Classical Physics

Lecture notes from the first-year undergraduate course *Klasična fizika*, taught by prof. dr. Marko Mikuž at the Faculty of Mathematics and Physics at the University of Ljubljana in the 2020-2021 academic year. The notes were transformed from recorded lectures into textbook form, translated to English, and expanded with additional material in the 2021-2022 academic year by Elijan Mastnak.

Note: This work is in its early stages and will inevitably contain some mistakes. If you find errors and feel like telling me, I will be grateful and happy to hear from you, even for the most trivial of mistakes. You can reach me by email, in English, Slovene, or Spanish, at ejmastnak@gmail.com.

Faculty of Mathematics and Physics, University of Ljubljana
Last update: December 13, 2021

Contents

1 Introduction

Measurements are the foundation of science. For our purposes, measurements are quantified observations of the natural world in which we associate *numerical* values with a physical phenomenon we wish to better understand. The essence of measurement in physics goes roughly as follows:

1. observe a process in the natural world,
2. make numerical measurements of the observed process, and
3. study the measurements for patterns and structure. On the basis of these patterns, formulate physical laws—essentially relationships between measured quantities—in the language of *mathematics*.

When a well-defined, higher-level structure emerges from the jumble of experimentally-determined observations and patterns, the resulting conclusions are called a *physical model*. As a model is fine-tuned and continually stands the test of experiment—meaning the model's theoretical predictions are found to agree with experimental observations—the model becomes a *physical theory*.[1]

Meanwhile, if, when testing the model, we find situations in which experimental results consistently deviate from the model's theoretical predictions, we have two options:

1. reject the theory entirely, or
2. conclude that the theory is only valid within the limits of a specific regime.

Example: A short history of Newtonian gravitation

As a concrete example, we now give a quick history of the theory of Newtonian gravitation that illustrates all of the important steps listed above. Newtonian gravitation's roots trace back to early astronomers observing the sky and recognizing patterns and structure in distribution of stars. These observations eventually took on a quantitative nature (e.g. times of orbit, when certain stars appear in the sky, estimates of distances between the Earth and Sun, etc.) and eventually evolved into the Copernican model of planets orbiting the Sun and the current solar system model. These early astronomical models were later polished and summarized mathematically by Kepler's laws of planetary motion, which eventually evolved into Newton's theory of gravitation.

Later measurements with more advanced 20$^{\text{th}}$ century instruments revealed slight but consistent discrepancies between experiment and the predictions of Newtonian gravitation.[2] These discrepancies were later explained by the 20$^{\text{th}}$ century theory of general relativity, and Newtonian gravitation was eventually deemed to hold in the limits of small gravitational fields and low speeds.

[1] Note that the definitions of and borders between concepts like "model" and "theory" are not rigidly defined. Don't worry if the distinctions seem vague—they often are, and you will gain intuition with experience. For now it suffices to remember that a theory is more general and better tested than a model.

[2] For example discrepancies involving the perihelion precession of Mercury or the angular deflection of light around massive objects

1.1 The landscape of physical theories

Many theories exist to describe the natural world; the most important of these are:

1. **Classical physics** (For our purposes[3] classical physics is the material covered in this course, i.e. Newtonian mechanics, classical thermodynamics, and classical electrodynamics.)

2. **Special relativity** (covered in the second-year course *Moderna fizika 1*)

3. **Quantum mechanics** (covered initially in *Moderna fizika 1* and in more detail in the aptly-named third-year course *Kvantna mehanika*)

4. **General relativity**, **quantum field theory**, and various theories of **quantum gravity** (all beyond the scope of undergraduate study at FMF)

Each theory is valid (and useful) in a specific physical regime; which theory applies in which regime depends largely on three physical quantities. These are:

1. speed,

2. distance (informally, size), and

3. gravitational field strength.

To say whether a physical system[4] is, say, "big" or "small", we have to compare it to something. Ideally, this reference value for the system's size should be something with fundamental physical meaning, such as a natural constant. Although somewhat ahead of our discussion of physical quantities in Section 1.3, the idea of comparison to meaningful physical constants motivates describing a physical system in terms of *dimensionless parameters* (dimensionless meaning without units). We construct these dimensionless parameters by multiplying or dividing dimensioned quantities characteristic of the system by fundamental constants with deeper physical meaning (these constants set a meaningful scale against which to compare the system's values) and choose a combination such that the net result is dimensionless.

Since that might sound rather abstract, we now show how to describe a system's speed, size, and the surrounding gravitational field strength in dimensionless form. Importantly, this exercise will give us a simple way to estimate which of the theories mentioned earlier applies in which physical regime.

- **Speed**: we first determine a good universal reference speed. As you might guess, a natural choice is the speed of light, *defined*[5] as of May 2019 as

$$c \equiv 2.997\,924\,58 \cdot 10^{8}\,\mathrm{m\,s^{-1}}. \tag{1.1}$$

  We then analyze a system with speed v in terms of the dimensionless ratio

$$\frac{v}{c}. \tag{1.2}$$

[3]Note that some authors also define special (and sometimes general) relativity as "classical physics", and only theories involving quantum physics are deemed "non-classical" or "modern".

[4]For orientation, this "physical system" could be, for example, a mass on a spring, a ball tossed in the air, a planet orbiting a sun, two protons colliding in a particle accelerator, a black hole...

[5]More on the (re-)definition of physical units coming soon in Section 1.3.1

The quantity v/c is less than one for any physical system (since the speed of light c is a universal speed limit), and how much less than one gives a physically meaningful statement of the system's speed. Systems with $v/c \ll 1$ are accurately described by classical physics, while systems with $v/c \lesssim 1$ require special relativity or more advanced theories.[6]

- **Size**: A fundamental distance characteristic of the boundary regime between classical and quantum physics is a quantity called the *electron Compton wavelength*, denoted by λ_C and equal to

$$\lambda_\mathrm{C} \equiv \frac{hc}{m_\mathrm{e}c^2} = 2.426\,310\,238\,67(73) \cdot 10^{-12}\,\mathrm{m}, \tag{1.3}$$

where h is the Planck constant and m_e is the electron mass. The Compton wavelength sets the scale for distances at the level of atoms and fundamental particles—you will hear more about it in *Moderna fizika 1.* We then represent the size of a system with characteristic length L in terms of the dimensionless parameter

$$\frac{\lambda_\mathrm{C}}{L}. \tag{1.4}$$

Systems with $\lambda_\mathrm{C}/L \ll 1$ (i.e. lengths much larger than λ_C) are accurately described by classical physics, while systems with $\lambda_\mathrm{C}/L \gtrsim 1$ require a quantum-mechanical treatment (although quantum mechanics is often relevant even at nanometer scales).

Alternatively, one could also separate classical and quantum mechanics with the Bohr radius

$$a_0 = 5.291\,772\,109\,03(80) \cdot 10^{-11}\,\mathrm{m} \tag{1.5}$$

and the associated dimensionless parameter a_0/L. Physically, the Bohr radius represents the expected distance between the electron and nucleus in the ground state of a hydrogen atom; you'll get plenty of experience with both λ_C and a_0 in *Moderna fizika 1.*

To be clear, the exact values of both λ_C and a_0 to the long strings of decimals given above are not important now; the takeaway here is that both are characteristic of distances at the atomic scale, which occurs at roughly $10^{-12}\,\mathrm{m}$ and is much smaller than anything familiar from everyday life (the width of a human hair, for instance, is of the order $10\,\mu\mathrm{m}$ to $100\,\mu\mathrm{m}$.)

- **Gravitational Field Strength**: Only for the sake of completeness—this predates our treatment of gravitation—a characteristic *dimensioned* quantity representing a system of a mass m a distance r from a massive body of mass M generating a gravitational field is the gravitational potential

$$\phi_\mathrm{g} = \frac{GM}{r}, \tag{1.6}$$

where $G = 6.674\,30(15) \cdot 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}$ is a universal constant called the *gravitational constant.* We then turn this into a *dimensionless* parameter by

[6] In case you haven't seen it before, the symbol $\ll$ means "much less than", while $\lesssim$ means "less than, but of the same order". For example, one might reasonably write $1 \ll 1000$ and $1 \lesssim 1.1$.

dividing through by c^2 and including the mass m of the body exposed to the gravitational field:

$$\frac{m\phi_{\mathrm{g}}}{mc^2} = \frac{GmM}{rmc^2}. \tag{1.7}$$

It might seem unproductive to include the mass m in both the numerator and denominator without simplifying, but this has an instructive physical interpretation: mc^2, as you might have recognized from the famous equation $E = mc^2$, is the rest energy inherently associated with the mass m, while $m\phi_{\mathrm{g}}$ is the gravitational energy associated with immersing the mass m in the gravitational field generated by the mass M. In other words, $m\phi_{\mathrm{g}}/mc^2$ is a ratio of two energies—one associated with the gravitational field and one inherent to the mass m—and energies are easier to interpret than an abstract quantity like the gravitational potential ϕ_{g} alone.

To a first approximation, systems with $m\phi_{\mathrm{g}}/mc^2 \ll 1$ are well-described by classical physics, while systems with $m\phi_{\mathrm{g}}/mc^2 \gtrsim 1$ require general relativity. For orientation, the Earth-Sun system, using $M_{\mathrm{sun}} \approx 2 \cdot 10^{30}$ kg and an Earth-Sun separation $r \approx 1.5 \cdot 10^{11}$ m, produces $m\phi_{\mathrm{g}}/mc^2 \sim 10^{-8} \ll 1$. In other words, the Earth-Sun system, as might be expected, is well-described by simple Newtonian gravitation. Don't worry if you don't understand the equation for ϕ_{g}—you are not expected too, since we have not yet covered gravitation.

Figure 1 shows where each of the important theories mentioned earlier falls in the "phase space"[7] of physical theories spanned by the parameters v/c, λ_{C}/L and $m\phi_{\mathrm{g}}/mc^2$.

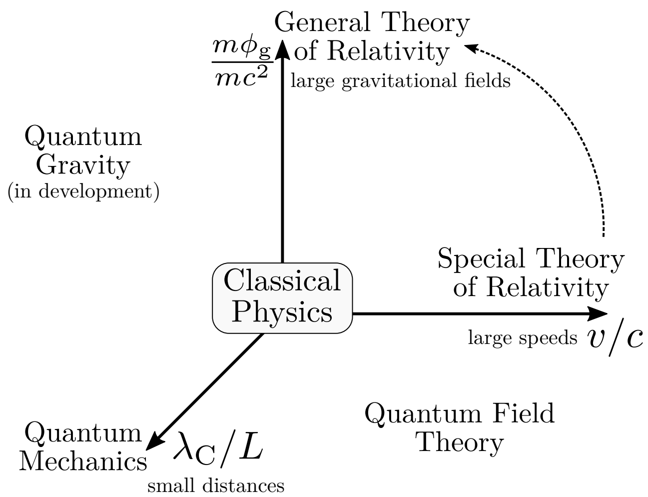

Figure 1: The landscape of physical theories. The appropriate theory to describe a given system depends on the system's speed, size, and the gravitational field to which it is exposed.

Validity of Classical Physics

[7] To not leave you in the dark, a physical system's *phase space* refers to the space of the system's possible states. This concept turns out to be useful in many branches of physics (you'll meet it, for example, in the second-year courses *Statistična termodinamika* and *Klasična mehanika*). But we use the term here only loosely to refer to the space of all existing physical theories.

Classical physics holds at "not too small" distances and "not too large" speeds and gravitational fields. More quantitatively, the theoretical predictions of Newtonian mechanics agree with experiment for physical systems with characteristic speed v, distance L and gravitational potential ϕ_{g} satisfying

$$\frac{v}{c} \ll 1, \qquad \frac{\lambda_{\mathrm{C}}}{L} \ll 1, \quad \text{and} \quad \frac{\phi_{\mathrm{g}}}{c^2} \ll 1. \tag{1.8}$$

The physical world we experience in our everyday lives generally falls in this regime.

Of course, we also wish to describe the physical world in the more exotic regimes shown in Figure 1, where classical physics fails. For the most part, each exotic regime has its own specialized theory; a very active subject of research involves connecting the exotic theories among themselves in a self-consistent manner. Following is a whirlwind tour of the physical regimes and theories shown in Figure 1.

- The first theory to transcend Newtonian mechanics was the **special theory of relativity** (STR), which holds at large speeds ($v/c \lesssim 1$) but only "everyday" distances and gravitational fields, i.e. at $\lambda_{\mathrm{C}}/L \ll 1$ and $\phi_{\mathrm{g}}/c^2 \ll 1$.

- **Quantum mechanics** holds at small distances but, in its basic form, only at everyday speeds and gravitational field strengths.

- The **general theory of relativity** (GTR) accurately describes the physical world at large gravitational field strengths. Happily, the general theory of relativity (as the name might suggest) generalizes the special theory of relativity, i.e. the GTR covers everything described by the STR, in addition to large gravitational fields.

- **Quantum field theory** (QFT) is a generalization of quantum mechanics and special relativity and accurately describes both small distances and large speeds.

- Theories of small distances and large gravitational fields are collectively called **quantum gravity**. This part of the physics landscape is still poorly understood and a topic of active research; there is not a yet a universally-accepted theory of quantum gravity.

  Development of quantum gravity is plagued by both (i) a lack of measurements due to its extreme physical conditions (subatomic particles and black holes are notoriously challenging to measure, let alone simultaneously) and (ii) the lack of a well-established mathematical formalism with which to describe the theory. In other words, we lack both the mathematics to formulate theory and the measurements to test it.

  To illustrate the experimental problem, one current theory of quantum gravity involves eleven dimensions; confirming this theory would require (somehow) a transition from 11 dimensional space to the three dimensional space of our everyday lives. As you can imagine, that poses quite an experimental challenge!

The main lesson here is that classical physics covers only a small regime of the vast and diverse physical phase space, plenty of which, particularly in the more exotic corners of small distances and large gravitational fields, remains poorly understood.

This course will cover only the "center" of the physical phase space shown in Figure 1, in which Newtonian mechanics, classical thermodynamics, and classical electrodynamics accurately describe the physical world. We will, however, (for example when studying electrical conduction) to some extent consider the connection between the microscopic and macroscopic worlds, and aim to describe physical processes in both worlds and find connections between them. We will only qualitatively mention the deviations of special relativity from Newtonian mechanics at large speeds, and leave the other exotic theories for future courses.

1.2 A tour of the mathematics used in this course

Before solving problems, we first summarize the mathematical and physical quantities with which we will operate. We will represent physical quantities mathematically mostly in terms of scalars and vectors. Vectors have both magnitude and direction in space and are well-suited to describing physical processes in the three-dimensional Euclidean space of classical physics. Both the magnitude and direction of a vector can *change* over the course of a physical process. Thus, we need a mathematical formalism describing change of both scalars and vectors.

The change of scalars is governed by basic differential and integral calculus, which forms the backbone of the Newtonian physics covered in this course. As an example, the speed $v(t)$ of an object is related to the total distance $s(t)$ traveled by the object according to

$$v = \frac{\mathrm{d}s}{\mathrm{d}t} \qquad \text{and} \qquad s = \int v \, \mathrm{d}t, \tag{1.9}$$

where the derivative and integral are performed with respect to time.

The change of vector quantities is governed by a branch of mathematics called *vector calculus*, which is the generalization of scalar differential and integral calculus to vector quantities. Vector calculus is covered formally in the second-year course *Matematika 3*, but we will dabble with it in this course as well. For orientation, here are some physical examples of calculus involving vector quantities:

- The momentum $\boldsymbol{p}(t)$ of an object of mass m with position vector $\boldsymbol{r}(t)$ is

$$\boldsymbol{p} = m\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}. \tag{1.10}$$

- The work[8] W done by a force $\boldsymbol{F}$ as the force's point of application moves along a curve in space from position $\boldsymbol{r}_1$ to $\boldsymbol{r}_2$ is given by the line integral

$$W = \int_{\boldsymbol{r}_1}^{\boldsymbol{r}_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{s}. \tag{1.11}$$

- The power P (energy per unit time) through a surface S through which flows heat current density $\boldsymbol{j}$ (energy per unit time per unit area) is given by the surface integral

$$P = \iint_S \boldsymbol{j} \cdot \mathrm{d}\boldsymbol{S}. \tag{1.12}$$

[8] I've denoted work by the symbol W, from the English word "work", because this is most common in the modern scientific literature. In Slovenia, at least in introductory courses, work is commonly denoted by A, from the German word "Arbeit".

Don't worry if you don't understand all of the equations quite yet—you'll gain experience with them in this course, and learn them formally when your mathematics courses catch up with our curriculum over the next few months. The goal here is just to give a taste of how calculus is used to formulate the concepts of physics.

1.3 Physical quantities

After our quick tour through the *mathematical* formalism of vectors and calculus, we now consider how *physical* quantities are described and what they represent. For our purposes, a physical quantity is a quantity that can be quantitatively measured (i.e. measured with numbers).

1.3.1 Base quantities and the 2019 redefinition of units

Physics involves seven *base quantities* from which all other physical quantities are derived. These base quantities are time, distance, mass, electric current, temperature, amount of substance, and luminous intensity. These are shown in Table 1. The final quantity—luminous intensity—occurs rarely outside of photometry, and we will not study it in this course.

Different physical quantities are distinguished by their units. The International System of Units (abbreviated "SI") established by an international organization called the International Bureau of Weights and Measures defines the SI *base units* for each of the base quantities; we will quote these units shortly. But first, a (recent) historical aside: in May 2019, all SI units were completely redefined in terms of only physical constants. Using physical constants is good! Namely, units defined in terms of physical constants are more fundamental and more stable than arbitrarily-defined prototypes of human creation.

Quantity	Unit	Unit symbol
Time	second	s
Distance	meter	m
Mass	kilogram	kg
Electric current	ampere	A
Temperature	kelvin	K
Amount of substance	mole	mol
Luminous intensity	candela	cd

Table 1: The seven base quantities and their SI base units (discussed shortly).

As an example of the problems associated with human-created standards, the International Prototype of the Kilogram (a cylinder made of a platinum-iridium alloy whose mass served as the definition of the kilogram before the 2019 redefinition) was found to have a fluctuating mass (of the order $\Delta m \sim 10\,\mu\mathrm{g}$); we still lack a universally accepted explanation why. As a more amusing example, the yard (an imperial unit roughly equal to the meter) was once defined as the length from the nose to the thumb of the outstretched arm of King Henry I of England. Not exactly a scientifically rigorous definition of distance!

1.3.2 The SI base units

We now summarize the base quantities, their units, and the natural constants on the basis of which the units are defined.

1. The SI base unit of **time** is the **second** and is defined in terms of the unperturbed, ground state hyperfine transition frequency of the cesium-133 atom. Concretely, the second (symbol s) is defined such that this transition frequency takes exactly

$$\Delta\nu_{\mathrm{Cs}} \equiv 9\,192\,631\,770\,\mathrm{s}^{-1}. \tag{1.13}$$

   The relevant physics will be more clear after *Moderna fizika 1*, but the takeaway here is that the second is based on a precise and fundamental natural process—the cesium atom's transition frequency is an extremely consistent timekeeper.

   Previously, the second was defined in terms of the Earth's rotation cycle over the course of a day. Aside from being rather anthropocentric, this definition was plagued by instabilities in the Earth's rotation.

2. The SI base unit of **distance** is the **meter**. The meter (symbol m) is based on the speed of light in vacuum c and the above-defined second. The meter is defined such that speed of light in vacuum is exactly

$$c \equiv 299\,792\,458\,\mathrm{m\,s}^{-1}, \tag{1.14}$$

   assuming the second is defined in terms of $\Delta\nu_{\mathrm{Cs}}$.

3. The SI base unit of **mass** is the **kilogram** (symbol kg) and is based on a universal constant called Planck's constant, denoted by h. The kilogram is defined such that Planck's constant is exactly

$$h \equiv 6.626\,070\,15 \cdot 10^{-34}\,\mathrm{kg\,m}^2\,\mathrm{s}^{-1}, \tag{1.15}$$

   assuming the second is defined in terms of $\Delta\nu_{\mathrm{Cs}}$ and the meter in terms of c. You'll hear more the Planck constant in second-year courses.

   Actually implementing the relationship between the kilogram and Planck's constant involves an extremely precise electromechanical scale called a *Kibble balance*. You can read more about this in the BIPM's publication "*Mise en pratique* for the definition of the kilogram" available on the BIPM's page on the practical realization of the definition of the base units.

4. The SI base unit of **electric current** is the **ampere** (symbol A) and is based on the elementary charge (the charge of single proton). The ampere is defined such that the elementary charge is exactly

$$e_0 \equiv 1.602\,176\,634 \cdot 10^{-19}\,\mathrm{A\,s}, \tag{1.16}$$

   assuming the second is defined in terms of $\Delta\nu_{\mathrm{Cs}}$.

5. The SI base unit of **temperature** is the **kelvin** and is based on a constant called the Boltzman constant, denoted by k_{B}. The kelvin (symbol K) is defined such that the Boltzman constant is exactly

$$k_{\mathrm{B}} \equiv 1.380\,649 \cdot 10^{-23}\,\mathrm{kg\,m}^2\,\mathrm{s}^{-2}\,\mathrm{K}^{-1}, \tag{1.17}$$

assuming the kilogram, meter, and second are defined in terms of natural constants as described above.

6. The SI base unit of **amount of substance** (e.g. the number atoms in a sample of carbon, the number of gas particles in closed chamber of gas, etc.) is the **mole** and is based on the Avogadro constant N_A. The mole (symbol mol) is defined such that Avogadro is exactly

$$N_{\mathrm{A}} \equiv 6.022\,140\,76 \cdot 10^{23}\ /\mathrm{mole}. \tag{1.18}$$

In other words, one mole of "stuff" contains exactly $6.02214076 \cdot 10^{-23}$ elementary entities.

7. For the sake of completeness, the SI base unit of **luminous intensity** (which no one really uses besides photometrists, and we won't mention further in this course) is the **candela**. If you feel inspired, you can read about its formal definition in the SI Brochure, or just check the Wikipedia article on the candela.

1.3.3 Derived quantities

Derived quantities are constructed by combining base quantities through multiplication and division. We already encountered some derived quantities above. For example, speed is, mathematically, the quotient of distance and time, and the speed of light was quoted above in meters per second, i.e. units of distance per units of time.

Derived quantities deemed important enough are assigned their own *derived unit*; for example, energy is assigned a unit called the *joule*. Derived units are convenient: they save us from writing out long strings of base units. For example, you will probably agree that it is more convenient to quote energy in joules than in the equivalent base unit combination $\mathrm{kg\,m^2\,s^{-2}}$. Some of the first derived quantities that we will meet in this course, together with their units and conventional symbols, are given in Table 2.

Quantity	Symbol	SI Base Unit	Derived Unit
Velocity	v	$\mathrm{m\,s^{-1}}$	-
Momentum	p	$\mathrm{kg\,m\,s^{-1}}$	-
Acceleration	a	$\mathrm{m\,s^{-2}}$	-
Force	F	$\mathrm{kg\,m\,s^{-2}}$	N (newton)
Energy	E	$\mathrm{kg\,m^2\,s^{-2}}$	J (joule)
Power	P	$\mathrm{kg\,m^2\,s^{-3}}$	W (watt)

Table 2: A few of the first derived physical quantities we will encounter in mechanics, together with their SI base units and derived unt, if applicable.

1.3.4 Measuring average and instantaneous quantities

We go directly to an example: measuring a quantity called *volume flow rate*. Physically, volume flow rate represents the volume of "stuff" (typically fluid) moving past a region of space (such as a given point in a hose) per unit time.

Suppose we wish to measure the volume flow rate of water out of a pipe's faucet. If our tools are limited to everyday objects, we could measure the VFR using a bucket of known volume[9] ΔV and a stopwatch. The measurement proceeds in two steps:

(a) Simultaneously start the stopwatch and place the (empty) bucket under the faucet and observe the rising water level in the bucket as the bucket fills.

(b) Stop the stopwatch when the bucket is full and record the time interval Δt between the stopwatch stop and start times.

This measurement gives us the time needed to fill the known bucket volume. From ΔV and Δt, we determine the *average volume flow rate* out of the faucet over the course of the measurement via

$$\overline{Q} = \frac{\Delta V}{\Delta t}. \tag{1.19}$$

We stress that $\overline{Q}$ is a single scalar quantity[10] representing the average volume flow rate over the course of the entire measurement. It contains no information about the volume flow rate at a specific time during the measurement process.

Alternatively, suppose we have an advanced measurement instrument capable of continuously measuring the small volume of water $\mathrm{d}V$ leaving the faucet every small time interval $\mathrm{d}t$ (e.g. of milli- or microsecond order). The continuous stream of $\{\mathrm{d}V, \mathrm{d}t\}$ measurements at each point in time until the bucket is full can be used to approximate the *instantaneous volume flow rate*

$$Q(t) = \frac{\mathrm{d}V}{\mathrm{d}t}. \tag{1.20}$$

Importantly, $Q(t)$ gives the VFR through the faucet as a *function of time* throughout the measurement, while $\overline{Q}$ from Equation 1.19 is a single scalar value giving the average VFR over the course of the entire measurement. To be clear, however, both $\overline{Q}$ and $Q(t)$ represent the same physical phenomena (i.e. flow of fluid volume with respect to time) and have the same units (volume per time). Thus, $\overline{Q}$ and $Q(t)$ still correspond to the same physical quantity, i.e. volume flow rate.

[9] If unknown, we could estimate the bucket's volume with a tape measure and geometrical measurements. Better yet, if we had a scale handy, we could measure the mass of water needed to fill the bucket and convert to volume using water's known (room temperature and pressure) density $\rho \approx 1\,\mathrm{g\,cm^{-3}}$.

[10] I have used the symbol Q to denote volume flow rate, which might seem rather arbitrary at first glance. I agree. But Q is the conventional symbol for VFR and I figure it is better to encounter it sooner than later.

2 Mechanics

Mechanics is the field of physics that describes and predicts the motion of bodies. Mechanics divides into two sub-fields:

1. *kinematics*, which *describes* the motion of bodies, and
2. *dynamics*, which *predicts* the motion of bodies.

Kinematics is observational (i.e. quantifies a body's *current motion*) while dynamics, which is more powerful, can predict a body's *future motion* given its current state.

2.1 Physical models

Real-life physical objects are complicated—they are asymmetric, anisotropic, deformable, have non-homogeneous mass distributions, and so on. Exactly describing and predicting their motion in all its detailed complexity is difficult—in fact, exact analytical solutions are often impossible. Instead, physicists make a compromise: we *approximate* bodies and physical systems with simple models that are relatively easy to analyze, at the expense of a perfectly exact prediction. A simplification of a physical system that...

1. dramatically improves analysis and also
2. preserves a result reasonably in agreement with real-life behavior...

is called a *physical model*. Models aim to preserve only those properties of an object that are essential to predicting its motion and remove secondary details that significantly complicate analysis but produce only small corrections in observed motion.

Our plan in this chapter is to first study the kinematics and dynamics of a few useful models in a theoretical sense, and then apply the developed theory to approximate the motion of real-life objects. Certain models are particularly well-suited to mechanics (and many other branches of physics, too), and you will encounter them in any standard physics course. These are:

1. the point mass,
2. the system of point masses,
3. the rigid body, and
4. the linearly deformable body.

We will define these models shortly. Each is a successively better approximation of real-life objects, but each is also more difficult to work with analytically. In this course we will begin with the mechanics of a point mass, which we will later generalize to a system of point masses. We introduce rigid bodies in our treatment of rotational mechanics, and briefly cover deformable bodies in the lectures on elasticity and deformation. Rigid bodies are covered in much more detail in the second-year course *Klasična mehanika*, while deformable bodies and continuous media are covered in the third-year elective course *Mehanika kontinuov*.

2.1.1 The point mass

The simplest and most fundamental model is the point mass: a hypothetical object with its *entire mass concentrated at a single point in space.* To specify a point mass's state, you need to describe only its mass and its position at a *single* spatial coordinate. Importantly, you *don't* have to worry about a point mass's size, geometry, orientation in space, mass distribution, and so on—a spatial coordinate and the mass are all you need, at least for kinematics applications. In the context of dynamics, if you also specify the point mass's velocity at any given moment in time (in addition to its mass and position), you can predict everything there is to know about its future state within the scope of Newtonian mechanics.

Physical point masses do not occur in nature in the sense that you cannot walk down the street and find a body with infinitesimal size but finite mass. The closest physical objects to point masses, within the scope of current scientific knowledge, are elementary particles with no (currently) known internal structure, such as electrons or quarks. But this might be a limitation of current experimental technology rather than a fundamental truth. For our purposes, it suffices to remember that perfect point masses do not occur in nature, although some elementary particles currently appear to come close.

Validity of the point mass model

A finite-sized object may be treated as a point mass if the object's size is insignificant in the analysis—this is often the case when the ambient physical system or the distance traveled by the object is much larger than object itself. For example, the Earth on the scale of the solar system, an electron on the scale of an atomic nucleus, or an ion on the scale of a particle accelerator are all excellent candidates for a point mass approximation.

A word of caution: The validity of a point mass (or any other physical model) in describing a physical object *depends crucially upon context and scale.* For example, the Earth is very well described as a point mass on the scale of the solar system (for example in the context of predicting planetary orbits using Newtonian gravitation), but the Earth is certainly nothing like a point mass from the perspective of a human being walking on its surface.[11] Note the difference in scales: the Earth's size is insignificant on the scale of the solar system but very large on the scale of a human being. A general statement like "we can model the Earth as a point mass" is not well defined—you must say "we can model the Earth as a point mass at XYZ scale in the context of ABC analysis".

To conclude, here are a few miscellaneous comments about point masses:

- Within the realm of experimental error, a point mass approximation and an exact analysis of a body may be indistinguishable.
- Spherically-symmetrical objects under the influence of inverse square laws (such as the gravitational and electrostatic forces) behave *exactly* as point particles.
- The terms "point particle" and "point mass" are often used interchangeably.

[11] Well, its gravitational field actually comes pretty close—more on this in the lectures on gravitation—but at least the Earth's geometry is not remotely point-like in this context.

2.1.2 Other physical models

In passing, we now briefly define the other common physical models mentioned in this chapter's introduction.

- A **system of point masses** is exactly what it sounds like—a set of multiple point masses. This model can be useful for describing mutually-interacting (often spherical) objects. The Earth-Moon-Sun system, for example, is well-described as a system of point masses (a point-mass approximation is valid because the distances between the planets are much larger than the planets themselves).

  Alternatively, when the number of points is large, systems of point masses can be used to model continuous objects, such as a ball, disk, or rod.[12] This approximation works particularly well if the object is homogeneous, and we will return to this concept in the lectures on center of mass, center of gravity, and moment of inertia.

- A **rigid body** is an object in which the relative distances between all constituent points are constant. In practice, this means a rigid body cannot deform when subject to outside forces—all points in the body retain their original orientation no matter how hard you push it. For orientation, on the scale of everyday stresses, a slab of concrete is very well-described as a rigid body, while human muscle or a piece of gelatin are poor examples of rigid bodies—both deform if you poke or squeeze them. Note that a rigid body and a system of point masses are not mutually exclusive—any system of point masses in which the distances between all points is fixed is also a rigid body.

  Rigid bodies are the canonical model in the elementary theory of rotational mechanics—we will return to them when studying rotation, and you will study them in more detail in the second-year course *Klasična mehanika*.

- A **linearly deformable body** is a generalization of the rigid body in which the relative distances between constituent points are allowed to vary in response to outside stresses, but the resulting deformation must be *linear* in the applied stress. Linearly deformable bodies are the canonical model in the elementary theory of elastomechanics, and are covered in detail, using a tensor formalism, in the third-year course *Mehanika kontinuov*. We will cover them only briefly in the lectures on deformation using scalars and vectors.

2.2 Kinematics

Kinematics is a quantitative description of the motion of bodies. After covering the *description* of motion using kinematics, we will be equipped to *predict* motion using dynamics.

[12] In practice, *numerical* computer simulations treat physical objects as systems of discrete points. Although the number of points can be very large on modern hardware, it is still finite, since no physical computer can store in memory the theoretically infinite number of points needed to describe a continuous body.

2.2.1 Position and trajectory

Position is a vector quantity, conventionally denoted by $\boldsymbol{r}$, that specifies a body's location in space at a given point in time. But we usually aren't satisfied with knowing a body's position at a just *single* point in time—we want to know the position for *all* times (or at least over an interval of time). In the language of kinematics, this desired quantity is called a *trajectory*, which is a body's position in space *as a function of time*.

Position, in the three-dimensional Euclidean space of everyday life and classical physics, is fully specified by a three-dimensional position vector $\boldsymbol{r} \in \mathbb{R}^3$. A trajectory, which describes how a body's position vector changes over time, is specified by a vector-valued function $\boldsymbol{r}(t)$. A trajectory associates with every time t (a scalar quantity $t \in \mathbb{R}$) a corresponding position vector $\boldsymbol{r}$ (a three-dimensional vector $\boldsymbol{r} \in \mathbb{R}^3$).[13]

Coordinate representation of position

After specifying a coordinate system and basis for the space $\mathbb{R}^3$, a position vector $\boldsymbol{r}$ can be represented with three coordinates, for example the (x, y, z) coordinates of the Cartesian coordinate system you are probably familiar with from high school. In a Cartesian coordinate system, a position vector and a trajectory would be represented in coordinate form as

$$\boldsymbol{r} = (x, y, z) \qquad \text{and} \qquad \boldsymbol{r}(t) = \big(x(t), y(t), z(t)\big). \tag{2.1}$$

For the purposes of this course, unless explicitly stated otherwise, we will work in three-dimensional Euclidean space and perform analysis in a Cartesian coordinate system using the standard basis. Translated to everyday language, this means we will continue using the (x, y, z) coordinate system you are familiar with from high school, and, if desired, you can forget about coordinate systems, bases, and basis vectors until *Matematika 2*.[14]

Note: distinguishing $\boldsymbol{r}$ and r

In your study of physics you will often encounter both the symbols $\boldsymbol{r}$ and r. These always represent different physical quantities:

- $\boldsymbol{r}$ is a vector quantity and, nearly universally, is used to represent a position in three-dimensional space.
- r is a scalar quantity usually used to represent a body's distance from a coordinate system's origin. In this usage, r would be the length (also "magnitude" or "norm") of a position vector, i.e. $r = |\boldsymbol{r}|$.[15]

[13] To make a connection to the formal notation of real analysis introduced in *Matematika 1*, you would define a trajectory as a vector-valued, single-variable function $\boldsymbol{r} : \mathbb{R} \to \mathbb{R}^3$ with $t \mapsto \boldsymbol{r}(t)$.

[14] In general, representing a vector in terms of coordinates requires specifying a basis and coordinate system for the ambient vector space, which in this course will always be the 3D Euclidean space familiar from everyday life—this space is called $\mathbb{R}^3$ in the notation of mathematics. You will learn about bases and coordinates formally in *Matematika 2* and get practical knowledge with common coordinate systems in *Proseminar A/B*. The point, for now, is to know that coordinate systems other than the (x, y, z) Cartesian system familiar from high school exist, and $\boldsymbol{r} = (x, y, z)$ is just a special case of writing a vector in coordinate form.

[15] The mathematically correct notation for a vector norm would actually be $\|\boldsymbol{r}\|$ (with $|\cdot|$ reserved for the magnitude of scalars) but physicists are sloppy and use the absolute value sign $|\boldsymbol{r}|$ for vector magnitudes, too. If you ever see the absolute value of a vector in a physics context, you can assume it denotes the vector's magnitude (i.e. the length or norm).

The symbol r can also denote the radial coordinate in a spherical coordinate system—you'll learn more about this in *Proseminar A/B*. Loosely, the net effect is the same: r represents the scalar distance from the origin.

Plotting trajectories

In one spatial dimension, or for a single coordinate, we conventionally show a trajectory by plotting the coordinate (on the ordinate axis) as a function of time (on the abscissa), while two-dimensional trajectories are often plotted in a coordinate plane with time as a parameter. Three-dimensional trajectories can also be plotted parametrically in three dimensions, but the result is often difficult to instructively interpret.

TODO: create example plots. Some ideas (for example):

- $z(t)$ for a particle in free fall
- $(x(t), y(t))$ parametrically for a particle in an EM field
- $\boldsymbol{r}(t)$ parametrically for the moon's orbit around the earth

2.2.2 Displacement

Consider a particle with initial position $\boldsymbol{r}_1$ and final position $\boldsymbol{r}_2$. The difference of these two vectors is called a *displacement* and is denoted by

$$\Delta\boldsymbol{r} = \boldsymbol{r}_2 - \boldsymbol{r}_1; \tag{2.2}$$

it is conventional to formulate displacement as final position minus initial position. Displacements have a very useful property: *a displacement between two positions is the same regardless of the coordinate system origin with respect to which the positions are measured.* This is because any relative offset of a coordinate system's origin is canceled in the difference of $\boldsymbol{r}_2$ and $\boldsymbol{r}_1$.

Differences and differentials

Consider a moving object traveling through space along some trajectory $\boldsymbol{r}(t)$, and imagine repeatedly observing the object's position. Now imagine measuring this position more and more frequently—over smaller and smaller time intervals. Of course, in practice you can't measure over arbitrarily small times—the smallest time interval can be as short, but no shorter, than the available experimental equipment can accurately measure. This experimentally bounded time interval between successive position measurements is a *finite*, measurable quantity, and is mathematically classified as a *difference*—we denote it by Δt. Associated with each interval Δt is the corresponding displacement $\Delta\boldsymbol{r}$ between the body's position at the, say, i-th and $(i+1)$-th measurement, i.e. the change in position between two subsequent measurements. Interpreted physically, $\Delta\boldsymbol{r}$ encodes the net direction and distance the body moved between two measurements. Like Δt, the difference $\Delta\boldsymbol{r}$ is finite and measurable and (assuming the body is not at rest) can only be as small, but no smaller, as the available experimental equipment allows.

But in *theory*—and this concept might be familiar from a differential calculus course from high school—the time interval between successive displacement measurements, and thus the corresponding displacement, can be made arbitrarily small. So small, in fact, that successive measurements are separated by nothing more than a hypothetical instant. This "infinitely small" time interval is mathematically classified as a *differential*, and we denote it by $\mathrm{d}t$. The associated differential change in position during the time $\mathrm{d}t$ is denoted by $\mathrm{d}\boldsymbol{r}$. Differentials are *infinitesimal*—they are so small that they exist only in theory, but are too small to be actually measured. Loosely, it may help to think of $\mathrm{d}t$ as the limit of the observation interval Δt as Δt approaches zero, i.e.

$$\mathrm{d}t = \lim_{\Delta t \to 0} \Delta t. \tag{2.3}$$

Note that if you are not familiar with limits and differentials from high school, it would be completely understandable if these concepts don't yet make sense on the basis of the above explanations alone. We are covering the material much faster than one would in a dedicated differential calculus course, since most students will have already seen the material in high school. If that is not the case for you, don't worry—you will catch up soon in *Matematika 1*, and for now just try to understand the general concepts.

Relating infinitesimal and differential quantities

Our goal in this section is to show how to analytically relate differences—which are finite—and differentials—which are infinitesimal.

- Let t_0 and t_N denote the time at the first and last measurement, and let $\boldsymbol{r}_0 = \boldsymbol{r}(t_0)$ and $\boldsymbol{r}_N = \boldsymbol{r}(t_N)$ denote the positions of the measured body at the initial and final times t_0 and t_N.
- Let $\Delta t = t_N - t_0$ and $\Delta \boldsymbol{r} = \boldsymbol{r}_N - \boldsymbol{r}_0$ denote the differences in initial and final time, and initial and final position, for the entire experiment.
- Let $\Delta t_i = t_i - t_{i-1}$ and $\Delta \boldsymbol{r}_i = \boldsymbol{r}_i - \boldsymbol{r}_{i-1}$ denote the differences in time and position between the i-th and $(i-1)$-th measurements.

We can then relate the net changes in time and position Δt and $\Delta \boldsymbol{r}$ for the entire experiment to the changes over individual measurements according to

$$\Delta t = \sum_{i=1}^{N} \Delta t_i \qquad \text{and} \qquad \Delta \boldsymbol{r} = \sum_{i=1}^{N} \Delta \boldsymbol{r}_i. \tag{2.4}$$

Next, consider the (theoretical) limit of infinitely frequent measurements, meaning that $N \to \infty$ and $\Delta t_i \to 0$ for all i. In this limit case (2.4) generalizes to

$$\Delta t = \lim_{N \to \infty} \sum_{i=1}^{N} \Delta t_i \overset{\text{(a)}}{\equiv} \int_{t_0}^{t_N} \mathrm{d}t \qquad \text{and} \qquad \Delta \boldsymbol{r} = \lim_{N \to \infty} \sum_{i=1}^{N} \Delta \boldsymbol{r}_i \overset{\text{(b)}}{\equiv} \int_{\boldsymbol{r}_0}^{\boldsymbol{r}_N} \mathrm{d}\boldsymbol{r}, \tag{2.5}$$

where in (a) and (b) we have written the infinite sums in a more convenient notation using *integral symbols*. Equation (2.5) is important—it provides a formalism for converting between finite quantities and differentials. Loosely, but instructively, we can interpret the relationship between differentials and differences in (2.5) as follows:

(a) First, for orientation, we stress that the differences Δt and $\Delta \boldsymbol{r}$ are *finite* quantities (i.e. macroscopic and measurable), while the differentials $\mathrm{d}t$ and $\mathrm{d}\boldsymbol{r}$ are *infinitesimal* quantities (i.e. so small they exist only in a theoretical sense).

(b) We then interpret (2.5) as formalizing the intuitive idea that a sum of infinitely many infinitesimally small quantities produces a finite quantity.

Mathematically, the procedure we have just performed is called Riemann integration—you will study it formally in *Matematika 1* in the chapters on integral calculus. Limits and infinite sums like in (2.5) raise questions of convergence, and in *Matematika 1* you will cover the conditions that the functions t and $\boldsymbol{r}$ must satisfy for the sums in (2.5) to be well-defined. Fortunately, physical quantities in Newtonian mechanics are well-behaved, and in this course we won't have to worry about convergence. For our purposes, for the time being, we are satisfied with the following ideas:

- We convert from between differentials and differences (e.g. between $\mathrm{d}\boldsymbol{r}$ and $\Delta\boldsymbol{r}$) through a sum of infinitely many infinitesimal quantities, which is formally called a Riemann integral.

- If a sum of individual differences (e.g. Δt_i or $\Delta\boldsymbol{r}_i$) converges to a finite net quantity (e.g. Δt or $\Delta\boldsymbol{r}$) for ever more frequent measurements (larger N) and continually smaller times and displacements, *within the scope of physical measurement*, we'll assume the sum's limit, as defined in (2.5), exists and is well-defined.

Total distance traveled

Suppose you travel from Ljubljana to Cambridge and then return to Ljubljana. Since the journey begins and ends in Ljubljana, the journey's displacement is zero, but—as anyone who has made the journey can immediately tell you—the total distance traveled is certainly not zero. In fact, the net displacement along any trajectory with the same start and end position (for example any closed loop) is zero, but the trajectory's total arc length could take on any value. Motivated by the wish to better distinguish such trajectories, we associate total distance traveled with a new physical quantity, typically denoted by s. Precisely, the total distance s traveled along a trajectory $\boldsymbol{r}(t)$ beginning at initial position $\boldsymbol{r}_1$ and ending at final position $\boldsymbol{r}_2$ is

$$s = \int_{\boldsymbol{r}_1}^{\boldsymbol{r}_2} |\mathrm{d}\boldsymbol{r}| \overset{\text{(a)}}{=} \int_{\boldsymbol{r}_1}^{\boldsymbol{r}_2} \mathrm{d}s, \tag{2.6}$$

where in (a) we have introduced the shorthand notation $\mathrm{d}s = |\mathrm{d}\boldsymbol{r}|$ to denote the magnitude of the vector differential $\mathrm{d}\boldsymbol{r}$.[16] Equation (2.6) formalizes a simple idea: to get the total distance traveled over an entire journey, divide the journey into many small steps $\Delta\boldsymbol{r}$, record the length $\mathrm{d}s = |\Delta\boldsymbol{r}|$ of each step as you go, and add the lengths together. Intuitively, imagine following a trajectory through space from $\boldsymbol{r}_1$ to $\boldsymbol{r}_2$ and maintaining a running sum of your step lengths as you go. We stress that s is a scalar quantity—this follows (i) intuitively because total distance traveled is just a single number and (ii) mathematically because integrating a scalar magnitude $|\mathrm{d}\boldsymbol{r}|$ produces a scalar result.

Computing total distance traveled

[16] The use of $\mathrm{d}s$ (instead of $\mathrm{d}r$) to represent the magnitude $|\mathrm{d}\boldsymbol{r}|$ is intentional, since $\mathrm{d}r$ can be confused with the differential of the radial coordinate in spherical coordinate systems, while $\mathrm{d}s$ is exclusively used for total distance traveled.

In Cartesian coordinates, we compute $\mathrm{d}s$ with the Pythagorean theorem,

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2, \tag{2.7}$$

where in this context $\mathrm{d}s^2$ is shorthand for $(\mathrm{d}s)^2$ and not for $\mathrm{d}(s^2)$. In other coordinate systems the expression for $\mathrm{d}s$ can be more complicated, and in two or more spatial dimensions practical computation of (2.6) involves a line integral, which is a subject we leave for *Matematika 3*. You won't be expected to compute line integrals in this course; the point for now is to introduce the *concept* of total distance traveled and show how one formalizes the idea, as in (2.6), in the language of vector calculus.

2.2.3 Velocity

Recall from Section 2.2.2 the idea of observing a body moving through space along some trajectory $\boldsymbol{r}(t)$, and continuously measuring the body's position over smaller and smaller time intervals Δt. Let $\Delta\boldsymbol{r}$ denote the body's change in position over a measurement interval Δt. Then associated with each measurement pair $(\Delta\boldsymbol{r}, \Delta t)$ is an *average velocity*, defined as

$$\overline{\boldsymbol{v}} = \frac{\Delta\boldsymbol{r}}{\Delta t}. \tag{2.8}$$

Interpreted physically, $\overline{\boldsymbol{v}}$ encodes in which direction and how fast a body moved between the two different positions and times associated with the differences $\Delta\boldsymbol{r}$ and Δt. From vector algebra, a vector's direction remains the same after multiplication or division with a scalar, so $\overline{\boldsymbol{v}}$ has the same direction in space as the displacement $\Delta\boldsymbol{r}$ (since Δt is a scalar). And because the time interval Δt carries a physical unit, i.e. time, the quotient $\Delta\boldsymbol{r}/\Delta t$ is a new physical quantity—which we have already defined as velocity—with units of distance per time (meters per second in SI units).

Instantaneous velocity

The question that naturally arises next is:

> If (2.8) gives a body's velocity between *two different* positions and times along a given trajectory, what is a body's velocity at *every* position and time?

The desired quantity is called *instantaneous velocity* and is conventionally denoted by $\boldsymbol{v}(t)$. To determine velocity at every point on a given trajectory, one observes the average velocity $\overline{\boldsymbol{v}} = \Delta\boldsymbol{r}/\Delta t$ over smaller and smaller time intervals—mathematically, in the limit as $\Delta t \to 0$. Intuitively, this just means observing the moving body's position at every possible instant in time. In any case, instantaneous velocity is defined as

$$\boldsymbol{v} = \lim_{\Delta t \to 0} \frac{\Delta\boldsymbol{r}}{\Delta t} = \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}. \tag{2.9}$$

In Cartesian coordinates, instantaneous velocity is represented component-wise as

$$\boldsymbol{v} = (v_x, v_y, v_z) = \left(\frac{\mathrm{d}x}{\mathrm{d}t}, \frac{\mathrm{d}y}{\mathrm{d}t}, \frac{\mathrm{d}z}{\mathrm{d}t}\right). \tag{2.10}$$

Convergence of the difference quotient

The limit in (2.9) raises the question of convergence. Loosely, the expression $\Delta\boldsymbol{r}/\Delta t$

as $\Delta t \to 0$ is essentially the quotient $\mathbf{0}/0$, which is undefined in general. But in the scope of physics, the limit in (2.9) exists and converges to well-defined velocity (at least on physical grounds this is obvious: any physical object's velocity is always finite—you cannot walk down the street and encounter a body with infinite velocity). On mathematical grounds, the limit defining $\boldsymbol{v}$ exists under the assumption that the trajectory $\boldsymbol{r}(t)$ is a continuously differentiable function of time.

You will study the existence of limits and difference quotients formally using the tools of real analysis developed in *Matematika 1*, but in classical physics there is not much to worry about. Because physical quantities are generally well-behaved, we can safely assume expressions like (2.9) are well-defined. That said, it is important to mention the general issue of convergence at least once.[17]

In the language of differential calculus, which you have probably encountered in high school and will soon cover in *Matematika 1*, the instantaneous velocity from (2.9) is called the *first derivative of position with respect to time*. Mathematicians and physicists have come up with various ways to write derivatives, all of which are equivalent. Here are some examples in the case of velocity:

$$\boldsymbol{v} = \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = \dot{\boldsymbol{r}}(t) = \boldsymbol{r}'(t). \tag{2.11}$$

In physics, a dot over a quantity, as in $\dot{\boldsymbol{r}}$, universally denotes the quantity's derivative with respect to time. You'll see plenty of this dot notation in the second-year course *Klasična mekanika*. The notation $\boldsymbol{r}'(t)$ is more common in mathematics—a primed function denotes the function's derivative with respect to its argument. When a function's argument is time, such as $\boldsymbol{r}(t)$, the notations $\dot{\boldsymbol{r}}(t)$ and $\boldsymbol{r}'(t)$ are equivalent.

Finally, we note that the direction of the instantaneous velocity vector $\boldsymbol{v}$ at time t is tangent to the corresponding trajectory $\boldsymbol{r}(t)$. Loosely, this is just a generalization of average velocity $\overline{\boldsymbol{v}}$ being parallel to the corresponding secant line $\Delta\boldsymbol{r}$; in the limit of instantaneous velocity, the secant line to the trajectory converges to the tangent line. More formally, you will show that velocity is tangent to trajectory in the lectures on space curves in *Matematika 1* and *Matematika 3*.

Velocity and differential equations

We now return to the concept, first introduced in (2.5), of converting between finite and infinitesimal quantities. Specifically, we will show how, given a particle's velocity $\boldsymbol{v}(t)$, it is possible to recover the particle's trajectory by solving a *differential equation* involving the infinitesimals $\mathrm{d}\boldsymbol{r}$ and $\mathrm{d}t$.

We begin with the definition of velocity, $\boldsymbol{v} = \mathrm{d}\boldsymbol{r}/\,\mathrm{d}t$, and, loosely, imagine multiplying through[18] by $\mathrm{d}t$ to get

$$\boldsymbol{v} = \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} \quad \longrightarrow \quad \mathrm{d}\boldsymbol{r} = \boldsymbol{v}\,\mathrm{d}t. \tag{2.12}$$

[17] In physics questions like these don't bother us too much, largely because they don't have too—physicists are blessed with quantities described by well-behaved functions. We are more interested in the practical use of given equations instead of the details of their existence. In fact, for better or worse, physicists often take pleasure in a tongue-in-cheek disregard for mathematical rigor. But this is of course tongue-in-cheek, and any serious physicist will have a strong training in mathematics.

[18] Formally, multiplication by differentials is not quite the same as regular scalar multiplication, even though the process appears analogous. But the details are beyond the scope of this course—you will discuss this in *Matematika 1*.

Caution: when manipulating equations involving differentials, the left- and right-hand sides of the resulting equality *must* be of the same multiplicity in the differentials. For example, $\mathrm{d}\boldsymbol{r} = v\,\mathrm{d}t$ is a valid expression (both sides involve first-order differentials) while $\mathrm{d}\boldsymbol{r} = v \cdot t$ is not (the left-hand side contains a first-order differential and the right-hand side does not contain any differentials). Remember that a differential is an infinitesimally small quantity; loosely, a differential on one side of an equation but not the other would be like equating a finite value to zero.

In any case, to recover a finite-valued trajectory $\boldsymbol{r}(t)$ from the differential equation $\mathrm{d}\boldsymbol{r} = \boldsymbol{v}\,\mathrm{d}t$, we must integrate both sides of the equation, written schematically as

$$\int \mathrm{d}\boldsymbol{r} = \int \boldsymbol{v}\,\mathrm{d}t. \tag{2.13}$$

Of course we must also specify the integration limits, and in this context there are two common choices:

(a) One could integrate over a *definite* interval, say from t_1 to t_2 for time and $\boldsymbol{r}_1 \equiv \boldsymbol{r}(t_1)$ to $\boldsymbol{r}_2 \equiv \boldsymbol{r}(t_2)$ for position. In this case both the upper and lower integration limits are fixed values.

(b) Alternatively, one integrates over an *indefinite* interval in which the lower limits, say t_0 and $\boldsymbol{r}_0 \equiv \boldsymbol{r}(t_0)$ are fixed, but the upper limits are allowed to vary arbitrarily and are written as the generic functions t and $\boldsymbol{r}(t)$.

Depending on the notation (we will comment on this shortly), these two choices could be written

$$\int_{\boldsymbol{r}_1}^{\boldsymbol{r}_2} \mathrm{d}\boldsymbol{r} = \int_{t_1}^{t_2} \boldsymbol{v}(t)\,\mathrm{d}t \tag{2.14a}$$

$$\int_{\boldsymbol{r}_0}^{\boldsymbol{r}(t)} \mathrm{d}\boldsymbol{\rho} = \int_{t_0}^{t} \boldsymbol{v}(\tau)\,\mathrm{d}\tau. \tag{2.14b}$$

In either case the limits of integrations over time and position must correspond. For example, in (2.14a) the position limit $\boldsymbol{r}_2 = \boldsymbol{r}(t_2)$ is the position at the time limit t_2, just like $\boldsymbol{r}(t_1)$ is the position at time t_1.

We now comment on the notation in (2.14b). When, as in (2.14b), the symbol for the variable of integration would clash with the symbol for a limit of integration, one should change the integration variable to a different symbol. Since that might sound vague, here is an example of what *not* to do:

$$\int_{\boldsymbol{r}_0}^{\boldsymbol{r}(t)} \mathrm{d}\boldsymbol{r} = \int_{t_0}^{t} \boldsymbol{v}(t)\,\mathrm{d}t. \qquad \text{(example of poor notation)} \tag{2.15}$$

Notice how the upper integration limits—$\boldsymbol{r}(t)$ and t—clash with the variables of integration, which are also written as $\boldsymbol{r}$ and t.[19] Although admittedly somewhat

[19]On the RHS the clash involving time t is immediately obvious; on the LHS, even though the limit $\boldsymbol{r}(t)$ and differential $\mathrm{d}\boldsymbol{r}$ don't match explicitly, we must recall that the differential $\mathrm{d}\boldsymbol{r}$ is a function of time and could really be written $\mathrm{d}\boldsymbol{r}(t)$, in which case the clash with the limit $\boldsymbol{r}(t)$ is clear. We generally don't write differentials' functional dependence explicitly, since equations would get too cluttered, and infer their dependence on other variables from context.

pedantic, this sort of conflict means the integral is not mathematically well-defined. To resolve this clash, one modifies the symbol used for the integration variable—two common choices are adding a tilde, e.g. $\widetilde{\boldsymbol{r}}$ to replace $\boldsymbol{r}$, or replacing the letter with its Greek equivalent, e.g. $\boldsymbol{\rho}$ for $\boldsymbol{r}$ or τ for t. The result would be

$$\int_{\boldsymbol{r}_0}^{\boldsymbol{r}(t)} \mathrm{d}\widetilde{\boldsymbol{r}} = \int_{t_0}^{t} v(\widetilde{t})\,\mathrm{d}\widetilde{t} \qquad \text{or} \qquad \int_{\boldsymbol{r}_0}^{\boldsymbol{r}(t)} \mathrm{d}\boldsymbol{\rho} = \int_{t_0}^{t} \boldsymbol{v}(\tau)\,\mathrm{d}\tau, \tag{2.16}$$

just like in (2.14b). Of course, since physicists are sloppy with notation, you will see plenty of examples of (2.15) during your course of study, but keep in mind that (2.16) is a better choice, since the integration variable and limit don't conflict. And, if you see strange Greek letters appearing in integral expressions, you now know why.

After commenting on notation, we now return to the initial goal of solving (2.14) for a trajectory $\boldsymbol{r}(t)$; for review, (2.14) reads

$$\int_{\boldsymbol{r}_1}^{\boldsymbol{r}_2} \mathrm{d}\boldsymbol{r} = \int_{t_1}^{t_2} \boldsymbol{v}(t)\,\mathrm{d}t \tag{2.17a}$$

$$\int_{\boldsymbol{r}_0}^{\boldsymbol{r}(t)} \mathrm{d}\boldsymbol{\rho} = \int_{t_0}^{t} \boldsymbol{v}(\tau)\,\mathrm{d}\tau, \tag{2.17b}$$

We treat the two cases separately:

(a) Integrating the left-hand side of (2.17a) gives

$$\boldsymbol{r}_2 - \boldsymbol{r}_1 = \int_{t_1}^{t_2} \boldsymbol{v}(t)\,\mathrm{d}t \implies \boldsymbol{r}_2 = \boldsymbol{r}_1 + \int_{t_1}^{t_2} \boldsymbol{v}(t)\,\mathrm{d}t \tag{2.18}$$

Interpreted physically, (2.18) expresses the final position on the trajectory $\boldsymbol{r}_2$ in terms of the initial position $\boldsymbol{r}_1$ and an integral of velocity over the interval $[t_1, t_2]$. By the way, (2.5), which we met earlier, is an example of this type of definite integral.

(b) Integrating the left-hand side of (2.14b) gives

$$\boldsymbol{r}(t) - \boldsymbol{r}_0 = \int_{t_0}^{t} \boldsymbol{v}(t)\,\mathrm{d}t \implies \boldsymbol{r}(t) = \boldsymbol{r}_0 + \int_{t_0}^{t} \boldsymbol{v}(t)\,\mathrm{d}t. \tag{2.19}$$

Interpreted physically, (2.19) gives a body's trajectory $\boldsymbol{r}(t)$ in terms of an initial position $\boldsymbol{r}_0$ and an integral of velocity from the initial time t_0 to the time t at which the trajectory is evaluated.

In either case, actually solving for $\boldsymbol{r}_2$ or $\boldsymbol{r}(t)$ requires evaluating the integral of velocity; this could be anywhere from trivial to analytically impossible, depending on the functional form of $\boldsymbol{v}(t)$.

2.2.4 Acceleration

Having just introduced velocity—the rate of change of *position* with respect to time—it is natural to ask how a body's *velocity* itself changes with time. The rate of change of velocity with respect to time is called *acceleration*, and is derived in an analogous manner to the derivation of velocity in Section 2.2.3—it might be helpful to revisit Section 2.2.3 now.

Average acceleration

As in Section 2.2.3 and Section 2.2.2 before it, imagine observing a body moving through space along some trajectory $\boldsymbol{r}(t)$. To derive acceleration, imagine repeatedly measuring the body's velocity $\boldsymbol{v}(t)$, and continuously computing the changes in the body's velocity $\Delta \boldsymbol{v}$ between sequential measurements. Once again, let Δt denote the difference in time between subsequent measurements. Then associated with each measurement pair $(\Delta \boldsymbol{v}, \Delta t)$ is an *average acceleration*, defined as

$$\overline{\boldsymbol{a}} = \frac{\Delta \boldsymbol{v}}{\Delta t}. \tag{2.20}$$

Interpreted physically, $\overline{\boldsymbol{a}}$ encodes in which direction, and how quickly, a body's *velocity* changed over the measurement interval Δt. Note that an immediate, intuitive interpretation of acceleration is often difficult for people, so don't worry if it takes some time to come to terms with the concept. This is probably because the majority of daily life—besides brief spurts of speeding up or slowing down—is carried out with constant velocity, or zero acceleration. In any case, you will soon become familiar with acceleration from your study of mechanics.

Instantaneous acceleration

To derive instantaneous acceleration, just like in the derivation of instantaneous velocity, one imagines computing average acceleration $\overline{\boldsymbol{a}} = \Delta \boldsymbol{v}/\Delta t$ over smaller and smaller time intervals—mathematically, in the limit as $\Delta t \to 0$. The resulting instantaneous acceleration is defined as

$$\boldsymbol{a} = \lim_{\Delta t \to 0} \frac{\Delta \boldsymbol{v}}{\Delta t} = \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}. \tag{2.21}$$

Acceleration is the first derivative of velocity with respect to time and has SI units of meters per second squared. Acceleration can also be interpreted as the second derivative of position with respect to time, which is clear if one writes

$$\boldsymbol{a} = \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} \overset{\text{(a)}}{=} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t}\right) \overset{\text{(b)}}{=} \frac{\mathrm{d}^2\boldsymbol{r}}{\mathrm{d}t^2}, \tag{2.22}$$

where in (a) we have used the definition of velocity $\boldsymbol{v} = \mathrm{d}\boldsymbol{r}/\,\mathrm{d}t$ and in (b) we have introduced the standard notation for higher-order derivatives, which you will soon encounter in *Matematika 1*. In Cartesian coordinates, instantaneous acceleration has the following, equivalent, coordinate representations:

$$\boldsymbol{a} = (a_x, a_y, a_z) = \left(\frac{\mathrm{d}v_x}{\mathrm{d}t}, \frac{\mathrm{d}v_y}{\mathrm{d}t}, \frac{\mathrm{d}v_z}{\mathrm{d}t}\right) = \left(\frac{\mathrm{d}^2x}{\mathrm{d}t^2}, \frac{\mathrm{d}^2y}{\mathrm{d}t^2}, \frac{\mathrm{d}^2z}{\mathrm{d}t^2}\right). \tag{2.23}$$

Decomposition of acceleration

Consider a body moving along a trajectory through space with velocity $\boldsymbol{v}(t)$ and acceleration $\boldsymbol{a}(t)$. For the decomposition of acceleration (which we will introduce shortly) to make sense, let $\hat{\mathbf{v}}$ denote the dimensionless unit vector[20] in the direction

[20] The hat notation (e.g. $\hat{\mathbf{v}}$) is conventionally used to denote dimensionless direction vectors of unit norm—for example, you might have seen $\hat{\mathbf{x}}$ or $\hat{\mathbf{\imath}}$ used to denote the direction in space of the x axis.

of the body's instantaneous velocity,

$$\hat{\mathbf{v}}(t) = \frac{\boldsymbol{v}(t)}{|\boldsymbol{v}(t)|} \overset{\text{(a)}}{\equiv} \frac{\boldsymbol{v}(t)}{v(t)}, \tag{2.24}$$

where in (a) we have introduced the common shorthand notation $v(t) = |\boldsymbol{v}(t)|$ for the magnitude of the body's velocity—this is easier to write than $|\boldsymbol{v}(t)|$. In everyday language, $\hat{\mathbf{v}}(t)$ is just the direction in which the body is moving at time t. More precisely, $\hat{\mathbf{v}}$ is the unit tangent vector to the body's trajectory and would usually be written as $\mathbf{T}$ in the mathematical context of space curves.

Having introduced $\hat{\mathbf{v}}$, it is often useful to decompose the body's acceleration $\boldsymbol{a}$ into two independent components:

1. a component $\boldsymbol{a}_{\|}$ *parallel* to the body's instantaneous velocity, i.e. parallel to $\hat{\mathbf{v}}$, and

2. a component $\boldsymbol{a}_{\perp}$ *perpendicular* to the body's instantaneous velocity, i.e. perpendicular to $\hat{\mathbf{v}}$.

One could then decompose a body's total acceleration into the vector sum

$$\boldsymbol{a} = \boldsymbol{a}_{\|} + \boldsymbol{a}_{\perp}. \tag{2.25}$$

Both components have instructive physical interpretations:

- The component of acceleration parallel to velocity changes the *magnitude* of a body's velocity. In everyday language, $\boldsymbol{a}_{\|}$ is responsible for a body speeding up or slowing down.

- The component of acceleration perpendicular to velocity changes the *direction* of velocity. In everyday language, $\boldsymbol{a}_{\perp}$ is responsible for making a body's path twist, turn, or bend as the body moves through space.

Our goal here is just to gain some insight into what acceleration physically does, and we introduced the $\boldsymbol{a} = \boldsymbol{a}_{\|} + \boldsymbol{a}_{\perp}$ decomposition in the hope that changing a body's speed and direction of motion are two easily-understood concepts. Note that actually computing the components $\boldsymbol{a}_{\|}$ and $\boldsymbol{a}_{\perp}$ for the general case of arbitrary three-dimensional motion is more complicated and requires the study of space curves, a branch of vector analysis. You will analyze this general case in the lectures on the Frenet-Serret formulas in *Matematika 1* and in a dedicated chapter on space curves in *Matematika 3*. In this course we will return to the above decomposition of acceleration in the next lecture, but we will restrict ourselves to the special case of circular motion.

2.2.5 Kinematics of circular motion

Circular motion is exactly what is sounds like—a body is said to be in circular motion when the body's position traces out a circular trajectory in space. *Uniform* circular motion is circular motion in which the body travels at constant speed, i.e. with $v = |\boldsymbol{v}| = \text{constant}$.

Note: Choosing a coordinate system

Consider a body in circular motion at speed $v(t)$ around a circle of radius R. To proceed, we must first define a coordinate system in which to perform the analysis. Since choosing coordinate systems will come up again and again in your studies, now is a good opportunity to offer some general advice.

- *In theory*, the choice of coordinate system used to analyze any physical problem is arbitrary. Coordinate systems are a human construction used to facilitate analysis—physical objects have no knowledge of them and behave the same way regardless of the choice of coordinate system. A tossed ball, for example, will fall back down to the Earth in whatever coordinate system you choose to analyze it. You'll get the same physical results in any coordinate system, it will just be *much easier in a well-chosen one*. This brings us to the second point.

- *In practice*, a good choice of coordinate system can dramatically simplify the mathematical analysis needed to solve a physical problem. There is often a particular coordinate system well-suited to a particular problem. As your studies progress, you will find that good coordinate systems tend to reflect a problem's symmetries, preferential directions, or other special properties.

Circular motion has a clear special property—the distance to the center of rotation is constant. To make use of this property, we will analyze the problem in a *polar coordinate system*—a two-dimensional, planar system in which a body's position is specified by its distance r from the origin and angle φ with respect to a reference axis, which is usually chosen to align with the Cartesian x axis. In polar coordinates a general position vector $\boldsymbol{r}$ reads

$$\boldsymbol{r} = (r, \varphi), \qquad \text{(in polar coordinates)} \tag{2.26}$$

where r and φ are called the *radial* and *angular* coordinates, respectively. To analyze circular motion, we choose a polar coordinate system with the origin aligned with the center of rotation. The body's radial coordinate r is then constant and equal to the circle's radius R, and the only varying quantity is the body's angular position $\varphi(t)$. Thus, by choosing a polar coordinate system, we have simplified planar circular motion from a two-dimensional problem into a one-dimensional problem, which is easier to analyze.

Measuring angular displacement

In physics, angles are most commonly measured in radians, and not in degrees. Radians should be familiar from high school trigonometry, but we briefly review them now just in case. Radians are defined such that the angle in radians φ between two points on a circle of radius R is the ratio of the arc length s between the two points and the circle's radius. In equation form,

$$\varphi = \frac{s}{r}\ [\text{rad}]. \tag{2.27}$$

The result is dimensionless, i.e. does not carry any of the seven SI base quantities defined in Section 1.3.2. Because radians are dimensionless, it is common practice to omit the "rad" unit, but keep in mind that angular kinematics quantities are always measured in radians, even if the radian unit is not explicitly written. One can convert an angle θ measured in degrees to an angle φ measured in radians using the formula

$$\varphi\,\mathrm{rad} = \frac{2\pi\,\mathrm{rad}}{360^\circ}\theta^\circ. \tag{2.28}$$

Note that angular position φ is periodic with period 2π, in the sense that $0, 2\pi, 4\pi$, etc. (or equivalently 0°, 360°, 720°, etc.) correspond to the same angular position in space. But for angular displacement to be well-defined for changes of more than one full revolution, angular position φ must increase by 2π for each revolution, instead of being counted modulo 2π. For example, if one starts at the origin and makes one and a half full revolutions around a circle, the physically correct new angular position is $\varphi = 3\pi$, not $\varphi = \pi$.

Angular kinematics quantities

We now introduce angular displacement and angular velocity in analogous way to their linear counterparts $\Delta\boldsymbol{r}$ and $\boldsymbol{v}$ in Sections 2.2.2 and 2.2.3—it might be a good idea to review those sections now. *Angular displacement*, denoted by $\Delta\varphi$, is the difference between two angular positions. Just like in (2.2) for the linear displacement $\Delta\boldsymbol{r}$ between two position vectors, we define the angular displacement $\Delta\varphi$ between the angular positions φ_2 and φ_1 as

$$\Delta\varphi = \varphi_2 - \varphi_1. \tag{2.29}$$

Note that $\Delta\varphi$ (at least for motion in a plane) is a scalar, not a vector like $\Delta\boldsymbol{r}$. Conveniently, $\Delta\varphi$ is independent of the choice of the angular reference axis, since any constant offset of the reference axis is canceled in the difference of φ_1 and φ_2. This property mirrors linear displacement's independence of the choice of origin.

Caution: angular displacement is well-defined only if angular positions are taken to increase by 2π after a full revolution rather than being counted modulo 2π, as discussed following (2.28). For example, the correct angular displacement of a body making two and a half revolutions around a circle from $\varphi_1 = 0\,\mathrm{rad}$ to $\varphi_2 = 5\pi\,\mathrm{rad}$ is $\Delta\varphi = 5\pi\,\mathrm{rad}$, but if φ_2 were measured modulo 2π, the angular displacement would incorrectly be $\pi\,\mathrm{rad}$.

We now define angular velocity in analogy to the definition of linear velocity in Section 2.2.3. Consider a body in circular motion, and imagine repeatedly measuring its angular position φ and calculating the angular displacements $\Delta\varphi$ between each position. Let Δt be the difference in time between two successive measurements of angular position. Then associated with each $(\Delta t, \Delta\varphi)$ pair is an *average angular velocity* defined as

$$\overline{\omega} = \frac{\Delta\varphi}{\Delta t}. \tag{2.30}$$

As for angular displacement, average angular velocity is only well-defined if angular positions are counted so as to increase by 2π after a full revolution. *Instantaneous angular velocity* is the limit of $\overline{\omega}$ as the observation period Δt approaches zero, i.e.

$$\omega = \lim_{\Delta t \to 0} \frac{\Delta\varphi}{\Delta t} = \frac{\mathrm{d}\varphi}{\mathrm{d}t}. \tag{2.31}$$

Like φ, angular velocity, at least for our present purposes, is a scalar quantity. In everyday language, angular velocity describes how fast something is spinning or turning about an axis of rotation—it makes sense to speak of the angular velocity of a car wheel about its axle, for example. In the context of circular motion, a body's angular velocity ω describes how quickly the body is rotating around the circle. Note that instantaneous angular velocity is well-defined whether φ is counted modulo 2π or not, since ω is defined in terms of infinitesimal angular displacements $\mathrm{d}\varphi$ that never exceed a full revolution of 2π.

Finally, we define angular acceleration in analogy to the definition of linear acceleration in Section 2.2.4. Consider a body in circular motion, and imagine repeatedly measuring its angular velocity ω and calculating the changes in its angular velocity $\Delta\omega$ between each measurement. Let Δt be the difference in time between two successive measurements of angular velocity. Then associated with each $(\Delta t, \Delta\omega)$ pair is an *average angular acceleration* defined as

$$\overline{\alpha} = \frac{\Delta\omega}{\Delta t}. \tag{2.32}$$

Physically, $\overline{\alpha}$ encodes how quickly a body's angular velocity has changed with respect to time over the measurement period Δt. *Instantaneous angular acceleration* is the limit of $\overline{\alpha}$ as Δt approaches zero, i.e.

$$\alpha = \lim_{\Delta t \to 0} \frac{\Delta\omega}{\Delta t} = \frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{\mathrm{d}^2\varphi}{\mathrm{d}t^2}. \tag{2.33}$$

Physically, α encodes how quickly a body's angular velocity is changing with respect to time at any given instant in time.

Frequency

Frequency is a physical quantity used to describe how quickly periodic phenomena repeat themselves. Consider a periodic process that repeats itself over a time period t_0. The process's *frequency*, denoted by ν or f, is then defined as

$$\nu = \frac{1}{t_0}, \tag{2.34}$$

where t_0 is the period. Frequency is measured in units of cycles per unit time. Like the radian, cycles are a dimensionless quantity and are often left implicit, but writing them explicitly reminds us of a quantity's cyclically-repeating nature. The SI unit of frequency is the *hertz*, denoted by Hz and equal to one cycle per second.

Frequency can be naturally applied to a body in uniform circular motion if one takes the periodic process to be the body passing some fixed reference point on the circle. In this context frequency measures the number of revolutions per unit time. For *uniform* circular motion, angular velocity ω and frequency ν are related according to

$$\omega = 2\pi\nu = \frac{2\pi}{t_0}, \tag{2.35}$$

where the period t_0 corresponds to the time of one full revolution.

Note that frequency does *not* make sense in the context of *non-uniform* circular motion. Keep the following distinction in mind: angular velocity measures the rate of change of angular position, while frequency measures how often a periodic process repeats itself. Non-uniform circular motion is not periodic, so a description in terms of frequency is not well-defined. Angular velocity is the canonical quantity for describing the rate of circular motion, and applies to any type of circular motion, uniform or not. Although frequency can indeed describe *uniform* circular motion, frequency is a separate quantity applicable to general periodic processes.

Linear velocity in circular motion

We now consider the following question: what is the *linear velocity* $\boldsymbol{v}$ of an object in circular motion? There are multiple ways to answer this—we offer a geometric argument here that treats magnitude and direction separately, and provide a more general formulation in terms of unit vectors in Appendix A.1 for interested readers.

Consider a body in circular motion around a circle of radius R, and let $\boldsymbol{r}(t)$ denote the body's instantaneous position with respect to the circle's center. From geometric considerations, the magnitude of the linear displacement $|\mathrm{d}\boldsymbol{r}|$ associated with a small angular displacement $\mathrm{d}\varphi$ along the circle is

$$|\mathrm{d}\boldsymbol{r}| = |\boldsymbol{r}|\,\mathrm{d}\varphi. \tag{2.36}$$

Meanwhile, from the general formulation of speed as $v(t) = |\boldsymbol{v}(t)| = |\mathrm{d}\boldsymbol{r}/\,\mathrm{d}t|$, the change $|\mathrm{d}\boldsymbol{r}|$ for a body moving with speed $v(t)$ is

$$|\mathrm{d}\boldsymbol{r}| = v(t)\,\mathrm{d}t. \tag{2.37}$$

We then combine (2.37) with $|\mathrm{d}\boldsymbol{r}| = |\boldsymbol{r}|\,\mathrm{d}\varphi$ for the specific case of circular motion and get

$$|\mathrm{d}\boldsymbol{r}| = v(t)\,\mathrm{d}t = |\boldsymbol{r}|\,\mathrm{d}\varphi. \tag{2.38}$$

Finally, we solve for $v(t)$ to get

$$v(t) = |\boldsymbol{r}|\frac{\mathrm{d}\varphi}{\mathrm{d}t} \overset{\text{(a)}}{=} |\boldsymbol{r}|\omega(t) \overset{\text{(b)}}{=} R\omega(t), \tag{2.39}$$

where (a) uses the definition of instantaneous angular velocity $\omega = \mathrm{d}\varphi/\,\mathrm{d}t$ and (b) uses the fact that the distance from the center of rotation is constant and equal to R in all circular motion. Equation (2.39) gives the *magnitude* $v = |\boldsymbol{v}|$ of the velocity of a body in circular motion. Note (2.39) applies to general circular motion, since the derivation does not rely on the assumption that ω is constant.

Meanwhile, the *direction* of velocity in circular motion is tangent to the circle, since instantaneous velocity always points tangent to trajectory (see the second paragraph following (2.11)), and circular motion involves a circular trajectory by definition. In polar coordinates, the tangent to the circle is given by the unit vector $\hat{\mathbf{e}}_\varphi$, in terms of which the vector velocity of body in general circular motion reads

$$\boldsymbol{v}(t) = v(t)\,\hat{\mathbf{e}}_\varphi = R\omega(t)\,\hat{\mathbf{e}}_\varphi. \tag{2.40}$$

Linear acceleration in uniform circular motion

All circular motion, even uniform, is accelerated motion, because the direction of the velocity vector constantly changes as a body travels around the circle. Here we consider only uniform circular motion, while the general case of arbitrary circular motion is derived in Appendix A.1.

Consider a body in uniform circular motion, and imagine measuring the body's velocity $\boldsymbol{v}$ at two closely-spaced points in time. Let $\Delta\boldsymbol{v} = \boldsymbol{v}_2 - \boldsymbol{v}_1$ denote the vector difference of velocities and let $\Delta\varphi$ denote the angular displacement between the two velocity vectors. Under the assumption of uniform circular motion, the magnitude of both velocities is equal: $|\boldsymbol{v}_2| = |\boldsymbol{v}_1| \equiv |\boldsymbol{v}|$. From geometric considerations and the Pythagorean theorem, the two velocity vectors are related according to

$$\sin\frac{\Delta\varphi}{2} = \frac{1}{2}\frac{|\Delta\boldsymbol{v}|}{|\boldsymbol{v}|} \implies |\Delta\boldsymbol{v}| = 2|\boldsymbol{v}|\sin\frac{\Delta\varphi}{2}. \tag{2.41}$$

We then substitute (2.41) into the general definition of acceleration to get

$$|\boldsymbol{a}(t)| = \lim_{\Delta t\to 0}\left|\frac{\Delta\boldsymbol{v}}{\Delta t}\right| = \lim_{\Delta t\to 0}\left|2|\boldsymbol{v}|\frac{\sin(\Delta\varphi/2)}{\Delta t}\right| \tag{2.42}$$

$$\overset{\text{(a)}}{=} 2v\lim_{\Delta t\to 0}\frac{\sin(\Delta\varphi/2)}{\Delta t}, \tag{2.43}$$

where in (a) $\boldsymbol{v} = v$ is constant under the assumption of uniform circular motion. In the limit as $\Delta t \to 0$, we correspondingly have $\Delta\varphi \to 0$, and using the Taylor series approximation $\sin x \to x$ as $x \to 0$, (2.43) simplifies to

$$\begin{aligned}|\boldsymbol{a}|_{\text{ucm}} &= 2v\lim_{\Delta t,\Delta\varphi\to 0}\frac{\sin(\Delta\varphi/2)}{\Delta t}\\ &\overset{\text{(a)}}{=} 2v\lim_{\Delta t\to 0}\left(\frac{\Delta\varphi/2}{\Delta t}\right) = v\lim_{\Delta t\to 0}\frac{\Delta\varphi}{\Delta t}\\ &= v\frac{\mathrm{d}\varphi}{\mathrm{d}t} \overset{\text{(b)}}{=} v\omega,\end{aligned}$$

where (a) uses $\sin x \to x$ as $x \to 0$ and (b) uses $\omega = \mathrm{d}\varphi/\,\mathrm{d}t$. Using $v = R\omega$ from (2.39), the result $|\boldsymbol{a}| = v\omega$ for uniform circular motion can be rewritten

$$a_{\text{ucm}} = v\omega = \omega^2 R = \frac{v^2}{R}, \tag{2.44}$$

where R is the circle's radius.

Vector formulation of planar circular motion

For rotation in a *plane*, angular displacement, velocity, and acceleration can be thought of as vector quantities along the axis of rotation, with a sign determined by the right-hand rule convention. (For example, for counterclockwise circular motion in the xy plane, angular displacement, velocity, and acceleration would point along the positive z axis.) A vector formulation of arbitrary three-dimensional rotation is more complicated.[21] We will return to this problem briefly in the lectures on rotational mechanics and relegate a more thorough treatment to the second-year courses *Klasična mehanika* and *Matematična fikzika 1*.

Our goal for now is just to develop some experience in simple vector algebra using the special case of planar circular motion. Consider a body in planar circular

[21] Summarizing considerably, finite rotations $\Delta\boldsymbol{\varphi}$ do not satisfy the commutativity under addition required for elements of a vector space and thus cannot be described as vectors—you can confirm this yourself with the textbook rotation experiment usually prescribed in this context. However—and this can be surprising to the uninitiated—*infinitesimal* rotations $\mathrm{d}\boldsymbol{\varphi}$, and thus the angular velocity $\boldsymbol{\omega}$ and acceleration $\boldsymbol{\alpha}$ ultimately derived from them, *are* valid vector quantities in three dimensional space.

motion of radius R. Using the center of rotation as the origin, the body's x and y Cartesian coordinates are

$$x = R\cos\varphi \qquad \text{and} \qquad y = R\sin\varphi, \tag{2.45}$$

where φ denotes the angle between the body's instantaneous position $\boldsymbol{r}$ the x axis. The body's position vector then reads

$$\boldsymbol{r} = R(\cos\varphi, \sin\varphi). \tag{2.46}$$

Next, we express angular position φ in terms of angular velocity ω using $\varphi = \omega t$ (compare to $x = vt$ for linear motion) and substitute this into (2.46) to get

$$\boldsymbol{r} = R(\cos\omega t, \sin\omega t). \tag{2.47}$$

We now find the body's velocity with component-wise differentiation of $\boldsymbol{r}$:

$$\begin{aligned}
\boldsymbol{v} = \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} &\overset{\text{(a)}}{=} \frac{\mathrm{d}}{\mathrm{d}t}\big[R(\cos\omega t, \sin\omega t)\big] && (2.48)\\
&= R\left(\frac{\mathrm{d}}{\mathrm{d}t}\cos\omega t, \frac{\mathrm{d}}{\mathrm{d}t}\sin\omega t\right) && (2.49)\\
&\overset{\text{(b)}}{=} \omega R(-\sin\omega t, \cos\omega t), && (2.50)
\end{aligned}$$

where in (a) we have substituted in (2.47) and in (b) ω is treated as constant under the assumption of uniform circular motion. Equations (2.46) and (2.50) confirm that position and velocity are *perpendicular* in circular motion, i.e. $\boldsymbol{v} \perp \boldsymbol{r}$ or $\boldsymbol{v} \cdot \boldsymbol{r} = 0$. We can check this explicitly with the calculation

$$\boldsymbol{v} \cdot \boldsymbol{r} = \omega R^2\big[-\cos(\omega t)\sin(\omega t) + \cos(\omega t)\sin(\omega t)\big] = 0 \implies \boldsymbol{v} \perp \boldsymbol{r}. \tag{2.51}$$

To find the acceleration $\boldsymbol{a}_{\mathrm{ucm}}$ of a body in uniform circular motion, we differentiate (2.50) to get

$$\begin{aligned}
\boldsymbol{a}_{\mathrm{ucm}} = \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} &= \omega R\frac{\mathrm{d}}{\mathrm{d}t}\big[(-\sin\omega t), \cos\omega t\big] && (2.52)\\
&= -\omega^2 R(\cos\omega t, \sin\omega t) && (2.53)\\
&\overset{\text{(a)}}{=} -\omega^2\boldsymbol{r}, && (2.54)\\
&\overset{\text{(b)}}{=} -\omega^2 R\,\hat{\mathbf{e}}_r, && (2.55)
\end{aligned}$$

where in (a) we have recognized and substituted in $\boldsymbol{r}$ from (2.47) and in (b) we have written $\boldsymbol{r}$ in terms of the radial unit vector $\hat{\mathbf{e}}_r$ as $\boldsymbol{r} = R\,\hat{\mathbf{e}}_r$. The results $\boldsymbol{a} \parallel -\boldsymbol{r}$ from (2.54) and $\boldsymbol{v} \perp \boldsymbol{r}$ from (2.51) tell us that acceleration in uniform circular motion points radially inward and is perpendicular to $\boldsymbol{v}$; the decomposition of acceleration for uniform circular motion is thus

$$\boldsymbol{a} \overset{\text{(a)}}{=} \boldsymbol{a}_{\mathrm{c}} + \boldsymbol{a}_{\mathrm{t}} = -\omega^2\boldsymbol{r}; \quad \boldsymbol{a} = \boldsymbol{a}_{\mathrm{c}} = -\omega^2\boldsymbol{r}, \quad \boldsymbol{a}_{\mathrm{t}} = \mathbf{0}, \tag{2.56}$$

where (a) is the general decomposition, from (2.25), of acceleration into components parallel and perpendicular to a body's instantaneous velocity $\boldsymbol{v}$. The only difference compared to (2.25) is a change in notation: we use $\boldsymbol{a}_{\mathrm{c}}$ for $\boldsymbol{a}_{\perp}$ and $\boldsymbol{a}_{\mathrm{t}}$ for $\boldsymbol{a}_{\parallel}$. The subscripts stand for "centripetal" and "tangential", respectively, and are more common in the context of circular motion.

2.2.6 Kinematics of one-dimensional free fall

In this section we analyze the kinematics of a point mass in free fall in a uniform gravitational field. Physically, imagine a point-like body launched straight upward or straight downward with initial speed v_0, then left to fall under the influence of gravity. We will find that the object's height, say $z(t)$, changes quadratically with time according to

$$z(t) = z_0 + v_0 t - \frac{g}{2}t^2, \tag{2.57}$$

where z_0 and v_0 are the initial height and velocity, and g is the magnitude of the gravitational acceleration. For simplicity, we begin with a one-dimensional analysis and neglect air resistance, and return to two-dimensional free fall in the next section.

Note: Choosing a coordinate system

Recall from Section 2.2.5 on uniform circular motion that choosing a coordinate system that leverages a physical problem's symmetries or preferential directions can considerably simplify analysis. The problem of a free-falling body has a clear preferential direction—the direction of the gravitational acceleration, which we represent with the vector quantity $\boldsymbol{g}$. We will solve this problem in a Cartesian coordinate system whose z axis points opposite the direction of $\boldsymbol{g}$ and thus perpendicularly upward from the Earth's surface. In everyday language, this just means the z axis points in the direction we think of as "up", while $\boldsymbol{g}$ points "down". Since we consider only one-dimensional motion along the axis of free fall, there is no need to bother defining either the x or y axes.[22]

Solving for the falling body's velocity

A free-falling object, in the absence of air resistance, accelerates with the acceleration of the ambient gravitational field, which we have called $\boldsymbol{g}$. The current problem has a *uniform* gravitational field, which means that $\boldsymbol{g}$ is constant—this considerably simplifies analysis. Let $g \equiv |\boldsymbol{g}|$ denote the magnitude of the gravitation acceleration; for orientation, the value of g on the Earth's surface is approximately $g \approx 9.8\,\mathrm{m\,s^{-2}}$.

In this problem's Cartesian coordinate system, $\boldsymbol{g}$ can be written in either vector or component form as

$$\boldsymbol{g} = -g\,\hat{\mathbf{e}}_z \qquad \text{or} \qquad \boldsymbol{g} = (0, 0, -g), \tag{2.58}$$

where $\hat{\mathbf{e}}_z$ is the unit vector in the direction of the z axis. The minus sign in corresponds to the fact that $\boldsymbol{g}$ points opposite $\hat{\mathbf{e}}_z$ or, in everyday language, the fact that $\boldsymbol{g}$ points "down" while $\hat{\mathbf{e}}_z$ points "up". Since a free-falling body accelerates with the acceleration of the ambient gravitational field, the body's acceleration $\boldsymbol{a}$ is just

$$\boldsymbol{a} = \boldsymbol{g} \qquad \text{or, in component form,} \qquad (a_x, a_y, a_z) = (0, 0 - g). \tag{2.59}$$

The equation's z component, which corresponds to vertical motion, is

$$a_z = -g. \tag{2.60}$$

[22] More precisely, one might say that the problem of one-dimensional free fall is *invariant under rotation* about the axis of free fall. This means that as long as the z axis points upward, our analysis will be the same, and produce the same result, regardless of how we rotate the x and y axes with respect to the z axis.

In fact, the simple equation $a_z = -g$ encodes everything there is to know about a free-falling particle in one dimension. Our goal is to solve (2.60) for the falling body's position and velocity as a function of time. Like in (2.12) in the discussion of velocity and differential equations, we first turn $a_z = -g$ into a differential equation:

$$a_z \overset{\text{(a)}}{=} \frac{\mathrm{d}v_z}{\mathrm{d}t} = -g \implies \mathrm{d}v_z = -g\,\mathrm{d}t, \tag{2.61}$$

where in (a) we have used the definition of acceleration $a_z = \mathrm{d}v_z/\,\mathrm{d}t$ from (2.23). For shorthand, let $v \equiv v_z$ and $a \equiv a_z$, since there is only one dimension involved and the subscripts would only clutter the problem. Dropping the subscripts, (2.61) reads

$$\mathrm{d}v = -g\,\mathrm{d}t. \tag{2.62}$$

Along the same lines as (2.14b) in the section on velocity and differential equations, we first let $v_0 \equiv v(t_0)$ denote the free-falling body's vertical velocity at some initial time t_0, then integrate (2.62) from t_0 to an arbitrary time t to get

$$\int_{v_0}^{v(t)} \mathrm{d}\nu = v(t) - v_0 = \int_{t_0}^{t} (-g)\,\mathrm{d}\tau = -g\int_{t_0}^{t} \mathrm{d}\tau = -g(t - t_0). \tag{2.63}$$

After clearing things up and rearranging we have

$$v(t) - v_0 = -g(t - t_0) \implies v(t) = v_0 - g(t - t_0). \tag{2.64}$$

Equation (2.64) is the general result for the velocity of a free-falling object in one dimension as a function of time. If we choose to begin counting time at t_0, so that in this problem $t_0 = 0$, we recover the familiar kinematics formula

$$v(t) = v_0 - gt. \tag{2.65}$$

Solving for the falling body's position

Next, we will solve (2.65)[23] for the free-falling particle's position $z(t)$. As in (2.61), we first write (2.65) as a differential equation using the definition of velocity to get

$$v(t) = \frac{\mathrm{d}z}{\mathrm{d}t} = v_0 - gt. \tag{2.66}$$

We then rearrange, let $z_0 \equiv z(0)$, and integrate (recalling the choice $t_0 \equiv 0$) to get

$$\int_{z_0}^{z(t)} \mathrm{d}\zeta = \int_0^t v_0\,\mathrm{d}\tau - \int_0^t g\tau\,\mathrm{d}\tau. \tag{2.67}$$

(We have used the Greek letters zeta and tau for integration variables, as discussed in the context of (2.14b).) Solving the integral produces

$$z(t) - z_0 = v_0 t - \frac{g}{2}t^2 \implies z(t) = z_0 + v_0 t - \frac{g}{2}t^2. \tag{2.68}$$

Equation (2.68) gives the position of a free-falling particle in one dimension as a function of time, assuming time is counted so that the particle began falling at $t_0 = 0$.

[23] One could also solve the more general (2.64), which leaves t_0 arbitrary, but for our present purposes dragging along the extra t_0 just muddles the message without introducing any new physics.

At what time is the body at a given height?

We now ask the following question: if a body is launched upward at time $t_0 = 0$ from an initial height z_0, how much time does it take to reach an arbitrary height z? To answer the question, we first rearrange (2.68) into a standard-form quadratic equation in time, i.e.

$$\frac{g}{2}t^2 - v_0 t + z - z_0 = 0. \tag{2.69}$$

Our idea is to solve for t as a function of z. To do this, we apply the quadratic formula with $a = g/2$, $b = -v_0$ and $c = (z - z_0)$ to get

$$t = \frac{v_0 \pm \sqrt{v_0^2 - 2g(z - z_0)}}{g}. \tag{2.70}$$

By the fundamental theorem of algebra, (2.70) is guaranteed to have two (in general complex) solutions. But before diving right into mathematics, it's more instructive to think about the problem physically. Forgetting equations for a moment, imagine an object launched vertically upward at $t_0 = 0$ with initial velocity v_0 from initial height z_0. Everyday intuition holds good here: the object will initially rise, reach a maximum height $z_{\max} > z_0$ at which it stops rising, and then begin falling back down. We can immediately recognize the following: the object...

(a) ...won't ever reach heights above $z_{\max}$, so (2.70) shouldn't have real-valued solutions for $z > z_{\max}$.

(b) ...reaches heights in the range z_0 to $z_{\max}$ twice—once coming up and once coming down. We thus expect two solutions to (2.70) for $z \in (z_0, z_{\max})$, both at times after the object was launched.

(c) ...reaches its initial height z_0 twice—once exactly at launch time (at $t = 0$) and once at a later time when falling back down.

(d) ...reaches heights less than the initial height z_0 after passing z_0 on the way down, i.e. for times larger than $2v_0/g$.

After examining the problem physically, we can follow up with a mathematical analysis of each point predicted in the heuristic analysis above

(a) Equation (2.70) has no real solutions when the discriminant is zero, which occurs for height z larger than

$$v_0^2 - 2g(z - z_0) < 0 \implies z > \frac{v_0^2}{2g} + z_0. \tag{2.71}$$

The maximum height mentioned in point (a) above is thus

$$z_{\max} = \frac{v_0^2}{2g} + z_0. \tag{2.72}$$

Jumping ahead somewhat, in the future lectures on energy we will find that the object cannot pass $z_{\max}$ if its initial kinetic energy is too small to overcome the gravitational potential energy at $z_{\max}$.

(b) It is a straightforward exercise in grade school algebra to show that (2.70) has two positive solutions only for z in the range $(z_0, z_{\max})$. In this problem positive times are those after the object was launched, so this mathematical result agrees with the physical intuition of point (b) above.

(c) Similarly, (2.70) has two solutions for $z = z_0$, once at $t = 0$ (launch time) and once at $t = 2v_0/g$, in agreement with point (c) above.

(d) Careful here—(2.70) has *two* solutions for any $z < z_0$, one negative and one positive. Suppose you solve (2.70) for a height $z_- < z_0$ get the two solutions t_- and t_+. Physically, the positive time corresponds to the falling-down phase predicted in point (d) above.

What the negative solution tells you is this: if the object had been traveling along the trajectory determined by its initial height and velocity for *all times*, and had passed through the launch point with velocity v_0 at time $t_0 = 0$, it would have passed through the height z_- (on the way up) t_- seconds *before* passing reaching the launch point z_0.

It is tempting to throw this answer out with an argument like "negative times are non-physical", but that's not quite correct. *Complex-valued* times are non-physical—this is why the object never reaches heights above $z_{\max}$. But negative times can be perfectly physical—which times are negative and which are positive only depends on the arbitrary decision of when you begin to start counting time, just like which heights are negative and which are positive depends on where you define a coordinate system's origin.

Solving for final velocity

Finally, we answer the following question: if a body is launched upward at time $t_0 = 0$ from an initial height z_0, what is its velocity at an arbitrary height z? We begin with (2.65), $v(t) = v_0 - gt$, (the body's velocity as a function of time), and substitute in time as a function of height from (2.70). The result is

$$
\begin{aligned}
v &= v_0 - gt && (2.73)\\
&\overset{(a)}{=} v_0 - g\left(\frac{v_0 \pm \sqrt{v_0^2 - 2g(z - z_0)}}{g}\right) && (2.74)\\
&= \pm\sqrt{v_0^2 - 2g(z - z_0)}, && (2.75)
\end{aligned}
$$

where (a) uses (2.70). Like in the discussion in point (d) above, the two solutions correspond to a free-falling body reaching a given height twice—once going up and once going down. Note also that (2.75) has no real solutions for $z > z_{\max}$, which must be the case for consistency with point (a) above. It is also instructive to check that $v = 0$ at $z = z_{\max}$, which corresponds to a free-falling object stopping momentarily at its peak height $z_{\max}$ before falling back down again.

Additionally, to conceal questions of plus or minus, one can square (2.75), which produces the well-known kinematics formula

$$
v^2 = v_0^2 - 2g(z - z_0). \tag{2.76}
$$

A good lesson from (2.76) is that magnitude of final velocity v increases with an increasing distance $\Delta z = z - z_0$. Interpreted physically, this just means that the farther an object falls, the faster it gets.

In passing, is also possible, and instructive, to derive (2.76) using the chain rule from differential calculus. We begin with the definition of acceleration $a = \mathrm{d}v/\,\mathrm{d}t$ and apply the chain rule to get

$$a = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}t} = v\frac{\mathrm{d}v}{\mathrm{d}z} \implies v\,\mathrm{d}v = a\,\mathrm{d}z. \tag{2.77}$$

We then integrate (2.77), producing

$$\int_{v_0}^{v} \nu\,\mathrm{d}\nu = \int_{z_0}^{z} a\,\mathrm{d}\zeta \implies \frac{1}{2}(v^2 - v_0^2) = a(z - z_0), \tag{2.78}$$

which we rearrange to get

$$v^2 = v_0^2 + 2a(z - z_0). \tag{2.79}$$

Substituting in $a = -g$ for free fall recovers (2.76).

2.2.7 Projectile motion

In this section, we generalize the discussion of free fall from Section 2.2.6 to two-dimensional *projectile motion*, in which a launched or dropped body (called a *projectile* in this context) is allowed to move in both the horizontal and vertical directions. In everyday language, we are interested in what happens to objects launched *diagonally* upward, for example a tossed ball or an arrow shot from a bow. We neglect air resistance in the analysis, but we will qualitatively comment on the complications that arise when considering it at the end of the section.

Analysis and coordinate system

In principle, real-life projectile motion takes place in three dimensions simply because space is three-dimensional. But if you line yourself up with a projectile and observe its trajectory, you will notice that (in the absence of outside influences like wind) the projectile will only move forward/back and up/down, but not also side-to-side. The projectile's motion is restricted to a *plane*. By choosing a Cartesian coordinate system in which, say, the xz plane is aligned with the plane of motion, one can analyze projectile motion completely generally in only two dimensions instead of three. This is what we will do in this problem, and is yet another example of how a well-chosen coordinate system can simplify mathematical analysis.

Consider a projectile at initial position $\boldsymbol{r}_0$ launched with initial velocity $\boldsymbol{v}_0$ at an angle ϕ with the horizontal, and let $\boldsymbol{g}$ denote the gravitational acceleration. We will use a Cartesian coordinate system with the z axis point opposite $\boldsymbol{g}$ (as for one-dimensional free fall) and the x axis in the direction of the projectile's horizontal motion. A free-falling body has the same acceleration $\boldsymbol{a}$ as the ambient gravitational acceleration, so $\boldsymbol{a} = \boldsymbol{g}$, just like for one-dimensional free fall. In this coordinate system, the relevant vector quantities have the component representations

$$\boldsymbol{r}_0 = (x_0, 0, z_0), \qquad \boldsymbol{v}_0 = (v_{x_0}, 0, v_{z_0}), \qquad \boldsymbol{a} = \boldsymbol{g} = (0, 0, -g). \tag{2.80}$$

Letting $v_0 \equiv |\boldsymbol{v}_0|$ denote the magnitude of the projectile's initial velocity, we can relate v_0 to the launch angle ϕ according to

$$v_{x_0} = v_0 \cos\phi \qquad \text{and} \qquad v_{y_0} = v_0 \sin\phi. \tag{2.81}$$

Equations (2.80) and (2.81) completely specify the projectile's motion—we just have to solve them. One solves multi-dimensional problems of this type by writing equations for each dimension separately. In this problem only the x and z components are relevant—because of our choice of coordinate system, $y(t)$, $v_y(t)$ and $a_y(t)$ are all conveniently zero and we will not consider them further. We first write out the projectile's horizontal and vertical acceleration:

$$a_x = 0 \qquad \text{and} \qquad a_z = -g. \tag{2.82}$$

We then solve the horizontal and vertical components separately.

Horizontal component of motion

With experience, one would immediately recognize that

$$a_x = 0 \implies v_x(t) = v_{x0}, \tag{2.83}$$

i.e. that in the absence of a horizontal acceleration, a projectile will continue moving with its initial horizontal velocity v_{x_0} for all time.[24] But since this is our first time and some students might not have experience with integral calculus, we also show how to reach the result $v_x(t) = v_{x_0}$ formally. We begin with $a_x = 0$, which write as differential equation in the form

$$a_x = \frac{\mathrm{d}v_x}{\mathrm{d}t} \implies \int_0^t a_x(\tau)\,\mathrm{d}\tau = \int_{v_{x0}}^{v_x(t)} \nu_x\,\mathrm{d}\nu_x. \tag{2.84}$$

We then apply $a_x = 0$ and evaluate the integrals to get

$$0 = v_x(t) - v_{x0} \implies v_x(t) = v_{x0}, \tag{2.85}$$

in agreement with (2.83). To solve for the projectile's horizontal position, we integrate (2.85) over time to get

$$x(t) = x_0 + v_{x0}t. \tag{2.86}$$

Vertical component of motion

This problem has a very nice property: the horizontal and vertical components of motion are *independent*, i.e. y quantities do not feature in the x equations, and vice versa. This means that analyzing the vertical component of 2D projectile motion is just as easy as analyzing purely vertical free fall, which we have already done in Section 2.2.6. After beginning with $a_z = -g$ and following an analogous analysis to that in Section 2.2.6, which we won't re-write here, the results are

$$v_z(t) = v_{z_0} - gt, \tag{2.87}$$

$$z(t) = z_0 + v_{z_0}t - \frac{g}{2}t^2. \tag{2.88}$$

[24]Of course, because of air resistance, physical projectiles do not continue with their initial horizontal velocity ad infinitum—we return to this problem at the end of the section.

The only difference compared to (2.65) and (2.68) is the initial velocity v_0 in 1D motion generalizing to the *vertical component* v_{z_0} in 2D motion.

Relating horizontal and vertical motion

We now answer the following question: What is a projectile's height z at a given horizontal position x? The relevant equations are (2.86) and (2.88); for review these read

$$x - x_0 = v_{x_0} t, \tag{2.89a}$$
$$z - z_0 = v_{z_0} t - \frac{g}{2} t^2. \tag{2.89b}$$

To simplify notation, we first re-write (2.89) in terms of the new variables $\Delta x \equiv x - x_0$ and $\Delta z \equiv z - z_0$, producing

$$\Delta x = v_{x_0} t, \tag{2.90a}$$
$$\Delta z = v_{z_0} t - \frac{g}{2} t^2. \tag{2.90b}$$

We then solve (2.90a) for time, $t = \Delta x / v_{x_0}$, and substitute the result into Δz to get

$$\Delta z = \frac{v_{z_0}}{v_{x_0}} \Delta x - \frac{g}{2} \left(\frac{\Delta x}{v_{x_0}} \right)^2. \tag{2.91}$$

Finally, we use $v_{x_0} = v_0 \cos\phi$ and $v_{z_0} = v_0 \sin\phi$ from (2.81) to get the final expression

$$\Delta z = \Delta x \tan\phi - \frac{g}{2} \frac{(\Delta x)^2}{v_0^2 \cos^2\phi}. \tag{2.92}$$

The result is a parabola, meaning that a projectile's height z, in the absence of air resistance, varies parabolically with its horizontal position x. (Note that z also varies parabolically with t, but that is a separate concept.)

Flight time and range

We conclude by considering the following two questions:

(a) How much time t does a projectile take to reach a given height z? This is called the projectile's *flight time*.

(b) How far has the projectile traveled horizontally after reaching a given height z? This horizontal distance is called the projectile's *range*.

One finds flight time by solving (2.88) for t as a function of z using the quadratic formula. The results and discussion are analogous to the case of one-dimensional motion in (2.70), with v_0 from (2.70) replaced by the vertical component v_{z_0}. The result is

$$t = \frac{v_0 \pm \sqrt{v_{z_0}^2 - 2g(z - z_0)}}{g}. \tag{2.93}$$

In particular, a projectile falls back to the height it was launched from (found by setting $z = z_0$ in (2.93)) after a time

$$t_0 = \frac{2v_{z_0}}{g} \overset{\text{(a)}}{=} \frac{2v_0}{g} \sin\phi, \tag{2.94}$$

where in (a) we have used $v_{z_0} = v_0 \sin\phi$. This time is also twice the time to the highest point—try confirming this yourself by maximizing (2.88) with respect to time. To find the body's range, i.e. the horizontal distance Δx it has traveled after reaching a height z, one should

(i) solve for the time(s) t required to reach z using (2.93), then

(ii) substitute these times into (2.86) to find the range Δx.

The two times—and the resulting ranges Δx—correspond to a projectile reach a given height once on the way up and once on the way down (see also the discussion in point (d) from page 35). In particular, the body's range Δx when falling back to its initial height, using the time t_0 from (2.94), is

$$\Delta x = x - x_0 = v_{x_0} t_0 \overset{\text{(a)}}{=} v_{x_0} \cdot \frac{2v_{z_0}}{g} \tag{2.95}$$

$$\overset{\text{(b)}}{=} 2\frac{v_0^2}{g} \sin\phi \cos\phi \tag{2.96}$$

$$\overset{\text{(c)}}{=} \frac{v_0^2}{g} \sin 2\phi \tag{2.97}$$

where in (a) we have substituted in (2.94), in (b) we have used $v_{x_0} = v_0 \cos\phi$ and $v_{z_0} = v_0 \sin\phi$ and in (c) we have used the trigonometric identity $2\sin\phi\cos\phi = \sin 2\phi$.

Comment: Look for simple answers first

Equation (2.97) holds the answer to an interesting question—at what launch angle ϕ is a projectile's range maximized? There are two ways to go about this: one could blindly charge in with the full force of the differential calculus machinery and maximize the function $\Delta x(\phi)$ by testing the zeros of the derivative $\mathrm{d}(\Delta x)/\mathrm{d}\phi$. The mathematical result would be

$$\phi = \frac{\pi}{4} + \pi k, \quad k \in \mathbb{Z}. \tag{2.98}$$

You would then have to apply physical reasoning to conclude that $\phi = \pi/4$, i.e. 45° degrees relative to the vertical, is the physically sensible solution.

Alternatively, one could recall that the sine function is maximized when its argument equals $\pi/2$ (and that v_0^2/g is a constant and won't affect the result). You then solve the equation

$$2\phi = \frac{\pi}{2} \implies \phi = \frac{\pi}{4}. \tag{2.99}$$

Same result, much simpler procedure. There are two lessons here:

(i) (Neglecting air resistance), a projectile will travel the furthest horizontal distance at a given launch speed if it is launched at $\pi/4 = 45°$ relative to the horizontal.

(ii) More importantly, before applying the full force of the available mathematical machinery to solve a physical problem, check for straightforward routes to the same solution—this can be particularly relevant when maximizing, integrating, or averaging sinusoids in a physical context.

Comments on air resistance

Air resistance, also called drag, greatly complicates projectile motion. Although somewhat ahead of our discussion of dynamics, drag forces act in the opposite direction of velocity $\boldsymbol{v}$ and increase with increasing magnitude $v = |\boldsymbol{v}|$. Interpreted physically, this just means that drag forces have the effect of slowing a body down, and are stronger the faster a body is moving. At large Reynolds numbers (loosely, at high speeds), drag is proportional to squared speed, and a general form of a drag-induced acceleration might be

$$\boldsymbol{a}_{\mathrm{d}} = -C_{\mathrm{d}} v^2 \frac{\boldsymbol{v}}{|\boldsymbol{v}|} \overset{\text{(a)}}{\equiv} -C_{\mathrm{d}} v^2 \cdot \frac{1}{v}(v_x, v_y) = -C_{\mathrm{d}} v \cdot (v_x, v_y) \tag{2.100}$$

where in (a) we have written $\boldsymbol{v}$ in component form and used the shorthand $v = |\boldsymbol{v}|$. The term C_{d} is a constant drag coefficient that encodes the geometry of the projectile and properties of the surrounding air flow; its exact value is irrelevant in this discussion, and we include it only for completeness. Using $v = \sqrt{v_x^2 + v_y^2}$, the x and y components of the drag acceleration $\boldsymbol{a}_{\mathrm{d}}$ are

$$\begin{aligned} a_{\mathrm{d}_x} &= -C_{\mathrm{d}} \sqrt{v_x^2 + v_y^2} \cdot v_x, \\ a_{\mathrm{d}_y} &= -C_{\mathrm{d}} \sqrt{v_x^2 + v_y^2} \cdot v_y. \end{aligned}$$

Here is the problem: both the x and y components of acceleration now include both the x and y components of velocity. Unlike in (2.82), the components of motion in the horizontal and vertical dimensions are now *dependent*, or *coupled.* Concretely, means that a changing velocity in one dimension affects the acceleration (and thus velocity) in the other dimension, and vice versa. This interdependence considerably complicates analysis, so much so that problems involving drag in real-world applications are usually solved numerically. Numerical solutions to coupled systems of differential equations are introduced in the first year elective course *Računalniška orodja v fiziki*, and covered in more detail in the later courses *Numerične metode* and *Matematično-fizikalni praktikum.* Note that the goal here is just to offer a qualitative idea of how air resistance complicates real-life projectile motion—you won't be expected to actually compute such motion in the scope of this course.

2.2.8 Kinematics of relative motion and inertial frames

In this section we consider the problem of two different observers watching the same physical process from two different positions in space. This could be as simple as two students (the two observers) sitting on either side of a lecture hall watching the professor perform an experiment (the observed physical process). The question we will answer is:

> How does one convert between two different descriptions of the same physical process given the relative separation (and velocity) of the two observers?

Frame of reference

What, precisely, do we mean by an "observer"? To describe a physical process, an observer must specify the *position* and *time* of everything occurring in the scope of the process. Specifying position requires a coordinate system and origin, while specifying time requires a consistent means of keeping time, which we will abstractly call a *clock*. The combination of a coordinate system, origin, and a clock is called a *frame of reference*. What we really mean by an observer, then, is a frame of reference. In precise language, what we're after in this section is a formalism for converting between descriptions of the same physical process observed from different frames of reference.

Consider two observers, whom we will call A and B, watching the same physical process. Suppose each observer uses a coordinate system with the origin centered at the observer's current position, and, to be general, suppose their clocks are offset by some arbitrary interval t_0. Let S and S' denote A and B's frame of reference, respectively. Because the origins and clocks are different, the values of the times and coordinates used by A and B will be different. How, on the basis of A's measurements, can A predict the values of the times and coordinates that will appear in B's measurements? In more precise language, what is the *transformation* of space and time from the frame of reference S to the frame of reference S'?

Transformation for time

In the scope of classical physics, time passes at the same rate in all frames of reference.[25] Different clocks might have constant relative offsets, but all run *at the same rate*. This makes clock synchronization in classical physics very easy—one simply subtracts the time offset: if t is the time in S, t' is the time in S', and the clock in S is ahead of the clock in S' by a time t_0, then

$$t' = t - t_0. \tag{2.101}$$

This is no more difficult than someone calculating the current time in New York by subtracting six hours from the current time in Paris. But in nearly all practical cases in classical physics, two frames of reference will use the same clock, i.e. $t = t'$ with $t_0 = 0$, and you won't have to worry about synchronizing clocks at all. We have mentioned the general case involving an offset only for the sake of completeness. Meanwhile, synchronizing position (and its time derivatives, velocity and acceleration) is a little more involved, and we discuss this next.

Transformation for position

To make any progress, we must first specify the location of the S and S' origins relative to each other: let the vector $\boldsymbol{r}_0$ point from the origin of S to the origin of S', and, analogously, let $\boldsymbol{r}_0'$ point from the origin of S' to the origin of S. In this case, as is most easily seen from a figure, positions measured in S and S are related according to

$$\boldsymbol{r} = \boldsymbol{r}_0 + \boldsymbol{r}' \implies \boldsymbol{r}' = \boldsymbol{r} - \boldsymbol{r}_0. \tag{2.102}$$

The result $\boldsymbol{r}' = \boldsymbol{r} - \boldsymbol{r}_0$ is called the *Galilean transformation of position* from the frame S to S'—it uses the positions $\boldsymbol{r}$ and $\boldsymbol{r}_0$ measured in S to express the position $\boldsymbol{r}'$ measured in S'. The reverse Galilean transformation from S' to S, which is again best seen from a figure, is

$$\boldsymbol{r} = \boldsymbol{r}' - \boldsymbol{r}_0'. \tag{2.103}$$

[25]In the theories of special (and general) relativity, time passes at different rates in different frames of reference, but that is a topic for *Moderna fikzika 1*.

In this case, quantities measured in S' are used to express quantities in S.

Importantly, one can convert between between (2.102) and (2.103) simply by switching the primed and unprimed quantities. There is a subtle but powerful idea at work here—loosely, the fact that you can interchange primed and unprimed quantities and get the same equations means that observations in the two systems are equivalent. Which system we call S and which one we call S' is an arbitrary human construction, and a physical process will behave the same way and lead to the same results when described in either frame of reference. The principle that the equations describing any physical process take the same form regardless of the frame of reference in which the process is observed is called the *principle of relativity.*

Transformations of velocity

We now aim to find a relationship between observed velocities in S and S'. This is straightforward—we first differentiate (2.102) with respect to time and get

$$\boldsymbol{v}' = \frac{\mathrm{d}\boldsymbol{r}'}{\mathrm{d}t} = \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} - \frac{\mathrm{d}\boldsymbol{r}_0}{\mathrm{d}t} \overset{\text{(a)}}{\equiv} \boldsymbol{v} - \boldsymbol{v}_0, \tag{2.104}$$

where in (a) we have used $\boldsymbol{v}_0 = \mathrm{d}\boldsymbol{r}_0/\,\mathrm{d}t$ to denote the velocity of the S' origin relative to the S origin. Equation (2.104) is called the *Galilean transformation of velocity* from the frame S to S'. To find the inverse transformation from S' to S, one simply interchanges the primed quantities in (2.104) to get

$$\boldsymbol{v} = \boldsymbol{v}' - \boldsymbol{v}_0', \tag{2.105}$$

where $\boldsymbol{v}_0' = \mathrm{d}\boldsymbol{r}_0/\,\mathrm{d}t$. Note that one could reach the same result by differentiating (2.103) directly—this must be the case by the principle of relativity.

Relative acceleration and inertial systems

The natural question to ask next is how accelerations transform between frames of reference. Again the answer straightforward, one differentiates (2.104) with respect to time and gets

$$\boldsymbol{a}' = \boldsymbol{a} - \boldsymbol{a}_0, \tag{2.106}$$

where $\boldsymbol{a}_0 = \mathrm{d}\boldsymbol{v}_0/\,\mathrm{d}t$ is the acceleration of the S' origin relative to the S origin. In principle, we could keep playing this game ad infinitum, taking higher and higher derivatives of (2.106) and getting more and more transformation laws. But for reasons that will be more clear in the coming lectures on dynamics, it is conventional, and in fact physically sensibile, to *stop* this process at the level of acceleration.[26] We then distinguish between two types of reference frames:

(a) frames with zero relative acceleration, i.e. $\boldsymbol{a}_0 = \boldsymbol{0}$, and

(b) frames with non-zero relative acceleration, i.e. $\boldsymbol{a}_0 \neq \boldsymbol{0}$.

Any set of frames with zero relative acceleration, $\boldsymbol{a}_0 = \boldsymbol{0}$, is called a *family of inertial frames*. Inversely, any set of reference frames with non-zero relative acceleration, $\boldsymbol{a}_0 \neq \boldsymbol{0}$, is called a family of *non-inertial* reference frames. Because their origins move relative to each other with varying velocity, non-inertial reference frames exhibit

[26] Skipping ahead somewhat, acceleration is a natural stopping point because Newton's second law relates force to acceleration (and not, say, velocity, jerk, or some other derivative of position).

some counter-intuitive properties that we will return to in the lectures on system forces. For now, we focus only on inertial reference frames—these are the frames in which Newton's laws of motion hold, and they are most relevant for describing the physics of everyday life.

Absolute frames of reference

Caution: so far we have *not* defined the requirements for an *absolute* inertial frame—doing requires concepts from dynamics. In the scope of kinematics, one can only say if *two or more* frames are *mutually* inertial, but not if an individual frame is *absolutely* inertial. The problem is that an inertial reference is defined as having zero absolute acceleration. Trying to specify absolute acceleration in the scope of kinematics requires *another, external* non-accelerated reference frame with respect to which one measures the first frame's acceleration, and you are back at the problem you started with!

In the scope of kinematics, you *cannot* specify if a frame of reference is absolutely inertial. You can measure the acceleration of some frame S' relative to some other frame S, but if you measure zero relative acceleration $\boldsymbol{a}_0 = \mathbf{0}$, you cannot tell if both systems are non-accelerated or if both are accelerating with the same acceleration. We will resolve this problem using Newton's laws of motion; loosely, we will define inertial frames as those frames in which the observed motion of bodies agrees with the motion theoretically predicted by Newton's laws.

Inertial frames in practice

Inertial reference frames, like physical models, are an *idealization.* Within the scope of modern experiments, everything, even at cosmological scales, seems to be rotating (and thus accelerating) relative to something else. Yes, the Earth's surface is reasonably inertial, the center of the solar system is a better approximation of an inertial frame than the Earth, the center of our galaxy is better than the solar system, and so on, but as far as we can measure, everything seems to be at least *somewhat* accelerating relative to something else.

And that is perfectly fine! In practice, the relevant question is not "is this frame perfectly inertial?" but instead "*within the limits of experimental error*, does the physical behavior observed in my (technically non-inertial) experiment agree the behavior I would theoretically expect in an inertial frame of reference?" Very often, the answer is yes, and in this case it is perfectly acceptable, and indeed advisable, to perform analysis in an inertial frame of reference, simply because it is easier.

For the purposes of this course, we will generally use the Earth's surface as an inertial frame of reference. Of course the Earth's surface is *not* perfectly inertial—this can be seen from the behavior of cyclones, ocean currents, the motion of Focault's pendulum, and so on. We will explore these topics in the lectures on system forces, and you will study non-inertial reference frames in more detail in the second-year course *Klasična mehanika.*

2.3 Dynamics

In this section we have two goals:

1. to answer what causes motion, and
2. once we know the cause of motion, to develop a formalism for *predicting* motion.

2.3.1 Forces and the environment

The concept of the environment

So far, in the scope of *kinematics*, we have specified and analyzed only a *single* body. In *dynamics*, a single body is not enough. To speak of the dynamics of an observed body, you must bring *other* bodies into the picture—these other bodies are what ultimately cause the observed body to move. There are thus two actors at play in dynamics:

1. an observed body, whose motion we wish to predict, and
2. all other bodies in the observed body's surroundings that can cause the observed body to move. These other bodies are collectively called the observed body's *surroundings* or *environment*.

In principle, a body's dynamical environment is everything else in the universe. In practice things are much simpler—in any given situation there are usually only a few relevant bodies in the environment, and these become easy to identify with practice. The goal of dynamics is to *quantify* the effects of this environment on the motion of an observed body or physical system.

Forces

The motion-causing interactions between an observed body and the surrounding bodies in its environment are called *forces*. We immediately emphasize an important point: a force can arise only from the interaction between two *concrete*, *physical* bodies—bodies you point to, look at, or touch. There are two lessons here:

- Forces do not amorphously arise from mystical sources—they are the consequence of physical objects you can see, touch, or otherwise identify. Forces can act through contact or from a distance, but even long-range forces always involve an interaction between two *well-defined* bodies. Your weight, for example, arises from the interaction between your body and the Earth.
- The second point is more subtle: the correct way to think about forces is as a *mutual interaction* between *two* bodies. It does not make sense to talk about the force on a body without specifying the other body involved in the interaction. Even the popular formulation of one body "causing" a force and a second body "receiving" is misleading—there is no fundamental concept of "source" or "recipient" built into the Newtonian formalism, only a mutual interaction between two bodies. (We will formalize this idea in the context of Newton's third law.)

Of course, in practice, when we focus an analysis on an observed body it is common to speak only of the "force applied to the body" or the "force on the body", but even here it should be implicitly understood that the force of the environment on the body is an interaction intrinsically related to the force of the body on its environment.

Units of force

The SI unit of force is the *newton*. One newton is defined as the amount of force needed to accelerate a one-kilogram mass at one meter per second squared (in an inertial frame of reference). For orientation, one newton is roughly equal to the weight of a small apple on the Earth's surface.

Mass and inertia

Two relevant concepts in the context of dynamics are a body's inertia and mass. A body's *mass* is a fundamental physical quantity, introduced in Section 1.3.2, that measures the amount of matter making up the body. A body's *inertia* is a dynamical property that encodes the body's resistance to acceleration under the influence of an external force. In other words, a body with more inertia will accelerate less in response to a given force than a body with less inertia.

It turns out that, for the purposes of Newtonian dynamics, a body's inertia and mass are completely equivalent—in fact, the mass m in Newton's second law was originally called "inertia" in Newton's time. If you write Newton's second law (which we will formally define shortly) in the form

$$\boldsymbol{a} = \frac{\boldsymbol{F}}{m}, \tag{2.107}$$

it is mathematically clear that a body's acceleration in response to a given force is inversely proportional to its mass, i.e. mass causes a body to resist acceleration. This should also be intuitively reasonable—the more "stuff" making up a body, the harder it is to accelerate it.

To some extent, the one-to-one correspondence between inertia and mass makes inertia seem redundant—by specifying a body's mass, you have also fully specified its inertia. It might be helpful to think of inertia as being a dynamical *property* of mass. Mass tells you how much of a body there is, which in turn determines the body's resistance to acceleration under the influence of forces.

The vector nature of forces

Forces behave as vector quantities. In practice, this is relevant for the following reason: a body under the influence of multiple forces acts as if it were under the influence of a single force equal to the vector sum of the original forces, which in this context is called the *net force*. This property is called the *principle of superposition*, and it forms the backbone of Newtonian dynamics (and other branches of physics, too).

Comment: Distinguishing mass and weight

We now briefly comment on the difference between mass and weight, since the two are sometimes confused in everyday language. Mass is an intrinsic property measuring the amount of material making up a body. Weight, meanwhile, is a force—it measures the gravitational interaction between a body and its surroundings. The key difference is this: because weight is a force, a body's weight depends on both the body itself *and on its surroundings*, while a body's mass depends *only on the body itself.* In this light, it should be clear why bodies are weightless in outer space—there is nothing else *in* outer space to cause weight.

Confusion can arise from the combination of two factors:

(i) Weight happens to be *proportional* to mass—the formula for a body's weight is

$$\boldsymbol{F}_{\text{weight}} = m\boldsymbol{g}, \tag{2.108}$$

where m is the body's mass and $\boldsymbol{g}$ is the surrounding gravitational acceleration.

(ii) In everyday life, the Earth's gravitational influence (i.e. the "surroundings" in the language of dynamics, or the $\boldsymbol{g}$ in (2.108)) is essentially constant. Because the surroundings are constant, everyday life does not instill an intuitive understanding that weight also depends on the environment. Although a body's weight indeed depends on two things—its mass and its surroundings—only mass varies appreciably in everyday life, so it is understandable, albeit incorrect, that people use mass and weight interchangeably in the scope of everyday life.

(Finally, for the mathematically-inclined, mass is a scalar while weight is a vector, which makes the distinction, at least mathematically, quite obvious.)

2.3.2 Inertial frames and Newton's laws

Absolute inertial frames

In Section 2.2.8, we put off defining an absolute inertial frame until we began dynamics; it might be a good idea to review the section on absolute frames of reference now. What we were missing in kinematics was the concept of *forces*, i.e. motion-causing interactions between an observed body and its environment. Having introduced forces, we now define inertial frames:

> An *inertial frame of reference* is a frame in which a particle with zero net force acting on it travels at constant velocity.

The recipe for determining if a frame of reference is absolutely inertial is thus:

1. In the frame you wish to test, choose a test particle and ensure that the net force on the particle is zero.

2. Observe the particle's motion. If the particle moves with constant velocity (relative to the tested reference frame), then the frame is inertial.

Note that this experiment is well-defined only under the assumption that an observer can identify with certainty if the net force on the particle is or is not zero. In practice doing so is usually straightforward, although we will have a few more words to say about the theoretical validity of this assumption towards the end of this section.

Newton's first law

Newton's first law, in its most commonly quoted form, states:

> A body with zero net force acting on it moves at constant velocity.

There is also an alternate, perhaps more instructive formulation of Newton's first law that we will return to shortly, but we stick to the above interpretation for now. The first law implies that a body moving at constant velocity will continue to move at the same velocity forever, as long as the net force on the body is zero. In other words, the natural state of bodies left to themselves is *constant-velocity motion* (and *not* rest, as in the Aristotelian view).

This immediately raises the question of why moving bodies in everyday life generally come to rest when left to themselves, which seems to contradict the first law. The answer is simple—everyday moving objects are *not* left to themselves!

They are subject to friction, air (fluid) resistance, and other resistive forces, and as such do not meet the zero net force requirement for the first law to apply.

Newton's second law

To predict a body's motion, one must know the body's mass, the forces acting on the body, and a *law of motion* relating the two. The canonical law of motion in classical physics is precisely Newton's second law. The general form of Newton's second law states that the mass m, velocity $\boldsymbol{v}$, and net force $\boldsymbol{F}$ acting on a point-like body are related according the equation

$$\frac{\mathrm{d}(m\boldsymbol{v})}{\mathrm{d}t} = \boldsymbol{F}. \tag{2.109}$$

The quantity $m\boldsymbol{v}$ is called the body's *momentum* (we will return to momentum in a few lectures), and the second law says that the rate of change of a body's momentum with respect to time equals the net force acting on the body. Very often, since most bodies have constant mass, the derivative in Newton's second law reduces to

$$\frac{\mathrm{d}(m\boldsymbol{v})}{\mathrm{d}t} = m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = m\boldsymbol{a} = \boldsymbol{F}, \tag{2.110}$$

which is the $\boldsymbol{F} = m\boldsymbol{a}$ form probably familiar from high school and the popular science literature. We have three comments to make here:

- Newton's second law is valid only in *inertial frames of reference*. If you apply Newton's second law to a body in a non-inertial frame, your theoretical predictions will disagree with the body's observed motion, i.e. your predictions will be wrong! A formulation of dynamics in non-inertial frames is possible, but requires the introduction of *system forces*; we will return to this shortly.

- Newton's second law of motion applies to *point masses*. In practice this is not really a problem, since it turns out that the second law is easily generalized to systems of point masses, which are generally good approximations of real-life bodies.

- A body's acceleration and subsequent motion are the *result* of the force acting on the body. This is why we have written Eqs. 2.110 and 2.109 with acceleration/momentum (and not force) on the left-hand side—we wish to emphasize that accelerated motion is the result of an applied force. This is the physically instructive way to think about the second law, while the popular $\boldsymbol{F} = m\boldsymbol{a}$ can be misleading by giving the impression that force is the result of acceleration.

  Of course, mathematically, it makes no difference whatsoever if you write $\boldsymbol{F} = m\boldsymbol{a}$ or $m\boldsymbol{a} = \boldsymbol{F}$. Our goal here is just to emphasize that force and acceleration are not quite on equal footing—forces cause a body to accelerate, and not vice versa.

Finally, we stress that we have *not* derived Newton's second law, but simply quoted it. And that is the best physics currently knows how to do. One cannot derive the second law from fundamental principles because the second law *is* the fundamental principle. We take its validity as a fact of nature, because that is how an overwhelming body of experimental evidence shows that nature behaves. Keep in mind, as suggested in the very first sentence of this book, that *experiment* is the highest authority in physics.

On the necessity of Newton's fist law

Newton's second law raises an interesting question: If $\boldsymbol{F} = m\boldsymbol{a}$, which implies $\boldsymbol{a} = \boldsymbol{0}$ and thus $\boldsymbol{v} =$ constant when $\boldsymbol{F} = \boldsymbol{0}$, is Newton's first law redundant? In other words, if the first law is just a special case of the second, what is the point of the first? The answer is that the first law is not *just* a special case of the second. If we temporarily adopt a more mathematically-leaning, theoretical approach, Newton's first law is a sort of *existence theorem* for inertial frames. In this light, Newton's first law can be formulated as follows:

> If the net force on a body is zero, there exists an inertial frame of reference in which the body moves with constant velocity.

This then, is the purpose of the first law: Newton's *second* law is only valid in inertial frames, and Newton's *first* law guarantees that these frames exist. In more intuitive language, Newton's first law sets the stage in which the second law is valid.

On the well-definedness of inertial frames

Instead of sweeping it under the rug, now is a good opportunity to explicitly state the following:

> The key assumption underpinning the well-definedness of inertial frames, as we have defined them in this text, is the assumption that it is possible to identify all forces acting on a body, and to state with certainty that the net force on a body either is or is not zero.

Since we have defined forces as a body's interaction with well-defined, physical objects in its environment—things you can see or touch, or otherwise identify—the assumption that you can reliably identify forces is more or less reasonable. If you place a tested body far enough from everything else (and this is possible at least in principle), then the force on the tested body will be zero, simply because there is nothing else in the environment. Alternatively—and this is much easier in practice—one could arrange an experiment so as to ensure the *net* force on a tested body is zero (this is possible by the principle of superposition as long as one can identify the individual forces on the body) and thus determine if a frame is inertial or not. But the definition of inertial frames is still not as rigorous as one might like. For example, an Einstein quotation often cited in this context reads:

> "The weakness of the principle of inertia lies in this, that it involves an argument in a circle: a mass moves without acceleration if it is sufficiently far from other bodies; we know that it is sufficiently far from other bodies only by the fact that it moves without acceleration." [1]

Although this quote is somewhat out of context here (it comes originally from a lecture on general relativity, which dispenses with the Galilean concept of global inertial frames entirely) it captures well the struggles many students—the author included—have had at one point or another during their studies with pinpointing what exactly is a fault of their own understanding and what is a gray area of Newtonian theory. Our hope is that, from the student's perspective, it is helpful rather than disheartening to be clearly told that the assumption underlying our definition of inertial frames is that it must be possible to identify with certainty if forces do or do not act on an observed body.

In any case, once we transition from axiomatic questions of theoretical existence into practical application, the situation is much clearer. It turns out that in essentially all practical cases, it is quite straightforward to identify if the net force on a body is or is not zero, and to perform calculations in the scope of Newtonian mechanics (accounting for system forces if necessary) that perfectly agree with observed behavior within the realm of experimental error. The relevant question to ask is not "can Newtonian mechanics be formulated in a perfectly axiomatic, self-consistent manner?" but instead "while understanding its theoretical limitations, can we use Newtonian mechanics to predict the behavior of real-life objects in a manner that agrees with what is observed in nature?". And the answer to the second question is a resounding yes! Ultimately physics is an experimental science, and any attempts to fully axiomatize it in the style of mathematics or logic, where everything is self-consistent and well-defined, will probably fall short.

Newton's third law

Newton's third law is commonly quoted as something along the lines of "every action has an equal and opposite reaction". But that is not a particularly instructive way to think about the third law. A better formulation of Newton's third law might be:

(a) Forces arise from the *mutual interaction* between two bodies.

(b) In a dynamical interaction between, say, body A and body B, the force with which body A acts on body B is equal in magnitude and opposite in direction to the force with which body B acts on body A.

In particular, there is no concept of "first" and "second", "source" and "recipient", or even "action" and "reaction" built into Newton's third law. There are only two bodies, and that is that. Which body we call A which we call B is an arbitrary label of human construction, and interchanging the names does not change the underlying physics. Like Newton's first two laws, the third law is experimental—it is not derived from fundamental principles but assumed to hold based on an overwhelming body of experimental evidence that indicates that it does.

Two examples with Newton's third law

We now offer two real-life examples of force pairs in the context of Newton's third law:

(i) Consider a diver jumping into a swimming pool. As the diver falls towards the pool, the diver's weight (i.e. the gravitational force of the Earth on the diver) and the gravitational force of the diver on the Earth are an equal and opposite force pair. The diver attracts the Earth with exactly the same magnitude with which the Earth attracts the diver.

The reason, of course, that the diver seems to fall towards the Earth instead of the Earth rising up to meet the diver is the enormous difference in masses. The mass of the Earth is roughly $m_{\mathrm{E}} \sim 6 \cdot 10^{24}\,\mathrm{kg}$, while the mass of a diver might be closer to $m_{\mathrm{d}} \sim 60\,\mathrm{kg}$. The corresponding ratio of accelerations is

$$F_{\mathrm{E}} = F_{\mathrm{d}} \implies m_{\mathrm{E}} a_{\mathrm{E}} = m_{\mathrm{d}} a_{\mathrm{d}} \implies \frac{a_{\mathrm{E}}}{a_{\mathrm{d}}} = \frac{m_{\mathrm{d}}}{m_{\mathrm{E}}} \sim 6 \cdot 10^{-23}. \tag{2.111}$$

In the given time it takes the diver and Earth to come together, the Earth will have traveled a distance that is a factor $6 \cdot 10^{-23}$ smaller than the distance traveled by the diver. The distance traveled by the Earth is thus completely imperceptible.[27]

(ii) Now consider a block resting on a table. This situation is more subtle, because it implicitly involves three bodies: the block, the table, and the Earth. Here are four examples of correct third-law force pairs:

1. The contact force of the table's surface on the block and the reciprocal contact force of the block's bottom on the table's surface. (This type of contact force between two surfaces is called a *normal force*, and we will formally define it shortly.)
2. The block's weight and the corresponding gravitational force of the block on the Earth.
3. The normal force of the Earth's surface on table's legs and the reciprocal normal force of the table's legs on the Earth's surface.
4. The table's weight and the corresponding gravitational force of the table on the Earth.

And here is an example of an *incorrect* force pair from the perspective of Newton's third law:

~~The block's weight and the normal force of the table on the block.~~ (Incorrect!)

This example can be tricky for students the first time they encounter it, probably because as long as the table and block are at rest (as they usually are in everyday life) the block's weight and the normal force of the table on the block are indeed equal in magnitude and opposite in direction. But that is misleading. And if you think about forces in terms of interactions between two bodies, it becomes clear that the block's weight and the table-block normal force are unrelated by Newton's third law—the block's weight arises from its interaction with the Earth, and has nothing whatsoever to do with the table. But there is a more exciting way to reach the same conclusion: pick up the table, with the block still on it, and throw it out the window! You will find that in free fall, the block's weight is just the same as when the block was at rest, while the normal force of the table on the block is now zero. Any third law force pair must *always* be equal in magnitude and opposite in direction, so the block's weight and the table-block normal force are an incorrect force pair.

2.3.3 Some common forces from everyday life

We now list some of the more common mechanical forces from our daily lives, such as weight, friction, the normal force, etc., briefly discuss their mechanism, and provide relevant formulae for their practical computation.

[27] Of course there also countless other objects on the Earth's surface with mass similar to the diver that move the Earth every which way by similarly imperceptibly small amounts, so even if we had a measurement instrument that could measure distances to one part in $6 \cdot 10^{-23}$, we would probably be unable to isolate the effect of the diver.

Weight

Weight arises from the gravitational interaction between massive bodies, meaning that only bodies with mass can experience weight. In practice, the word "weight" refers to the gravitational force of a very massive body on another body with comparatively very little mass. For example, it makes sense to speak of the weight of a person on the Earth's surface or the weight of a spacecraft on the Moon, since both the person and spacecraft are many orders of magnitude less massive than the Earth or Moon. In such cases, it makes sense to attribute to the massive body a gravitational field $\boldsymbol{g}(\boldsymbol{r})$, which gives the gravitational acceleration $\boldsymbol{g}$ felt by the light body as a function of its position $\boldsymbol{r}$ from the massive body's center. We will discuss how to predict $\boldsymbol{g}(\boldsymbol{r})$ in the lectures on gravitation; for now we take its existence for granted.

Having introduced the concept of a planet's gravitational field $\boldsymbol{g}$, we can now give an expression for weight: the weight $\boldsymbol{F}_{\mathrm{g}}$ of a body of mass m under the influence of a gravitational acceleration $\boldsymbol{g}$ is

$$\boldsymbol{F}_{\mathrm{g}} = m\boldsymbol{g}. \tag{2.112}$$

In practice, m would represent something small, like the person, while the effects of the massive body (the planet, say) are encoded in the gravitational acceleration $\boldsymbol{g}$. For now we will study only the weight of everyday objects on the Earth's surface, in which case $\boldsymbol{g}$ is constant[28] and an object's weight is easy to calculate, since it depends only the object's mass m. (This is why people can confuse weight and mass, since in everyday situations a body's weight is determined only by its mass.)

Example: An object in free fall

For a body in free fall in the absence of air resistance, the only force on the body is its weight. The net force on the body is thus $\boldsymbol{F}_{\mathrm{net}} = \boldsymbol{F}_{\mathrm{g}}$, and Newton's second law for the free-falling body reads

$$m\boldsymbol{a} = \boldsymbol{F}_{\mathrm{net}} = \boldsymbol{F}_{\mathrm{g}} \overset{\text{(a)}}{=} m\boldsymbol{g} \implies \boldsymbol{a} = \boldsymbol{g}. \tag{2.113}$$

Recall that $\boldsymbol{a} = \boldsymbol{g}$ was the starting point for our kinematic analysis of free fall in Eqs. 2.59 and 2.80; we can now justify the use of $\boldsymbol{a} = \boldsymbol{g}$ on dynamical grounds.

Normal force

First, a note on terminology: the direction perpendicularly outward from a surface is called the *normal to the surface* in mathematical contexts. It is *this* use of the word normal that gives the normal force its name (and not the everyday use of the word normal to mean something regular, ordinary, or natural).

Normal force is the contact force exerted by a surface on a body in contact with the surface. As the name suggests, normal force always points perpendicularly outward from the surface exerting it, while its magnitude is such as to exactly cancel the net force applied by the body in direction normal to the surface. The normal force is a consequence of solids resisting deformation under applied stress, and our simple formulation of the normal force assumes the surface exerting it is perfectly rigid. While that language might sound complicated, the underlying concept is quite

[28]The magnitude of $\boldsymbol{g}$ on the Earth's surface is roughly $g = |\boldsymbol{g}| = 9.8\,\mathrm{m\,s^{-2}}$, and its direction is radially inward towards the Earth's center of mass (i.e. "down" in everyday language). We will often use the convenient approximation $g \approx 10\,\mathrm{m\,s^{-2}}$.

intuitive—in everyday terms, the normal force is what prevents you from falling through the floor or a book from falling through the desk it rests on.

Three examples involving the normal force

Here are three (hopefully) instructive examples, all involving a block resting on a flat surface.

1. Consider a block resting on a flat horizontal surface. As long as the block and the surface are at rest with respect to the Earth and no forces act on the block besides its weight...

   (a) The normal force of the block on the surface is equal in magnitude *and* in direction to the block's weight.

   (b) The normal force of the surface on the block is equal in magnitude and *opposite* in direction to the block's weight.

   The block-surface normal forces form a Newton's third law force pair (but the block's weight and either of the two normal forces do not—recall the second example with Newton's third law above).

2. Now consider a block resting on a flat surface inclined at an angle θ relative to the horizontal. In this case only a *component* of the block's weight is normal to the surface; this component, which we denote by $F_{\mathrm{g}_\perp}$, is equal to

   $$F_{\mathrm{g}_\perp} = mg\cos\theta. \tag{2.114}$$

   The magnitude of the normal force of the surface on the block is thus

   $$F_{\mathrm{n}} = F_{\mathrm{g}_\perp} = mg\cos\theta, \tag{2.115}$$

   while the direction, by definition, is normal to the surface. (The weight component parallel to the surface is $F_{\mathrm{g}_\parallel} = mg\sin\theta$, but only the component perpendicular to the surface is relevant in the context of the normal force.)

3. Finally, suppose someone puts a block on a flat table, then drops the block and table from the top of a tall building. As long as the block-table system is in free fall with acceleration $\boldsymbol{g}$, both the normal force of the block on the table and the normal force of the table on the block are zero. Loosely, the table is falling from under the block just as fast as the block is falling into the table, so the two do not exert any contact forces on each other.

   Of course, a large flat table will not fall with acceleration $\boldsymbol{g}$ but will experience considerable air resistance. In this more realistic scenario, the normal force of the table on the block would equal the upward force of air resistance on the falling table.

Friction

Friction is a resistive force that arises from the relative motion between two surfaces in contact.[29] Friction always acts so as to oppose (or informally, "slow down") the

[29] We will use the word *friction* in the context of one solid moving relative to another solid and the term *fluid resistance* in the context of a solid moving through a fluid, such as air or water. However, you might also hear fluid resistance referred to as "fluid friction", especially in engineering circles.

relative motion of two surfaces. For example, a book sliding across a table comes to a halt because of the friction force of the table on the book.

Friction arises largely from two sources: (i) chemical adhesion between the two surfaces and (ii) mechanical interlocking between the microscopic bulges, ridges, holes, and other imperfections on the two surfaces. In truth, the simplified view of friction taught in introductory mechanics courses is just a convenient macroscopic abstraction for a complex microscopic phenomenon that is still a subject of active research and not fully understood. We will study two types of friction, *kinetic* friction and *static* friction, both only in the scope of the macroscopic abstraction.

Kinetic friction

Kinetic friction is the force of friction between two *moving* surfaces. For the purposes of our course, the magnitude of kinetic friction f_{kin} on a body sliding along a surface is

$$f_{\mathrm{kin}} = k_{\mathrm{kin}} F_{\mathrm{n}}, \tag{2.116}$$

where F_{n} is the normal force of the surface on the body and k_{kin} is a dimensionless constant, called the *coefficient of kinetic friction*, that depends on the material properties of both the body and the surface. Note that the magnitude of kinetic friction is proportional to the magnitude of the normal force, which is what we might expect from everyday life—the harder you press two surfaces together, the more difficult it is to slide them past each other.

In this course we will treat the coefficient of kinetic friction as independent of velocity. Although this approximation holds good in most cases, kinetic friction does have a weak dependence on the relative velocity of the two surfaces. For orientation, typical values of k_{kin} for contacts between everyday materials range from 0.1 to 1.0.

The direction of the kinetic friction force on a body is always opposite the body's instantaneous velocity relative to the surface causing the friction; in symbols this could be written $\boldsymbol{f}_{\mathrm{kin}} \parallel -\boldsymbol{v}$. The vector form of kinetic friction on a body moving at velocity $\boldsymbol{v}$ relative to a surface exerting normal force F_{n} is thus

$$\boldsymbol{f}_{\mathrm{kin}} = -k_{\mathrm{kin}} F_{\mathrm{n}} \frac{\boldsymbol{v}}{|\boldsymbol{v}|} \equiv -k_{\mathrm{kin}} F_{\mathrm{n}}\, \hat{\mathbf{v}}, \tag{2.117}$$

where $\hat{\mathbf{v}} \equiv \boldsymbol{v}/|\boldsymbol{v}|$ is the unit vector in the direction of the body's instantaneous velocity (discussed more in the context of (2.24)) and k_{kin} is the coefficient of kinetic friction between the body and the surface.

Static friction

Static friction opposes the *potential* relative motion of two surfaces *at rest*. In the macroscopic approximation, the magnitude of static friction f_{s} on a body resting on some surface is

$$f_{\mathrm{s}} \leq k_{\mathrm{s}} F_{\mathrm{n}}, \tag{2.118}$$

where F_{n} is the normal force of the surface on the body and k_{kin} is called the *coefficient of kinetic friction* and depends on the material properties of both the body and the surface. Often, but not always, $k_{\mathrm{kin}} \lesssim k_{\mathrm{s}}$ and thus $f_{\mathrm{kin}} \lesssim f_{\mathrm{s}}$ for a given normal force F_{n}. In everyday language $f_{\mathrm{kin}} < f_{\mathrm{s}}$ means that it takes less force to keep a moving body moving than it does to put a static body into motion.

The direction of $\boldsymbol{f}_{\mathrm{s}}$ is such as to oppose the hypothetical direction in which a body would move in the absence of static friction. In this sense, the effect of $\boldsymbol{f}_{\mathrm{s}}$ is to cancel out the component of the net force on a body *parallel* to a surface.[30] One then interprets the inequality $f_s \leq k_s F_n$ as follows: f_s will be just large enough, but *no larger than is necessary* (hence the inequality), to stop whatever forces act on the body parallel to the surface from making the body slip. Of course, this can only go on so long; at some point the disruptive forces will reach $k_s F_n$, beyond which f_s cannot increase, and the body slips. Just at the slipping point (2.118) becomes $f_s = k_s F_n$, and after the body slips, kinetic friction replaces static friction.

Example: Measuring static friction

One can measure the coefficient of static friction between two materials (we will use steel and wood for concreteness) with the following experiment. Place the steel block on a wooden ramp with an adjustable incline, and let θ denote the ramp's incline relative to the horizontal.

In small increments, increase the ramp's incline, and measure the angle θ_{max} at which the block first slips off the ramp. Just as the block slips, the magnitude of static friction f_s and the block's weight are related according to

$$f_s = F_{g_\parallel} = mg \sin \theta_{\text{max}}, \tag{2.119}$$

where $F_{g_\parallel}$ is the weight component parallel to the ramp's surface. Meanwhile, the static friction and normal force on the block are in general related by the inequality $f_s \leq k_s F_n = mg \cos \theta$, where (assuming the block and ramp are at rest) the normal force on the block is $F_n = F_{g_\perp} = mg \cos \theta$, i.e. the block's weight component normal to the ramp's surface. Just as the block slips, this simplifies to the *equality*

$$f_s = k_s F_n = k_s \cdot (mg \cos \theta_{\text{max}}). \tag{2.120}$$

We then equate the expressions for f_s in Eqs. 2.119 and 2.120 to get

$$mg \sin \theta_{\text{max}} = k_s mg \cos \theta_{\text{max}} \implies k_s = \tan \theta_{\text{max}}, \tag{2.121}$$

where θ_{max} is the ramp incline angle at which the block first slips. Note that a coefficient of static friction larger than one is perfectly possible on physical grounds—this corresponds to a value of θ_{max} larger than $45°$.

Spring force

For purposes of this course, the spring force $\boldsymbol{F}_{\text{spring}}$ produced by a spring compressed or extended by a displacement $\Delta \boldsymbol{x}$ relative to the spring's equilibrium position is

$$\boldsymbol{F}_{\text{spring}} = -k \Delta \boldsymbol{x}, \tag{2.122}$$

where k, called the *spring constant*, is a property of the spring encoding the spring's stiffness. Equation 2.122 is called *Hooke's law* in honor of the English scientist Robert Hooke, who made the first recorded observations of the linear relationship between stress and strain in elastic materials.

Because of the minus sign, the spring force is opposite in direction to the displacement $\Delta \boldsymbol{x}$. In everyday language, this is why a spring will push your hands apart when you compress it and pull your hands to together when you stretch it.

[30] Just like the normal force cancels the force component *perpendicular* to a surface

Note that Hooke's law does not hold for arbitrarily large displacements; if nothing else, this is obvious from daily life, since a spring will eventually snap if stretched too far. We will return to the limits of linear elasticity in the lectures on deformation.

Tension forces and pulleys

Ropes, strings, cables, and so on, often in combination with pulleys, can be used to transfer, redirect, and otherwise manipulate a force applied to one of the rope's ends. In this context, the force exerted by the rope on the objects attached to it is called a *tension force*. Ropes and pulleys commonly feature in introductory mechanics courses because they are relatively straightforward to analyze and offer instructive practice with Newton's laws of motion.

We will not have much more to say about rope-pulley systems in these lectures (you will see them more in the *Exercises* portion of this course) and mention them here only for completeness. We will return to ropes, in a different context, in the lectures on mechanical waves.

2.3.4 Dynamics of uniform circular motion

Recall from Section 2.2.5 that a body in circular motion is accelerating, since the direction of the body's velocity continuously changes as the body moves around the circle. Acceleration is possible only under the influence of a force, and in this context the force responsible for a body's circular motion gets a special name: the *centripetal force*. Centripetal force is just a label given to an existing force responsible for causing circular motion; it makes sense to say, for example, that weight is the centripetal force for a satellite orbiting the earth, or that tension is the centripetal force for a ball tied to a string that you are whirling in a circle around your head. The centripetal force is *not* a new type of force, just a special name to remind you of which *existing* force is responsible for causing circular motion.

For review from the end of Section 2.2.4, in general a body's acceleration can be decomposed into components parallel and perpendicular to the body's instantaneous velocity $\boldsymbol{v}$ in the form

$$\boldsymbol{a} = \boldsymbol{a}_\perp + \boldsymbol{a}_\parallel \qquad \text{or} \qquad \boldsymbol{a} = \boldsymbol{a}_\mathrm{c} + \boldsymbol{a}_\mathrm{t}, \tag{2.123}$$

where $\boldsymbol{a}_\mathrm{c} = \boldsymbol{a}_\perp$ and $\boldsymbol{a}_\mathrm{t} = \boldsymbol{a}_\parallel$ are just two equivalent notations for the acceleration components perpendicular and parallel to $\boldsymbol{v}$, respectively. The centripetal component $\boldsymbol{a}_\mathrm{c}$ is responsible for causing uniform circular motion. For review from (2.55), the component $\boldsymbol{a}_\mathrm{c}$ is

$$\boldsymbol{a}_\mathrm{c} = -\omega^2 \boldsymbol{r}, \tag{2.124}$$

where ω^2 is the body's angular velocity and $\boldsymbol{r}$ is the body's position relative to the center of rotation. Any force acting as the centripetal force for a body of mass m in circular motion must then be equal to

$$\boldsymbol{F}_\mathrm{c} = m\boldsymbol{a}_\mathrm{c} = -m\omega^2 \boldsymbol{r}. \tag{2.125}$$

Like $\boldsymbol{a}_\mathrm{c}$, the centripetal force points to the center of rotation. Using the various equivalent expressions from (2.44), the centripetal force can also be written

$$\boldsymbol{F}_\mathrm{c} = -m\omega^2 \boldsymbol{r} = -mv\omega\,\hat{\mathbf{e}}_r = -m\frac{v^2}{|\boldsymbol{r}|}\,\hat{\mathbf{e}}_r, \tag{2.126}$$

where in this context $\hat{\mathbf{e}}_r = \boldsymbol{r}/|\boldsymbol{r}|$ is the unit vector in the direction of $\boldsymbol{r}$. In everyday language, $\hat{\mathbf{e}}_r$ is just the direction from the center of the circle to the body's current position on the circle's circumference, while the minus sign in (2.126) means that $\boldsymbol{F}_\mathrm{c}$ points radially inward toward the center of rotation, opposite the direction of $\hat{\mathbf{e}}_r$.

Example: Car turning a bend on a flat road

Consider a car, which we will treat as a point mass of mass m, driving in a circle of radius R on a flat road. In this case the force responsible for the car's centripetal acceleration is the static friction[31] of the road surface on the car's wheels. Using (2.118), the magnitude of this static friction is

$$f_\mathrm{s} \leq k_\mathrm{s} F_\mathrm{n}, \tag{2.127}$$

where k_s is the coefficient of static friction between the car wheels and the road surface, and F_n is the normal force of the road on the car. Assuming the car is at rest in the *vertical* direction, the normal force and the car's weight are equal in magnitude, i.e. $F_\mathrm{n} = mg$, so

$$f_\mathrm{s} \leq k_\mathrm{s} F_\mathrm{n} = k_\mathrm{s} mg. \tag{2.128}$$

The static friction is the only force acting as the centripetal force in this context (there are no other horizontal forces on the car, neglecting air resistance) and is thus equal in magnitude to

$$|\boldsymbol{F}_\mathrm{c}| \overset{\text{(a)}}{=} \frac{mv^2}{|\boldsymbol{r}|} \overset{\text{(b)}}{=} \frac{mv^2}{R} = f_\mathrm{s} \leq k_\mathrm{s} mg. \tag{2.129}$$

where (a) uses the final expression in (2.126) and (b) assumes a constant radius of rotation $|\boldsymbol{r}| = R$. We then simplify the expression and rearrange to get the condition

$$\frac{mv^2}{R} \leq k_\mathrm{s} mg \implies v^2 \leq k_\mathrm{s} gR. \tag{2.130}$$

Our analysis predicts that the car will round the curve without slipping as long as its squared velocity v^2 is less than $k_\mathrm{s} gR$. How does this prediction compare to everyday observations? To make things exciting (and make the calculation turn out nicely), let's consider a 50 m radius of curvature on a race track, where the coefficient of static friction between race car tires and the specialized asphalt used for racing circuits can reach $k_\mathrm{s} = 0.8$. Taking $g \approx 10\,\mathrm{m\,s^{-2}}$, the car will round the curve without slipping for velocities

$$v^2 \leq k_\mathrm{s} gR = 0.8 \cdot 10\,\mathrm{m\,s^{-2}} \cdot 50\,\mathrm{m} = 400\,\mathrm{m^2\,s^{-2}}.$$

The maximum permissible velocity is

$$v_\mathrm{max} = \sqrt{400\,\mathrm{m^2\,s^{-2}}} = 20\,\mathrm{m\,s^{-1}} \approx 70\,\mathrm{km\,h^{-1}}.$$

[31] As long as the car is not skidding, the friction between the wheels and road is static and not kinetic. From a physical perspective, kinetic friction can do just as good a job causing circular motion. But unless you are a race car driver, you probably don't want to be in a car that is rounding a curve under the influence of kinetic friction—that means the car is skidding and probably out of control! We will discuss rolling without slipping in the lectures on rotational mechanics.

The result, $v_{\max} \approx 70\,\mathrm{km\,h^{-1}}$, is a reasonable speed for a race car around a large bend. The main lesson here is that even a basic analysis (using a point mass model, neglecting air resistance and slipping, etc...) agrees to a first approximation with behavior observed in everyday life.

There are a few ways to increase the maximum speed $v_{\max}$ beyond the result calculated above. In automotive racing, there exist a whole host of engineering tricks, such as airfoils, fans, specialized chassis geometry, and so on, that redirect airflow so as to produce a downforce on the car. This downforce increases the normal force of the road on the car beyond the car's weight, thereby increasing the friction between the car and road. This is called "ground effect" in the language of automotive engineering. Meanwhile, in both racing circuits and in civil engineering (for example in highway exits), it is common to use banked curves to increase safe turning speeds. On a banked curve, the component of the normal force on the car that points radially inward toward the center of rotation can act as a centripetal force. You will study banked curves in the *Exercises* portion of this course.

A Topics from mechanics

A.1 Velocity and acceleration in general circular motion

In this section we generalize the results of Section **??** and derive the linear velocity $\boldsymbol{v}(t)$ and acceleration $\boldsymbol{a}(t)$ of a body in general circular motion of radius R with time-varying angular velocity $\omega(t) = \mathrm{d}\varphi / \mathrm{d}t$. Note that this material is completely optional—it belongs in a second year course and you are not required to understand it now. We include it only for more advanced students interested in a more formal treatment of circular motion.

We use a two-dimensional polar coordinate system with the origin at the center of rotation and work in terms of the polar unit vectors $\hat{\mathbf{e}}_\varphi$ and $\hat{\mathbf{e}}_r$. In terms of the standard unit vectors $\hat{\mathbf{e}}_x$ and $\hat{\mathbf{e}}_y$, the polar unit vectors $\hat{\mathbf{e}}_r$ and $\hat{\mathbf{e}}_\varphi$ read

$$\hat{\mathbf{e}}_r = \cos\varphi\,\hat{\mathbf{e}}_x + \sin\varphi\,\hat{\mathbf{e}}_y \qquad \text{and} \qquad \hat{\mathbf{e}}_\varphi = -\sin\varphi\,\hat{\mathbf{e}}_x + \cos\varphi\,\hat{\mathbf{e}}_y, \tag{A.1}$$

where φ is the body's angular coordinate, measured with respect to the x axis. For later use, the unit vectors and their derivatives are related according to

$$\frac{\mathrm{d}\,\hat{\mathbf{e}}_r}{\mathrm{d}\varphi} = -\sin\varphi\,\hat{\mathbf{e}}_x + \cos\varphi\,\hat{\mathbf{e}}_y = \hat{\mathbf{e}}_\varphi \tag{A.2}$$

$$\frac{\mathrm{d}\,\hat{\mathbf{e}}_\varphi}{\mathrm{d}\varphi} = -\cos\varphi\,\hat{\mathbf{e}}_x - \sin\varphi\,\hat{\mathbf{e}}_y = -\hat{\mathbf{e}}_r. \tag{A.3}$$

In terms of the polar unit vectors, the position vector $\boldsymbol{r}(t)$ of a body in circular motion is

$$\boldsymbol{r}(t) = r(t)\,\hat{\mathbf{e}}_r(t) \overset{\text{(a)}}{=} R\,\hat{\mathbf{e}}_r(t), \tag{A.4}$$

where using $r(t) = R$ in (a) follows from radial distance from the origin being constant in circular motion.

Velocity in general circular motion

Using (A.4) to express position $\boldsymbol{r}(t)$, the body's velocity is

$$\boldsymbol{v}(t) = \frac{\mathrm{d}\boldsymbol{r}}{\mathrm{d}t} = \frac{\mathrm{d}R}{\mathrm{d}t}\,\hat{\mathbf{e}}_r(t) + R\frac{\mathrm{d}\,\hat{\mathbf{e}}_r}{\mathrm{d}t} \overset{\text{(a)}}{=} R\frac{\mathrm{d}\,\hat{\mathbf{e}}_r}{\mathrm{d}t}, \tag{A.5}$$

where (a) uses the fact that $R = \text{constant} \implies \mathrm{d}R / \mathrm{d}t = 0$. Using the chain rule, we then make the auxiliary calculation

$$\frac{\mathrm{d}\,\hat{\mathbf{e}}_r}{\mathrm{d}t} = \frac{\mathrm{d}\,\hat{\mathbf{e}}_r}{\mathrm{d}\varphi}\frac{\mathrm{d}\varphi}{\mathrm{d}t} \overset{\text{(a)}}{=} \frac{\mathrm{d}\varphi}{\mathrm{d}t}\,\hat{\mathbf{e}}_\varphi \overset{\text{(b)}}{=} \omega(t)\,\hat{\mathbf{e}}_\varphi, \tag{A.6}$$

where (a) uses $\hat{\mathbf{e}}_\varphi = \mathrm{d}\,\hat{\mathbf{e}}_r / \mathrm{d}t$ from (A.2) and (b) uses $\omega(t) = \mathrm{d}\varphi / \mathrm{d}t$. Substituting (A.6) into (A.5), the body's velocity is thus

$$\boldsymbol{v}(t) = R\frac{\mathrm{d}\,\hat{\mathbf{e}}_r}{\mathrm{d}t} = R\omega(t)\,\hat{\mathbf{e}}_\varphi, \tag{A.7}$$

in agreement with the geometrically-derived result in (2.40). Similarly, the magnitude of the body's velocity is

$$|\boldsymbol{v}(t)| = v(t) = R\omega(t), \tag{A.8}$$

in agreement with the result in (2.39).

Acceleration in general circular motion

Using (A.8) to express the velocity $\boldsymbol{v}(t)$, the body's acceleration is

$$\boldsymbol{a}(t) = \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = R\frac{\mathrm{d}\omega}{\mathrm{d}t}\,\hat{\mathbf{e}}_\varphi + R\omega(t)\frac{\mathrm{d}\,\hat{\mathbf{e}}_\varphi}{\mathrm{d}t}. \tag{A.9}$$

Using the chain rule, we then make the auxiliary calculation

$$\frac{\mathrm{d}\,\hat{\mathbf{e}}_\varphi}{\mathrm{d}t} = \frac{\mathrm{d}\,\hat{\mathbf{e}}_\varphi}{\mathrm{d}\varphi}\frac{\mathrm{d}\varphi}{\mathrm{d}t} \overset{\text{(a)}}{=} -\frac{\mathrm{d}\varphi}{\mathrm{d}t}\,\hat{\mathbf{e}}_r \overset{\text{(b)}}{=} -\omega(t)\,\hat{\mathbf{e}}_r, \tag{A.10}$$

where (a) uses $\hat{\mathbf{e}}_r = -\,\mathrm{d}\,\hat{\mathbf{e}}_\varphi/\,\mathrm{d}t$ from (A.3) and (b) uses $\omega(t) = \mathrm{d}\varphi/\,\mathrm{d}t$. Substituting (A.10) into (A.9), the body's acceleration is thus

$$\boldsymbol{a}(t) = R\frac{\mathrm{d}\omega}{\mathrm{d}t}\,\hat{\mathbf{e}}_\varphi - R\omega^2(t)\,\hat{\mathbf{e}}_r. \tag{A.11}$$

For uniform circular motion $\mathrm{d}\omega/\,\mathrm{d}t = 0$, in which case (A.11) simplifies to

$$\boldsymbol{a}(t) = -R\omega^2\,\hat{\mathbf{e}}_r, \qquad \text{(uniform circular motion)} \tag{A.12}$$

in agreement with the result derived in (2.55) for the special case of uniform circular motion. Finally, in the language of decomposed acceleration $\boldsymbol{a} = \boldsymbol{a}_\mathrm{c} + \boldsymbol{a}_\mathrm{t}$, from Equations (2.25) and (2.56), the tangential and centripetal acceleration components for general circular motion are

$$\boldsymbol{a}_\mathrm{t} = R\frac{\mathrm{d}\omega}{\mathrm{d}t}\,\hat{\mathbf{e}}_\varphi \qquad \text{and} \qquad \boldsymbol{a}_\mathrm{c} = -R\omega^2\,\hat{\mathbf{e}}_r, \tag{A.13}$$

which follows directly from (A.11) together with $\boldsymbol{v} \parallel \hat{\mathbf{e}}_\varphi$ and $\hat{\mathbf{e}}_r \perp \hat{\mathbf{e}}_\varphi$.

References

[1] Albert Einstein. “The General Theory of Relativity”. In: *The Meaning of Relativity; Four lectures delivered at Princeton University*. Project Gutenberg, May 1921, p. 62. URL: https://www.gutenberg.org/files/36276/36276-pdf.pdf.