would accelerate in this direction and it would not be a stable state.
2. Inside the body there is no electric field. Otherwise again charge would be accelerated what contradicts to the assumption of the stable state.

3. Since the electric field inside the body is zero the potential all over the body is the same. Because if we look at two points of the body and want to calculate the potential difference we have to apply equation (8.10) and since $\vec{E} = 0$ we get that the potential difference is zero.

4. Inside the body there is no net charge (this means positive and negative charge have the same density). Otherwise it is not possible to have everywhere in and on the body the same potential. Therefore all charge is on the surface of the body.

### 8.2.6 Capacity

Let’s take two bodies made of conducting material which have no net charge on each of them. We call them electric neutral (so the negative and positive charge have the same amount). If we take some charge from the first body and put it on the second body there will be a potential difference $\Delta U$ between this bodies. Taking a very small charge $q$ (so small that it does not influence the potential or the electric field) and moving it from the one to the other we recognize that no mater which way we take, the absolute value of the energy is always $|\Delta Uq|$. Therefore the charge on each body is distributed in a very particular way such that the voltage between two points on the two bodies is always $\Delta U$. If we put more charge from the first to the second body the voltage will be bigger but the energy gain displacing $q$ from one body to the other is still independent on the path. This means that the way the charge is distributed is the same but with a bigger amount of charge. Therefore the direction of the electric field does not depend on how much charge was put from the first on the second body, only the strength of the electric field depends on this amount and the sense (if it is pointing in one or the other direction, depends on the sign of the charge we displaced). Since the electric field is proportional to the charge $\Delta Q$ displaced from the first body to the second, we have a proportional dependence between the charge on the bodies and the voltage between them:

$$C \cdot \Delta U = \Delta Q$$

where $C$ is the proportional constant which is called capacity. The capacity only depends on the geometry of the two bodies.

### 8.2.7 Example

Let’s look at some examples to get used to the theory.

**Plate capacitor**

Putting two metallic plates with area $A$ each and both parallel to the y-z plane in a distance $d$ ($d \ll A$) we get a plate capacitor (see figure 8.5). Let us take the charge $Q$ from the
left plate and put it on the right plate. Since \( d \ll A \) we model the electric field of this setup as one of infinite spread plates. This means the electric field between the plates has everywhere the same strength and the same direction\(^2\). Such a field is called homogeneous field. Additionally we define the charge density \( \sigma \) by \( \sigma = \frac{Q}{A} \).

![Plate capacitor](image)

Figure 8.5: Plate capacitor [36].

We know from chapter 8.1.6 the electric field is perpendicular to the plates and has everywhere the same strength. Therefore we get the voltage \( U \) between the two plates by

\[
U = \int_{\text{left plate}}^{\text{right plate}} E_{\text{tot}} \cdot d\vec{s}
\]

\[
= \int_{\text{left plate}}^{\text{right plate}} E_{\text{left plate}} \cdot d\vec{s} + \int_{\text{left plate}}^{\text{right plate}} E_{\text{right plate}} \cdot d\vec{s}
\]

\[
= E_{\text{left plate}} \cdot d - E_{\text{right plate}} \cdot d
\]

\[
= \frac{Q}{A} \cdot d - \frac{-Q}{A} \cdot d = \frac{Qd}{A\epsilon_0}
\]

Therefore the capacity \( C \) is given by

\[
C = \frac{Q}{U} = \frac{A\epsilon_0}{d}.
\]

\(^2\)This is because the plates are infinite spread out. Therefore the electric field must everywhere look the same (translation and rotation symmetry).
This problem was easy to solve because of the homogeneous electric field. Because in case of an homogeneous electric field the voltage between two points $\vec{r}_1$ and $\vec{r}_2$ is simply given by $U = \vec{E} \cdot (\vec{r}_2 - \vec{r}_1)$. If the connection line of the two points lies parallel to the electric field this simplifies even more to $U = |\vec{E}|d$ where $d$ is the distance between the two points.

Additionally we can calculate the energy $E_{\text{pot}}$ stored in the capacitor. In principle the energy is given as $E_{\text{pot}} = QU$. But we have to pay attention because if we load the capacitor, the voltage and the charge changes and we have to add the potential energy of the different stages of the charging capacitor by taking the integral

$$E_{\text{pot}} = \int_0^Q U(q) \, dq$$

$$= \int_0^Q \frac{q}{C} \, dq = \left[ \frac{q^2}{2C} \right]_0^Q = \frac{Q^2}{2C}$$

where we used the relation between the voltage and the charge given by $C = \frac{Q}{U}$. Instead of integrating with respect to the charge we can also do it with respect to the voltage and we get

$$E_{\text{pot}} = \int_0^U Q(u) \, du$$

$$= \int_0^U Cu \, du = \frac{C}{2} U^2 = \frac{Q^2}{2C}.$$ 

There is another approach to calculate the energy stored in the capacitor. Assume we have two charged plates with charge $Q$ on each. Assume they are separated by a small distance $\delta$. The force from one plate of the other is $F = EQ = \frac{Q}{2\epsilon_0 A} Q$ where $A$ is the area of the plates. We now can compute the energy to pull one plate to the distance $d$ to the other plate. This energy is given as

$$E_{\text{pot}} = \int_{\delta}^d F \, ds$$

$$= \int_{\delta}^d \frac{Q^2}{2\epsilon_0 A} \, ds = \frac{Q^2}{2\epsilon_0 A} (d - \delta).$$

If we now set $\delta = 0$ and use $C = \frac{\Delta \epsilon \pi}{d}$ we recover the stored energy from above: $E_{\text{pot}} = \frac{Q^2}{2C}$. 

189
Potential of an infinitely long wire

As we have seen in chapter 8.1.6 the electric field of an infinite long wire with charge density $\lambda$ is $\vec{E} = \frac{\lambda}{2\pi\epsilon_0 r} \hat{r}$ with $\hat{r}$ the unit vector pointing radial away from the wire. To calculate the voltage between two distances $R_1$ and $R_2$ we integrate radially from $R_1$ to $R_2$ (as a consequence $\vec{E}$ and $d\vec{r}$ are pointing parallel).

$$\varphi(r) = \int_{R_1}^{R_2} \vec{E} d\vec{r} = \int_{R_1}^{R_2} E dr = \int_{R_1}^{R_2} \frac{\lambda}{2\pi\epsilon_0 r} dr = \frac{\lambda}{2\pi\epsilon_0} [\ln(r)]_{R_1}^{R_2} = \frac{\lambda}{2\pi\epsilon_0} (\ln(R_2) - \ln(R_1))$$

8.3 Current and magnetic field

Current is basically moving charge. A current produces an additional field, called magnetic field which can be measured. The existence of the magnetic field can be predicted using relativity but the calculation is far beyond the stuff for this chapter so we make a phenomenological approach.

8.3.1 Current and conservation of charge

If charge is moving one speaks of a current. The unit of the current $I$ is Ampere A. The precise definition of the current is $I = \frac{dQ}{dt}$ Where $Q$ is the charge that passes at a certain point.

An important concept of electrodynamics is the conservation of charge. This means that the total charge of a closed system can not change. It is for example possible to have a charge neutral atom and take away an electron. But then the atom is positively charged and the total charge is still zero, i.e. the sum of both charges, the electron and atom. Therefore the charge at a point can only change if a current flows to that point. On the other hand a current starts and ends at points where the charge changes or the current is a closed circuit.
8.3.2 Magnets

Everybody has already seen a magnet, which looks like small pieces of metal. This magnets are called permanent magnets since they are always magnetic. There exist also magnets that work with current and which are called electro magnets (see 8.3.3). A magnet produces a magnetic field which is somehow similar to the electric field but has some very important differences. A magnet has always two poles, this are the parts of the magnet where the magnetic field leaves or enters the magnet. The north pole is the part where the magnetic field leaves the magnet and the south pole is where the magnetic field enters the magnet (see figure 8.6). Inside the magnet the field lines go from the south to the north pole, they build therefore closed filed lines (see chapter 10.1.4 and equation (8.14)). The names of the poles come from the fact that the earth has also a magnetic field and the north pole of a magnet is attracted by the geographic north pole and the south pole of the magnet is attracted by the geographic south pole.

As we know from electrostatics field lines want to be as short as possible and they repel each other. If we now put two magnets together (see figure 8.7), we see from the total magnetic field that the same poles repel each other and two different poles attract each other. Because in the first case the field lines repel each other and as a consequence also repel the two magnets. In the second case the field lines can build nice closed loops which want to get shorter and therefore attract the magnets.

We want to formalize the magnetic field a bit. What we called magnetic field is formally the magnetic flux density $\vec{B}$ with unit Tesla T. The connection to other SI-units is $1T = 1kg \cdot A^{-1} \cdot s^{-2} = 1W \cdot s \cdot A^{-1} \cdot m^{-2}$. One Tesla is a pretty strong field, for example the magnetic field of the earth is about $5 \cdot 10^{-5}T$ and a usual magnet produces a field in the order of $0.1T$. 

![Figure 8.6: Magnet with the magnetic field which leaves the magnet on the right, side where the north pole is, and enters the magnet on the left side, where the south pole is][38].
One important difference to the electric charge and field is that there exist no magnetic monopoles. This means that it is not possible to separate the north from the south pole, they build always an inseparable pair (see chapter 10.1.4). But this also means that magnetic field lines have no start and no end, therefore the magnetic field is a source free rotational field. If we adapt to Gauss law in electrostatic (see chapter 8.1.5) we find the law

\[ \oint_{\partial V} \vec{B} \cdot d\vec{S} = 0 \] (8.14)

### 8.3.3 Magnet and electric current

If one puts a permanent magnet which is freely moveable near a wire where current flows one can observe that the permanent magnet orient itself in a particular way which is shown in figure 8.8.

If we now think the magnetic field being such that the magnets are tangent to the field we get that the magnetic field looks like in figure 8.9. The rule is the following: If one takes the right hand and places the thumb in the direction of the current then the other four fingers show the direction of the field.
8.3. CURRENT AND MAGNETIC FIELD

8.3.4 Lorentz force

Putting a wire in a homogeneous magnetic field and letting flow current through the wire, the homogeneous magnetic field and the magnetic field of the wire superpose to a total magnetic field which is shown in figure 8.10.

From the picture it is obvious that on the left side of the wire the magnetic field is pushed more together than on the right side. Since magnetic field lines repel each other there is a force pushing on the left side of the wire. On the other hand the field lines on the right side are not straight lines but a bit curved. Since field lines want to be as short as possible they attract the wire. This force on the wire in the right direction is called Lorentz force $\vec{F}$. Formally it is described by the formula

$$\vec{F} = I\vec{l} \times \vec{B}$$  \hspace{1cm} (8.15)

where $\vec{l}$ is the direction of the current flowing in the wire and $\vec{B}$ is the magnetic field. If we want to calculate the force on a single charge $q$ we use the mathematically not really precise but intuitively correct equation
Figure 8.10: The homogenous field and the field from the wire are indicated by the dashed line. The total field is indicated by the continuous lines. The cross on the wire shows that to current flows into the page. [42].

\[ \mathbf{I} = \frac{dq}{dt} \frac{\mathbf{l}}{dt} = \frac{d}{dt} \mathbf{q} \frac{\mathbf{l}}{dt} = \mathbf{q} \frac{\mathbf{v}}{dt} \]

where \( \mathbf{v} \) is the velocity of the charge. Therefore we get that the force on a moving charge is

\[ \mathbf{F} = q \mathbf{v} \times \mathbf{B} \] (8.16)

It is important that a charge which is not moving has no Lorentz force which is obvious since only a moving charge has a magnetic field and a magnetic field can only interact with an other magnetic field and not with an electric field. If one looks at electric and magnetic fields which change in time then a changing electric field produces a magnetic field and a changing magnetic field produces an electric field. This is described by the Maxwell equations which are too complicated to be treated here. Therefore an electric field can only interact with a magnetic field if the electric field changes with time since then the electric field produces a magnetic field which can interact with an other magnetic field, or the magnetic field changes and produces an electric field which can interact which the electric field. Since we (nearly) always look at static systems (which do not change with time) we will not need the Maxwell equations.
Chapter 9

Direct current circuits

Besides the theoretical treatment of electric and magnetic fields there is also a very practical use. All the electronic devices base on the laws of electrodynamics. Since the calculation with the basic equations of electrodynamics is often very complicated one simplifies the calculation of electric circuits by introducing electrical components. The laws of electrodynamics then define the behaviour of the components (overview see chapter 9.4). We will now look at electrical circuits which have a constant voltage (and most also a constant current). This sort of circuits are called direct circuits. There exist also alternating current where the voltage and the current change with time but this is not treated here.

9.1 Ohm’s law

Applying a variable voltage $U$ to a body the current $I$ through the body might depend on many influences as temperature or humidity. The quotient $\frac{U}{I}$ is called resistance of a body. The simplest (not trivial) dependence is the proportional dependence: The current is proportional to the voltage with proportionality constant $R$ with $U = RI$. $R$ is called ohmic resistance. The linearity is just a model which is valid for many materials and bodies. Mathematically it is not completely wrong to describe the dependence between $U$ and $I$ by a linear function because any (nice) function can be approximated by a linear function. But there exist also components which have a non linear dependence as the light bulb or the diode. The light bulb is a typical example of the dependence of the temperature on the resistance of a material. For metallic materials the resistance is higher if the metal is warmer and since the wire of a light bulb gets hotter if a higher voltage is applied one can recognise that the resistance is higher at the higher voltage.
9.2 Equivalent circuit

Sometimes it is possible and useful to describe a collection of components by a single component. Since there are many possible combinations we want to look at the most important.

9.2.1 Wire

A real wire usually has a little resistance. Since this resistance is spread all over the wire it is cumbersome to describe the wire as a chain of little ohmic resistors. Instead one adds to an ideal wire (with no resistance) a single ohmic resistor which describes the total resistance of the wire.

Since two wires also have a capacity one could additionally add a capacitor to a closed circuit. As one can imagine it is nearly impossible to describe all effects and it is often not necessary to taking all into account but to consider the relevant ones.

9.2.2 Series circuit

If two ohmic resistors are connected one behind the other one talks about a serial circuit (see figure 9.1). Since the current $I$ trough $R_1$ and $R_2$ is the same, the voltage drop over both resistors is $U = R_1 I + R_2 I = I(R_1 + R_2)$. Therefore the resistance of both resistors is $R = R_1 + R_2$.

By the same argument one can calculate the total resistance $R_{\text{tot}}$ of an arbitrary number of resistors which are all connected in series. For $n$ resistors $R_1, R_2, \ldots, R_n$ the total resistance is given by $R_{\text{tot}} = \sum_{j=1}^{n} R_j$.

![Figure 9.1: Left: serial circuit of two resistors. Right: parallel circuit of two resistors.](image-url)
9.2.3 Parallel circuit

If we put two resistors parallel to each other we get a parallel circuit (see figure 9.1). The voltage $U$ over both resistors is the same and the total current that flows through the two resistors is $I_{\text{tot}} = \frac{U}{R_3} + \frac{U}{R_4}$. Therefore the total resistance is

$$R = \frac{U}{I_{\text{tot}}} = \frac{1}{\frac{1}{R_3} + \frac{1}{R_4}}$$

Similarly the resistance of $n$ parallel connected resistors is given by

$$R_{\text{tot}} = \left( \sum_{j=1}^{n} \frac{1}{R_j} \right)^{-1}$$

9.2.4 Voltage source

An ideal voltage source is a device where independent of the current the voltage is always the same. Since already the wires leaving the ideal voltage source have a resistance one has to add an additional resistance $R_S$ in series to the voltage source to describe a real voltage source. $R_S$ is usually very small one can often neglect it. Taking $R_S$ into account is only important if one wants take out a lot of energy from the voltage source or if the rest of the electrical circuit has a very low resistance in the order of $R_S$.

9.3 Electric power

The voltage describes how much energy one Coulomb gets if it passes the voltage. A current describes how much charge passes per unit of time. Therefore the product of a voltage and a current describes how much charge gets an energy by the voltage per unit time, therefore the product tells us the power that the current performs.

$$P = UI$$

where $P$ is the power that the current $I$ emits over the voltage drop $U$. 
CHAPTER 9. DIRECT CURRENT CIRCUITS

9.4 Electric components

Table 11.1 gives an overview over the most common symbols.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>![image]</td>
<td>ideal wire (with no resistance)</td>
</tr>
<tr>
<td>![image]</td>
<td>switch to open and close the electric circuit</td>
</tr>
<tr>
<td>![image]</td>
<td>ohmic resistor</td>
</tr>
<tr>
<td>![image]</td>
<td>variable ohmic resistor</td>
</tr>
<tr>
<td>![image]</td>
<td>(ideal) voltage source</td>
</tr>
<tr>
<td>![image]</td>
<td>ground (reference voltage in infinity), connected to the earth</td>
</tr>
<tr>
<td>![image]</td>
<td>capacitor (to store charge)</td>
</tr>
<tr>
<td>![image]</td>
<td>inductor (inductive resistor, often a coil)</td>
</tr>
<tr>
<td>![image]</td>
<td>light bulb</td>
</tr>
<tr>
<td>![image]</td>
<td>LED (light emitting diode)</td>
</tr>
<tr>
<td>![image]</td>
<td>diode (lets current only flow in direction of the arrow)</td>
</tr>
<tr>
<td>![image]</td>
<td>horn</td>
</tr>
</tbody>
</table>

Table 9.1: Overview over different symbols [45]

9.5 Kirchhoff’s circuit law

There are two very important rules which can be used to determine the current and voltage through any circuit.

9.5.1 Current law

Because there is conservation of charge we have a restriction on the currents: At any point where no charge is stored the sum of all currents must be zero. It is important to chose the currents that flow to that point with one sign (so positive or negative) and the currents that flow away with the other sign (negative or positive) (see figure 9.2).
Figure 9.2: Knot where many currents flow together. The sum $I_1 - I_2 + I_3 - I_4 - I_5$ must be zero. The currents that flow to the knot have positive sign, the currents that flow away negative sign.

### 9.5.2 Voltage law

As we have seen in chapter 8.2.4 the sum of all voltages in a closed electric circle is zero. Therefore we define a direction of summation (in clock wise or anti clock wise) and take the sum over all electric components of the circuit (see figure 9.3)

![Diagram of electric circuit with voltages](image)

Figure 9.3: Here the sense of summation is clock wise. For all the components (which does not necessarily have to be ohmic resistors) we take the potential difference between the potential at the end of the arrow minus the potential on the start of the arrow. This gives us the voltage in direction of the summation. The sum of all these voltages must be zero $U_1 + U_2 + U_3 + U_4 + U_5 = 0$ [43].
9.5.3 Applying Kirchhoff’s law

A knot is a point in the electrical circuit where more than two currents flow. For every knot we apply the current law. This gives us for $k$ knots $k$ independent equations. For every closed circuit one applies the voltage law which gives for every independent closed circuit one more equation. One has to pay attention on the independence of the circuits (see figure 9.4).

Figure 9.4: There are three closed circuits: circuit 1: ACDB, circuit 2: AEFB and circuit 3: CEFD. But the three circuits are not independent because circuit 3 is basically circuit 2 minus circuit 1. Therefore the voltage law must only be applied on two of the three circuits (it does not depend on which two). To be sure not doing anything wrong: only take the closed circuits which are the smallest possible circuits. For example circuit 3 can be contracted to circuit 2 by making a shortcut from B to A.[44].

If one has now $n$ equations (from the current and voltage law) one has to express the voltages by the currents or vice versa. Then one should have $n$ equations with $n$ variables which should be solvable.

This is a very useful method for complicated circuits. For easier circuit one can apply other methods, for example simplify complicated systems of resistors by a single resistors (see 9.2) and then apply the Kirchhoff’s voltage law by saying that the voltage over the voltage source must be equal to the voltage over the single resistor.
9.6 Examples

9.6.1 Maximal power from a real power source

Let’s look at a real voltage source with voltage $U_0$ and connect it to an ohmic resistor (see figure 9.5). Since the voltage source has an ohmic resistance itself it is not possible to get infinite power out of the source.

We want to calculate the maximal power that one can use at the resistor. The power $P$ on the resistor $R$ is $P = UI = I^2R$. The current is given by $I = \frac{U_0}{R + R_S}$. Therefore the power is $U_0^2 \frac{R}{(R + R_S)^2}$. To get the maximal power we consider the power as function of $R$ and set the derivative zero:

$$0 = \frac{dP}{dR} = U_0 (R + R_S)^2 - 2(R + R_S)R(R + R_S)^3$$

$$= R + R_S - 2R$$

$$R = R_S$$

Therefore one can take the maximal power out of a real voltage source if the resistor is equal to the interior resistor of the voltage source. Of course the same amount of power as one can use at the resistor $R$ is heating up the voltage source because of $R_S$.

9.6.2 Charging a capacitor

A capacitor with capacity $C$ is connected in series with a resistor with resistance $R$ to a voltage source with voltage $U_0$. At the time $t = 0$ the switch closes the circuit and the capacity starts to charge (see figure 9.6).
We want to calculate the voltage over the capacity $V_C$ as a function of time. We apply Kirchhoff’s law: $U_0 = U_R + U_C$ with $U_R = RI$ the voltage over the resistor and $U_C = \frac{Q}{C}$ with $Q$ the charge in the capacitor. Taking the time derivative we get

\[
0 = R \frac{dI}{dt} + \frac{I}{C} = RC \frac{dI}{dt} + I
\]

\[
-\frac{1}{RC} dt = \frac{dI}{I}
\]

\[
-\frac{1}{RC} \int_0^t dt' = \int_{I_0}^{I(t)} \frac{dI}{I}
\]

\[
-\frac{1}{RC} t = \ln\left(\frac{I(t)}{I_0}\right)
\]

\[I(t) = I_0 e^{-\frac{t}{RC}}\]

Where $I_0$ is the current when the switch is closed and it is given by $I_0 = \frac{U_0}{R}$ because at the first moment no voltage drops over the capacitor (since it is empty) and therefore all the voltage drops over the resistor. Therefore we get for the voltage over the capacitor

\[U_C(t) = U_0 - RI(t) = U_0(1 - e^{-\frac{t}{RC}})\]

This contains some expected properties: At the beginning when there is no charge in the capacitor there is no voltage drop over it. For $t$ very big nearly all voltage drops over the capacitor which is also clear since the capacitor is an interruption of the circuit for a constant voltage.
Chapter 10

Electrodynamics 2

In this chapter we continue electrodynamics. First we look at magnetism from a more formal point of view, which leads to Ampère’s Law and Biot Savart’s Law. Then we look at time dependent fields and phenomena related to them, for example induction and displacement current. Putting all these equations together leads to the famous Maxwell’s equations. In a second part we will investigate the behavior of electromagnetism in the presence of matter. At the very end, energy considerations of the electromagnetic field are discussed.

10.1 Magnetism

In electrodynamics 1 we looked at magnetism more from a phenomenological point of view without formal relations. We will catch up on this now. First we will have a closer look at the magnetic field $\vec{B}$ itself and introduce the magnetic flux. Then we will investigate an important property of the magnetic field: Remember that a current (flowing through a wire) creates a magnetic field around the wire (see 8.3.3). It is a circular field which needs some new formalism to describe it. And finally we look at the magnetic field of a point charge and some more examples.

10.1.1 Magnetic Field and Flux

In section 8.3.2 we gave the magnetic field $\vec{B}$ already a unit, namely the Tesla. To be more precise $\vec{B}$ is the magnetic field density\(^1\). In a field line picture, the field density indicates how many field lines pass through a certain area. For a given area $A$, we can therefore define a quantity which corresponds to the total number of field lines through that area. This quantity is called flux $\Phi$ and it is defined as

$$\Phi = \iint_A \vec{B} \cdot d\vec{A}$$

where the differential $d\vec{A}$ is a small area element pointing perpendicular to the surface, see figure 8.2. The scalar product $\vec{B} \cdot d\vec{A}$ indicates the flux through $dA$ which,

\(^1\)There is also a magnetic field strength, analogous to the electric field strength $\vec{E}$. This plays a less important role in physics, see also 10.5.4
summed/integrated up for all small elements, gives the total flux. If the area is not curved and the magnetic field is homogeneous through the whole surface the formula simplifies to

\[ \Phi = \vec{B} \cdot \vec{A} \]

where \( \vec{A} \) is the surface vector pointing perpendicular to the surface and its absolute value is equal to the area of the surface.

### 10.1.2 Ampere’s Law

Similar to Gauss’s law (see chapter 8.1.5) where the flux at the surface of a volume only depends on the charge inside, we can formulate a law with currents. But there is a very important difference between the (static) electric and the (static) magnetic field: The (static) electric field starts and ends on charges. Since there are no "magnetic charges", the magnetic field has no starting and no ending point. Therefore the magnetic field lines are always closed and we can quantify them using this property. Instead of the flux though a closed surface we look at the magnetic field along a closed path (enclosing a surface \( S \)). From an intuitive point of view we have to relate the current through a wire with the magnetic field around the wire\(^2\). If we have a wire where a current \( I \) flows through and a surface \( S \) (that is not closed) we have that

\[ \oint_{\partial S} \vec{B} \cdot \vec{d}l = \mu_0 I_{in} = \iint_S \mu_0 \vec{j} \cdot d\vec{S} \] (10.1)

where \( \partial S \) is the boundary line of the surface \( S \) and \( \vec{d}l \) is a infinitesimal short tangent vector on that boundary line (see figure 10.1). \( I_{in} \) is the current that flows through the surface \( S \) and \( \vec{j} \) is the current density (current per area) which points in the direction of the current flow. The direction of \( \vec{d}l \) is given by the right hand rule: if the thumb of the right hand shows in the direction of the current then the other four fingers show the direction of \( \vec{d}l \) and \( d\vec{S} \) shows the direction of the current (see also figure 8.9).

It is again not important being able to calculate the integrals above for arbitrary cases but it is very useful to understand the formula.

The important thing about equation (10.1) is that there exists a quantity connected to the magnetic field (namely the integral on the left hand side) which only depends on the current. In symmetric cases this is very useful, see example 10.1.4.

\(^2\)This in analogy to Gauss’law where we relate the charge in a volume to the flux through the surface of the volume. Here we relate the charge flowing through a surface with the magnetic field around the surface.
10.1.3 Magnetic Field of a Moving Point Charge

An other way to calculate the magnetic field of a current is to look at the explicit dependence of the magnetic field on a moving charge. Assume a point charge with charge \( q \) at the position \( \vec{r}' \) is moving with velocity \( \vec{v} \). We want to understand the formula of the magnetic field at a point \( \vec{r} \) of this point charge. The formula is

\[
\vec{B}(\vec{r}) = k \frac{q}{r^2} \vec{v} \times \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|} \tag{10.2}
\]

where \( \vec{r} - \vec{r}' \) is the vector pointing from the point charge to the point \( \vec{r} \) and \( k \) is a constant. In SI units \( k = \frac{\mu_0}{4\pi} \) when \( \mu_0 = 4\pi \cdot 10^{-7}\text{V}\cdot\text{s}\cdot\text{A}^{-1}\cdot\text{m}^{-1} \). This formula seems extremely complicated at first but looking a bit more precisely and comparing it with the electric field of a point charge, it gets a lot simpler. The \( \vec{B} \) field has a \( \frac{q}{|\vec{r} - \vec{r}'|^2} \) dependence as the \( \vec{E} \) field in the coulomb law. This is pretty reasonable since as mentioned above the electric field and the magnetic field have a strong connection\(^3\). This means that both should have the same dependence on \( q \) and on the distance \( |\vec{r} - \vec{r}'| \). The factor \( 4\pi \) is part of the constant and is separated with a similar reason as in the electric case, where \( 4\pi r^2 \) is the area of the sphere with radius \( |\vec{r}| \) around the charge. Additionally some other laws take a nice form, see equation (10.1). The big difference between the formula of the electric and the magnetic field is the vector product. But the vector product fulfils exactly the properties of the magnetic field we stated in chapter 8.3.3: The vector product ensures that the magnetic field always points tangent to a circle around the direction.

Figure 10.1: The black point with the circle around it denotes the current which flows perpendicular out of the sheet. The dashed line symbolizes the magnetic field. The shaded area is \( S \) where the current flows through (note it does not matter where it flows through). A small piece of the integral along the boundary of \( S \) is denoted by \( d\vec{l} \).

\(^3\)This gets obvious in relativity.
of the current (which here is $\vec{v}$) and since the magnetic field should be zero for $\vec{r}$ parallel to $\vec{v}$ the angle between $\vec{v}$ and $\vec{r} - \vec{r}'$ plays also a role and is respected in the vector product.

The formula above is not correct with respect to relativity because in the formula it is assumed that the moving particle has an immediate influence at the position $\vec{r}$ which is not possible due to relativity. But if we consider $|\vec{v}| \ll c$, when $c$ is the speed of light, the mistake is negligibly small.

To calculate the magnetic field of a current flowing through a wire we use equation (10.2) and redefine some quantities. The charge $dq$ in a short part of the wire with length $dl$ is $dq = \rho dl$ where $\rho$ is the charge density per unit length. Assume that charge is moving with a (mean) speed $v$. To pass the length $dl$ the time $dt$ is needed. We therefore have a current $I = \frac{dq}{dt} = \frac{\rho dl}{dt} = \rho v$. Hence we have $dqv = \rho dlv = I dl$. Since the charge $dq$ is assumed to be small, and located at a small spot, we can use equation (10.2) to calculate the magnetic field $d\vec{B}$ caused by the current at $\vec{r}$ through the small piece of wire $dl(\vec{r}')$. We turned the path element $dl$ into a vector in order to calculate the vector product. The vector has to point in the direction of the current. The formula is then given as

$$d\vec{B}(\vec{r}) = \frac{\mu_0 I}{4\pi} \frac{dl(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3}.$$ 

This formula is called Biot-Savart’s law. The total magnetic field at the point $\vec{r}'$ is then the integral of all the $d\vec{B}$ of all the wire, namely

$$\vec{B} = \int_{\text{wire}} d\vec{B} = \int_{\text{wire}} \frac{\mu_0 I}{4\pi} \frac{dl(\vec{r}') \times (\vec{r} - \vec{r}')}{|\vec{r} - \vec{r}'|^3}.$$ 

This integral does basically nothing else than sum up all the contributions of the different parts of the wire. If the current through the wire is constant, the equation above is also relativistically correct, because the magnetic field is constant in time.

10.1.4 Example

Magnetic Field of an Infinitely Straight Long Wire

Consider a wire along the x-axis with a current $I$ flowing in the positive x-direction, see also figure 10.2. Since the wire is infinitely long there is no component of the magnetic field pointing in the x-direction. Additionally the problem is rotational symmetric around the x-axis. Therefore the strength of the magnetic field at a point only depends on the distance to the wire. We imagine a circle around the x-axis with radius $R$. As the magnetic field also makes circles around the wire (see figure 8.9) the line vector of the boundary line of the circle and the $\vec{B}$ field point in the same direction and therefore $\vec{B} ds = B ds$ where $B$ and $s$ are the absolute values of the respective fields. Therefore equation (10.1) leads to
\[ \mu_0 I = \oint_{\partial S} \vec{B} \, ds \]
\[ = \oint_{\partial S} B \, ds \]
\[ = B \oint_{\partial S} \, ds \]
\[ = B 2\pi R \]
\[ B = \frac{\mu_0 I}{2\pi R}. \]

\( B \) can be taken out of the integral since \( B \) is constant at a constant distance \( r \) from the wire.

\[ \vec{B} \]

Figure 10.2: Infinitely long wire along the \( x \) axis. The magnetic field (at radius \( R \)) is indicated by the dashed line, the path of integration is indicated by the solid line.

**Force on Two Parallel Infinitely Long Wires**

Let us take a wire along the \( x \)-axis and one parallel to the \( x \)-axis through the point \( y = r > 0 \). Assume the wires only lie in the \( xy \) plane. Assume that currents \( I_1 \) and \( I_2 \) flow through the first and second wire, respectively. We take them to be positive if they flow in \( +x \)-direction. According to the example above the first wire produces a magnetic field at the position of the second wire

\[ \vec{B} = \frac{\mu_0 I_1}{2\pi r} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \]
Therefore the force on a length $l$ on the second wire is due to the Lorentz force

$$\vec{F} = I_2 \begin{pmatrix} l \\ 0 \\ 0 \end{pmatrix} \times \vec{B} = \frac{\mu_0 I_1 I_2}{2\pi r} \begin{pmatrix} l \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \frac{\mu_0 I_1 I_2 l}{2\pi r} \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}.$$ 

As a consequence the two wires attract each other if both currents flow in the same direction and repel each other if the two currents flow in opposite direction.

The formula above is also used to define the SI-unit for electromagnetism which is the current:

"The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 metre apart in vacuum, would produce between these conductors a force equal to $2 \cdot 10^{-7}$ newton per metre of length."

[37]

This definition is also the reason that $\mu_0 = 4\pi \cdot 10^{-8}$ V·s·A$^{-1}$·m$^{-1}$ is a precise constant.

Magnetic Field of a Coil

If we wind a wire to circles and place them close together we get a coil. As we have seen above, the magnetic field of a current flowing through a straight wire has a very small impact$^4$. This is different in a coil, which can be viewed from different aspects: As the windings are close to each other, the total magnetic field at a point is the superposition of the magnetic field produced by each winding. Having many windings, the magnetic field becomes much larger. Equivalently one could say that through a current flows each winding and therefore the total current corresponds to the current through the wire multiplied by the number of windings. Therefore coils play an important role when considering magnetism$^5$.

We now look at a coil whose diameter is much smaller than its length. To calculate the magnetic field of a coil we consider a rectangle with width $l$ and length $L$ which we place as seen in figure 10.3.

Let $n$ be the number of windings per unit length. By Amperes law (see equation (10.1)) we get

$$\mu_0 I_{\text{tot}} = \oint_{\partial S} \vec{B} \cdot d\vec{s}$$

$$\mu_0 I n l = l B$$

$$B = \mu n I$$

$^4$With $10^{-7}$ N one obviously can not make an electromagnet.

$^5$In fact one usually neglects the effect of the magnetic field of all devices except coils or coil-like devices.
when $I_{tot} = nLI$ is the total current flowing through the rectangle and $I$ is the current trough the wire. The step from the first to the second line is because if we look a the sides of the rectangle the magnetic field is almost perpendicular to the sides of the rectangle (using $l \ll$ length of coil) and as a consequence $\vec{B} \cdot d\vec{s} = 0$ for the sides. Additionally we make $L$ very large so that at the top of the rectangle the magnetic field is very weak and therefore the contribution can be neglected.

The magnetic field of a coil is similar to the one of a permanent magnet. Therefore one could explain the magnetic field of a permanent magnet by assuming to have little circular currents in the magnet. This explanation is not really true since in quantum mechanics there exist also magnetic fields which do not origin from currents. But one can imagine how the magnetic field looks inside a permanent magnet where the field lines go from the south to the north pole. The field lines inside and outside the permanent magnet build closed lines as it is stated by equation (8.14).
10.2 Induction

Induction is an important phenomenon in electrodynamics because it gives a (first) connection between the electric and magnetic field. It basically tells us that a magnetic field that changes with time produces an electric field\(^6\). After introducing induction, we will look at two examples, namely the generator and the transformer.

10.2.1 Approach and Definition

To approach induction we consider a metallic pipe that is falling down. In that pipe a magnet is kept fixed (see figure 10.4). In the metal there are charges (positive nucleus and negative electron gas). On these moving charges the Lorentz force is acting\(^7\) and accelerates the electrons. They start circling in the pipe such that the magnetic field they produce opposes the change of the external magnetic field\(^8\). (see also figure 10.5).

\(\text{Figure 10.4: Left: frame of the magnet, the pipe is falling and the magnet is kept fixed. The white circles with the cross or the point indicate the direction of the current due to Lorentz force. The circle with point indicates that the current flows out of the sheet and the circle with the cross that it flows into the sheet. Right: frame of the pipe: The pipe stands still and the magnet is moving. The white circles again correspond to the current in the pipe.}\)

In the rest frame of the pipe the pipe itself is not moving. As a consequence there is no moving charge and therefore no Lorentz force acting. But the current needs to be independent of the frame. The explanation of this is induction: Since the magnet is moving (in the frame of the pipe), the magnetic field at a certain position of the pipe is changing with time. This change causes an electric field, which causes the movement of the charges.

\(^6\)Also the opposite is true, see 10.3.
\(^7\)Since the nuclei are much heavier than the electrons the influence of the magnetic field is much smaller.
\(^8\)External means here outside the metallic pipe, i.e. the magnetic filed of the magnet.
10.2. INDUCTION

Figure 10.5: Top of the pipe. The pipe is the gray ring, the current (and the electric field) is indicated as a circle by arrows.

If we integrate the electric field along a closed path (for example the one indicated in figure 10.5) we get a voltage $u_{\text{ind}}$. This voltage is connected to the magnetic field by

$$\oint \vec{E} \, ds = u_{\text{ind}} = -\frac{d\Phi}{dt}$$

(10.3)

where $\Phi$ is the flux that flows through the closed path. The obtained equation (10.3) has some very important properties:

- The negative sign corresponds to the fact that the current is such that it opposes a changing of the magnetic field. It therefore represents energy conservation. Would there be no negative sign, the current would amplify the magnetic field which then also would lead to a bigger current. This self-amplification would lead to an infinite current (in absence of resistance) which is of course not physical and in particular would violate energy conservation.

- Opposite to the electrostatic case, the electric field we obtained above has no starting and no ending point. The work required to afford to change the position of a charge depends on the path, therefore one cannot find a potential which is equal to the potential energy of a particle.

10.2.2 Self induction

In most of the electrical components and in most cases there is a weak interaction between the component and the magnetic field. Nevertheless there is one component that couples very strong to the magnetic field. This component is the coil. Since the different windings of a coil are close to each other and the same current flows through each winding, there is a lot of charge moving on a small spot. This current produces a strong magnetic field (see section 10.1.4). Since the magnetic field is proportional to the current $I$ and since

9In concrete cases one has to chose some conventions as the positive direction of current. Due to this conventions it might happen that the induction law does not contain a negative sign (see section 11.2.3). Be aware about this and check at the end if the result is meaningful or leads to non-physical behavior.

10This is not always the case, for example if the magnetic field goes through a ferromagnetic material. There, so called saturation can occur.
the cross section of a coil does not change with time, the total flux $\Phi$ through the coil is proportional to the current $\Phi = LI$. The proportionality constant $L$ is called inductance of the coil.

If we apply an alternating current to a coil, the magnetic field through the coil (caused by the current) is also alternating. Therefore induction happens. The induced voltage in the coil is

$$u_{\text{ind}} = -\frac{d\Phi}{dt} = -L\frac{di}{dt}.$$ 

The coil therefore opposes a change of the current by inducing a voltage. For more details also see the AC impedance of an inductor (see section 11.2.3). This phenomenon is called self induction.

In order to get a feeling for the inductance, let’s compute the inductance of a very long coil (see also 10.1.4). The magnetic field of such a coil is $B = \mu n I$ where $n$ is the number of windings per length $l$ and $I$ the current through the coil. Assume that we have a coil with cross section area $A$. Then the magnetic flux is $\Phi = BA = \mu n I A$. Applying an AC voltage with angular frequency $\omega$, the amplitude of the induced voltage in each winding is $U_{\text{ind}} = \omega \Phi = \omega \mu n I A$. Therefore the total voltage for all $N = ln$ windings is $U = \omega \mu n I A N = \omega \mu N^2 A l I = L \omega I$, where we got the inductance

$$L = \frac{\mu N^2 A}{l} = \frac{N^2 \mu}{R_m}$$

where $R_m$ is the magnetic resistance of the material around the coil\(^{11}\). It obviously scales with $N^2$. The first factor $N$ comes from the fact that the total current density scales with $N$, meaning that if we double $N$, the amount of moving charge is also doubled (not the current through the wire itself, but the same current passes twice as many times). The second factor comes from summing up the voltage at each winding. This means the ratio between the (amplitude of the) voltage and the current scales as $N^2$, but the magnetic field scales as $N$.

### 10.2.3 Generator

The biggest part of electricity is produced with generators\(^{12}\). The principle is in most cases the same and uses induction: Some external energy (as water or hot steam) drives a rotation (e.g. a turbine). This rotation causes a magnet to turn and leads to a changing magnetic field. This changing field induces an electric field in a coil or equivalently a voltage.

To have a more formal look, assume we have a fixed coil and a magnet turning near the coil, see also figure 10.6. For simplicity we assume that the magnet and the coil are very

---

\(^{11}\)One can think similarly to electric circuits about magnetic circuits. If the magnetic field has two possible "paths" to "flow", the total resistance corresponds to a parallel circuit. Similarly if the magnetic field is forced to take a longer path, one has to add up the resistances of the paths.

\(^{12}\)Only photovoltaic produces it differently.
close to each other such that the magnetic field is homogeneous and constant over the area of the coil. As the magnet is rotating, the magnetic field at the coil changes periodically. The magnetic field is

$$\vec{B} = B \begin{pmatrix} \sin(\omega t) \\ \cos(\omega t) \end{pmatrix}$$

where $B$ is the amplitude of the field (at the coil) and $\omega$ is the angular frequency of the rotation.

![Figure 10.6: Magnet rotating near a coil where a voltage $U_{\text{int}}$ is induced.](image)

Therefore the flux through the coil is $\Phi = B_x A$ where $A$ is the cross-section area of the coil and $B_x$ is the magnetic field component pointing towards the coil (here along the $x$-axis). In each winding, the voltage

$$u_{\text{ind}} = -\frac{d\phi}{dt} = -BA\omega \cos(\omega t)$$

is induced and as a consequence the total voltage\(^{13}\) is $U_{\text{tot}} = Nu_{\text{ind}}$. There is another very common setting where the magnet is fixed and the coil is rotating. The disadvantage of this configuration is that one has to make a connection between the rotating coil and the static consumer. This is usually done by brushes. The advantage is that it is also possible to generate (pulsed) DC voltages.

\(^{13}\)One could also include the factor $N$ in another way. Namely by saying that the area where the magnetic field goes through is $N$ times larger than the cross-section area of the coil.
10.2.4 Transformer

Another very important application is the transformer. This device allows to change the voltage of an alternating current (AC) circuit. It consists of a loop of iron and two coils wound on this loop, see also figure 10.7.

![Figure 10.7: The piece of iron is gray. The two coils are placed at the top and bottom part of the iron. \( N_p \) is the number of windings of the primary coil and \( N_s \) the number of windings of the secondary coil.](image)

One of the coils is connected to an AC source and is called primary coil. The other is connected to a consumer, this coil is called secondary coil. As we have seen above (see example of section 10.2.2), the magnetic field scales with the number of windings. So the magnetic field in the iron is proportional to \( N_p \), the number of windings of the primary coil. In addition, the induced voltage in the secondary coil is proportional to \( N_s \), its number of windings. Therefore we have that

\[
\frac{U_s}{U_p} = C \frac{N_s}{N_p}
\]

where \( C \) is a proportionality constant we do not know yet from the above considerations. And here the role of the iron comes in: If there would be no iron, the magnetic field of the primary coil would not necessarily pass through the secondary coil. As iron extremely amplifies the magnetic field (factor of about 5000), most of the magnetic field produced by the primary coil "flows" inside the iron and therefore passes through the secondary coil. As a consequence all the magnetic properties in both coils are the same and therefore \( C = 1 \). To derive this assume we apply an alternating voltage with angular frequency \( \omega \) and amplitude \( U_p \) at the primary coil. The voltage then is of the form \( u_p(t) = U_p \sin(\omega t) \), see also chapter 11. The magnetic field in the iron is related to the primary voltage as
\[ u_p(t) = N_p \frac{dB(t)A}{dt} \]
\[ B(t) = \frac{1}{N_pA} \int u(t) \, dt = -\frac{U}{N_pA} U_p \cos(\omega t) \]

where \( A \) is the cross-section area of the iron. Note that in the above equation there is no minus sign. This is discussed in more detail in section 11.2.3 and is not important in this example as we are only interested in the amplitudes and not in the phase relation. You might be puzzled why we use the induction law but there is no external magnetic field inducing a voltage. The point is that the self-induced voltage of the primary coil must be equal to the applied voltage (assuming there is no resistance). If this is not the case, there would be a voltage difference over the primary coil causing a larger current. This current would lead to a larger magnetic field until there is no voltage difference left.

On the other hand, the induced voltage in the secondary coil is given by (once again neglecting the minus sign in the induction law)

\[ u_s(t) = N_s \frac{dBA}{dt} = \frac{N_s A}{N_p A} u_p(t) \]

where we used that the magnetic field through the secondary coil is the same as through the primary coil.

We therefore get the famous equation for the ideal transformer

\[ \frac{U_s}{U_p} = \frac{N_s}{N_p} = \frac{I_p}{I_s} \]

where \( I_p \) and \( I_s \) are the currents in the two coils. Their relation can easily be found by power conservation: \( P_p = P_s \). Note: This equation is only true if the magnetic field of the two coils is strongly connected, i.e. the magnetic field though both coils is the same.

For example if the consumer on the secondary side takes out a lot of current, this current produces a magnetic field that opposes the one from the primary coil. As a consequence the primary coil needs more current to sustain the magnetic field\(^\text{14}\). On the other hand, the magnetic field "looks for" an alternative way to avoid the secondary coil. Therefore the two coils do not have anymore the same magnetic field and the above equation is not valid.

This breaking down is related to the construction of the transformer and in particular how the coils are placed. For example if the coils are on top of each other, their magnetic field is stronger connected than if they are aside of each other as in the picture 10.7 above.

\(^{14}\)Remember: the magnetic field and the voltage of the primary coil are connected to each other by the induction law. So for a give voltage, there must pass a certain magnetic field through the coil.
10.3 Displacement current

In the previous section we discussed how a time dependent magnetic field causes an electric field. In this section we will look at the opposite, namely a time dependent electric field causing a magnetic field.

There is a very common way to introduce this topic. Consider a (plate) capacitor that gets loaded with a current $I$ (see figure 10.8). The current causes a magnetic field $B$ around the wire. If we only consider the current, the magnetic field between the plates would be smaller than outside the capacitor, because the magnetic field is stronger near the wire. As there is no wire between the plates, the contribution from the current itself is smaller. But if we would measure the magnetic field around the plates, it would be the same as around the wires\footnote{At least outside of the plates and only with respect to the same distance from the wire.}. The current that loads the capacitor changes the electric field $E$. This changing electric field also causes a magnetic field. Outside of the capacitor, one cannot tell from the magnetic field whether there is a current or a changing electric field creating that magnetic field.

![Figure 10.8: A plate capacitor (grey planes) gets charged by a current $I$. The current causes a magnetic field $B$. Between the plates there is no current, but the changing electric field $E$ causes also a magnetic field.](image)

To get a formal description consider a plate capacitor with capacity $C = \epsilon \frac{A}{d}$ where each plate has cross section area $A$ and the plates are separated by $d$. From the basic equation for capacitors we can relate the current and the changing electric field by

$$\frac{\delta E}{\delta t} = \frac{1}{C} \frac{\delta I}{\delta t}$$
where we used that the electric field $E$ and the voltage $U$ (for a homogeneous electric field) depend on each other as $U = Ed$. This "current" $I$ between the plates is called displacement current and it has the same effect on the magnetic field as a usual current.

As a consequence we have to take the displacement current also in account in Ampère’s law (see section 10.1.2). This then leads to

\[
\oint_{\partial S} \mathbf{B} \, d\mathbf{l} = \mu_0 I_{\text{tot}} = \mu_0 \left( I + \epsilon A \frac{dE}{dt} \right) = \int_S \left( \mu_0 j + \mu_0 \epsilon \frac{dE}{dt} \right) \, dS.
\]

The last line is the most general description where we assume an arbitrary area $S$ with boundary $\partial S$ and a current density $\mathbf{j}$. If the current density $\mathbf{j}$ is zero, the formula above is very similar to the induction formula: The change of the electric flux through an area $S$ is proportional to the integral of the magnetic field around a closed circle (see also section 11.2.3).

This displacement current might look a bit irrelevant and not very useful for applications. But it is of great theoretical importance as it predicts/ensures conservation of charge.

10.4 Maxwell’s equations and their conclusions

10.4.1 Maxwell’s equations

In this section we summarize the basic equations that we found in electrodynamics. They will form a set of integral equations\textsuperscript{16} that describe the behavior of the electric field $\mathbf{\bar{E}}$ and the magnetic field $\mathbf{\bar{B}}$. This set is called Maxwell’s equations. In principle these equations combined with the Lorentz force describe everything about electromagnetism. Nevertheless it is far too complicated to solve these equations in many configurations, in particular when many charges are involved like in materials. So one simplifies/adapts these Maxwell equations for an electromagnetic field in materials which we will do in the next section 10.5.

\textsuperscript{16}One can also write down equivalent differential equations but they are more complicated.
We use the following notation: With $V$ we denote a volume and $\partial V$ is the surface\textsuperscript{17} that confines the volume $V$. A small element of the volume is $dV$. To integrate over the surface of $V$ we need to split $\partial V$ into small pieces denoted by $d\vec{S}$ where the area of the small piece is equal to the absolute value $|d\vec{S}|$ and the direction of $d\vec{S}$ points perpendicular to the surface outside the volume.

With $A$ we denote an area and the border of the area\textsuperscript{18} is denoted by $\partial A$ which is a closed line/loop. The area is again split into small pieces $d\vec{A}$. We also divide the closed line into small pieces denoted $d\vec{l}$. The length of these pieces is equal to the absolute value $|d\vec{l}|$ and it points tangent to the line. The direction of the small area pieces $d\vec{A}$ and the small curve pieces $d\vec{l}$ need to fulfill the right hand rule: If $d\vec{A}$ points in the direction of the thumb of the right hand, the other fingers of that hand indicate the direction of $d\vec{l}$.

With this notation Maxwell’s equations are given as

\[
\oint_{\partial V} \vec{E} \cdot d\vec{S} = \iiint_V \frac{1}{\epsilon_0} \rho \, dV \quad \text{Gauss’s law},
\]

\[
\oint_{\partial V} \vec{B} \cdot d\vec{S} = 0 \quad \text{no magnetic monopoles exist},
\]

\[
\int_{\partial A} \vec{E} \cdot d\vec{l} = -\frac{d}{dt} \int_A \vec{B} \cdot d\vec{S} \quad \text{Faraday’s law (Induction)},
\]

\[
\int_{\partial A} \vec{B} \cdot d\vec{l} = \mu_0 \left( \int_A \vec{j} \cdot d\vec{A} + \epsilon_0 \frac{d}{dt} \int_A \vec{E} \cdot d\vec{A} \right) \quad \text{Ampère’s law},
\]

where $\rho$ is the charge density and $\vec{j}$ the current density. The circles at the integrals on the left side denote that the surface or the path is closed (therefore the surface of a volume or the boundary of an area). The number of integral signs denotes the dimensionality of the integral, meaning the dimension of the space the integral has to be taken over.

Additionally to Maxwell’s equations one has to mention the Lorentz force in order to describe the interaction between charges and fields. The Lorentz force is given as

\[
\vec{F} = q(\vec{E} + \vec{v} \times \vec{B})
\]

where $q$ is the charge of a particle and $\vec{v}$ its velocity.

With these five equations it is basically possible to describe any problem involving charges and the electromagnetic field.

\textsuperscript{17}Take surface $\partial V$ as notation for the surface and not as partial derivative or small piece of $V$.

\textsuperscript{18}This area has not to be closed (as surface of a volume), this is why we use another variable than above.
10.4.2 Electromagnetic wave

Maxwell introduced his equations in 1865. These equations predict the existence of electromagnetic waves which then were experimentally measured by Heinrich Hertz in 1886. It is impressive how Maxwell managed to predict this waves only due to theoretical considerations.

We now want to deduce the electromagnetic waves and in particular some important properties. As the derivation is pretty tedious, you will not need to know it, but the results are pretty important. So we state them first and give the proof(s) afterwards.

We are going to prove that a time dependent electric and magnetic field lead to a wave satisfying the wave equations.

\[
\frac{\partial^2 E_y}{\partial x^2} = \epsilon_0 \mu_0 \frac{\partial^2 E_y}{\partial t^2} \\
\frac{\partial^2 B_z}{\partial x^2} = \epsilon_0 \mu_0 \frac{\partial^2 B_z}{\partial t^2}
\]

where we choose the coordinate system such that the electric field points in the direction of the $y$ axis and the magnetic field in the direction of the $z$ axis. Do not get confused by the $\partial$ sign, these are simple derivatives and the $\partial$ only indicates that $\vec{E}$ and $\vec{B}$ depend on multiple variables ($x, y, z, t$).

From this wave equations we can read off the speed of light $c$ which is given by

\[
c = \frac{1}{\sqrt{\epsilon_0 \mu_0}}.
\]

As the spatial derivative is along the $x$ axis, we note that the wave propagates in the $x$ direction meaning the $\vec{E}$ and $\vec{B}$ field are perpendicular to the direction of propagation.

**Proof:**

To apply Maxwell’s laws we have to use the third and fourth law, meaning we have to consider different areas where we perform the integrations. For this we consider a small cuboid with small side length $dx$, $dy$ and $dz$, see also figure 10.9.
We start by considering a time and space dependent magnetic field $\vec{B}$ pointing in the $y$ direction of the coordinate system. Since the sides are small we can assume $\vec{B}$ being constant along the $y$ axis of the cuboid but we have to take into account the change in the $x$ direction. We consider the square with the four arrows as our area where we have to integrate around to get the left side of Ampère’s law and we have to calculate the electric flux through the square for the right side. The left side of Ampère’s law then reads

$$\oint_{\partial A} \vec{B} \cdot d\vec{l} = \left( B_z + \frac{\partial B_z}{\partial x} \, dx \right) \, dz - B_z \, dz = \frac{\partial B_z}{\partial x} \, dx \, dz.$$ 

To calculate the electric flux, we assume the electric field is constant across the area. Since the area is parallel to the $xz$ plane, only the $y$ component of the electric field contributes to the flux so the right side of Ampère’s law is

$$\mu_0 \epsilon_0 \frac{d}{dt} \iiint_A \vec{E} \, d\vec{A} = -\mu_0 \epsilon_0 \frac{\partial E_y}{\partial t} \, dx \, dz$$

where the minus sign enters because due to the right hand rule, the surface vector of this area points in the $-y$ direction. Equating these two sides yields

$$\frac{\partial B_z}{\partial x} = -\mu_0 \epsilon_0 \frac{\partial E_y}{\partial t}. \quad (10.4)$$

Next we have to use the induction law which is in absence of charges or currents structurally analogous to Ampère’s law. With the same argumentation applied to the area $dx \, dy$ we get

$$\frac{\partial E_y}{\partial x} = -\frac{\partial B_z}{\partial t}.$$ 

Taking the derivative with respect to $x$ on both sides and inserting the first equation (10.4), we find

$$\frac{\partial^2 E_y}{\partial x^2} = -\frac{\partial}{\partial x} \left( \frac{\partial B_z}{\partial t} \right) = -\frac{\partial}{\partial t} \left( \frac{\partial B_z}{\partial x} \right) = \epsilon_0 \mu_0 \frac{\partial^2 E_y}{\partial t^2}.$$
This is nothing but the wave equation we looked for. Taking the derivative with respect to \( t \) instead of \( x \) and eliminate \( E_y \) would yield the equation for \( B_z \).

## 10.5 Electro-magnetic field in Materials

All materials are built of protons, neutrons and electrons. These particles are charged and therefore interact with the electromagnetic field. As a consequence the presence of a material influences the electromagnetic field. We will discuss this influence and its description in this section. Since the influence on the electric field is more intuitive, we will start with the electric field and treat it more precisely and then claim a similar behavior for the magnetic field.

### 10.5.1 Polarizability and dielectric constant

The case of a conductor in an electric field was already discussed in section 8.2.5\(^{19}\). We now want to look at insulators. In an insulator, charge cannot move freely, nevertheless the electric field influences the charge distribution in two different ways.

**Molecular polarization**

Being an insulator does not mean that the electrons cannot move at all. It only means that the binding of an electrons to its atom is strong enough that it cannot hop from one atom to the next. But it still can move slightly such that there is more negatively charge on one side of the atom than on the other, see also figure 10.10. As a consequence the negative electron cloud moves on one side and builds a dipole with the positive molecule.

![Figure 10.10](image)

**Figure 10.10:** Polarization of a molecule: The external electric field shifts the electron (little black dot) cloud towards the positive charge (causing the external electric field). Therefore the molecule gets a dipole moment.

\(^{19}\)The electrons get redistributed such that there is no electric field in the conductor.
**Oriental polarization**

Some molecules are already polarized. For example water molecules, where the electrons are stronger attracted by the oxygen than the hydrogen and therefore the region around the oxygen is negative "charged" with respect to the region near the hydrogen atoms. When an electric field is aligned, the molecules get rotated such that the positive part of the molecule points in the direction of the electric field.

Independent of how the electric field influences the insulator the effect is always the same: The polarization of the insulator leads to an electric field $\vec{E}_p$ that opposes the external field $\vec{E}_0$. This situation gets very obvious when we look at the situation drawn in figure 10.11.

![Figure 10.11: An electric field origins from two charged plates (left and right). This electric field points from the left to the right and is called external field as it acts from outside on the insulator. A polarizable insulator is placed between the two plates. The positive part of each insulator molecule moves towards the negative charged plate and vice versa, see left picture. Inside the insulator the positive and negative moments compensate each other but at the edge of the insulator the positive moment is not compensated on the left side and the negative not on the right side. This leads to an electric field in the insulator pointing opposite the external one, see right picture.](image)

This means the effective electric field between the plates $\vec{E} = \vec{E}_0 + \vec{E}_p$ is smaller than the one applied to the plates. For not too large electric fields, this polarization can be assumed to be proportional to the external field\(^{20}\), we call the proportionality constant electric permittivity $\epsilon_r$.

\[
\begin{align*}
\vec{E}_0 &= \epsilon_r \vec{E}, \\
\vec{E}_p &= \vec{E} - \vec{E}_0 = (1 - \epsilon_r)\vec{E} = \chi \vec{E}
\end{align*}
\]  

\(^{20}\)An attentive reader might argue that when all molecules are rotated, the polarization saturates. But this is only the case with very strong fields, otherwise thermal fluctuations and other effects oppose this alignment.
where \( \chi \) is called the electric susceptibility. For vacuum (and approximately air) we get \( \epsilon_r = 1 \) and therefore \( \chi = 0 \). Many materials have a permittivity between \( \epsilon = 1 \) and \( \epsilon = 10 \) but there are some in the region of hounded or even thousand.

In our consideration so far we always kept the external field \( \vec{E}_0 \). If we think of the two charged plates as a plate capacitor, this is equivalent to a constant charge on the plate capacitor. In most cases the situation is slightly different, i.e. the voltage applied to a capacitor is given and not the charge. As we have discussed in chapter 8.2.7, applying a voltage \( U \) to a plate capacitor leads to an electric field \( E = \frac{U}{d} \) where \( d \) is the distance between the plates. If we now insert an insulator between the plates. It opposes the electric field. As the voltage and as a consequence also the electric field is fixed, more charge needs to flow on the plates to create a stronger electric field. Assume the insulator has permittivity \( \epsilon_r \) and it fills the whole space between the plates (else see 10.5.3). Then the field produced by the charge on the plates must be

\[
E_0 = \epsilon_r E = \frac{\epsilon_r U}{d}.
\]

The charge (surface) density on the plates must therefore be

\[
\sigma = \epsilon_0 E_0 = \frac{U \epsilon_r \epsilon_0}{d}
\]

and the capacity is therefore given by \( C = \frac{A \epsilon_r \epsilon_0}{d} \) where \( A \) is the area of the plates. Obviously the capacity of a capacitor can be increased by inserting an insulator with high permittivity.

### 10.5.2 Electric Displacement and Polarisation

When we introduced the electric field its properties were characterized by Gauss’ law, see 8.1.5. Taking into account the polarisation effect described in the previous chapter, we need a new version of Gauss’ law. This gets for example obvious when we look at the plate capacitor described above when the insulator does not fill the whole space between the plates. Then the electric field between the plates inside and outside of the insulator is not the same but should be according to Gauss (see also the next chapter 10.5.3). This is because the insulator creates an electric field without being charged.

To get this problem fixed, we introduce a new field \( \vec{D} \) called displacement field. To understand the conceptual difference between the \( \vec{E} \) and \( \vec{D} \) field, let’s go one step back and have a closer look how we introduced the \( \vec{E} \) field. We started with the Coulomb force and deduced that a force \( \vec{F} \) acts on a charged particle \( q \) in an electric field due to \( \vec{F} = q \vec{E} \). Hence the electric field \( \vec{E} \) is primarily related to the force and only related to the charge via the dielectric constant \( \epsilon_0 \). And this \( \epsilon_0 \) "causes" trouble in case of polarized insulators meaning we have to introduce a new constant \( \epsilon_r \) renormalizing \( \epsilon_0 \). For the new field \( \vec{D} \), we want to approach from the opposite side, meaning we want to relate it primarily to the charge and then somehow to the force. As the two fields should be connected, we expect them not to differ too much. In fact if a point like charge \( q \) is placed in empty space the \( \vec{D} \) field is given by
\[ \vec{D} = \frac{q}{4\pi r^2} \vec{E} r \]

which only differs from the \( \vec{E} \) field by the missing factor \( \epsilon_0 \). With the same argumentation as for the electric field we can deduce Gauss’ law. For the \( \vec{D} \) field it looks like

\[ \iiint_{\partial V} \vec{D} \, d\vec{S} = \iiint_{V} \rho_{\text{free}} \, dV. \]

In this law we find another small difference between the \( \vec{E} \) and \( \vec{D} \) field. In case of the \( \vec{D} \) field we only want to consider a free (movable) charge \( \rho_{\text{free}} \) and not small charge separations due to polarisation\(^{21}\). For this also have a look at figure 10.12.

Figure 10.12: Gauss’ law is applied to the same situation but with two different volumes, indicated by the dashed line. On the left side, the volume only contains the right plate. We can apply Gauss’ law as usual. For both, the \( \vec{E} \) and \( \vec{D} \) field the result is correct. Looking at the right side, the volume also contains a part of the insulator. In case of the electric field, we know that Gauss’ law for the \( \vec{E} \) field does not work. This is because the insulator produces an additional field but the charge is the same. For the \( \vec{D} \) field, we only take into account the charge on the plate. As the insulator has no free charge, it does not at all contribute to the \( \vec{D} \) field, therefore Gauss’ law is valid, see also the next chapter.

\(^{21}\)The origin of the problem with Gauss’ law and the \( \vec{E} \) field lies in this polarisation: The polarisation produces an electric field without really separating charges but only shift charges slightly.
In most materials the two fields are connected to each other by the permittivity

\[ \vec{D} = \epsilon_0 \epsilon_r \vec{E}. \]

Using this equation and the linear relation between the applied and effective field (see equation 10.5) we can define a new quantity called polarization \( \vec{P} \). It is defined as

\[ \vec{P} = \vec{D} - \epsilon_0 \vec{E} = \chi \epsilon_0 \vec{E}. \]

### 10.5.3 Continuity equations at interfaces

We now want to have a closer look at what happens at the interface of two insulators (or vacuum). There are two cases we have to examine: The case where the field is perpendicular to the surface and where it is parallel. For all other cases we can split the field in these two components and then use the corresponding relations.

In our example of the plate capacitor, we always had perpendicular fields so let’s first have a look at them. Consider the situation shown in figure 10.12. We assume the usual plate capacitor such that we assume the field to be perpendicular to the plates. Applying Gauss’ law to the situation on the right side of figure 10.12, we get

\[ D \cdot 2A = Q_{\text{free}} \]

where \( A \) is the surface area of the plate. As a consequence, the electric displacement caused by one plate is

\[ D = \frac{Q_{\text{free}}}{2A}. \]

The other plate contributes the same amount such that the total \( D \) field is

\[ D = \frac{Q_{\text{free}}}{A}. \]

Considering the left side of figure 10.12, we also get \( D \cdot 2A = Q_{\text{free}} \) for one plate. This is because the insulator has no free charge and the considered surface is the same. Therefore we get the same electric displacement \( D = \frac{Q_{\text{free}}}{A} \). We see that the electric displacement \( D \) is continuous at the surface of the insulator. This is obviously different to the electric field in case of an insulator with \( \epsilon_r \neq 0 \), because in the insulator, the electric field is

\[ E_{\text{in}} = \frac{D}{\epsilon_0 \epsilon_r} \]

whereas outside the insulator it is

\[ E_{\text{out}} = \frac{D}{\epsilon_0} \neq \frac{D}{\epsilon_0 \epsilon_r}. \]

To get the behaviour of the \( \vec{D} \) and \( \vec{E} \) field parallel to the surface we have to go back to Maxwell’s induction law. Assuming to have static fields, the time derivative of the magnetic field is zero, hence also the induced voltage. Consider the situation shown in figure 10.13 where an interface of an insulator and air (or vacuum) is drawn. To apply the induction law we have to consider a surface \( S \) where we would have to calculate the magnetic flux through. As the magnetic field is constant, its time derivative is zero anywhere, so we

---

22. Distance between the plates much smaller than the length and width of the plates.

23. Remember the factor 2 enters because the field passes through the left and right side of the surface of the volume.

24. In fact we should first prove that Maxwell’s induction law is still valid before we can use it here. We will have a look at it in section 10.5.5.
do not need to calculate it. The other side of the equation is the integration of the electric field along the path confining the area. As we only consider the electric field parallel to the interface, the scalar product of horizontal boundaries of the surface $S$ is zero and we only have to look at the vertical ones. Assume that the area is enough small so that the electric field is constant along the vertical sides which are $\vec{E}_r$ in the insulator and $\vec{E}_a$ in the air. We then have

$$u_{\text{ind}} = \vec{E}_r \cdot \vec{e}_z l - \vec{E}_a \cdot \vec{e}_z l = 0$$

$$\vec{E}_r = \vec{E}_a$$

where $\vec{e}_z l$ is the vector of the left boundary. For the second line we used that the electric field and the path are parallel and that the path on the right boundary points in the opposite direction of the electric field, leading to the minus sign. We see that the parallel electric field is continuous across the interface.

![Figure 10.13: Interface of an insulator with dielectric permittivity $\varepsilon_r$ (gray) and air. To apply the induction law we consider the area $S$ and integrate along its boundary (rectangle with arrows).](image)

10.5.4 Magnetic field

Very similar considerations can be done in case of the magnetic field which we will not repeat here. Nevertheless the most important facts and relations are summarized.

**Distinction $\vec{H}$ and $\vec{B}$ field**

As in the case of the electric field, we have to introduce a new field, called magnetic field $\vec{H}$. But the correspondence is not as expected. The discussed $\vec{B}$ field corresponds to the $\vec{D}$ field and the new $\vec{H}$ field corresponds to the $\vec{E}$ field. To be precise, the $\vec{B}$ field is the magnetic flux density.\(^{25}\)

\(^{25}\)As long as only the $\vec{B}$ field is involved, it is simply called magnetic field. In this section we will explicitly call it flux density to distinguish it from the magnetic field $\vec{H}$.\)
Magnetisation $\vec{M}$ and its relation to $\vec{H}$ and $\vec{B}$ field

Similarly to the polarisation $\vec{P}$ in case of the electric field, we can introduce a magnetisation $\vec{M}$, also called magnetic polarisation. This magnetisation accounts how much little magnetic blocks in materials are aligned or anti aligned (pointing in opposite direction) to the magnetic field. We then find the relations

$$\vec{B} = \mu_0 \mu_r \vec{H} = \mu_0 \vec{H} + \vec{M},$$
$$\vec{M} = \chi_m \vec{H}.$$

where similarly to the electric case we introduced a magnetic permeability $\mu_r$ and a magnetic susceptibility $\chi_m$.

Dia-, Para- and Ferromagnetism

In case of the electric field, the charge gets always redistributed such that $\epsilon_r \geq 1$. This corresponds to para- and ferromagnetism, where the small magnetic pieces in a body align along the magnetic field such that they amplify the magnetic field outside the magnet. In case of the Paramagnetism, this effect is small, usually $1 < \mu_r \lesssim 1.1$ and it immediately vanishes if the external magnetic field is turned off. This is different in ferromagnetism, where $\mu_r \approx 100 - 10000$ is possible. The values depend strongly on the temperature, the external magnetic field and other effects. This effect is so strong that after switching of the external magnetic field, a magnetization remains. The only everyday-ferromagnetic materials are iron, nickel and cobalt. Since magnetism is a more complex phenomenon which involves many different aspects including quantum mechanics it is possible to have $\mu_r < 1$. This is called diamagnetism which weakens the magnetic field outside the material. The most important diamagnetic material is water, but also some metals are diamagnetic. As we see, except the case of ferromagnetism, the influence of a material on the magnetic field is very small, usually $0.99 \leq \mu_r \leq 1.01$ such that the effect can be neglected and one does not need to consider the $\vec{H}$ field.

Continuity equations at interfaces

Across the surface of a body, the flux density $\vec{B}$ perpendicular to the surface is continuous, whereas the magnetic field $\vec{H}$ is continuous parallel to the surface. This corresponds to the expected analogy with the electric field.
10.5.5 Maxwell’s equations in Materials

As motivated above, the interaction of a material and the electromagnetic field requires to introduce the $\vec{D}$ and $\vec{H}$ fields. As a consequence we have to adapt Maxwell’s equations. Since the whole influence of the material can be absorbed in the constants $\epsilon = \epsilon_0 \epsilon_r$ and $\mu = \mu_0 \mu_r$, we want to reformulate Maxwell’s equations such that they do not appear any more. This leads to the equations

$$\iint_{\partial V} \vec{D} \, d\vec{S} = \iiint_V \rho_{\text{free}} \, dV,$$
$$\iint_{\partial V} \vec{B} \, d\vec{S} = 0,$$
$$\oint_{\partial A} \vec{E} \, d\vec{l} = -\frac{d}{dt} \iint_A \vec{B} \, d\vec{S},$$
$$\oint_{\partial A} \vec{H} \, d\vec{l} = \left( \iint_A j \, dA + \frac{d}{dt} \iint_A \vec{D} \, dA \right),$$

together with the connection of the fields

$$\vec{D} = \epsilon_0 \epsilon_r \vec{E},$$
$$\vec{B} = \mu_0 \mu_r \vec{H}.$$

10.5.6 Electromagnetic waves

In section 10.4.2 we have derived the existence of electromagnetic waves and some of their properties. The derivation in the most general case also involving the fields $\vec{D}$ and $\vec{H}$ is very similar and we will not repeat it. Nevertheless in this proper derivation one would use the $\vec{H}$ field instead of the $\vec{B}$ field. Because when also taking into account the propagation through a surface separating two materials, one needs the continuity equation for the field parallel to the surface when using the path integrals. And the $\vec{H}$ field is the continuous one, as discussed in 10.5.4. In addition one has to include $\epsilon_r$ and $\mu_r$. So one gets the two wave equations

$$\frac{\partial^2 E_y}{\partial x^2} = \epsilon_0 \epsilon_r \mu_0 \mu_r \frac{\partial^2 E_y}{\partial t^2},$$
$$\frac{\partial^2 H_z}{\partial x^2} = \epsilon_0 \epsilon_r \mu_0 \mu_r \frac{\partial^2 H_z}{\partial t^2}.$$

From these equations we find some interesting facts:
10.5. ELECTRO-MAGNETIC FIELD IN MATERIALS

Speed of light

The speed of light is modified by $\epsilon_r$ and $\mu_r$. Therefore it reads as

\[
c = \frac{1}{\sqrt{\frac{\epsilon_0 \mu_0}{\epsilon_r \mu_r}}} = \frac{c_0}{n},
\]

\[
n = \frac{1}{\sqrt{\epsilon_r \mu_r}}
\]

where $c_0$ is the speed of light in vacuum and we introduced the refractive index $n$, which we know from optics.

Impedance

The amplitude of the two waves is not arbitrary, they are related to each other. To see this, consider the following form of the $\vec{E}$ and $\vec{H}$ field:

\[
E_y = E_0 \sin \left(\omega \left(t - \frac{x}{c}\right)\right),
\]

\[
H_z = H_0 \sin \left(\omega \left(t - \frac{x}{c}\right)\right),
\]

where we assumed an angular frequency $\omega$ and the amplitudes $E_0$ and $H_0$. In the derivation of the wave equation we encountered the equations (adapted to our situation)

\[
\frac{\partial H_z}{\partial x} = -\epsilon_0 \epsilon_r \frac{\partial E_y}{\partial t}
\]

which applied to our ansatz leads to

\[
\frac{\omega}{c} H_0 \cos \left(\omega \left(t - \frac{x}{c}\right)\right) = \epsilon_0 \epsilon_r \omega \cos \left(\omega \left(t - \frac{x}{c}\right)\right)
\]

\[
H_0 = c_0 \epsilon_0 \epsilon_r E_0.
\]

Using the relation we found for the speed of light, we find the ratio between the $\vec{H}$ and the $\vec{E}$ field which is also called impedance $Z$.

\[
Z = \frac{E_0}{H_0} = \frac{|\vec{E}|}{|\vec{H}|} = \sqrt{\frac{\mu_0 \mu_r}{\epsilon_0 \epsilon_r}}
\]

This impedance is only valid for electromagnetic waves, in case of static charges it is obviously wrong.
10.6 Energy of the electromagnetic field

Now that we have introduced the $\vec{H}$ and $\vec{D}$ field, we can derive the energy density of the electric field in its full generality. In addition we can introduce a quantity which quantifies the energetic flux of the electromagnetic field.

10.6.1 Energy density of the electric field

As discussed in section 8.2.7, the stored energy of a capacitor is $W = \frac{1}{2}QU$ where $Q$ is the (free) charge and $U$ is the applied voltage. We choose this formula because there no capacity $C$ appears, i.e. it is does not contain the dielectric constant which might be affected by polarization. As we derived in section 10.5.3, the charge and the $D$ field are connected by

$$Q = DA$$

where $A$ is the area of the plates. In addition we can relate the electric field and the voltage by the usual formula

$$U = Ed$$

where $d$ is the distance of the plates. Inserting this in the equation above, we obtain

$$W = \frac{1}{2}AdED.$$  

As $Ad$ is the volume of the capacitor, the energy density $u$ of the electric field is

$$u = \frac{1}{2} \vec{E} \cdot \vec{D}$$

where we made the step to the most general formula using the scalar product of $\vec{E}$ and $\vec{D}$. In most cases, $\vec{E}$ and $\vec{D}$ are parallel and therefore the scalar product is not necessary.

---

26Remember the capacity of a plate capacitor is given by $C = \frac{\epsilon_0 A}{d}$, where $A$ is the area of a capacitor plate and $d$ the distance between them.

27This formula is still valid since the voltage is a measure for the energy per charge, and the energy is something like force times displacement. Since the electric field is defined via the force, its connection to the voltage is still valid.
10.6.2 Energy density of the magnetic field

With similar arguments we could derive the energy density of the magnetic field. Nevertheless, we will not do this, but use the strong similarity of the electric and magnetic field formalism. Using the most general form of the electric field density, we simply claim that the magnetic field density is

\[ u = \frac{1}{2} \vec{H} \cdot \vec{B}. \]

10.6.3 Poynting vector

The energy density of the electric field is

\[ u_e = \frac{1}{2} \varepsilon_0 \varepsilon_r E^2 = \frac{\mu_0 \mu_r}{2} H^2. \]

Considering an electromagnetic wave, we can use the impedance \( E = Z H \), where 

\[ Z = \sqrt{\frac{\mu_0 \mu_r}{\varepsilon_0 \varepsilon_r}}, \]

to find \( u_e = u_m \). For the total energy density we then find in scalar notation

\[ u = u_e + u_m = \varepsilon_0 \varepsilon_r E^2 = \mu_0 \mu_r H^2 = \frac{E H}{c}. \]

The energy density is related to the intensity \( I \) as

\[ I = uc \]

where \( c \) is the speed of light. To derive this equation, consider an electromagnetic wave transporting the energy density \( u \) moving with the speed of light, also see figure 10.14. The energy \( E \) passing through a surface of area \( A \) in a time \( \Delta t \) is the density times the volume that passes the surface

\[ E = u A c \Delta t. \]

The intensity is the energy per time and per surface, so it is simply \( I = uc \).

Inserting the energy density we found for the electromagnetic wave we get an intensity of
We can even go one step further and give the intensity a direction, namely the direction of propagation of the electromagnetic wave. The resulting vector is called Poynting vector \( \vec{S} \). As we know from the derivation of the electromagnetic wave, the direction of propagation, the \( \vec{E} \) and \( \vec{H} \) field are mutually orthogonal which leads to the formula

\[
\vec{S} = \vec{E} \times \vec{H}.
\]

The funny thing is that the Poynting vector is not restricted to electromagnetic waves but can also be applied to any configuration. It basically tells you in which direction how much electromagnetic power flows per area. Take for example a simple DC circuit with a voltage source and a resistor. In figure 10.15, the electric and magnetic field as well as the Poynting vector is schematically drawn. Obviously at the source, the Poynting vector points away as the electric energy is inserted in the circuit, so given away from the source. The opposite happens at the resistor, where electric energy flows to.
Chapter 11

Alternating current (AC)

When the electric voltage periodically oscillates there are a couple of phenomena which can be observed but do not appear when a constant voltage is applied. In the following chapter we will describe sinusoidal voltages and the behaviour of electric devices when they are connected to it. The main difference to constant voltages is that there exist devices whose resistances depend on the frequency of the voltage. In particular we will have a look at ohmic resistors, capacitors and inductors and combinations of them.

11.1 Describing alternating voltage and current

Since the voltage (and therefore also the current) oscillates with time the voltage is not simply a number but a function of time. There are a couple of possibilities to describe this and we will have a look at the most important ones in the next sections.

11.1.1 Fourier series

There is an important theorem in mathematics which states that any periodical function $f(t)$ can be decomposed in a sum $f_n(t)$ of harmonic oscillations\(^1\). This theorem is called Fourier’s theorem and the decomposition is called the Fourier series. Let’s formulate this a bit more mathematically: Let $f(t)$ be a periodic function with period $T$. This means that for any $t$, $f(T + t) = f(t)$. Then the theorem states that this function $f(t)$ can be written as

$$f(t) = \sum_{n=0}^{\infty} A_n \sin(\omega_0 nt) + B_n \cos(\omega_0 nt)$$

where $\omega_0 = \frac{2\pi}{T}$ is the angular frequency and $A_n$ and $B_n$ are constants which depend on the periodic function $f(t)$. The whole theory about Fourier series is too complicated to be treated here. The important point is that if we know the behaviour of a device for a sinusoidal voltage we can construct its behaviour on any other periodic voltage. So from now on we always consider the voltages to be sinusoidal unless explicitly mentioned.

\(^1\)Harmonic oscillations are nothing else than sinusoidal oscillations.
11.1.2 Usual real notation

A sinusoidal voltage takes the general form

\[ u(t) = U_0 \cos(\omega t + \varphi) \]

where \( U_0 \) is the amplitude of the oscillating voltage, \( \omega = \frac{2\pi}{T} \) is the angular frequency and \( \varphi \) is the phase. Of course we could also chose the sine instead of the cosine with a different phase but this description is more consistent with another description (see section 11.1.4). It is convenient in AC notation to write time dependent quantities with lower-case letters \((u(t))\) and time independent quantities with capital letters \((U_0, U_{\text{eff}})\).

We assume that the voltage oscillates already a long time. Therefore the whole physics does not change if we shift time. A shift in time corresponds to an additional phase. It is often easier for calculation to shift time such that the phase for the voltage or the current is zero. If we take a particular choice of the phase, it is not guaranteed that other quantities have the same phase.

11.1.3 Phasor

If the voltage oscillates harmonically, one can see the voltage \( u(t) \) at a certain time \( t \) as cathetus of a rectangular triangle (see picture 11.1) with hypotenuse length \( U_0 \). Let us now rotate the hypotenuse of the triangle with angular frequency \( \omega \) in the mathematical positive orientation, which is anti-clockwise. Then the cathetus of the triangle behaves exactly like the alternating voltage. This means the whole information of the alternating voltage is encoded in the vector in picture 11.1: The amplitude, the angular frequency and also the phase are given. Therefore we can imagine alternating current as a vector which rotates with constant frequency, and by looking at its projection on the \( x \)-axis we get the familiar real notation.

11.1.4 Complex notation

A phasor corresponds to a two dimensional vector which has its starting point at zero and its endpoint is an \( x \) and \( y \) coordinate. We can now associate this two dimensional vector to a complex number \( z = x + jy \in \mathbb{C} \) where \( j \) is the imaginary unit\(^2\) and \( x, y \in \mathbb{R} \) are real. Therefore a phasor can be associated to a complex number with radius \( r = \sqrt{zz} = \sqrt{x^2 + y^2} \) with a turning phase. This can easily be written as (see section ??)

\[ z = re^{j(\omega t + \varphi)} = r \cos(\omega t + \varphi) + jr \sin(\omega t + \varphi). \]

In the case of \( z \) representing a voltage, \( r \) is the amplitude. Therefore \( r = U_0 \). The voltage is the projection of the phasor on the \( x \)-axis.

\(^2\)Usually \( i \) is the imaginary unit. Since \( i = i(t) \) is already the time dependent current, \( j \) is used.
11.1. DESCRIBING ALTERNATING VOLTAGE AND CURRENT

Figure 11.1: Phasor $u(t)$ with hypotenuse length $U_0$. The phasor rotates around the origin with the angular frequency $\omega$. The projection on the real axis corresponds to the measured quantity (for example the voltage).

Therefore it is the real part of $z$:

$$u(t) = \text{Re}(z) = U_0 \cos(\omega t + \varphi).$$

The complex notation might be a bit strange at the beginning because we use a non-real quantity (a complex number) to describe something real (for example a voltage). This is not really a problem because the complex number is just a notation. Nevertheless there are cases where one can really think about "turning voltages" and then the phasor and also the complex notation as a description of the phasor get some real properties (see 11.5.1).

In this script we use the complex notation, therefore we write a voltage as $u(t) = U_0 e^{j(\omega t + \varphi)}$ and keep in mind that the physical property is only the real part. The complex notation allows us also to add or subtract complex numbers because the real part of the sum of two complex numbers is equal to the sum of the real parts of the two numbers. Thus for $z_1 = a_1 + jb_1$ and $z_2 = a_2 + jb_2$ where $a_1, a_2, b_1, b_2 \in \mathbb{R}$ it follows
\[ \mathcal{R}(z_1 + z_2) = \mathcal{R}(a_1 + a_2 + j(b_1 + b_2)) = a_1 + a_2 = \mathcal{R}(a_1 + jb_1) + \mathcal{R}(a_2 + jb_2) = \mathcal{R}(z_1) + \mathcal{R}(z_2). \]

This property is important when we formulate Kirchhoff’s laws, see section 11.3.1. It allows us to perform all the (linear) calculations in the complex notation as long as we do not multiply voltage and current. This is the case when we look at the power\(^3\).

The complex notation is often used and it simplifies the equations a lot. Nevertheless it is also possible to do the whole alternating current theory without complex numbers. One then has always to think about the amplitude and the phase and examine the different phenomena according to both of these quantities. With the complex notation everything can be done with one number since the complex number contains the amplitude and the phase.

### 11.2 Impedance

In this section we examine the relation between the applied alternating voltage and current for different electrical components. The components we examine are the resistor, the capacitor and the inductor. Most other electrical devices behave like a combination of these three basic components, we discuss this in section 11.3.

The impedance is a generalisation of the resistance of an electrical device. The resistance \( R \) in a direct current circuit is the ratio of the voltage and the current: \( R = \frac{U}{I} \). Since we now examine alternating currents there are more parameters of freedom than only the "amount" of voltage or current which corresponds to the amplitude. We additionally have a phase shift between the applied voltage and the current. Therefore the impedance not only takes into account the ratio of the amplitudes of the voltage and current but also their phase shift. Since the voltage as well as the current are represented by a complex function, the impedance \( Z \) is usually a complex number\(^4\). The absolute value of \( Z \) corresponds to the ratio of the amplitude of voltage and current. The angle between the real axis and \( Z \) corresponds to the phase shift.

Assume that we have an electrical device and we measure a voltage \( u(t) = U_0 e^{j\omega t} \) and a current \( i(t) = I_0 e^{j(\omega t + \varphi)} \). Then the impedance is given as

\[
Z = \frac{u(t)}{i(t)} = \frac{U_0}{I_0} e^{-j\varphi}.
\]

From this impedance we conclude that the ratio of the amplitudes is given by \( |Z| = \frac{U_0}{I_0} \) and that the cosinus curve of the current follows the cosinus curve of the voltage (see picture 11.2).

\(^3\)In this case one has to first take the real part and then multiply the two quantities.

\(^4\)If one describes AC without complex notation one has to consider the ratio of the amplitudes and the phase shift separately as two real numbers.
11.2. IMPEDANCE

Figure 11.2: Sinusoidal voltage and current. The amplitude of the voltage is $U_0 = 9$ and the one of the current is $I_0 = 4.5$. The absolute value of the impedance is therefore $|Z| = 2$ and the phase shift is $\varphi = 1\text{rad}$. The angular frequency is $\omega = \frac{\varphi}{10}$, the period therefore is $T = \frac{2\pi}{\omega} = 20$.

11.2.1 Ohmic resistor

An ohmic resistor is a device where the current is always proportional to the voltage. Therefore there is no phase shift between the voltage and the current and the impedance of an ohmic resistor therefore is a real number $R$. Of course $R$ is exactly the ohmic resistance known from the direct current. This means

$$R = \frac{u(t)}{i(t)} = \frac{U_0}{I_0}.$$  

11.2.2 Capacitor

A capacitor is a device where charge can be stored (see electrodynamics 1 section 8.2.6). The basic equation of a capacitor with capacity $C$ is

$$C = \frac{Q}{U}$$  \hspace{1cm} (11.1)

where $Q$ is the stored charge and $U$ is the applied voltage. As we have seen, the capacity does not depend on the applied voltage. Therefore it is constant for an alternating voltage.
as well. If we now multiply equation (11.1) with the voltage \( u(t) = U_0e^{j(\omega t + \varphi)} \) and take the derivative with respect to time of the whole equation we get

\[
\begin{align*}
    u(t)C &= q(t) \\
    C \frac{du(t)}{dt} &= \frac{dq(t)}{dt} = i(t) \\
    C \frac{dU_0e^{j(\omega t + \varphi)}}{dt} &= CU_0 \omega e^{j(\omega t + \varphi)} = j\omega u(t) = i(t).
\end{align*}
\]

As a consequence we get the impedance \( Z_C \) of a capacitor by

\[
Z_C = \frac{u(t)}{i(t)} = \frac{u(t)}{Cj\omega u(t)} = \frac{1}{jC\omega} = -\frac{j}{C\omega}.
\]

In figure 11.3 the phasor and the time evolution of the voltage and the current are shown. The \(-j\) in the impedance means that the voltage is rotated 90° clockwise\(^5\). This is intuitively clear because the capacitor has to be charged in order to have a voltage and it is the current that charges the capacitor. Therefore the current is first and after that, when the capacitor is already charged a bit, one can measure a voltage. The angular frequency \( \omega \) in the denominator is intuitively also clear because if \( \omega \) is big, the capacitor gets charged and discharged fast. As a consequence the current is big and therefore the impedance small. With the capacity \( C \) in the denominator it is nearly the same, because a big capacity can store more charge, therefore the current is big.

### 11.2.3 Inductor

An inductor is an electrical device with a high inductance and an ideal inductor has no resistance and no capacity. The inductance \( L = \frac{\Phi}{I} \) of a device is the ratio between the magnetic flux \( \Phi \) and the current \( I \) through a device (causing the magnetic field), see also section 10.2.2. In AC there is an additional phenomenon, called self inductance. An alternating current causes an alternating magnetic field which induces again a voltage in that device. The induced voltage is such that it opposes an additional growth of the current.

\(^5\)Multiplying a complex number with \( j \) rotates that number by an angle of 90° in the positive orientation. Therefore the multiplication with \(-j\) rotates it 90° in the negative orientation, therefore clockwise.
11.2. IMPEDANCE

Figure 11.3: Phasor at time $t = 0$ and time diagram for a capacity. Be aware that the voltage or current is the projection on the $x$-axis.

From electro-magnetism we get the following equation:

$$u(t) = -u_{\text{ind}}(t) = L \frac{di(t)}{dt}$$

where $L$ is the inductance of the inductor. The switch of the sign between $u(t)$ and $u_{\text{ind}}(t)$ comes from energy consideration: If we look at an ohmic resistor, the resistance is bigger than zero. As a consequence $u(t)$ and $i(t)$ are in phase. If we look at a voltage source then the voltage and the current have a phase shift of $180^\circ$. This is basically a convention but it ensures that Kirchhoff’s laws hold$^6$. The induced voltage is like a source voltage, it therefore has opposite sign to the applied voltage $u(t)$. Assume that the current is $i(t) = I_0 e^{j(\omega t + \varphi)}$. Then the voltage is

$$u(t) = L \frac{di(t)}{dt} = LI_0 j \omega e^{j(\omega t + \varphi)} = jL \omega i(t).$$

The impedance of the inductor therefore is

$$Z = \frac{u(t)}{i(t)} = \frac{jL \omega i(t)}{i(t)} = jL \omega.$$ 

The phasor and the time dependence are shown in figure 11.4. The $j$ causes the voltage to be $90^\circ$ earlier than the current. This is clear: If one applies a voltage at an inductor, the

$^6$If there is no external alternating magnetic field through the circuit, the sum of all voltages over the devices of a closed path is zero. If we go in the direction of the current then it is obvious that the source and the consuming devices have different sign. See also 11.3.1.
growing current causes a growing magnetic field which opposes the current to grow. The current therefore grows slowly whereas the voltage over the inductor drops immediately.\footnote{This is different to the capacitor where the capacitor has to be charged in order to have a voltage drop}

The dependence on the angular frequency and the inductance is intuitively clear because a large inductance causes a stronger magnetic field and therefore the induced voltage, which opposes the current to grow, is also larger and as a consequence the current smaller. For high frequencies, the magnetic field changes fast. Since the induced voltage is proportional to the change of the magnetic field, the current is small and therefore the impedance is high.

---

Figure 11.4: Phasor at time $t = 0$ and time diagram for an inductor. Be aware that the voltage or current is the projection on the $x$-axis.
11.3 Combinations of R, C and L

In this chapter we have a look at the most important combinations of basic electrical elements. For this the usefulness of the complex notation gets obvious. Nevertheless we calculate one example with real impedances as well in order to show how this works.

11.3.1 Kirchhoff’s laws

We can nearly adapt Kirchhoff’s laws with some small modifications from direct current (DC). The first modification is that we use complex currents, voltages and impedances. Since charge cannot be created or destroyed, we get the continuity equation: For any point in an electrical circuit and for all time the sum of all currents $i_k(t)$ flowing to a particular point has to be equal to the change of charge $q(t)$ at that point:

$$\sum i_k(t) = \frac{dq(t)}{dt}.$$ 

Assuming no charge is stored, as it is the case for (nearly) all electrical elements except the capacitor, we get that the sum of all currents flowing to a point has to be zero, $\sum i_k(t) = 0$. This is one of Kirchhoff’s laws.

From electrodynamics we know that for any time, the sum of all voltages $u_k(t)$ around a closed path has to be equal to the change of the external magnetic flux $\Phi_{ext}$:

$$\sum u_k(t) = u_{ind} = -\frac{d\Phi_{ext}}{dt}.$$ 

Assuming there is no time dependent magnetic field, the equation simplifies to the known Kirchhoff’s law $\sum u_k(t) = 0$. Be aware that the voltages are complex quantities, which means that the vectorial sum of all the voltage phasors has to be zero.

11.3.2 Serial and parallel circuit

With the same argument as in the case of DC one gets formulas for the total impedance of electric components connected in series or parallel.

If two or more components are connected in series, the current $i(t)$ through the components is the same everywhere. Multiplying each component by its impedance, one gets that the voltage drop $u_k(t) = i(t)Z_k$ over each element. The sum of all these voltage drops must be equal to the applied voltage $u(t) = \sum u_k(t)$\(^8\). The total impedance is then given by

$$Z = \frac{u(t)}{i(t)} = \frac{i(t)\sum Z_k}{i(t)} = \sum Z_k$$

\(^8\)Here the voltage over the source is measured in the opposite direction, which causes a change of the sign. It is then compatible with the notation in the Kirchhoff’s laws.
which is exactly what we expect from DC with the difference, that the impedances $Z_k$ are complex.

With the analogous argumentation for components which are connected parallel to the voltage source, one gets that the total impedance is given as

$$Z = \left( \sum \frac{1}{Z_k} \right)^{-1}$$

which is also exactly what we expect from DC.

### 11.3.3 High pass filter

One important AC circuit is the high pass filter. A high pass filter has a high impedance for low frequencies and a low impedance for high frequencies. The high pass filter is often used as an audio filter to keep low frequencies away from the tweeters. There are different realisations of a high pass filter with different characteristics. One of the easiest is the RC circuit as shown in figure 11.5.

![Figure 11.5: High pass circuit. An alternating voltage $u_{in}(t)$ is applied to a capacitor and a resistor and the voltage $u_{out}(t)$ over the resistor is measured.](image)

The total impedance is given as

$$Z = Z_R + Z_C = R + \frac{-j}{\omega C},$$
$$\frac{1}{Z} = \frac{1}{R - \frac{j}{\omega C}} = \frac{\omega C}{R \omega C - j} = \frac{\omega C(R \omega C + j)}{(R \omega C)^2 + 1},$$
$$|Z| = \sqrt{R^2 + \left( \frac{1}{\omega C} \right)^2}.$$
where \( R \) is the resistance of the resistor, \( C \) the capacity of the capacitor and \( \omega \) the angular frequency of the voltage applied. If a voltage \( u(t) = U_0 e^{j\omega t} \) is applied to the high pass filter, the voltage drop over the resistor is given by

\[
u_R(t) = R i(t) = R \frac{u(t)}{Z} = R u(t) \frac{\omega^2 C^2 R + j\omega C}{(R\omega C)^2 + 1}.
\]

This leads to an amplitude \( U_{R_0} \) and phase \( \varphi \) of

\[
U_{R_0} = \frac{R U_0 \sqrt{\omega^4 C^4 R^2 + \omega^2 C^2}}{(R\omega C)^2 + 1} = \frac{R U_0 \sqrt{(R\omega C)^2 + 1}}{\sqrt{(R\omega C)^2 + 1}} \]

\[
\tan(\varphi) = \frac{\omega C}{R\omega^2 C^2} = \frac{1}{R\omega C}.
\]

This means that one measures a voltage which has an amplitude \( U_{R_0} \) with an angular frequency \( \omega \) and which leads the voltage of the AC source by the angle \( \varphi \). The voltage over the resistor is leading the voltage of the source: Because the voltage drop over the resistor is proportional to the current; and at the capacitor, the current leads the voltage. In the limit \( \omega \to \infty \) there is no phase shift and all the voltage drops over the resistor. This is also clear, because at very high frequencies, the capacitor has a very low impedance and therefore almost no influence on the circuit. In figure 11.6 the amplitude and the phase is plotted as function of \( \omega \).

![Figure 11.6: Amplitude \( U_{R_0} \) and phase \( \varphi \) of the high pass filter. As expected, the output voltage is higher for high frequencies. For the plot the following values were assumed: \( U_0 = 1V \), \( R = 1000\Omega \) and \( C = 10^{-6}F \). Be aware that the x-axis is a frequency \( f = \frac{\omega}{2\pi} \).](image-url)
Now let us once do the calculation in real notation in order to see how elegant the complex notation is. For the real notation there are different approaches, the easiest is to do the calculation with phasors by geometrically add impedances. But this is basically the same as we did above in the complex notation with vectors instead of complex numbers.

For the calculation with real quantities let the current be \( i(t) = I_0 \cos(\omega t) \). This does not mean that the voltage at the voltage source is of the form \( u(t) = U_0 \cos(\omega t) \), because there will be a phase shift between current and voltage. We now compute this phase shift as well as the relation between \( U_0 \) and \( I_0 \).

From Ohms law we know that the voltage over a resistor is always proportional to the current, therefore \( u_R(t) = Ri(t) = RI_0 \cos(\omega t) \). Additionally we know that the voltage over a capacitor is always lagging 90° and therefore \( u_C(t) = \frac{I_0}{\omega C} \sin(\omega t) \). The voltage over the voltage source is equal to the voltage over the RC combination and this is the sum of the two voltages

\[
 u_{in}(t) = U_R(t) + u_C(t) = I_0 \left( R \cos(\omega t) + \frac{1}{C \omega} \sin(\omega t) \right) = U_0 \cos(\omega t - \phi).
\]

Using some trigonometric theorems to get

\[
 U_0 = I_0 \sqrt{R^2 + \left( \frac{1}{\omega C} \right)^2}, \\
 \phi = \arctan \left( \frac{1}{\omega CR} \right).
\]

The absolute value of the total impedance therefore is \( \frac{U_0}{I_0} = \sqrt{R^2 + \left( \frac{1}{\omega C} \right)^2} \), the same as above in the complex notation. The phase shift is the same as well.

### 11.3.4 Resonant circuit

There are two different types of resonant circuits, the parallel circuit and the serial circuit. They behave differently but the equations to solve the problem are almost the same. So we only look at the serial circuit.

A serial resonant circuit is a circuit where a capacitor, a resistor and an inductor are connected in series. If there is no resistor (or its resistance is zero) the circuit is called an ideal serial circuit or an LC circuit. Since this is a special case we consider the general case where \( R \neq 0 \). Figure 11.7 shows the serial resonant circuit. We connect the resonant circuit with a voltage source with an angular frequency \( \Omega \).

---

9We could also assume a different phase because only the phase shift between current and voltage matters. So we choose the phase of the current to be zero because the current through both components is the same, so this simplifies the calculation.
Minimal resistance

The total impedance of the circuit is given as

\[ Z = Z_R + Z_C + Z_L = R + j \left( L\Omega - \frac{1}{\Omega C} \right) . \]

Obviously the absolute value of the impedance is minimal if the imaginary part vanishes. In this case the angular frequency is often called resonance frequency \( \Omega = \omega_0 \). It then holds

\[ L\omega_0 = \frac{1}{\omega_0 C} \Rightarrow \omega_0 = \frac{1}{\sqrt{LC}} . \]

In fact this is only the resonance frequency for the ideal LC circuit, because in the case of a not ideal circuit it is a bit smaller (see below). Since the impedance of the circuit is minimal, the current from the source through the circuit will be maximal at this frequency.

Natural frequency

Since the Resonant circuit is a very beautiful example of a harmonic oscillator we shortly repeat its properties. As it is an oscillator it can oscillate itself. Consider the following case: We open the circuit, charge the capacitor with a charge \( q_0 \) and close the circuit again. As soon as the circuit is closed, the overall voltage must be zero, and the sum of all voltages is given as

\[ \text{Resonance is the phenomenon where an oscillating system gets maximally excited by an external exci-} \]

\[ \text{tation. Since maximal excitation is not the same as minimal resistance, the frequency we discussed is often} \]

\[ \text{misleadingly called resonance frequency although the resonance frequency is something different.} \]

\[ 245 \]
0 = u_L(t) + u_R(t) + u_C(t) = L \frac{di(t)}{dt} + Ri(t) + \frac{1}{C} q(t) \\
= L \frac{d^2q}{dt^2} + R \frac{dq}{dt} + \frac{1}{C} q.

(11.2)

We therefore search a function \( q = q(t) \) which solves the equation above. We try the ansatz \( q(t) = q_0 e^{\lambda t} \). Inserting the ansatz in equation 11.2 we get a quadratic equation for \( \lambda \). Since we search for oscillating solutions, the discriminant is constraint by \( R^2 - 4 \frac{L}{C} < 0 \) and therefore the square root is complex. The general solution is

\[
q(t) = e^{-\delta t} \left( A e^{j \sqrt{\omega_0^2 - \delta^2} t} + B e^{-j \sqrt{\omega_0^2 - \delta^2} t} \right),
\]

\[
\delta = \frac{R}{2L}, \quad \omega_0 = \frac{1}{\sqrt{LC}}.
\]

The constants \( A \) and \( B \) depend on the conditions at \( t = 0 \). In our case we choose them such that the capacitor is maximal charged at \( t = 0 \) which leads to

\[
q(t) = q_0 e^{-\delta t} \cos \left( \sqrt{\omega_0^2 - \delta^2} t \right).
\]

This means that the circuit oscillates with a frequency \( \omega = \sqrt{\omega_0^2 - \delta^2} < \omega_0 \) which is called natural frequency. The oscillation is damped which is not surprising because energy gets dissipated at the resistor (see also section 11.4.1).

**Resonance (maximal current)**

If we apply a harmonic oscillating voltage to the circuit we can observe resonance. Assume we apply the voltage \( u(t) = U_0 e^{j \Omega t} \). According to Kirchhoff’s law the applied voltage is equal to the total voltage over the consuming devices

\[
U_0 e^{j \Omega t} = j L \Omega i(t) + Ri(t) - j \frac{1}{\Omega C} i(t) = -\frac{j L i(t)}{\Omega} \left( \omega_0^2 - \Omega^2 + 2j \delta \Omega \right),
\]

\[
\delta = \frac{R}{2L}, \quad \omega_0 = \frac{1}{\sqrt{LC}}.
\]

\[
i(t) = U_0 \frac{j \Omega}{L} \frac{1}{\omega_0^2 - \Omega^2 + 2j \delta \Omega} e^{j \Omega t},
\]

\[
\frac{1}{Z} = \frac{j \Omega}{L} \frac{1}{\omega_0^2 - \Omega^2 + 2j \delta \Omega}, \quad |
Z| = \frac{1}{L \sqrt{(\omega_0^2 - \Omega^2)^2 + (2 \delta \Omega)^2}}.
\]
If we search for the frequency with maximal current, we can take the derivative of $\frac{1}{|Z|}$ with respect to $\Omega$ and set it equal to zero\(^{11}\).

$$0 = \frac{d}{d\Omega} \left( \frac{1}{|Z|} \right) = \frac{1}{L} \left( D - \Omega \frac{1}{2D} \times \frac{D^2 (2(\omega_0^2 - \Omega^2)(-2\Omega) + 4\delta^2 2\Omega)}{D^2} \right),$$

where $D = \sqrt{(\omega_0^2 - \Omega^2)^2 + (2\delta\Omega)^2}$ is the nasty square root in the denominator of $\frac{1}{|Z|}$. Multiplying this equation with $LD^3$ one gets

$$0 = ((\omega_0^2 - \Omega^2)^2 + 4\delta^2\Omega^2) + 2\Omega^2(\omega_0^2 - \Omega^2) - 4\delta^2\Omega^2 = \omega_0^4 - \Omega^4,$$

$$\Omega = \pm \omega_0.$$

This result was expected because if the impedance is minimal, the current is maximal.

**Resonance (maximal voltage)**

Now let us have a look at the voltage drop over the capacitor:

$$u_C = -\frac{j}{\Omega C}i(t) = U_0 \frac{1}{LC} \left( \frac{1}{(\omega_0^2 - \Omega^2)} - 2j\delta\Omega \right) e^{j\Omega t},$$

$$U_{C_0} = U_0 \frac{1}{LC} \frac{\omega_0^2}{\sqrt{(\omega_0^2 - \Omega^2)^2 + (2\delta\Omega)^2}} = U_0 \frac{1}{\sqrt{(1 - \Omega^2)^2 + 4\delta^2 \Omega^2}}$$

where $U_{C_0}$ is the amplitude of $u_C(t)$. Once again let us have a look at the maximal voltage

$$0 = \frac{dU_{C_0}}{dt} = -U_0 \frac{1}{LC} \frac{2(\omega_0^2 - \Omega^2)(-2\Omega) + 4\delta^2 2\Omega}{2D^3},$$

$$\Rightarrow \Omega^2 = \omega_0^2 - 2\delta^2.$$ (11.3)

This means that one can measure the highest voltage at the capacitor at the frequency $\Omega = \sqrt{\omega_0^2 - 2\delta^2}$. This frequency is usually called resonance frequency. Picture 11.8 shows the voltage over the capacitor for different frequencies depending on the dumping $\delta$. This frequency dependence is called resonance curve. Concerning the asymptotic behaviour of the resonance curve we get the expected result: For very small frequencies, the voltage drop over the capacitor converges toward the applied voltage and is exactly the applied voltage $U_0$ in case the of DC. This is because when applying a constant voltage, the capacity corresponds to an interruption in the circuit and therefore the whole voltage drops over it. For very high frequencies, the impedance of the capacitor goes towards zero and therefore the voltage drop also goes towards zero.
Figure 11.8: Voltage drop $U_{C_0}$ over the capacitor as a function of the applied frequency for different dumpings $\delta$. The $y$-axis is normed to the excitation and $RLC$ circuit. The $x$-axis and the dumping $\delta$ are normed with respect to $\omega_0$. The dotted line shows the position of the maxima depending on the resonance frequency (for a given $\Omega$ the $\delta$ was calculated according to equation 11.3).

Let us make a summery of the different frequencies:

<table>
<thead>
<tr>
<th>Name</th>
<th>Formula</th>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>natural frequency</td>
<td>$\sqrt{\omega_0^2 - \delta^2}$</td>
<td>Frequency at which the circuit oscillates without external excitation.</td>
</tr>
<tr>
<td>minimal impedance or maximal current</td>
<td>$\omega_0$</td>
<td>The two cases are the same since for minimal impedance the current gets maximal at a given voltage</td>
</tr>
<tr>
<td>resonance frequency</td>
<td>$\sqrt{\omega_0^2 - 2\delta^2}$</td>
<td>Maximal voltage drop over the capacitor</td>
</tr>
</tbody>
</table>

Table 11.1: Overview of different frequencies where $\omega_0 = \frac{1}{\sqrt{LC}}$ and $\delta = \frac{R}{2L}$.

\[\text{Since the first derivative vanishes at the extrema.}\]
11.4 Power consideration and effective values

Until now we always considered equations which were linear in the voltage or in the current. Therefore there was no problem when adding complex voltages or currents and then taking the real part to get the property one measures at the circuit. Now we look at the power and therefore multiply current and voltage. So we have to take the real value first and then multiply the two quantities.

11.4.1 Power

The power consumption at time $t$ of an electrical device is $p(t) = u(t)i(t)$ which is time dependent. In many cases the instantaneous power is not so important and one is more interested in the mean power consumption\(^{12}\). Since we look at periodic oscillations one only has to consider one period $T$ to calculate the mean value. Let’s apply a voltage $u(t) = U_0 \cos(\omega t)$ at a device which causes a current $i(t) = I_0 \cos(\omega t + \varphi)$ with phase shift $\varphi$ flowing through that device. The mean power then is

$$P = \frac{1}{T} \int_0^T u(t)i(t)dt$$

$$= \frac{U_0 I_0}{T} \int_0^T \cos(\omega t) \cos(\omega t + \varphi)dt$$

$$= \frac{U_0 I_0}{2} \cos(\varphi)$$

(11.4) (11.5)

where we use $\cos(\omega t + \varphi) = \cos(\omega t) \cos(\varphi) - \sin(\omega t) \sin(\varphi)$ and $\sin(\omega t) \cos(\omega t) = \frac{1}{2} \sin(2\omega t)$.

The first integral\(^{13}\) gives $\frac{T}{2}$ and the second one gives zero. This means that for constant current $I_0$ and voltage amplitude $U_0$ the mean power depends on their phase shift. To be more precise, the power we calculated is the one that flows into that device and gets converted into another energy, for example heat at the resistor or rotation in case of an electric motor. The power is maximal for $\varphi = 0$, so for example for a resistor. On the other hand the power is minimal for $\varphi = \pm 90^\circ$ which corresponds to a capacitor or an inductor. In case of the capacitor, the phase shift is $\varphi = 90^\circ$ (see figure 11.3). After the capacitor is completely charged (which is the case for voltage at the capacitor being maximal, i.e. $u(t) = U_0$) the capacitor gets discharged. This means energy flows back from the capacitor.

\(^{12}\)For example consider a resistor: There the mean heat dispersion is more relevant than its instantaneous power consumption because the resistor also has some heat capacity

\(^{13}\)One can calculate this using partial integration or using that the cosine and the sine are the same function up to a phase shift and therefore $\int_0^T \cos^2(\omega t)dt = \int_0^T \sin^2(\omega t)dt = A$ for a certain number $A$. Then $A = \int_0^T \cos^2(\omega t)dt = \int_0^T 1 - \sin^2(\omega t)dt = \int_0^T 1dt - A = T - A$ which leads to $A = \frac{T}{2}$
to the source until there is no charge left in the capacitor. This is the case when the voltage is zero, and then it gets charged again. While charging, energy flows from the source to the capacitor. In the time diagram, the discharging periods are those where the voltage and the current have opposite sign and as a consequence the instantaneous power is negative. Obviously the charging and discharging energy is the same which means that the capacitor has no mean power consumption. A similar argument can be made with a coil where the energy is stored in the magnetic field.

11.4.2 Effective values

In direct current, the power is simply $P = UI$ whereas in AC there are the factors $\frac{1}{\sqrt{2}}$ and $\cos(\varphi)$. One defines

$$U_{\text{eff}} = \frac{1}{\sqrt{2}} U_0,$$

$$I_{\text{eff}} = \frac{1}{\sqrt{2}} I_0,$$

as the effective voltage and effective current. Sometimes these quantities are called RMS (root mean square) of the voltage or current. Using $U_{\text{eff}}$ and $I_{\text{eff}}$ one gets the mean power of a device as

$$P = U_{\text{eff}} I_{\text{eff}} \cos(\varphi).$$

In everyday life most of the voltages and currents are indicated with their effective value instead of the amplitude. This has the advantage that one can simply multiply the voltage and the current in order to get the power consumption. As an example the voltage at the power sockets has $U_{\text{eff}} = 230\text{V}$ which means that the amplitude of the voltage is $U_0 = \sqrt{2} \cdot 230\text{V} \approx 325\text{V}$.

11.4.3 Active, reactive and apparent power

A closer look to the mean power in equation (11.5) allows an interesting view on the phase shift. If we apply a voltage $u = U_0 \cos(\omega t)$ to a device, we can rewrite the current as $i(t) = I_0 \cos(\omega t + \varphi) = I_0 (\cos(\omega t) \cos(\varphi) - \sin(\omega t) \sin(\varphi))$ which is basically a superposition of a sine and a cosine oscillation with amplitude $-I_0 \sin(\varphi)$ and $I_0 \cos(\varphi)$ respectively. The cosine oscillation is in phase with the voltage and therefore it corresponds to the part of the current which dissipates energy at the device. One calls this the active current, which means that is the part of the current that does (useful) work. The sine oscillation does not perform any work, it only causes energy to be transferred from the source to the device and back. This current is called reactive current.

\footnote{The factor $2$ is only valid for sinusoidal signals, for other signal forms one has to calculate it from equation (11.4).}

\footnote{Assuming to have no phase shift $\varphi = 0$. This is the case in most of the everyday life devices. Nevertheless at big electric motors $\cos(\varphi)$ is indicated.}
Analogously one can define active, reactive and apparent power. What we have defined as mean power is also called active power \( P = U_{\text{eff}} I_{\text{eff}} \cos(\phi) \) since this is the electric power that is used by the electric device. The apparent power \( S = U_{\text{eff}} I_{\text{eff}} \) is the total power that is transferred between the source and the device. One part of this total power is the past used as active power and the other part is transferred back to the source. The second part is the energy oscillation between the source and the device. It is called reactive power and defined as \( Q = U_{\text{eff}} I_{\text{eff}} \sin(\phi) \). Using \( \sin^2(\phi) + \cos^2(\phi) = 1 \) one gets the following relation between the three different powers:

\[
S^2 = P^2 + Q^2.
\]

Generally one wants to have small reactive currents because a reactive current does not transfer energy to the device but it causes a larger current than necessary which leads to more losses in the cables. There are different possibilities to avoid reactive current. The simplest is to connect a capacitor or an inductor parallel to the device such that the total impedance suffices \( \phi = 0 \). Then the reactive current only oscillates between the device and the capacitor or inductor and not through the cables from the power station to the device.

### 11.5 Three-phase electric power

The word wide power net, which also provides electricity at home\(^{16}\), is a bit more sophisticated version of AC than we have discussed until now. This version is called three-phase electric power and it has some nice applications we will look at.

#### 11.5.1 Definition and production

Usually, a three-phase power supply consists of three wires and to each wire an AC voltage is applied. Additionally there is one wire which is called neutral wire, which we consider later. The three wires with the voltage are called phases. The amplitude of the voltage in all of the three phases is the same but shifted by \( 120^\circ = \frac{2\pi}{3} \) which is shown in figure 11.9.

There are different ways to create three-phase power, the easiest is the three-phase generator. It works as an usual generator but the three coils placed around the rotating magnet are displaced by \( 120^\circ \) each, see figure 11.10. The magnet in the center rotates and therefore the magnetic field at the coils is changing. This induces a voltage in the coils. Since the coils are displaced by \( 120^\circ \), the induced voltage is also shifted by \( 120^\circ \). Be aware that the induced voltage at a coil reaches its maximum when the magnet changes its polarity because the induced voltage is proportional the variation of the magnetic field in time. This variation is maximal when the magnet changes its polarity.

\(^{16}\)The power is brought to you home with a three-phase system and then the three phases are split up. Therefore at an usual power socket there is only one phase together with the ground wire and the neutral wire. Only very few power sockets are three-phase sockets. They are usually only needed for machines that consume a lot of energy. These sockets (usually) have five poles.
Physics Olympiad Script

CHAPTER 11. ALTERNATING CURRENT (AC)

Figure 11.9: The three voltages of a three-phase system. The wires are labelled by $L_1$, $L_2$ and $L_3$.

The setup described above allows a visualisation of the phaser: Treat the phasor as a vector perpendicular to the North-South direction of the magnet such that the maximal voltage is induced in a coil when the phasor is pointing to that coil. Then the induced voltage in each coil is the projection of the phasor on the axis of the coil.

11.5.2 Star and Delta circuit

A big advantage of the three-phase power circuit is that there are two possibilities of connecting a device. A three-phase device is (usually) connected to all three phases and it (usually) consists of three independent loads which are drawn as resistors in figure 11.11. However, there are two possibilities how this can be done, see figure 11.11. Let $U_0$ be the amplitude of all the three phases, the first phase therefore has the voltage $u_1 = U_0 \cos(\omega t)$, the second phase has $u_2 = U_0 \cos(\omega t + 120^\circ)$ and the third one has $u_3 = U_0 \cos(\omega t + 240^\circ)$.

Star circuit

The intuitive easier one is the star circuit where each load is connected the same way as the coils in the three-phase generator. At the first load (connected to phase $L_1$) the voltage $u_1$ drops, as expected. Analogously the other loads. If all loads have the same impedance then there is no current flowing from the three loads to the generator through the neutral wire. This is because the sum of all three currents is zero\(^{17}\). Therefore it is not necessary to connect the common point of the three loads with the neutral wire, the loads need only to be connected at one point with each other. It is often useful to connect the common point with the neutral wire to stabilize the circuit, see also chapter 11.5.3.

\(^{17}\)This can easily be seen if one adds the currents as phasors. Since we assume that all the three impedances are the same, the three currents are also the same. The three phasors form an equilateral triangle, therefore the starting point is equal to the ending point and as a consequence the sum of the three currents is zero.

252
The other possibility is the Delta circuit where each load is connected to two phases. The advantage of this circuit is that the voltage drop over each load is higher. To understand this let’s have a look at the voltage $u(t)$ over one load, for example connected to $L_1$ and $L_2$:

\[
u(t) = u_1 - u_2 = U_0(\cos(\omega t) - \cos(\omega t + 120^\circ))
\]
\[
= U_0(\cos((\omega t + 60^\circ) - 60^\circ) - \cos((\omega t + 60^\circ) + 60^\circ)
\]
\[
= U_0(\cos(\omega t + 60^\circ)\cos(60^\circ) + \sin(\omega t + 60^\circ)\sin(60^\circ)
\]
\[
- \cos(\omega t + 60^\circ)\cos(60^\circ) + \sin(\omega t + 60^\circ)\sin(60^\circ))
\]
\[
= \sqrt{3}U_0\sin(\omega t + 60^\circ)).
\]

This means that the voltage drop over each load has an amplitude of $\sqrt{3}U_0$.

To start a big electric motor one can use both, the properties of the star and Delta circuit. When starting the motor it needs a lot of current. If one applies a smaller voltage, the
current is also smaller\textsuperscript{18}. So one starts the motor in the star circuit and as soon as it rotates with constant speed, one connects it as Delta circuit and the motor has more power because the load is connected to a higher voltage. In the past this was often used to start large motors and there were special switches which allowed to change easily from the star to the Delta circuit. Of course it is a bad idea to connect a motor which is not build for a Delta circuit to a Delta circuit as it might get damaged.

\subsection*{11.5.3 Advantage of a three-phase system}

There are different reasons why our power supply at home is a three-phase system. We now investigate some of them.

\textbf{Role of neutral wire}

As we have seen in the section about the star circuit (see section 11.5.2), there is no current flowing through the neutral wire as long as the impedances at the three phases are the same. But if the impedances of the three consumers are not the same, a current flows through the neutral wire. If the common point in the star circuit is not connected to the neutral wire, this common point has a non-zero voltage with respect to the neutral wire. This is because at the common point the total current has to be zero\textsuperscript{19}. This is only possible if the three currents are the same\textsuperscript{20}. Since not all impedances are the same, the voltage drop over the loads will not be the same and as a consequence the sum of the voltage will not be zero. This sum of the voltage drops is the same as the voltage between the common point and the neutral wire. To guarantee that the voltage drop over all the consumers is the same, one connects the common point with the neutral wire and then the neutral wire is like a

\textsuperscript{18}This might sound trivial but it is important for the fuse because it should not melt.

\textsuperscript{19}There is no capacitor and Kirchhoff’s law holds.

\textsuperscript{20}To distinguish: The three currents are the same if the common point is not connected to the neutral wire. They are not the same if the common point is connected to the neutral wire.
fixing point where current can flow away if there are different impedances. This is exactly the situation at home: The houses are connected to the three-phase net. In the house the different sockets are connected to different phases and to the neutral wire. If all the phases are loaded the same (connected with the same impedance), then there is no current flowing out of the house through the neutral wire. But if for example only one power socket is used then the current flows through the neutral wire out of the house.

Energy transfer

Concerning the energy transfer the three-phase system offers a very efficient way to transport electrical energy over a (long) distance. If the impedance at the three phases is the same then (nearly) no current flows through the neutral wire. Therefore we only need the three phases to transport the electrical energy whereas in a simple AC circuit we need two wires for a closed circuit which corresponds to one single phase. This means in the three-phase system the three wires transport as many energy as three simple AC circuit consisting of six wires.

This gets very important if one wants to transport a lot of energy over a long distance because a lot of energy means thick cables and long distances mean long cables. Hence, a lot of wire. The power net consists of three thick cables where the phases are connected and one thin cable with the neutral wire (to stabilize the net). The different phases are connected then cleverly to the different houses (or even different villages) such that the impedance on all three phases is approximately the same. In terms of whole villages it is not relevant if someone at home uses a lot of power from one single phase\textsuperscript{21}. In conclusion the impedances applied to the three phases can be considered the same.

Three-phase motor

The three-phase motor is in principle the same as a generator (see figure 11.10) but instead of turning the magnet and inducing a voltage in the coils, a voltage is applied to the coils which create a magnetic field which causes the magnet to turn around. Since the magnetic field of a coil is proportional to the current flowing through it and the current is proportional to the applied voltage\textsuperscript{22} and the voltage correspond to a rotating phasor, the magnetic field also rotates. But the direction of the rotation (whether it rotates clockwise or anti-clockwise) depends on how the three coils are connected to the three phases. Actually, it only depends on how the three phases are connected to the coils: in clockwise or anti-clockwise direction. This means if the three coils coil 1, coil 2 and coil 3 are connected to $L_1$, $L_2$ and $L_3$ respectively or $L_1$, $L_3$ and $L_2$. By swapping the phases of two coils the motor changes its direction of rotation.

This property seems pretty unspectacular but it allows to run a motor in a certain direction which is not that simple to achieve in a simple AC circuit: In a simple AC circuit there is no rotating magnetic field (because there is only one phase). There is only a magnetic field that always points in the same direction and changes its strength. As a consequence the motor

\textsuperscript{21}Additionally each phase at home is connected to a fuse, the power consumption is therefore limited. In terms of the whole village the consumption of one single house is small/negligible

\textsuperscript{22}Maybe with phase shift but this phase shift is the same for the three coils in a three-phase motor.
does not know in which direction it should turn so it starts turning in an arbitrary direction. Of course there are some tricks how one can force the motor of a simple AC circuit to turn in a specific direction but it is more complicated than in case of a three-phase system.
Chapter 12
Special relativity

Roughly 100 years ago there were some phenomena in different fields of physics that could not be explained by the theories of that time. Einstein recognised that (at least some of) these phenomena can be described using a modification of classical mechanics and taking into account how one observes an other reference frame. The complete theory of relativity is very complex and needs a lot of maths. Nevertheless there is a small part of it which gives already a lot of interesting insights and which gets along with high school maths. This easier part is called special relativity and the more general theory is called general relativity. As long as nothing else is mentioned in this chapter, we will always do special relativity and sometimes simply call it relativity. The main source of this chapter are [64] and [65].

In this chapter we will first have a look at classical mechanics. After we understand the most important properties of classical mechanics we will look first at the assumptions of special relativity and then at its conclusions and interpretations. In the end we will resolve some of the most common paradoxes. Before we start doing physics lets have a look at some historical milestones that lead to relativity.

12.1 Historical milestones

Before Einstein developed relativity there were different experiments and parts of the theory which were incompatible. In electrodynamics this inconsistency was especially big. Let’s have a look at the big questions physicists had roughly 150 years ago.

12.1.1 Aether and electromagnetic waves

In 1864, Maxwell elaborated the famous Maxwell’s equations which describe (nearly) everything in classical electrodynamics. These equations also predict electromagnetic waves which were then experimentally proven by Heinrich Hertz in 1887. If one thinks about waves in every day life one always imagines a medium where this waves propagate. For example sound waves consist of air (or an other medium) which gets compressed and stretched. But the important thing is that there is a medium in which the wave propagates, eg. if there is no medium, nothing can get compressed and stretched so no sound can propagate. Only in the reference frame where the medium is resting, the wave equations take the simple form,
for all other reference systems one has to add additional terms\(^1\).

Until the beginning of the 20th century physicists thought that light would also propagate in a medium called aether. But Michelson and Morley showed in their famous experiment that this can not be true. For this they measured the speed of light in the direction of the rotation of the earth and perpendicular to it. They thought that the aether should rotate a bit slower than the earth because it must somehow glue to the rest of the universe which is not rotating. Therefore the speed parallel and perpendicular to the direction of rotation should differ, which was not the case. Since light propagates as a wave it fulfills the wave equations. In the frame of the medium they take the simple form given in equation 6.1. If one calculates the transition from one to another (moving) reference frame using classical mechanics, the simple wave equation change. Since light does not travel in a medium, the simple wave equations must hold in all reference frames, which contradicts the calculation of classical mechanics. This was a clear hint that something was wrong. At the beginning of the 20th century, famous physicists as Lorentz and Poincaré figured out how one has to calculate the transition from one to the other reference frame. This lead to the Lorentz transformation, which is nowadays the basis of relativity. So the found the equations some years before Einstein and Einstein surely knew the equations already. But the big success of Einstein was to give the equations a physical meaning. Because Lorentz and Poincaré (and many others) didn’t realise that these equations contain a deep property of nature, they simply used these equations to resolve the contradiction in the theory.

12.1.2 Flying electron

We now want to look at an example where classical mechanics leads to a contradiction. Consider the case where a current flows through a wire and an electron is flying with a constant velocity parallel to the wire, see figure 12.1. The current in the wire creates a magnetic field around the wire. The Lorentz force acting on the moving electron pulls it toward the wire (depending of the direction of the current and movement of the electron). This is different if we change the reference system. Let’s chose the system moving with the electron. In this system the electron does not move. As a consequence there is no Lorentz force and therefore it will not be attracted by the wire.

This seems to be a paradox but considering special relativity, one can resolve it. We will have a look at this at the end of this chapter when we understand relativity better.

12.2 Galileo transformations

In order to understand the basic concepts of relativity such as transformation from one reference system to another, let’s first have a look at these concepts in classical mechanics. There the transformation from one reference system to another is called Galileo transformation. But before we start with the Galileo transformation we have to define what a reference

\(^{1}\)To show this mathematically one needs a bit more math, therefore we skip this here
12.2. GALILEO TRANSFORMATIONS

Figure 12.1: In the first reference frame (left) the electron (black dot) is moving parallel to the wire through which a current flows. The current creates a magnetic field which causes a force acting on the electron (Lorentz force). The system left moves with the electron. Since the electron is not moving, no Lorentz force is acting.

system, also called frame of reference, is and which kind of frames we will look at\(^2\).

12.2.1 Reference System

To describe physics it is often very useful to describe it in a reference system. A reference system consists in a choice of an origin and three axis. Usually one takes the axis pairwise orthogonal which is then a Cartesian coordinate system\(^3\). In this system each object has at every time \(t\) a certain position \(\vec{x}(t)\). The origin is usually fixed at a certain object such as the sun if one wants to describe the motion of the planets or one edge of a desk when describing the motion of a ball on the desk. As a consequence the object which is connected to the origin does not move in that frame of reference because it has for all time the position \(\vec{x}(t) = \vec{0}\). If we fix the origin at a certain object \(X\) (\(X\) is here the name of the object) we say that we describe the problem in the reference frame of \(X\) or in the rest frame of \(X\). For example in chapter 12.1.2 we described the problem in the reference system of the wire (left) and in the system of the electron (right)\(^4\).

Sometimes we will give our reference frames names, for example we will often denote two frames as \(\Sigma\) and \(\Sigma'\). One then says that a property, for example time or position, are measured in some (specific) frame. This sounds often pretty abstract and it is sometimes useful to visualize it. One convenient visualization is to think for example \(\Sigma\) as the rest frame of the earth. So measuring a quantity in \(\Sigma\) means that someone (for example you) is standing on the earth and measures this quantity. If there are two frames involved it is often easy to think about a space ship as the second frame \(\Sigma'\). For this assume that a colleague of you flies in a space ship and performs his own measurement.

There are different names for the most important and most common reference frames.

\(^2\)The description of the transformation between arbitrary reference frames is described by general relativity.
\(^3\)One could also take polar or spherical coordinates. The choice depends on the symmetry of the problem but in many cases Cartesian coordinates is a good choice (at least to start with).
\(^4\)In that example it was not necessary to chose concrete axes so we left it out.
Rest frame
As already defined above this is the system which is connected to a certain object. This object will then rest in that system, therefore it’s the rest frame of that object.

Laboratory frame of reference
This is usually the system where we start describing a problem. If one performs a measurement this is the rest frame of the experiment.

Centre-of-momentum frame
This is the frame where the total momentum is zero. If the momentum is conserved\(^5\), the total momentum will stay zero which often simplifies the calculation.

Inertial frame of reference
Since this is a very important type of reference frame let’s have a closer look at it in the next chapter:

12.2.2 Inertial frame of reference
Of particular importance is the inertial frame of reference. It is a frame where Newton’s laws are valid\(^6\). In particular this means that if no force is acting on an object, this object will not move or move with a constant velocity.
For example the earth is approximately an inertial frame\(^7\) because any force acting on an object (for example gravity) is proportional and parallel to the acceleration. It is not really a inertial frame because the earth is moving around the sun and rotating around its axis, therefore it’s accelerated and there are fictitious forces such as the centrifugal force or Coriolis force.
If we look at a space station, for example the ISS, orbiting around the earth we might think this is an inertial system. But in fact it is not because all the masses in the space station seem to be weightless (for example a pen floating), although there is a gravitational force. If we once found an inertial frame of reference then any system moving with a constant velocity to that system is also an inertial system. In Newtonian mechanics this is obvious.
Lets assume the frame \(\Sigma\) is an inertial system and that the system \(\tilde{\Sigma}\) is moving with velocity \(v(t)\) measured in \(\Sigma\). If a force \(F\) (measured in \(\Sigma\)) is acting on an object the object gets accelerated with \(a = \frac{F}{m}\), since it is an inertial frame of reference. If in \(\tilde{\Sigma}\) the same force \(\tilde{F}\) and the same acceleration \(\tilde{a}\) is measured, then \(\tilde{\Sigma}\) is also an inertial system. This gives us restrictions on the velocity \(v(t)\) because since

\[
a = \tilde{a} + \frac{dv}{dt} = \tilde{a}
\]

\(^5\)This means no external forces act on our system.
\(^6\)This means a force acting on an object is proportional to its acceleration: \(\vec{F} = m\vec{a}\).
\(^7\)Neglecting relativistic effects due to gravity.
where the first step is simply the transformation of the acceleration between two reference frames (in Newtonian mechanics) and the second step is the assumption that the two accelerations are equal. The consequence is that \( \frac{dv}{dt} = 0 \) and therefore the velocity is constant (including \( v = 0 \)). This means that \( \bar{\Sigma} \) is an inertial frame if and only if it is moving with constant velocity or it is displaced by a constant vector or rotated by a fix angle with respect to \( \Sigma \) (assuming \( \Sigma \) is an inertial system).

### 12.2.3 Galileo Transformation

Now we have enough definitions to look at the Galileo Transformation. This transformation describes how time and space transform between different inertial frames in classical mechanics. Let \( \Sigma \) and \( \Sigma' \) be two inertial frames of reference. In classical mechanics there is no reason why the time in one frame elapses slower or faster than in the other. This means the time in both frames is the same or mathematically \( t = t' \). This is the transformation of time between the two systems. To examine the transformation of space, assume that the axis in the two systems are parallel to each other and that \( \Sigma' \) is moving with a constant speed along the \( x \) axis of \( \Sigma \) (see figure 12.2). In \( \Sigma \) the origin of \( \Sigma' \) has therefore the coordinates

\[
O_{\Sigma'} = \begin{pmatrix} vt \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} x_0 \\ 0 \\ 0 \end{pmatrix}
\]

where \( x_0 \) is the position of the origin of \( \Sigma' \) at \( t = 0 \) measured in \( \Sigma \). This means that a point \( \vec{r}' \) measured in \( \Sigma' \) has the coordinates \( \vec{r} \) in \( \Sigma \)

\[
\vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} vt \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} vt \\ 0 \\ 0 \end{pmatrix} + \vec{r}'.
\]

This means that the \( y \) and \( z \) coordinates of the two systems are always the same, eg. \( y = y' \) and \( z = z' \). The \( x \) coordinate transforms as \( x = vt + x' \) or \( x' = x - vt \).

If we put all the transformations together we get

\[
t = t' \\
x = vt + x' \\
y = y' \\
z = z'
\]

where obviously time and space only mix at the \( x \) component. Sloppily said: time influences measurements of space but not the opposite. This will be different in relativity.

---

8We denote all quantities measured in \( \Sigma' \) with a prime ‘ and if measured in \( \Sigma \) without.

9They might have a different zero point of time and therefore the time differs by a constant. But this constant has no influence on the the physics so we can without loss of generality set it to zero.
12.3 Lorentz transformation

In this chapter we will describe the basic transformation in relativity. Further phenomena will be discussed in the next chapter.

12.3.1 Einstein’s Postulates

All the special relativity can be derived by two assumptions about nature. These assumptions are called Einstein’s postulates which are:

1. The laws of physics are the same in all inertial frames of reference.

2. The speed of light, denoted by $c$, has a finite value.\footnote{The value of $c$ is $c = 299'792'545.8$ m·s$^{-1}$.}

Sometimes the postulates are formulated in a slightly different way but the conclusions remain the same. In the regime of classical mechanics and considering the definition of inertial frames it is clear that the physics in all inertial frames is the same.\footnote{At least if the speed of light is considered infinite.} In relativity we have to state this explicitly because we cannot assume any more that the acceleration in two frames has the same value. The second postulate needs some more explanation. What is called speed of light is in fact a much more important speed, namely the speed of information. A precise discussion about this would go beyond this scope and is not important to understand relativity. So we only illustrate this at a small example. Consider two objects that interact with each other, for example by electromagnetic or gravitational forces. Then this interaction only happens with the speed of light, which means that one object "sees" or "feels" the other object where it was some time before. This is because the information where the objects are spreads out with the speed of light, so the objects have already moved a bit before one sees their position.

Without calculation we can already now state some interesting things.
• Since the physics is in all inertial frames the same and the speed of light is a physical property, the speed of light has the same value in all inertial frames, independent of the movement of the light source of the observer.

• Since the speed of light is the maximal possible speed\textsuperscript{12}, the addition of velocities will not behave the same as in classical mechanics.

In the following we will always look at inertial frames of reference because for all other frames one needs one more postulate which leads to general relativity.

### 12.3.2 Synchronisation of clocks

Since time will not be an absolute quantity\textsuperscript{13} in relativity we have to find a method how we can synchronise two clocks. Two clocks are synchronised if they "show" the same time, where "show" means that if we look at two distant clocks we might see two different times because the light from the clock further away needs longer to reach our eyes than from the nearer one. If the distance from the observer to each clock is the same then the observers sees the same time on the two synchronized clocks. In fact the "show" is defined more precisely by the process of synchronisation. For the synchronisation we of course neglect all effects that might happen in a real-life experiment such as imprecise clocks or delayed detectors and so on.

To synchronize two clocks we send out one light pulse at the time $t_0$ from one clock, we call it clock 1, to the other, clock 2. There it is (immediately) reflected and it returns to clock 1. The whole time from clock 1 to clock 2 and back to clock 1 we denote as $\Delta t$. Clock 2 is now set such that at the time the light pulse arrived it would have shown the time $t_0 + \frac{\Delta t}{2}$. We therefore have constructed the synchronisation such that two light pulses which are send out at the two clocks at the same time meet exactly in the middle of the two clocks. As we will see later, simultaneousness is not in all frames the same, therefore it does not make sense to synchronize two clocks in different frames of reference.

### 12.3.3 Time dilation

Let’s start examine how the different quantities transform from one inertial frame to another. We start with a time interval. Already now it shall be pointed out that the transformation of time itself behaves differently than the transformation of time intervals, which can be seen in section 12.3.6. To investigate the transformation of time intervals we need a clock or even more fundamental something that creates a periodic signal. The simplest clock is the so called light clock, which is shown in figure 12.3. A light clock consists of two mirrors between which a short light pulse is trapped. Assume we observe a light clock moving parallel to the mirrors with a constant velocity $v$. The frame we observe the moving clock

---

\textsuperscript{12}Because if something is faster than light it could deliver informations faster than light which is in contradiction to the fact that informations cannot be delivered faster than light

\textsuperscript{13}Absolute time means time does not depend on the position and the reference frame.

\textsuperscript{14}This looks like we have to set the time of clock 2 before we actually know which time we have to set. But this is not necessary because clock 2 can remember the time the light pulse arrived and then comparing with the time $t_0 + \Delta t$ at clock 1 one can calculate by which amount one has to change the time at clock 2.
we call $\Sigma$ and the rest frame of the clock $\Sigma'$. One period $\Delta t'$ in $\Sigma'$ is then the time the light needs to make one complete cycle, which is

$$\Delta t' = \frac{2L'}{c}$$

where $L'$ is the distance between the mirrors and $c$ is the speed of light, which is the same in all frames (and therefore there is no '$$').

![Diagram](image)

Figure 12.3: Left: the light clock in the rest frame of itself. The light pulse is moving up and down between the mirrors. Right: The same light clock observed from a frame which is moving with respect to the clock. The light is not moving straight up and down but makes a zigzag.

In our system we see the same distance between the mirrors\(^{15}\), so $L = L'$. In $\Sigma$, however, the light does not move perpendicularly to the mirrors, as the whole clock is moving. The Period is defined as the time the light needs to go from one mirror to the other and back. In Sigma this is

$$\Delta t = \frac{2L}{v_\perp}$$

where $v_\perp$ is the velocity perpendicular to the mirrors. Using Pythagoras one can calculate $v_\perp$ and one gets for $\Delta t$:

---

\(^{15}\)This would follow if one derives relativity more rigorously but this would need more math and would not give more insight.
\[ \Delta t = 2L \frac{1}{v_{\perp}}, \]
\[ = 2L \frac{1}{\sqrt{c^2 - v^2}}, \]
\[ = \frac{2L}{c} \frac{1}{\sqrt{1 - \beta^2}} = \gamma \Delta t', \]

where \( \beta \) and \( \gamma \) are often used short cuts defined as

\[ \beta = \frac{v}{c}, \]
\[ \gamma = \frac{1}{\sqrt{1 - \beta^2}}. \]

This means that if on a clock in \( \Sigma' \) one time interval passes, for example one second, in our frame \( \Sigma \) more time passes, for example one and a half seconds. Since both systems are inertial systems one cannot tell whether one is moving or the other. Therefore it does not depend whether we look at a light clock or a (Swiss) watch, all will show the same time and therefore show the phenomenon described above. This means we see the clocks in a moving frame going slower than ours. This phenomenon is called time dilatation.

### 12.3.4 Lorentz contraction

Similar to time we can have a look at distances. Since we don’t know yet how distances transform, we cannot simply compare an unknown distance with a meter stick in in two different systems\(^{16}\). Since the speed of light is constant in all systems, we can reduce this problem to measuring the time the light needs to pass a certain distance. First we notice that distances perpendicular to the direction of motion do not change\(^{17}\). This is because if the light moves perpendicular to the direction of motion there is no difference in time whether they move in one or the other direction (to understand this, keep on reading). A first (wrong) approach to get the transformation for distances parallel to the direction of motion might be \( \Delta L = c \Delta T \) which then would lead to \( \Delta L = \gamma \Delta L' \) which means that a moving object seems longer than in its rest frame (which is wrong). The mistake in our derivation happened because we did not take into account that in a moving system, the time the light needs for a certain distance depends weather the light propagates in the direction of flight or opposite (see figure 12.4). In fact, moving objects seem shorter than in the in their rest frame.

To get the right result we have to do the following: We send out a light pulse along the direction we would like to measure. At the other end of this distance we place a mirror

\(^{16}\)We don’t know how long a meter seems in a moving system.
\(^{17}\)We have already implicitly used this at the derivation of the time dilatation.
Figure 12.4: 1) Sending out a light pulse. 2) Arrival of the light pulse at the mirror after the time $\Delta t_1$. 3) Reflection of the light at the mirror (this happens immediately). 4) Arrival of the light pulse at the detector, the time the light needs from the mirror to the detector is $\Delta t_2$.

which reflects the light pulse. The time the light needs to pass the distance source-mirror we denote by $\Delta t_1$. The time from the mirror to the detector by $\Delta t_2$. Measuring the time for twice the distance ($\Delta t_1 + \Delta t_2$) and dividing it by $2c$ we get the right length. In the rest frame this is easy because the time to the mirror and back are the same. We denote this distance $\Delta L'$ (we call this the system $\Sigma'$ again since we want to measure it from our system $\Sigma$. Therefore $\Sigma'$ is the moving system with respect to our system $\Sigma$). The time for the distance in $\Sigma'$ is therefore $\Delta t' = \frac{\Delta L'}{c}$.

In $\Sigma$ we get for $\Delta t_1$ and $\Delta t_2$

$$\Delta t_1 = \frac{\Delta L}{c - v},$$

$$\Delta t_2 = \frac{\Delta L}{c + v}$$

where the denominators are given this way because the light pulse moves in the same or opposite the direction of flight. The fact that we simply add or subtract velocities is no contradiction to the statement above because there is no particle moving with $c + v$. The term $c + v$ appears because the light pulse and $\Sigma'$ are moving. From time dilatation we
know that $\Delta t_1 + \Delta t_2 = \gamma(\Delta t'_1 + \Delta t'_2) = 2\gamma \Delta t'$ and using some math we get

$$
\Delta t' = \frac{1}{2\gamma} (\Delta t_1 + \Delta t_2)
$$

$$
\frac{\Delta L'}{c} = \frac{1}{2\gamma} \left( \frac{\Delta L}{c-v} + \frac{\Delta L}{c+v} \right)
$$

$$
= \frac{2c\Delta L}{2} \sqrt{1 - \frac{v^2}{c^2}}
$$

$$
= \frac{\Delta L}{c} \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}
$$

$$
\Delta L' = \gamma \Delta L.
$$

Since $\gamma \geq 1$, a moving object seems shorter than it is in its rest frame.

### 12.3.5 Symmetry of time dilatation and Lorentz contraction

According to the first postulate, the equations (Time dilatation and Lorentz contraction) look the same in all inertial frames. Assume you are observing a space ship. According to time dilatation you see the clocks in the space ship going faster than yours with $\Delta t = \gamma \Delta t'$. A colleague in the space ship would observe the same with your clocks, namely that your clocks move slower than his. According to the colleague, the equation $\Delta t' = \gamma \Delta t$ holds. One might now argue that from this follow $\Delta t = \gamma \Delta t' = \gamma^2 \Delta t$ and therefore $\gamma^2 = 1$ which questions of course all we did (since $\gamma > 1$ if $v > 0$). This argumentation is of course wrong because we have to compare same things with same things, what we did not. Because observing a moving clock (which goes slower than yours) one has to measure the time at two different places (whereas the moving clock in its rest frame stays at the same position). This leads to an asymmetry in measuring the time one a clock.

### 12.3.6 Lorentz transformation

Similarly to the Galileo Transformations we can now calculate the transformation of space and time between moving reference frames. To do this we first have to be aware what we mean by transforming time and space. From classical mechanics we can keep the definition of a (spatial) reference frame as we have defined it in section 12.2.1 but since space and time will mix, we have to develop a new intuition for time\(^{18}\). For this we imagine to have at every point in space an (imaginary) clock and each inertial frame of reference has its own clocks. All the clocks belonging to one system show the same time in their rest frame (see also chapter 12.3.2 about synchronisation of clocks). We can then compare the time shown on the clocks of different frames at the same position. This is necessary because if we compare the time shown on clocks at different positions we would have to take into account

\(^{18}\)In classical mechanics time is absolute, which means that it is a given quantity for all frames (up to a constant shift in time).
the time the light needs from the clock to our eyes. In relativity, this union of space and
time is called spacetime\(^{19}\).

Let \(\Sigma\) and \(\Sigma'\) be two inertial frames of reference with their clocks showing the time \(t\) and \(t'\). In the two frames, a certain point is given as \(\vec{r} = (x, y, z)\) in \(\Sigma\) and \(\vec{r}' = (x', y', z')\) in \(\Sigma'\). For convenience we chose the two systems such that the corresponding axis are parallel to
each other and that \(\Sigma'\) is moving along the x-axis with the speed \(v\). Furthermore we assume
that the two origins at \(t = t' = 0\) coincide. Then the transformation is then given as

\[
\begin{align*}
t' &= \gamma \left( t - \frac{v}{c^2}x \right) \\
x' &= \gamma \left( x - vt \right) \\
y' &= y \\
z' &= z
\end{align*}
\]

where \(\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}\) is the factor introduced above. Before we look at a (rough) proof, we
can rewrite the transformation above in order make the two equations symmetric. If we
associate time \(t\) by the distance \(ct\) and rewrite the equations in terms of \(ct\), we get

\[
\begin{align*}
x' &= \gamma \left( x - \frac{v}{c}ct \right) \quad (12.1) \\
ct' &= \gamma \left( ct - \frac{v}{c}x \right) \quad (12.2)
\end{align*}
\]

To prove this transformations we basically have to repeat the steps we did deducing time
dilatation and Lorentz contraction. Hence there will be rather a sketch of proof instead of
a rigorous proof.

- The transformation of space is in principle the same as in classical mechanics (see
  Galileo transformation in chapter 12.2.3). The only difference is that we have to take
  into account Lorentz contraction which causes the \(\gamma\) factor in equation (12.1).

- For the transformation of time, we first look at the spatial dependence. This transfor-
mation depends only on the coordinate that is in the direction of movement of the two
frames. When we look at the synchronisation of clocks in a moving frame (as described
in chapter 12.3.2), we observe that the time the light pulse needs in one direction is
longer than in the other. We have already observed this in the section about Lorentz
contraction (section 12.3.6). As a consequence we observe for a moving frame that
the two clocks (in the moving frame at different positions) are not synchronous which
literally means they do not show the same time\(^{20}\). The time in the moving frame
therefore depends on the position.

\(^{19}\)Mathematically it is a 4 dimensional (vector) space but with an important difference to the usual 4
dimension (spatial) space: There is a different measure of distances, see chapter 12.4.

\(^{20}\)To avoid any misunderstanding: They show the same time in their rest frame but in our system this
frame is moving and then they do not show the same time any more.
When we look at the time dependence in the time transformation we would expect $t' = \frac{t}{\gamma}$ according to time dilatation (see section 12.3.3). But there we looked at a slightly different case than here: In section 12.3.3 we observed a time interval in the moving frame from our (rest) frame. Here we look at the corresponding clock that passes a certain point $\vec{r}$ at a certain time $t$. Therefore we have to calculate which clock is at the time $t$ at the position $\vec{r}$. It is the clock that was at time $t = 0$ at the position $\vec{r}_0 = \vec{r} - \vec{v}t$ for which we can perform a clock synchronisation at $t = 0$ (denoted by $t'_0$) and then calculate how much time in $\Sigma'$ is passed until this clock reached the point $\vec{r}$. The synchronisation at $t = 0$ leads to $t'_0 = x'_0 c^2 = \gamma x_0 c^2$ and the time passed is then $t' = \frac{t}{\gamma}$.

The time dependence of $\vec{r}_0$ and the time itself lead then to the given dependence in equation 12.2.

Before we finish this chapter about the Lorentz transformation let’s deduce the time dilatation from the Lorentz transformation. For this consider two points in time $t_1$ and $t_2$ in $\Sigma$ and $t'_1$ and $t'_2$ in $\Sigma'$. For the time dilatation we have to observe one clock in $\Sigma'$, which shall be located at $x_1$ at time $t_1$ and $x_2 = x_1 + v(t_2 - t_1)$ at $t_2$, whereas $x'_1 = x'_2$. We then get

$$\Delta t' = t'_2 - t'_1 = \gamma \left( t_2 - \frac{v}{c^2} x_2 \right) - \gamma \left( t_1 - \frac{v}{c^2} x_1 \right)$$
$$= \gamma \left( t_2 - \frac{v}{c^2} (x_1 + v(t_2 - t_1)) \right) - \gamma \left( t_1 - \frac{v}{c^2} x_1 \right)$$
$$= \gamma (t_2 - t_1) - \gamma \frac{v^2}{c^2} (t_2 - t_1)$$
$$= \frac{1 - \frac{v^2}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}} (t_2 - t_1) = \frac{1}{\gamma} \Delta t$$

as we had it in section 12.3.3.

### 12.4 Minkowski metric

In classical mechanics a distance in the 3 dimensional space is conserved by the Galileo transformations. This means a distance between two points has the same length in all inertial frames.

---

21This is because the two systems are equivalent according to the first postulate of Einstein and in $\Sigma'$ the velocity of $\Sigma$ is $-v$ instead of $v$.

22Since the $y$ and $z$-coordinate have no influence on the transformation, we omit them.
Similarly one can define something like a "distance" in the 4-dimensional time space which will be invariant under Lorentz transformations. This section will examine this "distance" and define so called four vectors.

### 12.4.1 Definition of Minkowski metric

Let $\Sigma$ and $\Sigma'$ be two frames of reference. In $\Sigma$ we have two points in space and time $t_1$, $r'_1 = (x_1, y_1, z_1)$ and $t_2$, $r'_2 = (x_2, y_2, z_2)$. These points are often called events because an event happens at a certain time at a certain place. In $\Sigma'$ we look at the same points which are given according to the Lorentz transformation (see section 12.3) and denoted by $t'_1$, $\vec{r}' = (x'_1, y'_1, z'_1)$ and $t'_2$, $\vec{r}'_2 = (x'_2, y'_2, z'_2)$. Then the "distance" in the 4 dimensional time space given by

\[
\Delta s^2 = c^2 \Delta t^2 - \Delta r^2
\]

is in both frames the same ($\Delta s'^2 = \Delta s^2$). This "distance" is called Minkowski metric\(^{23}\) and it has some properties that are similar to the distances we know from 3 dimensions.

To prove that the Minkowski metric is invariant under Lorentz transformations we assume for simplicity that $\Sigma'$ is moving along the x-axis of $\Sigma$ with velocity $v$ and that the corresponding axis of the two systems are parallel to each other. Therefore we can already state that $(y_2 - y_1)^2 + (z_2 - z_1)^2 = (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2$. So we only have to look at time and the x-coordinate. We get

\[
c^2 \Delta t'^2 - \Delta x'^2 = c^2 (t'_2 - t'_1)^2 - (x'_2 - x'_1)^2
\]

\[
= \gamma^2 [((c\Delta t_2 - \beta x_2 - c\Delta t_1 + \beta x_1)^2 - (x_2 - \beta c\Delta t_2 - x_1 + \beta c\Delta t_1)^2]
\]

\[
= \gamma^2 (c^2 \Delta t^2 + \beta^2 \Delta x^2 - 2\beta \gamma \Delta tx - \Delta x^2 + c^2 \beta^2 \Delta t^2 + 2\beta\gamma \Delta t \Delta x)
\]

\[
= \frac{1}{1 - \beta^2} c^2 \Delta t^2 (1 - \beta^2) + \frac{1}{1 - \beta^2} \Delta x^2 (1 - \beta^2) = c^2 \Delta t^2 - \Delta x^2
\]

### 12.4.2 Properties of Minkowski metric

The reason why the Minkowski metric is always denoted as "distance" with quotation marks is that there is one important difference to distances we know from 3 dimensions. This difference is not the dimensionality. The reason is the minus sign between the time and space part, if there would be a plus sign one would also call this a distance. The reason is exactly that minus sign. A distance is something that has always a positive value (or

\(^{23}\)In some books, the Minkowski metric is defined as $\Delta s^2 = \Delta \vec{r}^2 - c^2 \Delta t^2$. Obviously this definition does not change the fact that $\Delta s^2$ is invariant under Lorentz transformations and the properties are also the same (up to a sign)
zero). However using the Minkowski metric, it is also possible to get a negative value if \( c^2 \Delta t^2 < \Delta \vec{r}^2 \). This change of sign allows us to classify the "distance" of two events:

**light cone**

If two events have \( \Delta s^2 = 0 \) we say that they lie on the light cone. This means if at the time \( t_1 \) at \( \vec{r}_1 \) a light pulse would be send towards \( \vec{r}_2 \) it would reach \( \vec{r}_2 \) exactly at \( t_2 \). This property is obvious because \( \Delta s^2 = 0 \) leads to

\[
 c \Delta t = | \Delta \vec{r} |
\]

\[
 c = \frac{| \Delta \vec{r} |}{\Delta t}
\]

where the right side is the mean velocity of an object moving from \( \vec{r}_1 \) to \( \vec{r}_2 \) in the time \( \Delta t \). This velocity is the speed of light \( c \).

**timelike**

In case \( \Delta s^2 > 0 \) the "distance" is called timelike. This means the 3-dimensional distance of the two events is smaller than the time a light pulse needs to travel from the first to the second event. Therefore there exists a frame of reference where the two events happen at the same position. This frame is travelling with the speed \( v = \frac{| \Delta \vec{r} |}{\Delta t} < c \). Furthermore the condition \( \Delta s^2 > 0 \) means that the first event might have caused the second event. This is because if at \( t_1 \) a light pulse would be send out at \( \vec{r}_1 \), it would reach \( \vec{r}_2 \) before \( t_2 \) and might therefore causes an action at \( t_2 \).

**spacelike**

The opposite of timelike is spacelike, where \( \Delta s^2 < 0 \). This means light would have longer between the spatial positions of the two events than \( \Delta t \). Since the speed of light is the highest speed for interaction there is no possibility that the first event could influence the second one. But there is a frame of reference where the two events happen at the same time. Since in this frame \( \Delta t = 0 \) and \( \Delta s^2 \) independent of the frame, \( \Delta \vec{r}^2 \) must be minimal.

### 12.4.3 Four vectors

A compact notation for an event is the so called four vector. A four vector is a four dimensional vector which has as first component time multiplied by the speed of light \( c \) and the other four components are the space components. The four vector to the event at \( t_1 \) and position \( \vec{r}_1 \) is therefore

\[
 r_1 = \begin{pmatrix}
 ct_1 \\
 x_1 \\
 y_1 \\
 z_1
\end{pmatrix}
\]
where we use the convention that a four vector has no arrow above and a spatial (position) vector has an arrow. Usually it is clear from context if a symbol represents a four vector or a scalar. In this script we will not really use four vectors so you do not have to bother about the distinction of four vectors and scalars. The special feature about four vectors is that one can define a "scalar product" for them. We have already seen this "scalar product" and it is nothing else than the Minkowski metric. So the (relativistic) "scalar product" of two four vectors is defined as

\[
\begin{pmatrix}
ct_1 \\
x_1 \\
y_1 \\
z_1
\end{pmatrix} \cdot \begin{pmatrix}
ct_2 \\
x_2 \\
y_2 \\
z_2
\end{pmatrix} = ct_1 ct_2 - (x_1 x_2 + y_1 y_2 + z_1 z_2)
\]

12.4.4 Note on rigorous derivation

To derive relativity more rigorous one would start from the Minkowski metric. Since this derivation would need more math but would not give more intuitive insight into relativity we omit it. Nevertheless a very rough sketch of the derivation shall be given in order to show the importance of the Minkowsky metric and the use of four vectors. From Newtons laws we know that in a inertial frame of reference a free particle moves with a constant velocity. This must be true in all inertial frames because otherwise Einstein’s postulate number one would be violated. From this one can conclude that the transformations (with will lead to the Lorentz transformation) between the two frames must be a linear map. Additionally a light pulse should travel with the speed of light in all inertial frames, therefore we search for a transformation that leaves the Minkowski metric invariant. The task is therefore to find the transformations for a 4 dimensional (vector) space that leaves the Minkowski metric invariant. One then can directly deduce the Lorentz transformation as transformation of four vectors.

12.5 Velocities

In this chapter we will take a closer look at the physical quantity velocity. We will examine how velocities "add" in relativity which will prepare us for the dynamical part of relativity. Additionally we will have a look at the four vector of the velocity. Before we start let’s repeat an important property of the velocity between to frames of reference. Assume you are in an inertial frame and you observe a space ship passing with a constant velocity $v$. From the frame of the space ship you are also moving with the speed $-v$. This means the relative velocity between two frames is something constant and not relative.

---

24 If you cannot follow the sketch do not mind, you will see it in full detail if you study physics.
25 A free particle is a particle on which no force is acting.
26 Because the Minkowski metric fulfils exactly the condition with the light pulse.
12.5. VELOCITIES

12.5.1 Additon of parallel velocities

In $\Sigma$ you are observing another frame $\Sigma'$ which is moving with a constant speed $v$. Assume that with respect to $\Sigma'$ an object is moving in the same direction as $\Sigma'$ with a speed $u'$ (measured in $\Sigma'$). The question is, what is the speed of that object in $\Sigma$? As already mentioned, it cannot be $u = v + u'$ because this might be greater than the speed of light $c$. To calculate the right velocity we have to go back to the definition of the velocity and apply the Lorentz transformation. We denote by $x$ the position of the object, the velocity is then given as

$$u = \frac{dx}{dt} = \frac{\gamma (dx' + vdt')}{\gamma (dt' + \frac{v}{c^2}dx')} = \frac{\frac{dx'}{dt'} + v}{1 + \frac{v}{c^2} \frac{dx'}{dt'}} = \frac{u' + v}{1 + \frac{v}{c^2}}$$

where we reduced the fraction be $dt'$ and used the definition of $u' = \frac{dx'}{dt'}$. The obtained result has some interesting features:

- For small speeds ($v \ll c$ and $u' \ll c$) the denominator is nearly 1 and as a consequence we get the classical result.
- If $v = c$ or $u' = c$ we also get $u = c$.

The obtained result might lead to the conclusion that the relativistic addition of velocities is symmetric in $v$ and $u'$. But this is not the case, and would be obvious if one would look at velocities which are not parallel to each other. The symmetry happens because the $\gamma$ in the numerator and denominator drops.

12.5.2 Addition of perpendicular velocities

Assume nearly the same setup as above (section 12.5.1) with the difference that the object in $\Sigma'$ moves perpendicular to the relative motion between $\Sigma$ and $\Sigma'$. To make a distinction to the calculation above we denote the velocities that are measured perpendicular to $v$ with a subscript $\perp$, therefore $u_\perp$ and $u'_\perp$. Before we calculate the relation between $u_\perp$ and $u'_\perp$ lets state an important fact: Although the object is moving perpendicular to $v$ in the frame $\Sigma'$ it is not moving perpendicular in $\Sigma$ because there it also moves with the speed $v$ in the direction of the relative motion between $\Sigma$ and $\Sigma'$. Of course it also has a perpendicular component which is exactly the one denoted by $u_\perp$. This means $u_\perp$ is not the speed of the object in $\Sigma$, the speed would be given as $\sqrt{v^2 + u_\perp^2}$. Now let's calculate $u_\perp$ by following the same steps as above but using that there is no Lorentz contraction perpendicular to the direction of motion, so $x_\perp = x'_\perp$.

---

27 This might be written a bit theoretically. So assume the following example: You observe a space ship ($\Sigma'$) which moves with $v$. In that space ship some one throws a ball with speed $u'$ in the same direction as the space ship is moving.

28 Actually one need multidimensional analysis but since the Lorentz transformation is a linear map, the derivative "behaves" like the Lorentz transformation itself.
\[
\frac{d}{dt} = \frac{dx_\perp}{\gamma (dt' + \frac{v}{c^2} dx'_\parallel)} = \frac{u'_\perp}{\gamma (1 + \frac{v}{c^2} u'_\parallel)} = \frac{u'_\perp}{\gamma}
\]

where we used that the parallel component of the velocity if the object in \(\Sigma'\) is zero\(^{29}\). In case of \(u_\perp\) it is obvious that \(v\) and \(u'_\perp\) do not enter symmetrically the formula (\(v\) enters the formula through \(\gamma\)).

### 12.5.3 Velocity Four Vector

Similarly to the four vector defined in section 12.4.3 one can define four vectors for velocities. Let \(r = (ct, x, y, z)\) be the four vector of an object moving with a constant speed. This means at the time \(t\) that object is at the position \(\vec{r} = (x, y, z)\). A first (wrong) approach to a four vector for the velocity \(u\) might be

\[
u = \frac{d}{dt} \begin{pmatrix}
c t \\
x \\
y \\
z
\end{pmatrix} = \begin{pmatrix}
c \\
v_x \\
v_y \\
v_z
\end{pmatrix}
\]

where \(v_x\) is the \(x\) component of the velocity and similarly for \(y\) and \(z\). But this approach contradicts the idea of the absolute value of the four vector being invariant under Lorentz transformations. This is because the scalar product of \(u\) with itself is given as

\[u \cdot u = u^2 = c^2 - v_x^2 - v_y^2 - v_z^2.\]

Depending on the frame the velocities are different so \(|u|^2\) is not invariant.

To get the right result by an intuitive approach we have to go back at the definition of the derivative. A derivative is basically the fraction of two quantities \(\frac{\Delta x}{\Delta t}\) with the limes \(\Delta t\) going towards zero (and \(\Delta x\) too). If the numerator is a four vector (as we assume here) and the denominator is a quantity that is constant in all frames, then the whole fraction is again a four vector that has an absolute value independent of the frame. There is one special time, namely the proper time \(\tau\) which is independent of all frames. This is the time in the rest frame of the moving object. We obtain \(\Delta \tau\) by the time dilatation, which leads to \(\Delta \tau = \frac{1}{\gamma} \Delta t\). Taking the limit for the derivative we have to replace the \(\Delta\) by the \(d\) (roughly speaking) and we get

\[
u = \frac{d}{\frac{1}{\gamma} dt} \begin{pmatrix}
c t \\
x \\
y \\
z
\end{pmatrix} = \gamma \begin{pmatrix}
c \\
v_x \\
v_y \\
v_z
\end{pmatrix}.
\]

\(^{29}\)If the object would move in an arbitrary direction (not perpendicular) we would keep \(u'_\parallel\).
If our construction of the four vector was successful, \( u \cdot u \) should be invariant under Lorentz transformations. Let’s check this by considering \( u \cdot u \) in the rest frame of the object and in an arbitrary frame. In the rest frame the object does not move, therefore \( v_x = v_y = v_z = 0 \) and therefore \( u \cdot u = c^2 \). In an arbitrary frame it is moving and we get

\[
    u \cdot u = u^2 = \gamma^2 (c^2 - v_x^2 - v_y^2 - v_z^2)
    = \frac{1}{1 - \frac{v^2}{c^2}} c^2 (1 - v^2) = c^2
\]

where we used \( v_x^2 + v_y^2 + v_z^2 = v^2 \). Obviously this is invariant under Lorentz transformations and we succeeded in constructing a four vector for the velocity.

### 12.6 Dynamics

Until now we only did relativistic kinematics, this means we developed a formalism do describe the motion of an object and how this description changes by a Lorentz transformation. Now we will look at dynamics. A proper description of dynamics in special relativity as we know in classical mechanics (Newton’s laws) is more difficult. Nevertheless we can define quantities as momentum or energy which will be conserved and which allow to calculate a lot of examples (using these conservation laws). To derive this we start with the 3-dimensional momentum and then search for the four vector of the momentum. From the four vector of momentum one can derive energy considerations including Einstein’s famous formula \( E = mc^2 \). At the end of this section we will briefly look at acceleration and forces.

#### 12.6.1 Momentum

Looking at the following example one can see that the momentum cannot simply be \( \vec{p} = m \vec{v} \) where \( m \) is the (rest) mass of the object: We observe a space ship that moves with velocity \( v \) along the \( x \)-axis. In the space ship two balls with speed \( v \) are moving perpendicular to the \( x \)-axis towards each other (see figure 12.5). When they meet the bounce away such that they move in the \( x \) direction. From energy and momentum conservation we can deduce that they move also with velocity \( v \) along the \( x \)-axis in the frame of the space ship. From our system (where the space ship is moving), the total momentum of the two balls is \( 2p \) where \( p \) is the momentum in the \( x \)-direction of each ball (since the balls are moving towards each other, the total momentum perpendicular to the \( x \)-axis is zero). After the collision the total momentum must be conserved. But the ball moving to the left does not move in our frame because in the space ship it moves with \( -v \) and the space ship itself moves with \( v \), so the overall movement of this ball is zero. The second ball moving to the right has according to the relativistic addition of velocities a speed of

\[
    v_2 = \frac{v + v}{1 + \frac{v^2}{c^2}} = \frac{2v}{1 + \frac{v^2}{c^2}}
\]

If the momentum would be \( p = mv \) this leads to a contradiction since
$p_{\text{before}} = 2mv \neq \frac{2v}{1 + \frac{v^2}{c^2}} m = p_{\text{after}}.$

Figure 12.5: Collision and scattering of two balls in the moving space ship. Initially, the two balls move with the velocity $v$ perpendicular to the direction of flight of the space ship. After the scattering they move in the same direction as the space ship.

Consequently the momentum cannot depend linear from the velocity. The right dependence is given as

$$p = \gamma mv = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} mv.$$ 

By a longer calculation one can show that the momentum defined this way fulfils the conservation of momentum. Since there is an other method showing that this is a good candidate for the momentum using four vectors we will look at four vectors.

One important note about the mass $m$: Sometimes a relativistic mass $m_{\text{rel}} = \gamma m$ is defined. This is an unsuitable definition (see section 12.6.4), when we talk about masses we always mean the (rest) mass, so the mass measured in the rest frame of the mass.

12.6.2 momentum four vector

Since $\vec{p} = m\vec{v}$ is wrong, the next approach might be taking the last three components of the four vector of velocity $v$ and multiply them with the mass. And indeed this leads to the right result. Encouraged from this success we define the four vector for the momentum as

$$p = mv = \gamma m \begin{pmatrix} c \\ vx \\ vy \\ vz \end{pmatrix}.$$
As the absolute value of the velocity four vector is invariant under Lorentz transformations and the mass is a constant, the absolute value of the momentum four vector is also constant with $p^2 = m^2c^2$. What we don’t understand yet is the first component of the momentum four vector, namely $\gamma mc$. This will clarify in the next section:

### 12.6.3 Energy

The (total) energy of an object can be calculated by

$$E = \int F \, ds = \int \frac{dp}{dt} \, ds \sim m\gamma c^2$$

where many steps were left out including some longer calculation. Looking at the momentum four vector, we see that the first component is exactly the energy divided by $c$. This means we can rewrite the four vector as

$$p = mv = \gamma m \begin{pmatrix} c \\ v_x \\ v_y \\ v_z \end{pmatrix} = \gamma \begin{pmatrix} \frac{E}{c} \\ p_x \\ p_y \\ p_z \end{pmatrix}.$$ 

This discovery leads to some interesting conclusions:

#### Rest Energy

In the rest frame of an object one has $\vec{p} = 0$ and $\gamma = 1$. Since $p^2 = m^2c^2$ we get for the rest frame

$$\frac{E^2}{c^2} = m^2c^2$$

$$E = mc^2$$

which is exactly Einstein’s famous mass-energy equivalence. This means the mass itself is energy and the minimal possible energy of an object is $E = mc^2$ the rest mass (times $c^2$) of that object.
Energy-momentum relation

For an arbitrary system the absolute value square of $p$ multiplied with $c^2$ leads to

\[ p^2 c^2 = m^2 c^4 = E^2 - \vec{p}^2 c^2 \]

\[ E^2 = \vec{p}^2 c^2 + m^2 c^4. \]  \hspace{1cm} (12.7)

This equation relates the momentum of an object with its energy. Of course this energy is the same as the one obtained above in equation (12.6) as the following calculation shows:

\[ E = \sqrt{\vec{p}^2 c^2 + m^2 c^4} = mc^2 \sqrt{\gamma^2 - 1} + 1 = mc^2 \sqrt{\frac{\beta^2 + 1 - \beta^2}{1 - \beta^2}}. \]

small speed limit

If we look at small velocities the Energy as written in equation (12.6) can be approximated by a Taylor expansion around $v = 0$ an one obtains

\[ E = \gamma mc^2 \approx \left( 1 + \frac{1}{2} \frac{d\gamma}{dv^2} \bigg|_{v^2=0} v^2 \right) mc^2 \]

\[ = \left( 1 + \frac{1}{2} \frac{v^2}{c^2} \right) mc^2 = mc^2 + \frac{1}{2} mv^2. \]

The first term is the rest energy, which is always there but the second term is interesting, since this is exactly the kinetic energy we know from classical mechanics. So in the limit for small velocities, the kinetic energy of an object is exactly the one from classical mechanics.

12.6.4 Acceleration and forces

According to Newton’s law of motion the force acting on a body is equal to the rate of change of the momentum $\vec{F} = \frac{d\vec{p}}{dt}$. Since $\vec{p} = m\gamma \vec{v}$ and $m$ constant, the derivative is only acting on $\gamma \vec{v}$. There are two simple cases one has to consider:

Radial acceleration

If the Force acts perpendicular to the direction of motion, as it is the case at circular motion, the absolute value of $\vec{v}$ does not change. Therefore $\gamma$ does not change and we get

\[ \vec{F} = \gamma m \frac{d\vec{v}}{dt} \]
Linear acceleration

If we accelerate in the direction of motion we also have to take the derivative of \( \gamma \) which is

\[
\frac{d\gamma}{dt} = \frac{\gamma^3 \vec{v} \vec{a}}{c^2 dt}.
\]

This leads then to

\[
\vec{F} = m \left( \frac{\gamma^3 v^2 \vec{v}}{c^2 dt} + \frac{\gamma \vec{v} \vec{a}}{dt} \right) = m\gamma^3 \frac{d\vec{v}}{dt}.
\]

Since these two cases which scale different with \( \gamma \) an arbitrary force is not parallel to the acceleration any more! That’s why it makes no sense to talk about a relativistic mass because a relativistic mass should be defined as proportionality constant between \( \vec{F} \) and \( \vec{a} \). But since these two quantities are not proportional any more (except in the two cases above), one cannot talk about a relativistic mass. So the only meaningful mass is the rest mass \( m \), as we have defined above (see section 12.6.1).

12.7 Paradoxes

In this section we will look at some paradox which might appear when mixing classical mechanics and relativity. Some can be resolved using special relativity, some need general relativity and we won’t be able to totally resolve them.

12.7.1 Ladder and barn

A farmer has a ladder and a barn where the ladder is slightly longer than the barn. The farmer wants to put the ladder in the barn but since the barn is shorter this does not work. His idea is to put the ladder on his high speed tractor and then moves with \( \approx 0.25c \) in the barn. Due to length contraction the ladder should be smaller and therefore fit into the barn (see figure 12.6. The son of the farmer, who read this script and therefore understood relativity very well, redoes the calculation and points out that in the frame of the moving ladder, the barn seems shorter and therefore there is absolutely no way how the ladder would fit into the barn. Father and son decide to risk the experiment, who is right?

The question "fit into the barn or not" is equivalent to the question weather the head of the ladder passes the second door before the back of the ladder passes the first door or not. This is therefore a question about simultaneousness, which is of course relative. Or to be more precise: the frame of the tractor and the frame of the barn use different clocks and the question when which part of the ladder passes which door depends on the time indicated on the clocks. So the father on the tractor effectively measures \( 30 \) that the ladder does not fit into the barn whereas the son standing outside shortly sees the ladder disappear in the barn (assuming the father drives with constant speed through the barn).

\[30\] Measure instead of see because to see something, the light needs some time to pass the corresponding distance
Since the son was a curious guy he suggests the following modification of the experiment:
He installs a photoelectric sensor at each door of the barn and puts a lamp in the middle of the barn. If the front of the ladder passes the photoelectric sensor at the second door before the end of the ladder passes the photoelectric sensor at the first door, the lamp gets on. This experiment should allow to take a decision whether the ladder fits into the barn or not. What will they observe?
They indeed observe that the light goes on. In the frame of the son (standing next to the barn) this is obvious since the ladder gets smaller as it moves. In the frame of the father on the tractor this is not that obvious and we have to consider how the signal from the photoelectric sensor gets to the lamp. Assume there is a cable between the photoelectric sensor and the lamp where the signal moves with the speed of light $c$ and that there are no effects that delay the signal. Since the speed of light is a natural constant, both the father and the son see it move with $c$. But in the frame of the father the lamp is moving with velocity $v$. Therefore the overall\footnote{This is the addition of two relative (and parallel) velocities measured in the same frame and not the relativistic velocity addition discussed in section 12.5.1.} effect of the second door moving with velocity $v$ and the signal moving with $c$ is the same as $c - v$. This is opposite at the first door, where the farmer sees an overall velocity of $c + v$. Although the distance between the lamp and each door is the same, the signal from the second door reaches the lamp first and as a consequence it shines.
In this paradox we see very nicely that there are examples in relativity which seem contradictory but at a closer look there is no mistake in the theory.
12.7.2 Twin paradox

The probably most famous paradox is the twin paradox: One of the two twins makes a space trip in a very fast space ship whereas the other twin stays at home. After several years the first twin returns and meets the other twin. The one who made the trip is much younger than the one stayed at home.

Not knowing relativity this is astonishing because time is usually considered as absolute which should not be influenced by movements. Knowing relativity this might be astonishing because each of the twins see the other twin move and therefore the other twin should stay younger. The point in this paradox is that one of the twins has to accelerate in order to return to the other twin. This acceleration breaks the symmetry of the problem and leads to the fact that one twin stays younger than the other.

The simplest way to calculate the paradox is to assume that the twin which makes his trip moves with constant velocity \( v \) until he changes to another space ship which moves with \( -v \) back to earth. The clocks of the two space ships are set such that they have the same time at the moment they meet and at the position they meet.

12.7.3 Solution to the flying electron problem

We now want to go back to the problem from the beginning, namely the flying electron (see section 12.1.2). For simplicity we assume that the electrons in the wire move with the same speed as the electron above the wire\(^{32}\). Furthermore we assume that in the lab frame (where the electron is moving) the wire is neutral, this means the distance between two neighbour positive atoms and the distance between two neighbour negative electrons is the same. Let’s denote this distance with \( L \).

First of all we have to state that from the point of the lab we see the distance of two neighbour moving electrons in the wire with a Lorentz contraction. Therefore their distance (in the direction of flight) in their own frame is \( L'_e = L \gamma \) where the \( \gamma = \frac{1}{\sqrt{1-\beta^2}} \) and \( \beta = \frac{v}{c} \) the velocity in terms of \( c \). If we now look at the frame of the moving electron, the electrons in the wire are not moving any more (by our assumption above), their distance is therefore \( L' \). Additionally the positive atoms are moving which leads according to the Lorentz contraction to a distance between neighbour atoms of \( L'_p = \frac{1}{\gamma}L' \). As a consequence the density of the positive atoms is bigger than the one of the negative electrons, the wire is therefore electrically positive charged. This positive charge leads to an attraction of the single electron, similarly we observe in the lab frame.

This example show how important relativity is in electrodynamics. Without relativity electrodynamics would not be a consistent theory, meaning that different observers would observe different result of one experiment. Nevertheless one needs a bit more math to precisely formulate electrodynamics with relativity.

\(^{32}\)One can do the whole calculation without this assumption which leads to the same result but needs a longer calculation respectively is less intuitive.
Chapter 13
Quantum Mechanics

Classical mechanics describes how particles or bodies behave in time. So it characterises in which way a particle moves. In Newtonian mechanics the main formula that characterises the way particles move is $\vec{F} = m\vec{a}$. One then assumes that the particle moves according the the formula above on a well defined path. At the end of the 19th century some phenomena did not fit to the predictions of classical mechanics. A new description was developed which is called quantum mechanics.

Quantum mechanics describes also how a particle behaves in time but the laws are very different of those of classical mechanics and intuitively not really understandable. The main idea is that a particle does not move on a well defined path but the probability to find it at a certain location is connected to wave properties. This chapter shows the phenomena which did not fit to classical mechanics and what changed in the description switching to quantum mechanics. Then some concepts of quantum mechanics are discussed. Since the general description of quantum mechanics is pretty complicated the concepts are discussed in a more intuitive way. At the end we will look how a rigorous calculation in quantum mechanics looks like. This rigorous calculation is surly not necessarily to know for the Olympiads (not even for the IPhO) but it should show how quantum mechanics really looks like. The main sources for this chapter are [?], [47], [48] and [49].

13.1 Experiments

We want to look at some experimental setups and discuss how quantum mechanics changed the understanding of physics and what the main statements are.

13.1.1 Black body radiation

A body with a certain temperature $T$ emits electromagnetic waves, called radiation. Depending on the surface, the body can emit electromagnetic waves for certain frequencies better than for others. If there are two bodies and one body is hotter than the other, energy can only go from the hotter to the colder body. Therefore for any frequency the capability of a body to absorb an electromagnetic wave is the same as the capability to emit a wave. A black body is a body which absorbs all electromagnetic waves. As a consequence it also emits waves very well. If a body absorbs an electromagnetic wave it gets heated up by the
energy of the wave. The emission of an electromagnetic wave is also due to the thermal energy of a body.

A surface which has the property of a black body can be achieved if one takes a closed cavity with a little hole (see figure 13.1). The area where the hole is behaves like a black body because at the spot where the hole is all the incident waves will go into the cavity and are therefore absorbed by the body.

![Figure 13.1: A cavity with a little hole](image)

Inside the body there are some other interesting properties: Since all the walls of the cavity have the same temperature there is no net flow of energy from one wall to the other. Therefore the electromagnetic waves inside the cavity have an intensity such that the emission and absorption at the walls is in equilibrium. We assume the hole to be very small, therefore at the hole the radiation is the same as inside the cavity. Since the hole behaves like a black body the electromagnetic waves inside the cavity look like they were emitted by a black body, although the walls of the cavity need not to be black. This is because the electromagnetic field and the walls are in equilibrium.

At the end of the 19th and beginning of the 20th century there was a big discussion about the question how the power that a black body radiates is distributed as a function of the frequency $f$. This means how much power is radiated between a small interval $f$ and $f + df$. The classical approach is to model the electric field inside the cavity as standing waves between opposite walls. According to thermodynamics any possible standing wave has the same amount of energy (equipartition theorem). But for higher frequencies also diagonal standing waves can happen therefore the number of standing waves per frequency is not fix. Since usually the dimension of such a cavity is in the order centimetre and the wavelength of light is in the order of 500 nanometres the discrete difference between two successive frequencies is very small and we treat the spectrum continuous. A rigorous calculation then gives that the number of standing waves between $f$ and $f + df$ is

$$g(f)df = \frac{4\pi f^2 V}{c^3}df$$
where $V$ is the volume of the cavity. We see that the number of standing waves of a certain frequency is proportional to the square of the frequency (for high frequencies).

The equipartition theorem from thermodynamics states that each standing wave has the mean energy $k_B T$ where $k_B$ is the Boltzmann constant and $T$ the temperature. Therefore we have an energy density (Energy per volume) inside the cavity of

$$\rho(f) df = k_B T \frac{8\pi f^2}{c^3} df$$

This law is called Rayleigh-Jeans law (see picture 13.2). The additional factor 2 appears because there are two possibilities of polarisation. The big problem is that the energy density for high frequencies goes to infinity. Therefore in the cavity should be infinite energy which is of course not possible.

In 1900, Max Planck introduces an other idea: He said that the energy of light (of a certain frequency) does not have an arbitrary amount. Instead he claimed that light at a certain frequency has a smallest amount of energy which is given by $E_0 = hf$ where $h$ is the Planck constant with $h = 6.626 \cdot 10^{-34}$ J·s. He did not believe that his assumption describes a real property of nature, he just tried to do the calculation a bit different. The calculation lead to the formula

$$\rho(f) df = \frac{8\pi hf^3}{c^3} \frac{1}{e^{\frac{hf}{k_B T}} - 1}$$

The additional exponential in the denominator causes the curve to get small for high frequencies. The curve is drawn for different temperatures in figure 13.2. The measurements confirm Planck's calculation.

Since for small frequencies: $e^{\frac{hf}{k_B T}} - 1 \approx 1 + \frac{hf}{k_B T} - 1 = \frac{hf}{k_B T}$ the Rayleigh-Jeans law is therefore an approximation for small frequencies.

This was then the beginning of a new theory, the quantum theory. The introduction of the smallest energy $E_0$ quantises the energy of an electromagnetic field. Therefore electromagnetic waves behave like waves but have also quantised property which suits more to the model of particles than waves. This particle property gets clearer in the next chapter.

### 13.1.2 Photoelectric effect

If one shines light on a charged metallic plate one can observe that the plate loses charge with time. To examine this effect let’s look at the experimental setup shown in figure 13.3.

We shine light with a frequency $f$ on the left metallic plate. We will see that this plate loses electrons therefore we call it cathode. Parallel to this plate we have an other plate where no light shines on, we call it anode. Between these two plates we connect a voltage source with variable voltage $U$. We define $U > 0$ if the cathode is connected to the plus pole of the
voltage source (so electrons flow from the cathode to the voltage source) and $U < 0$ if the cathode is connected to the minus pole. Additionally we put an ammeter on the electric circuit in order to measure the current $I$ that flows.

**Dependence of voltage $U$**

We see that for all voltages that are smaller than a certain voltage $U_t$ (also called threshold voltage) there is a current flowing from the anode to the cathode. Since metallic atoms cannot really leave the metallic plate we conclude that the electrons move. Since the electrons are negative charged and the current flows from the anode to the cathode, the electrons move from the cathode to the anode. The light seems to hit out electrons at the cathode and they travel to the anode. The existence of $U_t$ (for a certain intensity) is reasonable because for higher voltages the electrons need more energy to fly against the electric field between the anode and cathode.

**Dependence of intensity**

If the voltage is smaller than the threshold voltage $U < U_t$ we recognise that the current is proportional to the intensity. This seems also reasonable because if we shine with more light on the cathode there will be more electrons hitting out and reaching the anode (until saturation effects occurs).

For a certain intensity the existence of a maximal voltage $U_t$ is reasonable because if we raise the voltage the electrons need more energy to overcome the electric field. The remarkable

---

**Figure 13.2**: Radiated intensity for different temperatures. The x-axis is the wavelength and it is obvious that for high frequencies (small wavelength) the intensity according to Planck’s calculation gets small. For the classical approach the intensity goes towards infinity [51].
thing that the experiment shows is, that $U_t$ is independent of the intensity. This means that an electron does not collect energy from the light until it can leave the cathode and reach the anode. Instead it gets once a certain amount of energy and if the energy is big enough to leave the cathode and reach the anode the electron will do so. Therefore the energy from the light is concentrated in small packets. We call this energy packets photons.

**Dependence of frequency**

If we change the frequency of the light we see that the threshold voltage also changes. If we rise the frequency the threshold voltage also rises. The dependence seems to be linear function which is drawn in figure 13.4. Therefore the energy of a photon is proportional to the frequency. The y-intercept can be interpreted as the work that has to be applied to the electron in order to extract it from the metallic plate. The work to leave the plate is therefore $V_0 e$ where $e$ is the charge of an electron.

**Interpretation**

As we already stated, the energy of the light does not flow continuously but in small packets. We can read out the amount of Energy per frequency of figure 13.4 by taking the slope of the linear function. If we do this, we get that the slope is $h$, the Planck’s constant. Therefore the energy of such an energy packet is given by $E = hf$ as we already assumed in chapter 13.1.1.

The intensity of the light is only a measure of how many Photons arrive at the cathode per second.

Light is therefore on one hand a wave because we have wave phenomena like refraction and diffraction (see chapter 13.1.3). On the other hand it has particle-like properties because its
Figure 13.4: Dependence of the cut-off potential $U_t$ on the frequency $f$ [53].

energy (and also its momentum) travels in small packets. This shows the very unintuitive nature of quantum mechanics.

### 13.1.3 Double slit experiment

The double slit experiment is an experiment which points out the wave property of quantum objects. This experiment (see figure 13.5) consists of a light source which emits light with a wave length $\lambda$ and a wall which absorbs light except at two small slits (small compared to the wavelength) which are separated by a distance similar to $\lambda$. The light can go through these to slits and is measured on a screen.

Let’s now discuss what we measure on the screen and what happens if we change the setup slightly:

**Normal setup, no changing**

The light behaves as a wave and gets diffracted at the two slits. This means that we have to treat each slit as a source of new spherical waves $A_1$ and $A_2$ and then for each point on the screen add the amplitudes from each slit $A = A_1 + A_2$. The measured pattern is then proportional to $A^2$ because we measure the intensity. Therefore we get an interference pattern (the right pattern in figure 13.5).

**Closing one slit**

If we close one slit the light goes only through the other slit. There a spherical wave is caused which hits the screen making a pattern like $I_1$ (or $I_2$ if the first slit is closed).

**Measuring the path**

If we have both slits open we get a diffraction pattern as described above. Thinking the light to be a flow of photons which (should) behave like particles we might be interested in
measuring through which slit the individual photons pass. Therefore we install a detector which measures where the particles come from. As soon as the detector is installed and measures, the light behaves differently: The diffraction pattern disappears and we get a pattern as if the particle would fly through one or the other slit, therefore we get the left pattern in figure 13.5. This (process) is called collapse of the wave (function) because the wave property is disappeared. We state that the measurement influences the outcome of the experiment, which is typically for quantum mechanics.

**Very low intensity**

If we reduce the intensity until only one photon is emitted at a time we will also get the interference pattern if we wait long enough (see figure 13.6). Therefore each single photon interferes with itself.

The pattern is then no intensity distribution any more because single photons produce only small dots. The light dots show that light hits the screen as a small packet (in fact it is not possible to see a photon hitting an usual screen. But if one replaces the screen by a very sensitive detector it is possible to measure single photons). Therefore the pattern is the sum of all these photons and it is proportional to the probability that a photon interferes to that position.

We can also switch on the detector to measure through which slit a photon passes. If we do so we measure again a probability distribution like the left pattern in figure 13.5.
CHAPTER 13. QUANTUM MECHANICS

Figure 13.6: The light shines with very low intensity such that single photons reach the screen. On the left side the beginning of the experiment is shown where not many photons hit the screen. On the right side many photons hit the screen and an interference pattern is visible. [55].

Interpretation

Since the behaviour of the experiment does not change if we perform it with single photons we have to describe the observations by behaviour of single photons (and not by the interaction of many photons). We relate the fact that a photon is a light packet to a particle property. Therefore a photon is a particle which travels and if it hits the screen we see a dot like we would expect if a ball hits a wall. The interference pattern however is clearly caused by a wave property.

If we combine these two properties we get that light is a flow of particles and the probability to find a particle is connected to a wave like probability distribution. If we describe the light this way we can also explain intuitively the collapse of the wave when we perform the measurement of the path. Because in that moment we see the photon going through one slit the probability to find that photon in the other slit is zero. Therefore the photon starts its wave like behaviour from the slit we observe it and therefore it gets an other probability distribution.

13.2 Laws of Quantum Mechanics

As we have seen in the experiments above quantum systems behave pretty different to what we know from classical mechanics. We now want to examine some basic properties of quantum mechanics. Since the general description is too complicated for this level we will not be able to derive all laws and understand all connections.

13.2.1 Wavefunction and probability

As we have seen in the double slits experiment the probability $P$ is proportional to the square of the superposed light waves $P \propto (E_1 + E_2)^2$ where $E_1$ is the electric field coming from the first slit and $E_2$ from the second slit. Therefore the interference is due to the fact
that one takes the sum and then squares and not the other way round \((E_1 + E_2)^2 \neq E_1^2 + E_2^2\).

If we describe quantum systems generally we have a function \(\Psi(x)\) which is called wavefunction and which assigns to each place \(x\) a value \(\Psi(x)\). The probability \(P\) to find a particle at a certain place \(x\) is then given by \(P = |\Psi(x)|^2\). The absolute value is necessary if the function \(\Psi\) is a complex number.

The probability \(P\) is a probability distribution. This means that the probability to find the particle at exact the position \(x_0\) is zero for every \(x_0\). It’s like shooting a football on a target: It is impossible to hit the target exactly in the middle and if it seems one has hit it in the middle you have to look more precise (perform a more exact measurement) and you will find out that it was not the middle. Therefore the probability has to be understood a bit different: The probability \(P_{12}\) to find a particle between \(x_1\) and \(x_2\) is

\[
P_{12} = \int_{x_1}^{x_2} P(x)dx = \int_{x_1}^{x_2} |\Psi(x)|^2 dx
\]

Therefore one can understand a probability distribution (of the place) as \(Pdx\) is the probability to find a particle between \(x\) and \(x + dx\).

### 13.2.2 Measurement

If we perform a measurement the wave behaviour (of the measured quantity) of a quantum mechanic particle disappears and we get a concrete value (with a certain uncertainty, see chapter 13.2.4). If our wavefunction \(\Psi(x)\) is related to the probability to find the particle at a certain position \(x\) (to be more precise to find it between \(x\) and \(x + dx\)) and we measure the position we get a value of the position of the particle according to the probability distribution.

But if \(\Psi(x)\) is the wavefunction to find the particle at a certain position and we measure the momentum of the particle we will not get a momentum distribution according to \(|\Psi(x)|^2\). To get the probability distribution \(|\Phi(p)|^2\) for the momentum \(p\) with wavefunction \(\Phi(P)\) we have to make some more calculations which we do not treat here. But we recognise that performing a measurement of the momentum knowing the position distribution is similar to transforming the position distribution to the momentum distribution. Therefore measurements are associated to operators: The measurement of the momentum is like an operator \(\hat{p}\) which acts as \(\hat{p}(\Psi(x)) = \Phi(p)\Psi_p(x)\) where \(\Psi_p(x)\) is the wavefunction with \(|\Psi_p(x)|^2\) describes the probability distribution to find the particle with momentum \(p\) at the position \(x\) (for any momentum \(p\) we get a certain distribution \(\Psi_p(x)\)). If we measure the momentum we measure a concrete value \(p_0\) with wavefunction \(\Psi_{p_0}(x)\). If \(\Psi_{p_0}(x) \neq \Psi(x)\) we get an new wavefunction. This means that the measurement might changes the behaviour of the quantum system.

### 13.2.3 De Broglie hypothesis

Until now we only looked at the photon as quantum object. In 1924 De Broglie stated that any object is a quantum object which behaves according to quantum laws. Since quantum
physics works with waves we have to contribute a wavelength to each object. De Broglie stated that the wavelength $\lambda$ of any object is given by

$$\lambda = \frac{h}{p}$$

(13.1)

where $p$ is the momentum of the object and $h = 6.626 \cdot 10^{-34}\text{J}\cdot\text{s}$ the Planck constant. Therefore big objects with a high mass have a very small wavelength. This is also the reason why we do not experience quantum effects in every day life. Because to get the typical properties of a wave the wave has to hit structures which have a similar size as the wavelength (for example diffraction is nearly not observable if slits are very far away from each other). Since the wavelength of objects around us (like a football) is very small ($\approx 10^{-34}\text{m}$) there exist no slits or something similar with such a small distance (Atoms have a diameter of $\approx 10^{-10}\text{m}$). Therefore no interference phenomena occur with every day objects.

But one can take for example electrons or atoms and perform similar experiments as the double slits experiment (see chapter 13.1.3) and one get also the wave behaviour of those particles.

A comment to the photon. From relativity there is the equation

$$E^2 = c^2 p^2 + c^4 m^2$$

with $E$ the energy, $p$ the momentum, $c$ the speed of light and $m$ the mass (to be more precisely the rest mass). Since the photon has no mass, one gets that $E = pc$. If we put this in equation (13.1) we get $\lambda = \frac{h}{E}$ and using $\lambda f = c$ we have $E = hf$ as we stated at the black body radiation and the photoelectric effect (chapter 13.1.1 and 13.1.2).

### 13.2.4 Uncertainty Principle

If we want to measure the position of a quantum object we might use light. A limitation of the measuring precision is the angular resolution: To distinguish two objects which are separated by a distance $l$ we need light with a wavelength $\lambda < l$. Otherwise the interference patterns of the two objects in our eye (or on the film in a camera) are too close to each other to distinguish them. If we want to measure the position of a quantum object very precisely we need light with very low wavelength. But light with very low wavelength has a very high momentum and if the light of our measurement interacts with the quantum object, the quantum object might gain a lot of momentum. Therefore we know its position very well but we know nothing about its momentum.

Heisenberg stated that the uncertainty of the position $\sigma_x$ and the uncertainty of the momentum $\sigma_p$ are not independent. There is a natural bound which avoids that we can measure the position and the momentum very precisely. The Uncertainty Principle is

$$\sigma_x \cdot \sigma_p \geq \frac{h}{4\pi}$$
where \( h \) is again the Planck constant.
This inequality does not only hold if we measure with light, it is a fundamental uncertainty and independent of the measuring method.
There is also an uncertainty of the time and energy which is given by

\[
\sigma_{E} \cdot \sigma_{t} \geq \frac{h}{4\pi}
\]

where \( \sigma_{E} \) is the uncertainty about the energy and \( \sigma_{t} \) about time.

The uncertainty principle is an interpretation of the diffraction of a quantum object at a very small slit: The smaller the slit is the more precise we know where the particle is when it crosses the slit. But a small slit produces a wide diffraction pattern, therefore the smaller the slit is the less we know about the momentum perpendicular to the slit.

### 13.2.5 Schrödinger Equation

Until now we never discussed how a quantum system changes with time. This has its reason because the time evolution is described by the Schrödinger equation which is a pretty complicated differential equation. Therefore this chapter is more to give a complete overview over quantum mechanics and it is absolutely not relevant to know or understand the equations.

Let \( \Psi(x, t) \) be the wavefunction of a quantum particle along the x-axis which also depends on the time \( t \). The (time dependent) Schrödinger equation is given by

\[
i \frac{\hbar}{2\pi} \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{8\pi^2 m} \frac{\partial^2 \Psi}{\partial x^2} + V(x)\Psi
\]

(13.2)

where \( i \) is the imaginary unit, \( m \) the mass of the particle and \( V(x) \) the potential energy.

The \( \frac{\partial \Psi}{\partial t} \) is the derivative with respect to the time. The reason why we used \( \partial \) instead of \( d \) is because \( \Psi \) depends on the time \( t \) and the position \( x \) and by using \( \partial \) we indicate that we only differentiate according to time. \( \frac{\partial^2 \Psi}{\partial x^2} \) is the second derivative with respect to \( x \) and it is related to the kinetic energy of the particle.

The differential operator \( i \frac{\hbar}{2\pi} \frac{\partial \Psi}{\partial t} \) is the operator which is related to the energy distribution. If we have a constant energy \( E \) the energy operator gives us just the energy, therefore \( i \frac{\hbar}{2\pi} \frac{\partial \Psi}{\partial t} = E \Psi \). This leads to the time independent Schrödinger equation

\[
E \Psi = -\frac{\hbar^2}{8\pi^2 m} \frac{\partial^2 \Psi}{\partial x^2} + V(x)\Psi
\]

(13.3)

In a problem one often searches a function \( \Psi \) which solves the equation. As you can imagine this is pretty complicated.
13.3 Examples

We now want to look at some examples and apply the principles of quantum physics.

13.3.1 Bohr model

One big application of quantum physics is the description of atoms. The simplest atom is the hydrogen atom with one proton as nucleus in the middle and one electron "flying" around the proton. Since the precise calculation of the hydrogen atom is pretty laborious and needs much math we will derive the Bohr model. The Bohr model is a semi classical description of the hydrogen atom which is not really correct but gives some nice predictions. The Bohr model assumes the following properties:

- The electron orbits around the nucleus. Since an accelerated charge radiates energy, a classical orbit, as the motion of planets around the sun, is not possible. Instead we assume that the electron behaves like a wave with frequency $f$. This wave goes around the nucleus and has to be a standing wave (this means that the "start" and "end" point of the wave must meet each other). These standing waves are drawn in figure 13.7.

- If an electron changes the orbit with energy difference $\Delta E$ it emits a photon with frequency $\nu$ according to $\Delta E = h\nu$.

Figure 13.7: The nucleus in the middle and two possible electron orbits. The orbits are like closed standing waves [56].
13.3. EXAMPLES

The first assumption leads to the restriction, that circumference of the orbit must be a multiple of the wavelength of the electron wave. Therefore

\[ 2\pi r = n\lambda = \frac{nh}{mv} \]

where \( r \) is the radius of the orbit and \( n \) is an integer. We also used the de Broglie wavelength \( \lambda = \frac{h}{p} \) where the momentum is classically given by \( p = mv \) where \( m \) is the mass of the electron and \( v \) its velocity.

An other equation is given by the classical orbit equation where the centripetal force is due to the electric attraction. Therefore we get

\[ m \frac{v^2}{r} = \frac{Ze^2}{4\pi\epsilon_0 r^2} \]  \hspace{1cm} (13.4)

where \( Z \) is the number of protons in the nucleus (which is 1 for hydrogen, but we want to calculate it more generally).

If we combine these two equations we get

\[ r_n = n^2 \frac{\epsilon_0 h^2}{\pi m Z e^2} \]

where \( r_n \) labels the different orbits according to \( n \). For the smallest radius of the hydrogen atom we get \( r_1 = 5.3 \cdot 10^{-11} \text{m} \)

If we now calculate the energy of the electron going from an orbit with radius \( r_a \) to one with radius \( r_b \) we have to take into account two terms: The kinetic energy and the potential energy. The potential energy is given by

\[ \Delta E_{\text{pot}} = \frac{Ze^2}{4\pi\epsilon_0} \left( \frac{1}{r_a} - \frac{1}{r_b} \right) \]

The kinetic energy of the electron on an orbit \( r_n \) is given by

\[ E_{\text{kin}} r_n = \frac{1}{2} mv_n^2 = \frac{1}{2} \frac{Ze^2}{4\pi\epsilon_0 r_n} \]

where we used equation (13.4) to express the velocity by the radius.

The total energy difference between \( r_a \) and \( r_b \) is therefore

\[ \Delta E_{\text{tot}} = \Delta E_{\text{pot}} + E_{\text{kin}} r_n - E_{\text{kin}} r_a \]

\[ = \frac{1}{2} \frac{Ze^2}{4\pi\epsilon_0} \left( \frac{1}{r_a} - \frac{1}{r_b} \right) \]

\[ = \frac{1}{2} \frac{Ze^2}{4\pi\epsilon_0 r_1} \left( \frac{1}{a^2} - \frac{1}{b^2} \right) \]

295
where we used that $r_a = a^2 r_1$ and $r_b = b^2 r_1$ with $a$ and $b$ integers.

If $r_a > r_b$ the electron loses energy which causes a photon leaving the atom with the frequency $\nu = \frac{E}{h}$ respectively with the wavelength $\lambda = \frac{ch}{E}$. For transition from the second $a = 2$ to the first $b = 1$ orbit we get a wave length of 121nm which is clearly in the ultra violet. Since any other transition to the first orbit has more energy, all transitions to the first orbit are not visible. This is different for transitions to the second orbit $b = 2$ where we might get visible light. This is shown in figure 13.8. An important property of the calculated energies is that they have discrete values. Therefore a hydrogen atom has some certain frequencies which it can emit light. This frequencies correspond to the spectral lines of hydrogen.

If we calculate how much energy is needed to take the electron from the first orbit to infinity we get an energy of 13.6eV which agrees to the measurement.

The Bohr model is a first approach to describe the behaviour of atoms with only one electron. The problem with the model is, that it still does not really avoid the problem that accelerated charge radiates electromagnetic waves and therefore looses energy. Additionally it violates the Heisenberg’s uncertainty principle. To get the precise description of an atom one has to solve Schrödinger’s equation (see chapter 13.2.5).

### 13.3.2 Rigorous example

In this chapter we want to look how the calculation with the wave function looks like. This is not relevant for any selection round, and not even for the IPhO. Since the math is sometimes beyond this level, not anything can be derived.

The simplest case for a quantum system is when we have a particle in an interval $[0,a]$ where the potential is zero and infinite anywhere else (see figure 13.9).

We now search for static wave functions which means that they do not change with time. This simplifies the calculation a lot because we can use the time independent equation (13.3) instead of the time dependent (13.2). For $x < 0$ or $x > a$ we get therefore

$$-\frac{\hbar^2}{8\pi^2 m} \frac{\partial^2 \Psi}{\partial x^2} = (E - V) \Psi$$

Since $E - V$ is something like minus infinity we see that the particle would need infinite energy to be in this region which is impossible. Therefore the probability to find the particle there is zero and as a consequence the wave function too.
Figure 13.8: Transitions to the second orbit [57].

More interesting is the region \(0 < x < a\) because there we have to solve the equation

\[-\frac{\hbar^2}{8\pi^2m} \frac{\partial^2 \Psi}{\partial x^2} = E \Psi\]  

(13.5)

If the second derivative appears it is always a good idea to think of \(\sin(x)\) and \(\cos(x)\). We try to solve the equation (13.5) by attempting \(\Psi(x) = A \sin(kx) + B \cos(kx)\) where \(A\) and \(k\) are constants which have to be determined. If we formulate equation (13.5) a bit different and use the attempt we get
To get a restriction on $A$ and $B$ we have to consider the following: at the position $x = 0$ the wave must have amplitude zero because there the infinite potential begins. Therefore $B = 0$. Additionally we have also the restriction $\Psi(a) = 0$ which leads to the fact that the argument in the $\sin$ must be a multiple of $\pi$. This means $ka = n\pi$ where $n$ is an integer. But this can only be the case if the Energy $E$ has a certain value, namely

$$E_n = \frac{n^2\pi^2\hbar^2}{2ma^24\pi^2} = \frac{n^2\hbar^2}{8ma^2}$$

This means that only discrete energy levels are allowed in order to get time independent solutions. The subscript $n$ at $E_n$ indicates which energy we look at.

To get the $A$ we use the condition that the total probability to find the particle between 0 and $a$ must be 1.
\[ 1 = \int_{0}^{a} |\Psi(x)|^2 dx \]

\[ = \int_{0}^{a} A^2 \sin(kx)^2 dx \]

\[ = A^2 \left[ \frac{1}{2} \left( -\frac{1}{k} \sin(kx) \cos(kx) + x \right) \right]_{0}^{a} \]

\[ = A^2 \frac{a}{2} \]

\[ A = \sqrt{\frac{a}{2}} \]

Therefore the possible wavefunctions are superpositions of the \( \Psi_n(x) \) which are

\[ \Psi_n(x) = \sqrt{\frac{2}{a}} \sin\left( \frac{n\pi}{a} x \right) \]

where the \( \Psi_n(x) \) has the energy

\[ E_n = \frac{n^2 \pi^2 \hbar^2}{2ma^24\pi^2} = \frac{n^2 \hbar^2}{8ma^2} \]
The main goal of statistical methods is to make inferences based on data. There are three important steps in this process: Collecting the data, describing the data and analyzing the data. While we will focus on descriptive statistics in this introduction, it is important to mention that every step heavily relies on the previous step. If your experimental setup does not provide good data, you will not be able to draw any meaningful conclusions (Garbage in - Garbage out).

The main sources are [58], [59], [60], [61] and [62].

### 14.1 Location and Spread of a single Set of Data

Let $X = \{x_1, x_2, \ldots, x_n\}$ denote a set of $n$ data points. The mean $\bar{x}$ and the variance $\sigma^2$ of $X$ are defined as

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{14.1}
\]

\[
\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{14.2}
\]

By taking the square root of $\sigma^2$, we find the standard deviation, a common measure for how spread the data is, which has the nice property of having the same units as the original data. By dividing the standard deviation by the mean, we get the coefficient of variation, an indicator for the relative spread of the data.

\[
\sigma = \sqrt{\sigma^2} \tag{14.3}
\]

\[
CV = \frac{\sigma}{\bar{x}} \tag{14.4}
\]

Another way of better describing how data is located are quantiles. The idea is to divide the data into equally large groups and indicate where these cuts are. As an example, when calculating the lower and upper quartile, one quarter of the data points are smaller than the lower quartile and one quarter are larger than the upper quartile, with the remaining half being located between these two values. It is important to note, that for this calculation, the data must first be sorted smallest to largest (i.e. $x_{i-1} \leq x_i \leq x_{i+1}$). The general formula
for calculating quantiles is then given by (14.5) for a cut at the fraction \( p \) (i.e. \( p \) of the data points below the cut and \( 1 - p \) above the cut). As an example, for the lower and upper quartiles one would calculate \( Q(0.25) \) and \( Q(0.75) \).

\[
Q(p) = (1 - g) \cdot x_j + g \cdot x_{j+1}
\]

\[
j = \lfloor pn + \frac{1}{2} \rfloor, \quad g = \left( pn + \frac{1}{2} \right) - j
\]

(Note that \( \lfloor z \rfloor \) designates the floor of \( z \), i.e. the next lower integer)

Common quantiles used are quartiles \( (p = \frac{k}{4}) \), deciles \( (p = \frac{k}{10}) \), percentiles \( (p = \frac{k}{100}) \) and of course the median \( (p = \frac{1}{2}) \), for which we can simplify the calculation:

\[
\hat{x} = \begin{cases} x_{\lfloor \frac{n+1}{2} \rfloor} & n \text{ odd} \\ \frac{1}{2} \cdot (x_{\lfloor \frac{n}{2} \rfloor} + x_{\lfloor \frac{n}{2} \rfloor + 1}) & n \text{ even} \end{cases}
\]

14.1.1 Bivariate Analysis

If we have two sets of data \( X \) and \( Y \) collected in parallel so that each \( x_i \) is associated to its corresponding \( y_i \), it might be interesting to measure how the data in these sets might be connected. This can be quantified by calculating the covariance and correlation of these sets:

\[
\text{Cov}(X,Y) = \frac{1}{n - 1} \sum_{i=1}^{n} (x_i - \bar{x}) (y_i - \bar{y})
\]

\[
\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_x \cdot \sigma_y}
\]

Finally, if we suspect that there is a linear relationship between \( X \) and \( Y \), we can try to calculate a line \( y = a + bx \) that fits our data as much as possible. This is called linear regression and is often done using the least squares method. This method minimizes the square of the errors between the calculated line and the observed data points. The coefficients can be calculated as shown in (14.10) and (14.11). The derivation can be found in the appendix.

\[
b = \frac{\text{Cov}(X,Y)}{\sigma_x^2}
\]

\[
a = \bar{y} - b \bar{x}
\]

14.2 Uncertainty Propagation

When performing experiments one should always be aware of uncertainties in measurements. There are many sources of uncertainties: Noise, minimum resolution, misalignment, calibration, etc. To assess the impact of uncertain measurements on results of complex calculations, the propagation of uncertainty has to be analyzed.
14.2. Quantification of Uncertainty

There are two different representations of uncertainty for a measurement. Additive uncertainty is expressed as an absolute range by which the real value might be different from the measured value (e.g. $\pm 0.001\text{m}$), while relative uncertainty is expressed as a fraction of the measurement (e.g. $\pm 0.1\%$). They are easily convertible into one another:

$$\varepsilon_{\text{rel}} = \frac{\varepsilon_{\text{add}}}{|X|} \quad (14.12)$$

Since it is often impossible to find definitive upper and lower bounds for the error of a measurement, we usually express uncertainty in terms of the expected standard deviation of the measurement, $x_{\text{real}} = x_m \pm \sigma_{x_m}$, where $\sigma_{x_m}$ can be found either by performing a measurement multiple times or by carefully assessing the different possible sources of uncertainties (e.g. for a ruler with mm markings, it is reasonable to assume an uncertainty of $\pm 0.5\text{mm}$). For this reason, we will express uncertainties as additive uncertainties $\delta X$ in this document.

14.2.2 Propagation of Uncertainty

In this section, we will consider different types of operations $R$ being applied to uncertain variables $X \pm \delta X$, $Y \pm \delta Y$ and $Z \pm \delta Z$, and will quantify the uncertainty $\delta R$ of the result.

- **Addition of a constant**
  The addition of a constant will not affect additive uncertainty.

  $$R(X) = X + c \quad (14.13)$$
  $$\delta R = \delta X \quad (14.14)$$

- **Addition of uncertain variables**
  By adding multiple uncertain variables, the variance of the result is the combined variance of the individual elements. However, as we express uncertainties with standard deviation, we need to take the square root of the added variances.

  $$R(X, Y, Z) = X + Y - Z \quad (14.15)$$
  $$\delta R = \sqrt{(\delta X)^2 + (\delta Y)^2 + (\delta Z)^2} \quad (14.16)$$

- **Multiplication with a constant**
  By multiplying an uncertain quantity with a constant, the uncertainty is simply multiplied with the absolute value of the constant.

  $$R(X) = a \cdot X \quad (14.17)$$
  $$\delta R = |a| \cdot \delta X \quad (14.18)$$
• **Multiplication of uncertain variables** When multiplying uncertain variables, this corresponds to adding the relative variances to find the relative variance of the result.

\[
R(X,Y,Z) = \frac{X \cdot Y}{Z} \quad (14.19)
\]

\[
\delta R = |R| \cdot \sqrt{\left(\frac{\delta X}{X}\right)^2 + \left(\frac{\delta Y}{Y}\right)^2 + \left(\frac{\delta Z}{Z}\right)^2} \quad (14.20)
\]

• **General Operations**

For general operations, the formula can be derived by considering the variation of the result due to every variable and again adding them just like standard deviations. In fact, all previous rules are just special cases of this rule and can be derived easily.

\[
R(X,Y,\ldots) \quad (14.21)
\]

\[
\delta R = \sqrt{\left(\frac{\partial R}{\partial X} \delta X\right)^2 + \left(\frac{\partial R}{\partial Y} \delta Y\right)^2 + \ldots} \quad (14.22)
\]

### 14.3 Units

A good understanding of the different units and their respective dimensions can be very helpful to avoid careless mistakes.

#### 14.3.1 The International System of Units (SI)

The International System of Units (SI, *système international (d’unités)*), is the most widely used system of measurement due to its simplicity regarding unit conversions. The system comprises seven base units from which many other units can be derived (e.g. \(1\, N = 1\, \frac{kg \cdot m}{s^2}\)).

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Unit</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electric Current</td>
<td>Ampere</td>
<td>A</td>
</tr>
<tr>
<td>Temperature</td>
<td>Kelvin</td>
<td>K</td>
</tr>
<tr>
<td>Time</td>
<td>Second</td>
<td>s</td>
</tr>
<tr>
<td>Length</td>
<td>Meter</td>
<td>m</td>
</tr>
<tr>
<td>Mass</td>
<td>Kilogram</td>
<td>kg</td>
</tr>
<tr>
<td>Luminous Intensity</td>
<td>Candela</td>
<td>cd</td>
</tr>
<tr>
<td>Amount of Substance</td>
<td>Mole</td>
<td>mol</td>
</tr>
</tbody>
</table>

Table 14.1: The seven SI base units
14.3.2 Prefixes

Prefixes can be used to change the order of magnitude of a unit.

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Symbol</th>
<th>Factor</th>
<th>Prefix</th>
<th>Symbol</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>femto</td>
<td>f</td>
<td>$10^{-15}$</td>
<td>peta</td>
<td>P</td>
<td>$10^{15}$</td>
</tr>
<tr>
<td>pico</td>
<td>p</td>
<td>$10^{-12}$</td>
<td>tera</td>
<td>T</td>
<td>$10^{12}$</td>
</tr>
<tr>
<td>nano</td>
<td>n</td>
<td>$10^{-9}$</td>
<td>giga</td>
<td>G</td>
<td>$10^{9}$</td>
</tr>
<tr>
<td>micro</td>
<td>µ</td>
<td>$10^{-6}$</td>
<td>mega</td>
<td>M</td>
<td>$10^{6}$</td>
</tr>
<tr>
<td>milli</td>
<td>m</td>
<td>$10^{-3}$</td>
<td>kilo</td>
<td>k</td>
<td>$10^{3}$</td>
</tr>
<tr>
<td>centi</td>
<td>c</td>
<td>$10^{-2}$</td>
<td>hecto</td>
<td>h</td>
<td>$10^{2}$</td>
</tr>
<tr>
<td>dezi</td>
<td>d</td>
<td>$10^{-1}$</td>
<td>deca</td>
<td>da</td>
<td>$10^{1}$</td>
</tr>
</tbody>
</table>

Table 14.2: Metric Prefixes

14.3.3 Dimensional Analysis

We can use the fact that dimensions corresponding to the seven base units cannot be created from any other base dimensions to quickly check equations. If an equality does not have the same dimension on each side, then surely it cannot be true (however, the opposite cannot be said, even if the dimensions agree there could be a mistake in form of a dimensionless factor). To perform dimensional analysis on an equation, we replace all involved variables with their respective dimension. We then simplify both sides to see if the dimensions cancel each other. As an example, this technique is applied to Newton’s second Law:

\[
F = am
\]

\[
\frac{[M][L]}{[T]^2} = \frac{[L]}{[T]^2} \cdot [M]
\]

14.4 Graphs

14.4.1 Elements of good graphs

When presenting your results in form of graphs, there are some guidelines that you should respect to make your graph clear and understandable:

1. **Complexity**: A graph should not be more complex than the data it represents. Avoid irrelevant decoration, 3D effects and distortion.

2. **Scaling**:
   
   (a) The data should not be clumped in one section of the graph
   
   (b) The scale should not change along one axis
   
   (c) Your axes should include 0 and have no jumps
   
   (d) Use simple steps, one square or tick mark could represent 1, 2, 5, 10, etc..
3. **Title**: Your graphs should have a descriptive title that contains information about the origin of the data.

4. **Multiple Data Sets**: If your plotting multiple sets of data on the same graph, make sure they’re easily distinguishable and include a key/legend (Should not obstruct Data)

5. **Labeled Axes**: Label your axes with the name of the variable, its unit and the scale (Ticks and Numbers). There are multiple ways of including the units, however we suggest you use the ISO standard `variable_name /unit`.

6. **Readability**: When drawing graphs by hand, use a Ruler.

### 14.4.2 Logarithmic Plots

Logarithmic and Semi-Logarithmic Plots are useful tools to identify special types of relationship between variables. They are characterized by one or both axis being scaled logarithmically instead of linearly. For a logarithmic plot, this means that monomials of the form $y = ax^k$ appear as straight lines with slope $k$. This can be seen by applying a log function to both sides of the equation:

$$\log (y) = \log (ax^k) = \log (a) + k \log (x) \quad (14.23)$$

In a similar fashion, we can see that relations of the form $y = \lambda a^{\gamma x}$ appear as a line with slope $\gamma$ on a semi-log plot:

$$\log_a (y) = \log_a (\lambda a^{\gamma x}) = \gamma x + \log_a (\lambda) \quad (14.24)$$

Or using a base 10 log:

$$\log y = \log (\lambda a^{\gamma x}) = (\gamma \log (a))x + \log (\lambda) \quad (14.25)$$
14.4. GRAPHS

Figure 14.1: Different Monomials on a Logarithmic Plot.

Figure 14.2: Different Exponentials on a Semi-Logarithmic Plot.
Appendix A

Further derivations

A.1 Derivations of Statistics

A.1.1 Alternative formulations for Variance and Covariance

By applying the sum to individual elements and using $\sum_{i=1}^{n} x_i = n \bar{x}$, alternative formulations can be found that are sometimes more comfortable to apply.

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$
$$= \frac{1}{n-1} \sum_{i=1}^{n} (x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$
$$= \frac{1}{n-1} \left( \sum_{i=1}^{n} x_i^2 - 2\bar{x} \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2 \right)$$
$$= \frac{1}{n-1} \left( \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \right)$$
$$= \frac{1}{n-1} \left( \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right)$$
Cov(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})
= \frac{1}{n-1} \sum_{i=1}^{n} (x_i y_i - x_i \bar{y} - x \bar{y} + \bar{x} \bar{y})
= \frac{1}{n-1} \left( \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i - \bar{x} \sum_{i=1}^{n} y_i + \bar{y} \right)
= \frac{1}{n-1} \left( \sum_{i=1}^{n} x_i y_i - n \bar{y} \bar{x} - n \bar{y} \bar{x} + n \bar{y} \bar{x} \right)
= \frac{1}{n-1} \left( \sum_{i=1}^{n} x_i y_i - n \bar{y} \bar{x} \right)

A.1.2 Derivation of the Least Squares Coefficients
We want to generate a linear regression to predict y for a given x. We will call \( \hat{y} = a + bx \) our predictor for y. We define the sum of square errors as a function of the coefficients a and b:

\[ S(a, b) = \sum_{i=1}^{n} (y_i - \hat{y})^2 \]

\[ = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \]

Since we want to find a and b that minimize this function, we will set the partial derivatives with respect to a and b equal to zero and solve for a and b.

\[ \frac{\partial S}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0 \] (A.1)

\[ \frac{\partial S}{\partial b} = -2 \sum_{i=1}^{n} (x_i(y_i - a - bx_i)) = 0 \] (A.2)

We will start by simplifying the sum in (26):

\[ \sum_{i=1}^{n} (y_i - a - bx_i) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a - b \sum_{i=1}^{n} x_i \]

\[ = ny - na - bn \bar{x} \] (A.3)

\[ = n \bar{y} - na - bn \bar{x} \] (A.4)

In (26):

\[ -2(n \bar{y} - na - bn \bar{x}) = 0 \] (A.5)

\[ \bar{y} - a - b \bar{x} = 0 \] (A.6)

\[ a = \bar{y} - b \bar{x} \] (A.7)
We will use this expression for \( a \) and simplify the sum in (27):

\[
\sum_{i=1}^{n} (x_i(y_i - a - bx_i)) = \sum_{i=1}^{n} (x_i(y_i - \bar{y} + b\bar{x} - bx_i)) = \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \bar{y} + \sum_{i=1}^{n} x_i b\bar{x} - \sum_{i=1}^{n} b x_i^2 \tag{A.8}
\]

\[
= \sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i + b\bar{x} \sum_{i=1}^{n} x_i - b \sum_{i=1}^{n} x_i^2 \tag{A.9}
\]

\[
= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} + b \left( n\bar{x}^2 - \sum_{i=1}^{n} x_i^2 \right) \tag{A.10}
\]

\[
= (n - 1) \cdot \text{Cov}(X,Y) - b(n - 1) \cdot \sigma_x^2 \tag{A.11}
\]

In (27):

\[
-2 \left( (n - 1) \cdot \text{Cov}(X,Y) - b(n - 1) \cdot \sigma_x^2 \right) = 0 \tag{A.13}
\]

\[
\text{Cov}(X,Y) - b\sigma_x^2 = 0 \tag{A.14}
\]

\[
b = \frac{\text{Cov}(X,Y)}{\sigma_x^2} \tag{A.15}
\]
Appendix B

Tables

B.1 List of physical constants (in SI units)

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atomic mass unit</td>
<td>u</td>
<td>1.660 \cdot 10^{-27}</td>
<td>kg</td>
</tr>
<tr>
<td>Atomic mass unit</td>
<td>uc^2</td>
<td>931.49</td>
<td>MeV</td>
</tr>
<tr>
<td>Avogadro constant</td>
<td>N_A</td>
<td>6.022 \cdot 10^{-23}</td>
<td>mol^{-1}</td>
</tr>
<tr>
<td>Bohr radius</td>
<td>a_0</td>
<td>5.2917 \cdot 10^{-11}</td>
<td>m</td>
</tr>
<tr>
<td>Boltzmann constant</td>
<td>k_B</td>
<td>1.3806 \cdot 10^{-23}</td>
<td>J\cdot K^{-1}</td>
</tr>
<tr>
<td>Elementary charge</td>
<td>e</td>
<td>1.602 \cdot 10^{-19}</td>
<td>C</td>
</tr>
<tr>
<td>Vacuum permittivity / electric constant</td>
<td>\epsilon_0</td>
<td>8.8541 \cdot 10^{-12}</td>
<td>A\cdot V^{-1} \cdot m^{-1}</td>
</tr>
<tr>
<td>Gravitational acceleration (average)</td>
<td>g</td>
<td>9.807</td>
<td>m\cdot s^{-2}</td>
</tr>
<tr>
<td>Universal Gas constant</td>
<td>R</td>
<td>8.3145</td>
<td>J\cdot mol^{-1} \cdot K^{-1}</td>
</tr>
<tr>
<td>Gravitational constant</td>
<td>G</td>
<td>6.673 \cdot 10^{-11}</td>
<td>m^3\cdot kg^{-1} \cdot s^{-2}</td>
</tr>
<tr>
<td>Speed of light</td>
<td>c</td>
<td>2.9979 \cdot 10^8</td>
<td>m\cdot s^{-1}</td>
</tr>
<tr>
<td>Vacuum permeability / magnetic constant</td>
<td>\mu_0</td>
<td>4\pi \cdot 10^{-7}</td>
<td>V\cdot s\cdot A^{-1} \cdot m^{-1}</td>
</tr>
<tr>
<td>Normal pressure</td>
<td>p_0</td>
<td>101324</td>
<td>Pa</td>
</tr>
<tr>
<td>Planck constant</td>
<td>h</td>
<td>6.626 \cdot 10^{-34}</td>
<td>J\cdot s</td>
</tr>
<tr>
<td>Mass of electron</td>
<td>m_e</td>
<td>9.109 \cdot 10^{-31}</td>
<td>kg</td>
</tr>
<tr>
<td>Mass of neutron</td>
<td>m_n</td>
<td>1.675 \cdot 10^{-27}</td>
<td>kg</td>
</tr>
<tr>
<td>Mass of proton</td>
<td>m_p</td>
<td>1.673 \cdot 10^{-27}</td>
<td>kg</td>
</tr>
<tr>
<td>Rydberg constant</td>
<td>R_H</td>
<td>1.097 \cdot 10^4</td>
<td>m^{-1}</td>
</tr>
<tr>
<td>Stefan-Boltzmann constant</td>
<td>\sigma</td>
<td>5.670 \cdot 10^{-8}</td>
<td>W^2\cdot m^{-3} \cdot K^{-1}</td>
</tr>
<tr>
<td>Wave impedance of vacuum</td>
<td>Z_0</td>
<td>376.7</td>
<td>\Omega</td>
</tr>
</tbody>
</table>

Table B.1: List of physical constants [?]
B.2 List of named, SI derived units

<table>
<thead>
<tr>
<th>Unit</th>
<th>Symbol</th>
<th>Quantity</th>
<th>Equivalents</th>
<th>SI Equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>hertz</td>
<td>Hz</td>
<td>frequency</td>
<td>1/s</td>
<td>s⁻¹</td>
</tr>
<tr>
<td>radian</td>
<td>rad</td>
<td>angle</td>
<td>m/m</td>
<td>1</td>
</tr>
<tr>
<td>newton</td>
<td>N</td>
<td>force</td>
<td>kg·m/s²</td>
<td>kg·m·s⁻²</td>
</tr>
<tr>
<td>pascal</td>
<td>Pa</td>
<td>pressure, stress</td>
<td>N/m²</td>
<td>kg·m⁻¹·s⁻²</td>
</tr>
<tr>
<td>joule</td>
<td>J</td>
<td>energy, work, heat</td>
<td>N·m, W·s</td>
<td>kg·m²·s⁻²</td>
</tr>
<tr>
<td>watt</td>
<td>W</td>
<td>power</td>
<td>J/s, V·A</td>
<td>kg·m²·s⁻³</td>
</tr>
<tr>
<td>coulomb</td>
<td>C</td>
<td>electric charge</td>
<td>F·V</td>
<td>A·s</td>
</tr>
<tr>
<td>volt</td>
<td>V</td>
<td>voltage</td>
<td>W/A, J/C</td>
<td>kg·m²·s⁻³·A⁻¹</td>
</tr>
<tr>
<td>farad</td>
<td>F</td>
<td>capacitance</td>
<td>C/V</td>
<td>kg⁻¹·m⁻²·s⁻⁴·A²</td>
</tr>
<tr>
<td>ohm</td>
<td>Ω</td>
<td>resistance, impedance</td>
<td>1/S, V/A</td>
<td>kg·m²·s⁻³·A⁻²</td>
</tr>
<tr>
<td>siemens</td>
<td>S</td>
<td>conductance</td>
<td>1/Ω, A/V</td>
<td>kg⁻¹·m⁻²·s⁻⁴·A²</td>
</tr>
<tr>
<td>tesla</td>
<td>T</td>
<td>magnetic field strength</td>
<td>V·s/m²</td>
<td>kg·s⁻²·A⁻¹</td>
</tr>
<tr>
<td>henry</td>
<td>H</td>
<td>inductance</td>
<td>V·s/A, Ω·s</td>
<td>kg·m²·s⁻²·A⁻²</td>
</tr>
</tbody>
</table>

Table B.2: List of named, SI derived units

B.3 List of material constants
### Table B.3: Properties of different metals[63].

<table>
<thead>
<tr>
<th>Name</th>
<th>Density $\rho$/kg·m$^{-3}$</th>
<th>Speed of sound $c_s$/m·s$^{-1}$</th>
<th>Linear expansion coefficient $\alpha$/K$^{-1}$</th>
<th>Specific heat capacity $C$/J·kg$^{-1}$·K$^{-1}$</th>
<th>Melting temperature $T_m$/°C</th>
<th>Heat conductivity $\lambda$/W·m$^{-1}$·K$^{-1}$</th>
<th>Specific electric resistance (at 20°C) $\rho_e$/Ω·m$^{-1}$</th>
<th>Magnetic permeability $\mu_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminium</td>
<td>2700</td>
<td>5240</td>
<td>$23.8 \cdot 10^{-6}$</td>
<td>896</td>
<td>660.1</td>
<td>239</td>
<td>$2.82 \cdot 10^{-8}$</td>
<td>$1+2.1 \cdot 10^{-5}$</td>
</tr>
<tr>
<td>Lead</td>
<td>11340</td>
<td>1250</td>
<td>$31.3 \cdot 10^{-6}$</td>
<td>129</td>
<td>327.4</td>
<td>34.8</td>
<td>$2.2 \cdot 10^{-7}$</td>
<td>diamagnetic</td>
</tr>
<tr>
<td>Iron</td>
<td>7860</td>
<td>5170</td>
<td>$12.0 \cdot 10^{-6}$</td>
<td>450</td>
<td>1535</td>
<td>80</td>
<td>$1 \cdot 10^{-7}$</td>
<td>$\approx 5800$</td>
</tr>
<tr>
<td>Gold</td>
<td>19290</td>
<td>3240</td>
<td>$14.3 \cdot 10^{-6}$</td>
<td>129</td>
<td>1063</td>
<td>312</td>
<td>$2.2 \cdot 10^{-8}$</td>
<td>$1−3.4 \cdot 10^{-5}$</td>
</tr>
<tr>
<td>Copper</td>
<td>8920</td>
<td>3900</td>
<td>$16.8 \cdot 10^{-6}$</td>
<td>383</td>
<td>1083</td>
<td>390</td>
<td>$1.7 \cdot 10^{-8}$</td>
<td>$1−6.4 \cdot 10^{-6}$</td>
</tr>
<tr>
<td>Brass</td>
<td>8470</td>
<td></td>
<td>$18 \cdot 10^{-6}$</td>
<td>380</td>
<td>905</td>
<td>79</td>
<td>$7.8 \cdot 10^{-8}$</td>
<td></td>
</tr>
<tr>
<td>Silver</td>
<td>10500</td>
<td></td>
<td>$19.7 \cdot 10^{-6}$</td>
<td>235</td>
<td>860.8</td>
<td>428</td>
<td>$1.59 \cdot 10^{-8}$</td>
<td></td>
</tr>
</tbody>
</table>

### Table B.4: Properties of different fluids[63].

<table>
<thead>
<tr>
<th>Name</th>
<th>Density $\rho$/kg·m$^{-3}$</th>
<th>Speed of sound $c_s$/m·s$^{-1}$</th>
<th>Volume expansion coefficient $\alpha$/K$^{-1}$</th>
<th>Specific heat capacity $C$/J·kg$^{-1}$·K$^{-1}$</th>
<th>Melting temperature $T_m$/°C</th>
<th>Boiling temperature $T_b$/°C</th>
<th>Enthalpy of fusion $L_m$/J·kg$^{-1}$</th>
<th>Enthalpy of vaporization $L_b$/J·kg$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetone</td>
<td>792</td>
<td>1190</td>
<td>$1.49 \cdot 10^{-3}$</td>
<td>2160</td>
<td>$−94.86$</td>
<td>56.25</td>
<td>$9.8 \cdot 10^4$</td>
<td>$5.25 \cdot 10^5$</td>
</tr>
<tr>
<td>Benzol</td>
<td>879</td>
<td>1326</td>
<td>$1.23 \cdot 10^{-3}$</td>
<td>1725</td>
<td>5.53</td>
<td>80.1</td>
<td>$1.28 \cdot 10^4$</td>
<td>$3.94 \cdot 10^5$</td>
</tr>
<tr>
<td>Ethanol</td>
<td>789</td>
<td>1170</td>
<td>$1.1 \cdot 10^{-3}$</td>
<td>2430</td>
<td>$−114.5$</td>
<td>78.33</td>
<td>$1.08 \cdot 10^5$</td>
<td>$8.4 \cdot 10^5$</td>
</tr>
<tr>
<td>Oil</td>
<td>$\approx 900$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mercury</td>
<td>13546</td>
<td>1430</td>
<td>$1.84 \cdot 10^{-4}$</td>
<td>139</td>
<td>$−38.87$</td>
<td>356.58</td>
<td>$1.18 \cdot 10^4$</td>
<td>$2.85 \cdot 10^5$</td>
</tr>
<tr>
<td>Water</td>
<td>998</td>
<td>1483</td>
<td>$2.07 \cdot 10^{-4}$</td>
<td>4182</td>
<td>0</td>
<td>100</td>
<td>$3.338 \cdot 10^3$</td>
<td>$2.256 \cdot 10^6$</td>
</tr>
<tr>
<td>Name</td>
<td>Density $\rho$/kg·m$^{-3}$</td>
<td>Speed of sound $c_s$/m·s$^{-1}$</td>
<td>Molare heat capacity $(p$ constant) $C_p$/joule/mol·K</td>
<td>Melting temperature $T_m$/°C</td>
<td>Boiling temperature $T_b$/°C</td>
<td>Van-der-Waals constant $a$/N·m$^{-4}$·mol$^{-2}$</td>
<td>Van-der-Waals constant $b$/m$^3$·mol$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------</td>
<td>-------------------------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
<td>-------------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------</td>
<td></td>
</tr>
<tr>
<td>Aargon</td>
<td>1.784</td>
<td>-</td>
<td>20.9</td>
<td>$-77.7$</td>
<td>$-33.4$</td>
<td>0.425</td>
<td>$3.73 \cdot 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>Helium</td>
<td>0.1785</td>
<td>1005</td>
<td>20.9</td>
<td>-</td>
<td>$-268.94$</td>
<td>0.0034</td>
<td>$2.36 \cdot 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>Carbon dioxide</td>
<td>1.977</td>
<td>268</td>
<td>36.8</td>
<td>-</td>
<td>$-78.45$</td>
<td>0.366</td>
<td>$4.28 \cdot 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>Air</td>
<td>1.293</td>
<td>344</td>
<td>29.1</td>
<td>-</td>
<td>$-191.4$</td>
<td>0.135</td>
<td>$3.65 \cdot 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>Methan</td>
<td>0.717</td>
<td>445</td>
<td>35.6</td>
<td>-</td>
<td>$-191.4$</td>
<td>0.229</td>
<td>$4.28 \cdot 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>Neon</td>
<td>0.9</td>
<td>-</td>
<td>20.8</td>
<td>$-248.61$</td>
<td>$-245.06$</td>
<td>0.0217</td>
<td>$1.74 \cdot 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>Oxygen</td>
<td>1.429</td>
<td>326</td>
<td>29.3</td>
<td>$-218.79$</td>
<td>$-182.97$</td>
<td>0.138</td>
<td>$3.17 \cdot 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>Nitrogen</td>
<td>1.25</td>
<td>1310</td>
<td>29.1</td>
<td>$-210.0$</td>
<td>$-195.82$</td>
<td>0.137</td>
<td>$3.87 \cdot 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>Water vapour</td>
<td>-</td>
<td>-</td>
<td>33.6</td>
<td>0</td>
<td>100</td>
<td>0.553</td>
<td>$3.04 \cdot 10^{-5}$</td>
<td></td>
</tr>
<tr>
<td>Hydrogen</td>
<td>0.0889</td>
<td>1310</td>
<td>28.9</td>
<td>$-259.2$</td>
<td>$-252.77$</td>
<td>0.0248</td>
<td>$2.66 \cdot 10^{-5}$</td>
<td></td>
</tr>
</tbody>
</table>

Table B.5: Properties of different gases[63].
Bibliography


[19] https://courses.lumenlearning.com/cheminter/chapter/phase-diagram-for-water/
[25] Wikipedia; Tacoma Narrows Bridge (1940); https://en.wikipedia.org/wiki/Tacoma_Narrows_Bridge_(1940)
[27] Wikimedia; https://commons.wikimedia.org/wiki/File:Damping_1.svg
[30] Heribert Stroppe; Physik für Studierende der Natur- und Ingenieurwissenschaften 15., aktualisierte Auflage; Carl Hanser Verlag; 2012 München
[31] Joachim Grehn, Joachim Krause; Metzler Physik; 4. Auflage; Bildhaus Schulbuchverlage; 2011
[33] http://hyperphysics.phy-astr.gsu.edu/hbase/electric/gaulaw.html
[34] Youtube; https://www.youtube.com/watch?v=Jx0IdlBKYNk
[36] http://archive.cnx.org/contents/3e23d0f1-22c7-47a2-8185-c9a9fc23a9fd@2/capacitors-and-dielectrics


[40] Forphys; http://www.forphys.de/Website/student/bfeld.html


[44] https://commons.wikimedia.org/wiki/Category:Kirchhoff


[46] Roland Puntaier; http://www.texample.net/tikz/examples/bernoulli/; modified

[47] Joachim Grehn, Joachim Krause; Metzler Physik; 4. Auflage; Bildhaus Schulbuchverlage; 2011

[48] Gerthsen, Vogel; Physik; 17. Auflage; Springer Lehrbuch; 1993


[50] http://physical-easy.blogspot.ch


[56] https://www.tf.uni-kiel.de/matwis/amat/mw1_ge/kap_2/backbone/r2_1_2.html

[57] http://hyperphysics.phy-astr.gsu.edu/hbase/Bohr.html

[58] J-M. Helbling, Course «Probabilités et Statistique» (EPFL, Fall 2015)

[59] Michigan State University Online Lecture «Error Propagation»


[65] https://de.wikipedia.org/wiki/Geschichte_der_speziellen_Relativit%C3%A4tstheorie