Physics I:
Introduction to physics

PHYSIK INSTITUT
UNIVERSITÄT ZÜRICH

Prof. Ben Kilminster

Introduction to mechanics, waves, and fluid dynamics

This script is the first part of an undergraduate course in introductory physics. It is typically taught in the first semester, with part 2 often taught in the second semester. The level of material is appropriate for physics majors as well as those in the life sciences. The latter may not be expected to learn the full level of detail included that would be expected of physics majors, but may still benefit from the additional material in order to understand better. It is recommended that students should be already familiar with geometry, trigonometry, and also take or have taken a class in mathematics that covers vectors and calculus (derivatives and integrals).

Last edited on February 20, 2024
Typeset & figures by Izaak Neutelings

Contents

Introduction

Physics! What exactly is it and what is it useful for? Physics deals with matter, energy, the principles of motion for particles and waves, and their interactions. It aims to explain properties and behavior of things at the small level, such as molecules, atoms, nuclei and quarks; but also at the larger scale, such as gases, liquids, solids, but also plants, stars, solar systems and star clusters. Physics is the study of the fundamental laws of nature, and is necessary to understand chemistry, biology, astronomy, cosmology, etc. at a fundamental level. If you want to understand the basics of how things work, this is the place to be!

These lecture notes cover the basics of classical mechanics for first-year physics students. We will starts off with some basic concepts like units and dimension and how to describe motion with *kinematics*. After that, we will learn about the Newtonian laws that govern motion using forces. Then we will discuss fluid dynamics, and finally finish off with the study of waves mechanics.

A good reference and source of materials for the topics discussed here can be found in Part I of *Fundamentals of Physics* by David Halliday, Robert Resnick & Jearl Walker. It contains many more examples and exercises. This book is available in both English and German.

In case you find any errors or typos, please send an e-mail to `izaak.neutelings@uzh.ch` and `ben.kilminster@physik.uzh.ch`.

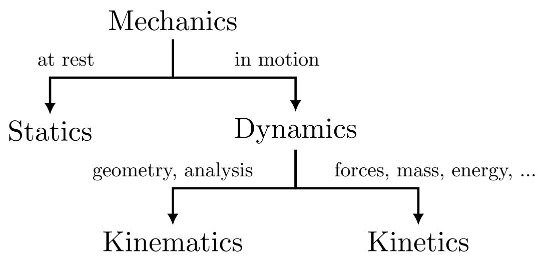


Figure 1: Mechanics and its branches.

Part I

Mechanics

Chapter 1

Units & Dimensions

Before we can start uncovering the underlying principles of Nature we need to understand the basic ingredients.

1.1 Fundamental definition of units

The basic units that are used in science are defined by the *International System of Units* (SI). The system has seven basic units for seven quantities, or *dimensions*:

- second (s) for time,
- meter (m) for distance or length,
- kilogram (g) for mass,
- Ampere (A) for electrical current,
- Kelvin (K) for temperature,
- mole (mol) for amount of substance, and
- candela (C) for luminous intensity.

This semester, we will focus only on time, distance and mass.

From all of these units, we can derive any other unit we need in physics. Table 1.1 contains some of examples of derived units we need this semester. Table 1.2 lists the official prefixes in the SI unit system.

Note that the choice of these seven dimensions and their units are in some sense arbitrary. Alien scientists on another planets might have chosen a different set of dimensions or units to base their alien physics on, although we still expect the laws of physics to be the same.

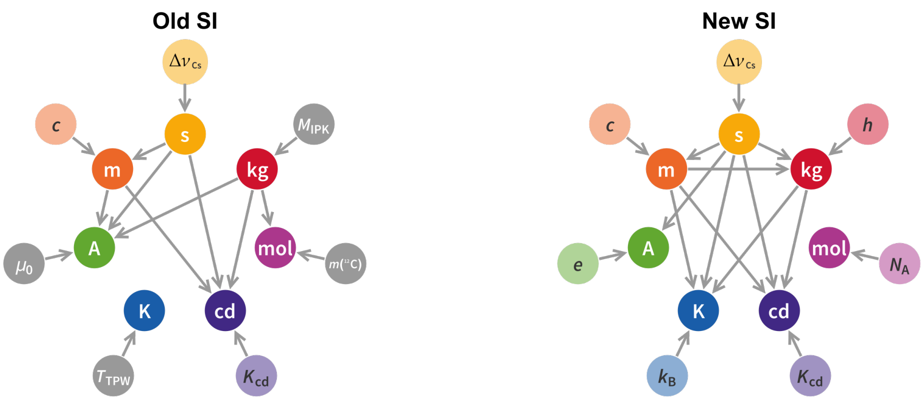
However, the definition of units needs to be rigorously standardized for scientific, technological, and financial reasons (such as trade). Before the world was as interconnected as is today, each village could have its own definition for length, mass or time. They were often based on the (average) length of some body part, like feet, hands or forearms. Besides the fact that the average foot length could be larger one village over, how can you convert between these? How many elbows are there in one foot? At some point, people tried to fix units more widely with some standards. In the case of the metric system, on which the SI units are based, this came about after the French Revolution. These standards would be some physical artefact, or *prototype*. It could be a bar that set the definition of the meter,

Table 1.1: On top are the basic units of distance, time, and mass. Below are units derived from these basic units. Careful: the formulas shown are simplified examples, and there may be a more precise formula needed for a particular problem.

Measurement	Symbol	Unit
Distance	x	meter (m)
Time	t	second (s)
Mass	m	kilogram (kg)
Velocity	$v = x/t$	m/s
Acceleration	$a = v/t$	$\mathrm{m/s^2}$
Momentum	$p = mv$	kg m/s
Force	$F = ma$	Newton ($\mathrm{N = kg\,m/s^2}$)
Energy	$E = Fx$	Joule ($\mathrm{J = kg\,m^2/s^2}$)

Table 1.2: Metric prefixes to indicated different orders of magnitude.

Name	Symbol	Base 10	Decimal
peta	P	10^{15}	1 000 000 000 000 000
tera	T	10^{12}	1 000 000 000 000
giga	G	10^{9}	1 000 000 000
mega	M	10^{6}	1 000 000
kilo	k	10^{3}	1000
hecto	h	10^{2}	100
deca	da	10^{1}	10
–	–	10^{0}	1
deci	d	10^{-1}	0.1
centi	c	10^{-2}	0.01
milli	m	10^{-3}	0.001
micro	µ	10^{-6}	0.000 001
nano	n	10^{-9}	0.000 000 001
pico	p	10^{-12}	0.000 000 000 001
femto	f	10^{-15}	0.000 000 000 000 001
atto	a	10^{-18}	0.000 000 000 000 000 001



(a) Old definitions. **(b)** New definitions since May 2019.

Figure 1.1: Definition of SI base units and their interdependence. The fundamental units are the meter (length), second (time), kilogram (mass), mol (amount of substance), candela (luminous intensity), Kelvin (temperature) and ampere (current). Taken from Wikipedia.

or a block of weight that set the kilogram. These would be securely stored in Paris, and other institutes could request a copy. However, over time this method proved unreliable.

A new approach was chosen, which took effect as of May of 2019. The base units of the SI system are now all defined to depend on the measurements of fundamental constants, which have been fixed to one value for ever. Figure Fig. 1.1 summarizes the definition of the base units with constants of Nature and their interdependence. We will look at some examples below.

1.1.1 Time

Time used to be defined by the orbit of the earth around the sun. Now, however, we use a more precise atomic clock. This "clock" is based on the transition rate of Cesium-133 atoms. When a liquid of Cesium isotopes is heated up inside an oven, it will emit a beam of Cesium atoms in either of two states. A microwave cavity is precisely tuned to switch between these two states based on the natural oscillation frequency of the Cesium, and magnets are used to select one of these states, such that when the frequency is perfectly matched, we get a Cesium beam that is maximally detected as being in one of these states. Therefore a second is defined exactly as the time it takes for 9 192 631 770 oscillations of the microwaves in the cavity. The relative precision of such an atomic clock is typically

$$\frac{\Delta T}{T} \sim \frac{1\,\mathrm{s}}{100\,\mathrm{My}}, \tag{1.1}$$

so 1 second in 100 million years.

1.1.2 Distance

The meter was defined such that distance from the equator through Paris to the north pole would be 10^7 meters. The circumference of the earth is about 40 000 km. Now, the meter is defined using the definition of the speed of light, c, and the definition of the seconds. One meter is defined as exactly the distance light travels in

$$\frac{1}{299\,792\,458}\,\text{seconds}, \tag{1.2}$$

such that

$$c = 299\,792\,458\,\mathrm{m/s}. \tag{1.3}$$

Notice that this definition of the meter relies on the definition of the second.

1.1.3 Mass

The kilogram used to be defined by a standard object, which was a metal object kept under vacuum in a vault in Paris. Since May of 2019, it is defined using Planck's constant h. This constant relates time and energy via

$$h = 6.626\,070\,15 \times 10^{-34}\,\mathrm{Js}. \tag{1.4}$$

In a later physics course, you may determine h from Einstein's equation for the energy E of a photon with frequency ν,

$$E = h\nu. \tag{1.5}$$

The kilogram then, is defined by seconds (from the Cesium clock), meters (from c), and Joules (from h).

1.2 Dimensional analysis

We know we cannot add seconds to meters, as they have different dimensions, time and length, respectively. What about

$$x = \frac{E}{F} + vt + \frac{a}{v}, \tag{1.6}$$

with distance x, energy E, force F, velocity v, time t and acceleration a? Is this a valid equation? From Table 1.1, we know that the units of the first term on the right-hand side is

$$\left[\frac{E}{F}\right] = \frac{[E]}{[F]} = \frac{\mathrm{kg\,m^2/s^2}}{\mathrm{kg\,m/s^2}}, \tag{1.7}$$

while the second term is

$$[vt] = \mathrm{m/s \cdot s = m}, \tag{1.8}$$

and the third term,

$$\left[\frac{a}{v}\right] = \frac{[a]}{[v]} = \frac{\mathrm{m/s^2}}{\mathrm{m/s}} = \frac{1}{s}. \tag{1.9}$$

Each term in Eq. (1.6) has different units, so the equation cannot possibly be valid.

This check is called *dimensional analysis*. Its one of the most powerful tools at the disposal of a physicist. If you came to a final results after a lot of algebra, it is one way to test if you made any mistakes along the way. It can also be useful to remember formulas. In fact, dimensional analysis is even used to guess the form of some unknown equation as *ansatz*, if one knows what the possible ingredients (i.e. the possible variables) and restrictions of your problem are.

Chapter 2

Measurement & Uncertainty

Let's say we want to measure some distance x. To get a more precise value, it's often better to make the measurement several times and take the *average*:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{2.1}$$

where x_i are n measurements of x. This is our data set. The symbol $\sum$ stands for sum. The n in $\sum_{i=1}^{n}$ means that we have n measurements to sum over, iteratively, starting with the first one that is labeled $i = 1$. The average is sometimes written as the *expected value* $\langle x\rangle = \overline{x}$ of x, written with brackets. For example, if we have 3 measurements: 2, 3, and 7, then n = 3, $x_1 = 2$, $x_2 = 3$, and $x_3 = 5$. The average would equal $\overline{x} = \frac{1}{3}\sum_{i=1}^{n=3} x_i = \frac{1}{3}(2 + 3 + 7) = 4$.

However, measurements in science are meaningless without some estimate of the *uncertainty*, or *error*, on the result. They are typically denoted with σ. There are two types of uncertainties that we will see now.

2.1 Statistical uncertainty

In statistics, we typically look at how spread out our data set is. If the real value is x, then the *standard deviation* is defined as

$$\sigma = \frac{1}{n}\sqrt{\sum_{i}^{n}(x - x_i)^2}. \tag{2.2}$$

However, we don't know the true value of x, since this is exactly what we are trying to measure. So instead we substitute the average, and the formula becomes

Standard deviation.

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i}^{n}(\overline{x} - x_i)^2}, \tag{2.3}$$

such that for $n \to \infty$, these two definitions are the same. This number gives you an idea of the "spread", or *variance*, of your measurements.

We can also estimate what the uncertainty in our measurement of the average $\overline{x}$ is:

Measurement error.

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)} \sum_{i}^{n} (\overline{x} - x_i)^2}. \tag{2.4}$$

This is our *measurement error* or *statistical error* in our measurement of $\overline{x}$. It is also called the *standard error of the mean*. Again, as our data set increases, our uncertainty becomes smaller.

Example 2.1: Consider the data set of measurements

$$\{21.4, 21.0, 20.9, 21.3, 21.0, 21.3, 21.2, 21.0\}. \tag{2.5}$$

The average is $\overline{x} = 21.14$, the standard deviation of the data set is $\sigma = 0.18$ and the uncertainty in, or standard error on, $\overline{x}$ is $\sigma_{\overline{x}} = 0.07$. These numbers are visualized in Fig. 2.1.

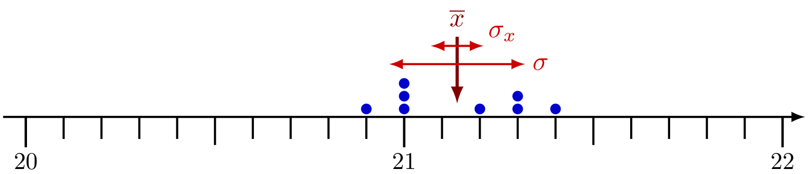


Figure 2.1: Visualization of the data set in Example 2.1.

This will be covered in a more detail if you take physics practica courses such as PHY112.

2.2 Systematic uncertainty

The *systematic uncertainty* is related to the limited knowledge you have of your measurement instruments. As a rule of thumb, this uncertainty is often estimated as the smallest unit of measurement. This is your *precision*. For example, the precision of most high school rulers is $\sigma_x = 1\,\text{mm}$, because this is the unit of the smallest lines can read off.

However, most instruments can also have some unknown "shift" from the real value called the *bias*. This is referred to as the *accuracy* of your instrument. By properly calibrating your device, the systematic uncertainty can be reduced.

For example, if you want to accurately measure the temperature outside, but the thermometer is located next to a light bulb that emits heat, then your thermometer will absorb this extra heat, and systematically measure the temperature higher than the actual temperature. This would be a systematic uncertainty, and you could estimate its effect by comparing your outside temperature to one from the weather app on your phone. You may find that the difference is always less than °C, so you define your systematic uncertainty to be ± 2°C. However, you could also do a test and find that the temperature is 1 degree warmer when the light is on than when it is off. Now you could correct for this systematic uncertainty and eliminate it.

In scientific studies, it is important to consider potential sources of systematic uncertainties and try to correct for the largest ones. A general rule of thumb is to assess and correct for systematic uncertainties if they are larger than your statistical uncertainties.

2.3 Propagation of errors

In class we may measure the speed of light by measuring the time t light travels over a distance x. The speed of light then, simply is

$$c = \frac{x}{t}. \tag{2.6}$$

Suppose we estimate the precision of the time and distance measurements as $\sigma_t = 2\,\text{ns}$ and $\sigma_x = 0.01\,\text{m}$, and now we want to estimate the uncertainty σ_c in our result c.

In general, if $f = f(x, y, z, ...)$, the uncertainty on the quantity f is given by "propagating" the uncertainties of its arguments.

Propagation of errors.

$$\sigma_f = \sqrt{{\sigma_x}^2 \left(\frac{\partial f}{\partial x}\right)^2 + {\sigma_y}^2 \left(\frac{\partial f}{\partial y}\right)^2 + {\sigma_z}^2 \left(\frac{\partial f}{\partial z}\right)^2 + ...} \tag{2.7}$$

Here, ∂ is the partial derivative, sometimes called the *del* symbol. It means that you take the derivative of the function f with respect to only one of its variables.

One hidden assumption in Eq. (2.7) is that the uncertainties in the variables are completely uncorrelated. More generally, this is not the case, and Eq. (2.7) will become more complicated by including terms that reflect the correlation between the uncertainties. In this course we will only consider fully uncorrelated uncertainties.

The partial derivative

$$\frac{\partial f}{\partial x} \tag{2.8}$$

is the derivative of f with respect to x, while keeping all other variables of f (i.e. y, z, ...) constant. Let's look at a few examples.

Example 2.2: For a simple sum function like

$$f(x, y) = x + y, \tag{2.9}$$

we find

$$\frac{\partial f}{\partial x} = 1 = \frac{\partial f}{\partial y}, \tag{2.10}$$

Since the function depends linearly on both x and y, Eq.2.7 simplifies to

Uncertainty in a sum or difference.

$$\sigma_f = \sqrt{\sigma_x^2 + \sigma_y^2}. \tag{2.11}$$

This also tells you that if you have multiple (uncorrelated) uncertainties $\sigma_{x,i}$ in some variable x, you have to add them in "quadrature":

$$\sigma_{x,\text{tot}} = \sqrt{\sigma_{x,1}^2 + \sigma_{x,2}^2 + ...} \tag{2.12}$$

Notice that the sign of x or y do not matter because they are added in quadrature, so e.g., Eq. (2.11) holds also for a difference $f(x, y) = x - y$.

Example 2.3: For a simple quotient function like

$$f(x,y) = \frac{x}{y}, \tag{2.13}$$

the total uncertainty is given by

$$\sigma_f = \sqrt{\left(\frac{\sigma_x}{y}\right)^2 + \left(-\frac{\sigma_y x}{y^2}\right)^2}, \tag{2.14}$$

or,

Uncertainty in a product or quotient.

$$\sigma_f = |f|\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}, \tag{2.15}$$

This is a good formula to remember. Notice that x and y only contribute via their *relative uncertainties* σ_x/x and σ_y/y, added in quadrature. It is easy to show that you would obtain the same formula for the product $f(x,y) = xy$.

Example 2.4: For something more complicated like

$$f(x,y,a,b) = K\frac{xy^n}{ab^m}, \tag{2.16}$$

with a constant K, we find after some algebra

$$\sigma_f = |f|\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{n\sigma_y}{y}\right)^2 + \left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{m\sigma_b}{b}\right)^2}, \tag{2.17}$$

which is left as an excellent exercise for home. Notice that the relative uncertainty of y and b contribute more to the overall uncertainty, as they have an extra factor n or m.

Example 2.5: In our measurement of the speed of light, we estimate $\sigma_x = 0.01\,\mathrm{m}$ to be the uncertainty in the distance x, and $\sigma_t = 2\,\mathrm{ns}$ in the travel time of light according to our oscilloscope. The uncertainty in our measurement of c is

$$\sigma_c = c\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_t}{t}\right)^2} = 0.06\,\mathrm{m/s}. \tag{2.18}$$

Example 2.6: Suppose we measure gravitational acceleration g from the formula $x = \frac{1}{2}gt^2$, so that $g = \sqrt{2x/t^2}$. If we measure $t = 0.6395 \pm 0\,0001s$, and $x = 2.00 \pm 0\,002m$, then our calculation of g will have an uncertainty according to the formula

$$\sigma_g = g\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{2\sigma_t}{t}\right)^2} = 0.01\,\mathrm{m^2/s}. \tag{2.19}$$

Notice that time is a more important uncertainty to measure well.

2.4 Scientific notation

If you would measure the height of an A4 paper sheet with a simple high school rulers, you could write your result for example as

$$h = 29.8 \pm 0.1\,\text{cm}.$$

Here, the three digits (2, 9 and 8) are your *significant digits*. The uncertainty is always written with just one non-zero significant digit (the zeros in front do not count). Typically you write the value with as many significant digits as your precision. This includes any trailing zeros, e.g. for the A4 width:

$$w = 21.0 \pm 0.1\,\text{cm}.$$

In science, results are often written in terms of powers of ten. This allows you to keep easier track of the order of magnitude of your results. For example, for our measurement of c:

$$c = (2.98 \pm 0.06) \times 10^8\,\frac{\text{m}}{\text{s}}.$$

With calculators the power of ten can be written with the "e" or "exp" or "E" button, as in `2.98e8`. Notice that the power depends on your choice of units; for our A4 sheet:

$$h = (2.98 \pm 0.01) \times 10^1\,\text{cm} = (2.98 \pm 0.01) \times 10^{-1}\,\text{m},$$
$$w = (2.10 \pm 0.01) \times 10^1\,\text{cm} = (2.10 \pm 0.01) \times 10^{-1}\,\text{m},$$

because $1\,\text{cm} = 1 \times 10^{-2}\,\text{m}$.

In the exercises on the homeworks and exams, it is often sufficient to write down only three significant digits if you have no uncertainty given.

Chapter 3

Vectors & Reference Frames

A vector is a representation of a quantity that has a certain *direction* and *magnitude*. Vectors are very useful for a couple of things: to indicate, for example, the velocity of an object ($\mathbf{v}$), or a force acting upon it ($\mathbf{F}$), but also its position with respect to some point ($\mathbf{r}$), or even differences like displacement ($\mathbf{\Delta r}$) or change in velocity ($\mathbf{\Delta v}$). Vectors are typically indicated with an arrow or overline, but in these notes, we will use the bold notation.

3.1 Vectors in coordinate systems

The location of an object can be described by a point P in 3D space. In a *Cartesian coordinate system*, P has some coordinates (x, y, z), as in Fig. 3.1a. This can also be written as a vector $\mathbf{r}$. Using *unit vectors* $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ that point in each of the three directions:

$$\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}. \tag{3.1}$$

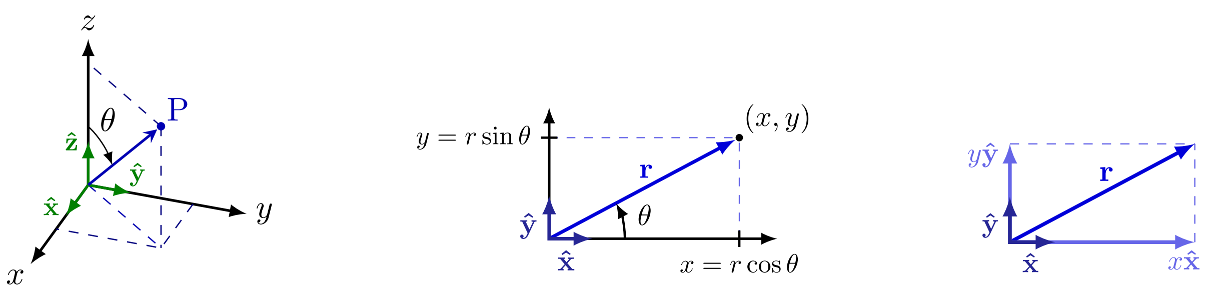
Some books also use the convention

$$\mathbf{r} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}},$$

or even

$$\mathbf{r} = x\hat{\mathbf{e}}_x + y\hat{\mathbf{e}}_y + z\hat{\mathbf{e}}_z,$$

but we will stick with Eq. (3.1). In two dimensions, $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}}$, and we simply ignore the z component.



(a) Position vector in a 3D Cartesian coordinate system.

(b) Position vector in a 2D Cartesian coordinate system.

(c) A vector can be broken down into its x and y vector components.

Figure 3.1: Postition vectors in two dimensions.

In terms of linear algebra, if $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$ is our (*linearly independent*) *basis*, then we can identify vectors by its components in a column vector:

$$\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{3.2}$$

3.2 Vector length

The *length*, also called the *magnitude* or *modulus*, of a vector is given by Pythagoras:

$$|\mathbf{a}| = a = \sqrt{a_x^2 + a_y^2 + a_z^2}. \tag{3.3}$$

Note some books use double bars notation. Unit vectors, denoted by point hats, represent directions, and have by definition lengths of 1, such as

$$|\hat{\mathbf{x}}| = 1. \tag{3.4}$$

If you have some vector $\mathbf{a}$, you can figure out its unit vector, by *normalizing* it, or dividing the vector by its magnitude.

$$\hat{\mathbf{a}} = \frac{\mathbf{a}}{|\mathbf{a}|}. \tag{3.5}$$

A hat (circumflex) always means the vector is normalized to 1.

3.3 Vector sum

Vector summing is very simple: you simply add them *component-wise*:

$$\begin{aligned} \mathbf{a} + \mathbf{b} &= (a_x\hat{\mathbf{x}} + a_y\hat{\mathbf{y}} + a_z\hat{\mathbf{z}}) + (b_x\hat{\mathbf{x}} + b_y\hat{\mathbf{y}} + b_z\hat{\mathbf{z}}) && (3.6) \\ &= (a_x + b_x)\hat{\mathbf{x}} + (a_y + b_y)\hat{\mathbf{y}} + (a_y + b_y)\hat{\mathbf{z}}. && (3.7) \end{aligned}$$

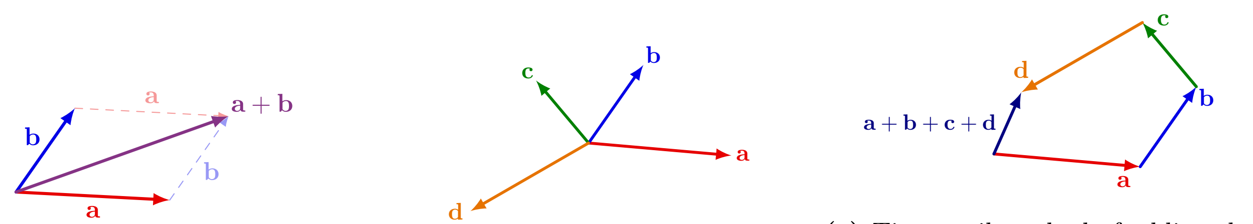
This is shown visually in Fig. 3.2a with the so-called tip-to-tail method. Figure 3.2a shows the same method for many vectors.

3.4 Scalar multiplication

Vectors can be multiplied by a *scalar*, or in other words, *scaled* by a real number $b \in \mathbb{R}$:

$$b\mathbf{a} = ba_x\hat{\mathbf{x}} + ba_y\hat{\mathbf{y}} + ba_z\hat{\mathbf{z}}. \tag{3.8}$$

So each component is scaled by the same number. It is easy to show that the length of the new vector is simply $|b\mathbf{a}| = |b|a$. This is the main effect: The length changes, but the new



(a) Summing two vectors. **(b)** A bunch of vectors. **(c)** Tip-to-tail method of adding the vectors of part (b).

Figure 3.2: Adding vectors.

(a) Scalar multiplication $b\mathbf{a}$ of vector $\mathbf{a}$ with scalar b, where $b = 2$ or -1.

(b) Scalar product $x = \mathbf{r} \cdot \hat{\mathbf{x}}$ is the projection of $\mathbf{r}$ onto the x axis.

Figure 3.3: Multiplication of vectors.

vector $b\mathbf{a}$ will be *parallel* to $\mathbf{a}$ (if $b \neq 0$), like in Fig. 3.3a. In case $b < 0$, it will "flip" the direction. A special case of this is where $b = -1$, such that

$$b\mathbf{a} = -a_x\hat{\mathbf{x}} - a_y\hat{\mathbf{y}} - a_z\hat{\mathbf{z}} = -\mathbf{a}, \tag{3.9}$$

which means that $b\mathbf{a}$ is parallel to $\mathbf{a}$, has the same length, but points in the opposite direction.

3.5 Scalar product

The *scalar product*, or *dot product*, is the componentwise product of two vectors:

$$\mathbf{a} \cdot \mathbf{b} = (a_x\hat{\mathbf{x}} + a_y\hat{\mathbf{y}} + a_z\hat{\mathbf{z}}) \cdot (b_x\hat{\mathbf{x}} + b_y\hat{\mathbf{y}} + b_z\hat{\mathbf{z}}) \tag{3.10}$$

$$= a_x b_x + a_y b_y + a_z b_z. \tag{3.11}$$

The result is a number with no direction, as opposed to a vector. Hence the name "scalar". A convenient relation is

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta = ab\cos\theta, \tag{3.12}$$

where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. From either of these formulas, it is trivial to show that the scalar product is *commutative*, i.e. swapping two vectors gives the same result:

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}. \tag{3.13}$$

The scalar product is also *distributive* with vector sum, which means that for any three vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$,

$$\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}. \tag{3.14}$$

Notice that a scalar product of a vector with itself gives you its magnitude squared,

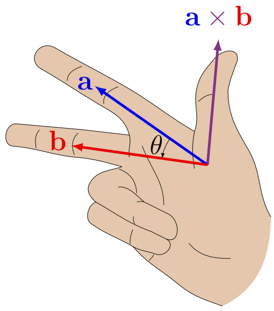
$$\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2. \tag{3.15}$$

The scalar product is also useful to calculate the projection of one vector onto another. This is a measure of a magnitude of their overlap. For instance, one can project a vector onto the unit vector for the x direction, e.g.

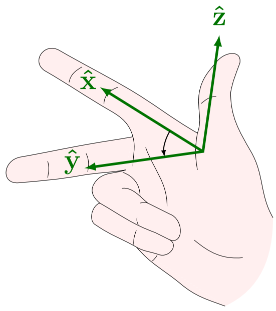
$$\mathbf{a} \cdot \hat{\mathbf{x}} = a\cos\theta, \tag{3.16}$$

where θ is the angle between the x axis and $\mathbf{a}$, as shown in Fig. 3.3b. This projection gives the x component of the vector $\mathbf{a}$.

Since the y and x axes are *orthogonal*, or *normal*, to each other, the $\hat{\mathbf{y}}$ unit vector has no component along $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = 0$. This holds for any two vectors that are orthogonal to



(a) The right-hand rule gives the direction of the vector product.



(b) Conventionally, we use a right-handed coordinate system.

Figure 3.4: Right-hand rule like on the CHF 200 bill: Point your flat hand in the direction of the first vector **a**. Keeping your index finger along **a**, point your middle finger along **b**, sweeping the smallest angle θ between **a** and **b**. Your thumb will point along $\mathbf{a} \times \mathbf{b}$.

each other. A different way to see this is by realizing the angle between them is $\theta = 90°$, and so $\cos\theta = 0$.

Using the rules of matrix multiplication, one can write the scalar product of two vectors using the *transpose*:

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^\mathrm{T}\mathbf{b} = \begin{pmatrix} a_x & a_y & a_z \end{pmatrix} \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} = a_x b_x + a_y b_y + a_z b_z. \tag{3.17}$$

3.6 Vector product

Lastly, there is the *vector product*, or *cross product*, defined as

$$\mathbf{a} \times \mathbf{b} = (a_y b_z - a_z b_y)\hat{\mathbf{x}} + (a_z b_x - a_x b_z)\hat{\mathbf{y}} + (a_x b_y - a_y b_x)\hat{\mathbf{z}}. \tag{3.18}$$

One way of remembering this formula is by creating the following matrix, of which you compute the *determinant*:

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix}. \tag{3.19}$$

This result of the vector product is again a vector, hence the name. The length of the vector products is given by

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin\theta. \tag{3.20}$$

This is the vector product's counterpart of Eq. (3.12).

Important to know is that this new vector is orthogonal to the original vectors. It is easy to show that

$$\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = 0 \tag{3.21}$$
$$\mathbf{b} \cdot (\mathbf{a} \times \mathbf{b}) = 0. \tag{3.22}$$

So $\mathbf{a} \times \mathbf{b}$ is orthogonal to **a** and **b**, but in which direction? There are two possibilities, but the convention is given by the right-hand rule, illustrated in Fig. 3.4. In fact, the Cartesian coordinate system we conventionally use in 3D is *right handed*:

$$\hat{\mathbf{x}} \times \hat{\mathbf{y}} = \hat{\mathbf{z}}. \tag{3.23}$$

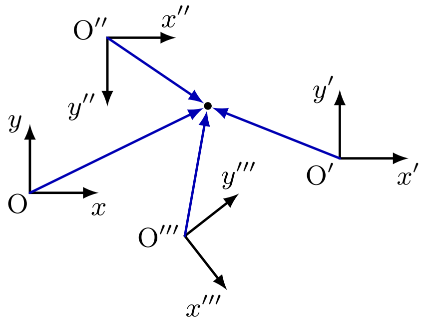


Figure 3.5: Here, the same point is shown in different reference frames, labeled as O, O', O", and O"'. Since the reference frames are shifted and rotated with respect to O, the position vector (in blue) that specifies the point is also different. Transformations between these reference frames are possible without changing the physics, like translation, reflection and rotation.

From Eq. (3.20) it is easy to see that the vector product of two non-zero vectors is zero if and only if $\sin\theta = 0$. This happens when they are aligned ($\theta = 0$) or back-to-back ($\theta = 180°$). This also makes sense from a geometrical standpoint, as two aligned vectors do not span a plane anymore, and the vector product has no unique direction to be orthogonal to them. In terms of linear algebra: The two vectors are not *linearly independent*.

Finally, unlike the scalar product, the vector product is not commutative:

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}, \tag{3.24}$$

but since swapping the vectors only introduces a minus sign, it is called *anticommutative*.

3.7 Reference frames

We will see many physics problems in the coming chapters and exercise classes. One important thing is that the choice of the coordinate system is arbitrary. Where you put the origin, and in what direction you point the axes is your choice, and the physics (i.e. the prediction) will be the same. That is not to say, there are no bad choices. Often there is a natural choice, such as x for the horizontal direction, along the ground, and y along the vertical. The right choice can spare you a lot of extra algebra.

You can do several types of *coordinate transformations* without changing the physics:

- Translation: $(x, y) \mapsto (x', y') = (x + a, y + b)$ for constants $a, b \in \mathbb{R}$.
- Rotation: $(x, y) \mapsto (x', y') = (x\cos\theta + y\sin\theta, x\sin\theta - y\cos\theta)$ for angle θ.
- Reflection; e.g. $(x, y) \mapsto (x', y') = (x, -y)$.

Some of these are illustrated in Fig. 3.5. In theoretical physics this fact is related to the deeper concept of *symmetry*: The physics does not change under these transformations, just like rotating a circle by any angle gives you the same shape, and therefore the same description. Namely, it is easy to show that for the equation of a circle after rotating the coordinates by some angle θ as shown above,

$$x^2 + y^2 = r^2 \xrightarrow{\text{rotation}} x'^2 + y'^2 = x^2 + y^2 = r^2. \tag{3.25}$$

In the same way the equations of motions will not change under coordinate transformation between reference frames, as long as they are inertial, which means they are not accelerating.

In Chapter 10 we will learn more about inertial and non-inertial reference frames.

3.8 Extra: Vector transformations

This is some extra information to make a connection to linear algebra for the interested reader.

Multiplying a vector by some matrix A,

$$f(\mathbf{a}) = A\mathbf{a}, \tag{3.26}$$

is called a *linear transformation* or *linear map.* It is easy to show it preserves vector addition,

$$f(\mathbf{a}+\mathbf{b}) = f(\mathbf{a}) + f(\mathbf{b}), \tag{3.27}$$

and scalar multiplication

$$f(b\mathbf{a}) = bf(\mathbf{a}), \tag{3.28}$$

hence the name *linear*. The reflection of the y coordinate we saw above has the transformation matrix

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.29}$$

such that

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}. \tag{3.30}$$

We will see another example of a linear transformation in the extra Section 10.3.2, namely rotation which has a transformation matrix

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{3.31}$$

A more general transformation includes a *translation*, which can be written as

$$f(\mathbf{a}) = A\mathbf{a} + \mathbf{b}, \tag{3.32}$$

with matrix A and some constant *translation vector* $\mathbf{b}$.

Chapter 4

Motion in One Dimension

Movement can be described by the change in coordinates as a function of time. In three dimensions, an object can move from a point $P_1 = (x_1, y_1, z_1)$ to point $P_2 = (x_2, y_2, z_2)$ like in Fig. 4.1. At some time t_1 it was at P_1, until it arrived at P_2 at some time t_2. Its path will be described by some continuous vector functions

$$\mathbf{r}(t) = x(t)\hat{\mathbf{x}} + y(t)\hat{\mathbf{x}} + z(t)\hat{\mathbf{x}}, \tag{4.1}$$

where its coordinates $x = x(t)$, $y = y(t)$ and $z = z(t)$ are again functions of time.

In this course we will study the behavior of motion with some relatively simple functions such as linear, quadratic, circular, and sinusoidal functions.

4.1 Uniform motion: constant velocity

Let's start with some basic examples in one dimension (1D). If a point is moving in some dimension, it means its position x depends on time t,

$$x = x(t). \tag{4.2}$$

Consider a car driving with a constant velocity v. Its position at any time is then given by

$$x(t) = vt. \tag{4.3}$$

This is called *uniform linear motion*, or *uniform motion* in one dimension. Fig. 4.2a shows that this looksizaa like a straight line through the origin of a position versus time graph.

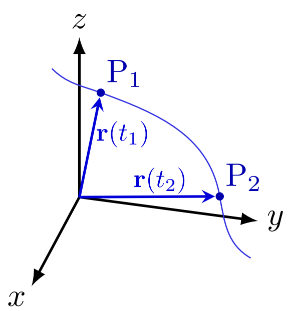
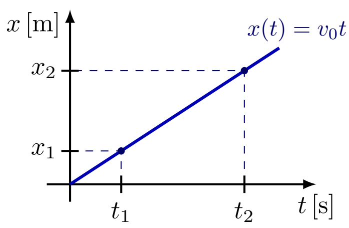
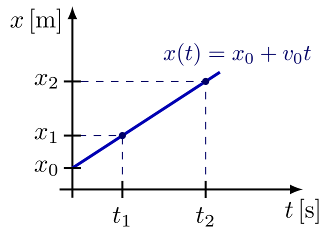


Figure 4.1: Motion of a point in a 3D. The position vector points to different points along the point's path at different times.



(a) Starting at $x(0) = 0$ at $t = 0$.



(b) Starting at an offset $x(0) = x_0$ at $t = 0$.

Figure 4.2: Uniform motion in one dimension with constant velocity.

The slope of the line is the velocity v. For example, between two points (t_1, x_1) and (t_2, x_2), the slope of the line is given by

$$v = \frac{x_2 - x_1}{t_2 - t_1} = \frac{\Delta x}{\Delta t}. \tag{4.4}$$

A trivial case is when $v = 0$: The point does not move at all in time and it stands still in the same position. To get a feeling for this: If the car started at $x(0) = 0$ at t and moved to $x(1\,\mathrm{s}) = 5\,\mathrm{m}$ after $t = 1\,\mathrm{s}$, the car has an average velocity of $5\,\mathrm{m/s}$.

Note that in Eq. (4.3) we assumed that the car started at $x(0) = 0$ at $t = 0$. In general, the car could have been at some position $x(0) = x_0$ at $t = 0$, as in Fig. 4.2b:

Uniform motion in one dimension.

$$x(t) = x_0 + vt. \tag{4.5}$$

4.2 Uniform acceleration

Velocity of course does not have to be constant. It can depend on time as well:

$$v = v(t). \tag{4.6}$$

For example, if velocity increases with time linearly, starting from some initial velocity v_0 then

$$v(t) = v_0 + at \tag{4.7}$$

with the *constant acceleration* a, the position $x(t)$ will now become a parabola in x-t space. As velocity has dimensions time over length, acceleration must clearly have dimensions length over time squared, so units $\mathrm{m/s^2}$.

4.2.1 Positive acceleration

Consider again a car which *starts from rest* (i.e. $v(0) = 0$) and speeds up with a constant acceleration, then its position is given by

$$x(t) = x_0 + \frac{at^2}{2}, \tag{4.8}$$

as illustrated in Fig. 4.3a. For simplicity, you could let the car start at $x(0) = x_0 = 0$ at time $t = 0$. We will understand in a few sections where this equation comes from better.

Now consider the car was moving with some initial velocity $v(0) = v_0$. Its position will be described by

$$x(t) = x_0 + v_0 t + \frac{at^2}{2}, \tag{4.9}$$

instead.

4.2.2 Negative acceleration

Another example is when you throw up a ball. Say we start the clock at $t = 0$ when you release it from your hand, and count this height as $y(0) = y_0 = 0$. The ball will have some initial velocity $v(0) = v_0$ that you gave it. At the same time, gravity is pulling it back down to Earth, and slows down the ball until it reaches its highest point, the *apex*, and comes, or "falls" back down for you to catch it. On the way back, it gained some velocity, but in the other direction. This is a school example of negative acceleration $-a < 0$:

$$y(t) = v_0 t - \frac{at^2}{2}. \tag{4.10}$$

This is graphed in Fig. 4.3b, where the ball reaches its apex at t_2. A typical value for the gravitational acceleration is $a = g \approx 9.8\,\mathrm{m/s^2}$, although for quick calculations we can round to $g = 10\,\mathrm{m/s^2}$. The velocity is given by the linear function

$$v(t) = v_0 - at, \tag{4.11}$$

graphed in Fig. 4.3c. You can see that it is positive at first, but becomes smaller until it reaches t_2. At t_2 the ball has reached its apex and comes to a standstill. After that the velocity becomes more and more negative, meaning it speeds up in the negative x direction, or, the ball falls as normal people would say.

Furthermore, notice that in Eq. (4.10) we had chosen the origin at the height where you release it: $y(0) = 0$, but this is somewhat arbitrary. We could have chosen an offset $y(0) = y_0$ at $t = 0$.

To summarize, constant, or *uniform*, acceleration in one dimension is given by a quadratic equation:

Uniform acceleration in one dimension.

$$\begin{cases} x(t) = x_0 + v_0 t + \dfrac{at^2}{2} \\ v(t) = v_0 + at \end{cases} \tag{4.12}$$

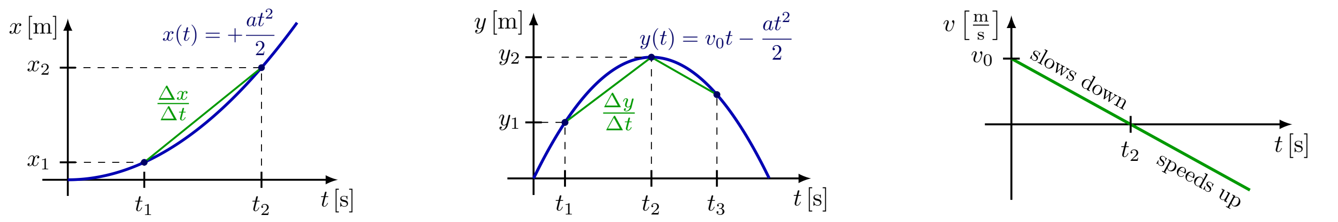
where x_0, v_0 and a can be some real number, that can be both positive or negative. If $a = 0$, this formula reduces to Eq. (4.5) again.

As another example, suppose you release a ball from rest from a height h, how long will it take to reach the ground? At time $t = 0$, the height is $y(0) = h$ and the velocity is $v(0) = v_0 = 0$, while for some time t, $y(t) = 0$. So we need to solve

$$y(t) = h - \frac{gt^2}{2} = 0 \tag{4.13}$$

for t:

$$t = \sqrt{\frac{2h}{g}}. \tag{4.14}$$



(a) Positive acceleration, starting with $v(0) = 0$.

(b) Negative acceleration, starting with $v(0) > 0$, and reaching its apex at t_2.

(c) Velocity as a function of time for a constant, negative acceleration.

Figure 4.3: Motion in one dimension with changing velocity due to constant acceleration is described by a parabola. The average velocity between two points is given by the slope of the line connecting them (green).

Notice that mathematically there is a negative and a positive solution, but we only choose the positive one. What will be the ball's velocity when it hits the ground? By substituting:

$$v(t) = \sqrt{2gh}. \tag{4.15}$$

4.3 Average velocity

Imagine you take the bike from home to Irchel. On your way you speed up several times, slow down when going uphill, and stop for a red light or pedestrian crossing the road. So your velocity changed quite a lot. The graph will look much more complicated than Fig. 4.3c.

In general, we can still define the *average velocity*, denoted as $\overline{v}$, or sometimes v_{ave}, between any two points can be written by:

Average velocity.

$$\overline{v} = \frac{\Delta x}{\Delta t} = \frac{x_2 - x_1}{t_2 - t_1}, \tag{4.16}$$

where Δx is the total distance traveled and Δt is the travel time.

The geometric interpretation of this formula is that the average velocity between two points (t_1, x_1) and (t_2, x_2) on the x-t curve is the slope of the straight line connecting them, as is shown in Figs. 4.3a and 4.3b.

4.3.1 Torricelli's equation

Here we derive what is known as Toricelli's equation. Consider the velocity under uniform acceleration:

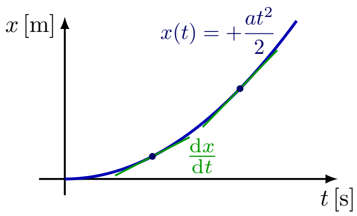
$$v(t) = v_0 + at \tag{4.17}$$

At some time $t = \Delta t$, the average velocity is given halfway between v_0 and $v = v(\Delta t)$,

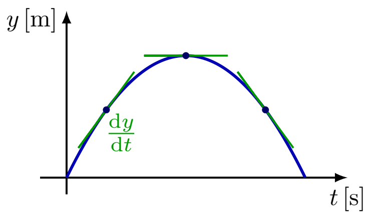
$$\overline{v} = \frac{v_0 + v}{2} = v_0 + \frac{a\Delta t}{2}. \tag{4.18}$$

Now, say, the object moved a distance Δx during this time, so from our definition of average velocity:

$$x = x_0 + \Delta x = x_0 + \overline{v}\Delta t. \tag{4.19}$$



(a) Positive acceleration.



(b) Negative acceleration.

Figure 4.4: Instantaneous velocity in some point is the slope of the tangent line to the x-t curve in that point. Compare to Figs. 4.3a and 4.3b and take $\Delta t = t_2 - t_1 \to 0$.

So we again arrive at

$$x = x_0 + v_0 \Delta t + \frac{a\Delta t^2}{2}. \tag{4.20}$$

This is one way of proving Eq. (4.12), but we will see a more general approach with derivatives and integrals in Section 4.5

We can also find the velocity as a function of displacement alone without explicit time-dependence. Rewriting Eq. (4.17) to

$$\Delta t = \frac{v(t) - v_0}{a}, \tag{4.21}$$

and plug it back into Eq. (4.20):

Torricelli's equation.

$$v^2 = v_0^2 + 2a\Delta x. \tag{4.22}$$

So, if you know the initial velocity v_0, and the constant acceleration a, then you can find the final velocity v for any displacement Δx using Eq. (4.22).

For example, if you let something fall from rest $v_0 = 0$, what will the object's velocity be after falling a height $\Delta x = h$? Using Torricelli's equation with $a = g$, we immediately find

$$v = \sqrt{2gh}. \tag{4.23}$$

This is consistent with our previous result in Eq. (4.15). There we had to solve two independent equations in Eq. (4.12) for t and v, but starting from Torricelli's equation (4.22) was quicker.

4.4 Instantaneous velocity

As time Δt gets shorter, we approach the *instantaneous velocity* v. As $\Delta t \to 0$ in Eq. (4.16), we find

Instantaneous velocity.

$$v(t) = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}. \tag{4.24}$$

This should remind you of the definition of the derivative. Instead of slope of the line connecting two points, this will be the slope of the tangent to the x-t curve in one single point, as is shown in Fig. 4.4. The instantaneous velocity is again a function of time.

The magnitude of the velocity, is called the *speed*,

Speed.

$$\textit{speed} = |v| = \left|\frac{\mathrm{d}x}{\mathrm{d}t}\right|. \tag{4.25}$$

Speed is a positive number, while velocity is more like a vector, as it indicates direction, which can also be negative. A negative velocity means that it points in the negative x direction.

Example 4.1: Let's calculate this the "old-fashioned" way with a numerical example. What is the velocity $v(t)$ at some time t for the following position function?

$$x(t) = 5t^2 \tag{4.26}$$

At some later time, $t + \Delta t$, the position is

$$x(t + \Delta t) = 5(t + \Delta t)^2 = 5t^2 + 10t\Delta t + 5\Delta t^2.$$

The change in position is

$$\Delta x = x(t + \Delta t) - x(t) = 10t\Delta t + 5\Delta t^2.$$

The average velocity then is

$$\overline{v} = \frac{\Delta x}{\Delta t} = 10t + 5\Delta t.$$

And we find the instantaneous velocity as

$$v = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = 10t.$$

Now, if you just take the derivative of Eq. (4.26), using rules for taking derivatives of polynomials, you get the exact same answer:

$$v = \frac{\mathrm{d}x}{\mathrm{d}t} = 10t.$$

Before we move on, yet another aside on notation: Here we used the Leibniz notation for differentiation $\mathrm{d}x/\mathrm{d}t$, but others used are Lagrange's notation $x'(t)$, or Newton's dot notation $\dot{x}(t)$. This last notation is almost exclusively used for time-derivatives in physics.

4.5 Instantaneous acceleration

The acceleration is the change in velocity with respect to time. This is just like velocity is the change in position with respect to time. And analogous to the instantaneous velocity Eq. (4.24), the instantaneous acceleration is

Instantaneous acceleration.

$$a(t) = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{\mathrm{d}v}{\mathrm{d}t}. \tag{4.27}$$

This is now the slope of tangent line to v-t graph.

Again, a negative acceleration means that it points in the negative x direction. Importantly, if the velocity and acceleration have opposite signs, there will be a *deceleration*. For instance, if the velocity is positive, $v > 0$, but the acceleration is negative, $a < 0$, then v becomes smaller. As illustrated in Figs. 4.3c and 4.5e, a negative acceleration means that the velocity becomes smaller if $v > 0$, or "more negative" if $v < 0$.

Example 4.2: To see an example of calculating the acceleration from a velocity function, we continue the previous Example 4.1,

$$\Delta v = v(t + \Delta t) - v(t) = 10(t + \Delta t) - 10t. \tag{4.28}$$

The average acceleration is $\overline{a} = 10\,\mathrm{m/s^2}$. This is the same as the instantaneous one:

$$\frac{\mathrm{d}v}{\mathrm{d}t} = 10\,\mathrm{m/s^2}, \tag{4.29}$$

which is to say, the acceleration is constant with time in this particular case.

Substituting Eq. (4.24), we see immediately that acceleration is the second time-derivative of position x with respect to time :

$$a(t) = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}^2x}{\mathrm{d}t^2}, \tag{4.30}$$

Or in other notations: $a(t) = x''(t) = \ddot{x}(t)$.

Remember that the general rule for finding derivatives with functions of the form $x(t) = Ct^n$ is

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Cnt^{n-1}. \tag{4.31}$$

We also know that integral is the opposite of derivation, and that the integral for a function of the form $a(t) = Ct^n$ is

$$\int a(t)\mathrm{d}t = C\frac{t^{n+1}}{n+1} + C', \tag{4.32}$$

with some integration constant C'. For a constant acceleration ($n = 0$), we find as a general solution

$$\begin{cases} x(t) = \displaystyle\int_{t_0}^{t} v(t')\mathrm{d}t' = x_0 + v_0(t - t_0) + \frac{1}{2}a(t - t_0)^2 \\ v(t) = \displaystyle\int_{t_0}^{t} a(t')\mathrm{d}t' = v_0 + a(t - t_0) \\ a(t) = a \end{cases} \tag{4.33}$$

Here, $x_0 = x(t_0)$ and $v_0 = v(t_0)$ are integration constants. We typically choose $t_0 = 0$ to keep these formula neater, as we did previously in Eqs. 4.8 and 4.12. Conversely, if we start from $x(t)$ and take derivatives with respect to time, we get $v(t)$ and $a(t)$:

$$\begin{cases} x(t) = x_0 + v_0 t + \dfrac{at^2}{2} \\ v(t) = \dfrac{\mathrm{d}x}{\mathrm{d}t} = v_0 + at \\ a(t) = \dfrac{\mathrm{d}v}{\mathrm{d}t} = \dfrac{\mathrm{d}^2 x}{\mathrm{d}t^2} = a \end{cases} \tag{4.34}$$

Notice that from dimensional analysis, these functions also make sense:

$$[v_0 t] = \frac{\mathrm{m}}{\mathrm{s}}\,\mathrm{s} = \mathrm{m} \tag{4.35}$$

$$\left[\frac{at^2}{2}\right] = \frac{\mathrm{m}}{\mathrm{s}^2}\,\mathrm{s}^2 = \mathrm{m} \tag{4.36}$$

$$[at] = \frac{\mathrm{m}}{\mathrm{s}^2}\,\mathrm{s} = \frac{m}{s}. \tag{4.37}$$

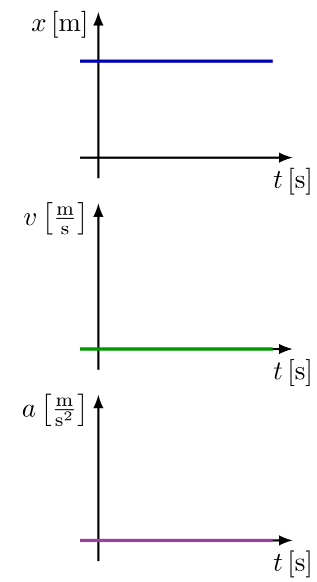
So everything is consistent.

The examples we have seen so far are polynomials. But position can be any other continuous function. It can also be a piece-wise function, like a train bouncing back and forth, which we saw in class. The train's position was given by Fig. 4.5f, where it had constant velocity, until it hit the end and quickly changed direction. Later, when we study springs and circular motion, we will also see sinusoidal movement as in Fig. 4.5g. The velocity and acceleration will also look like sinusoidal curves, but shifted by some phase:

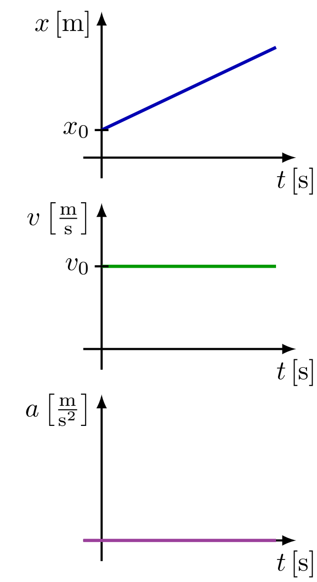
$$\begin{cases} x(t) = R\sin(\omega t) \\ v(t) = \dfrac{\mathrm{d}x}{\mathrm{d}t} = R\omega\cos(\omega t) \\ a(t) = \dfrac{\mathrm{d}v}{\mathrm{d}t} = -R\omega^2\sin(\omega t) \end{cases} \tag{4.38}$$

Here, ω is the *angular frequency*, which we will learn more about later in Section 5.3.

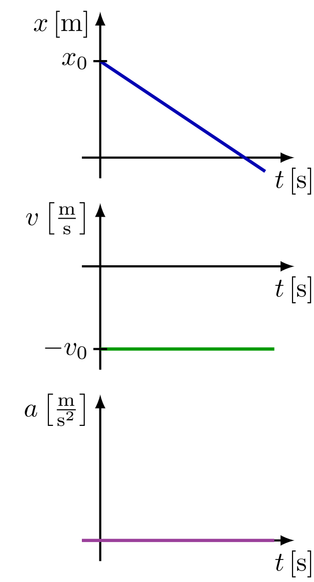
Figure 4.5 summarizes the different x-t curves we have seen, and its first and second derivatives, velocity $v(t)$ and acceleration $a(t)$.



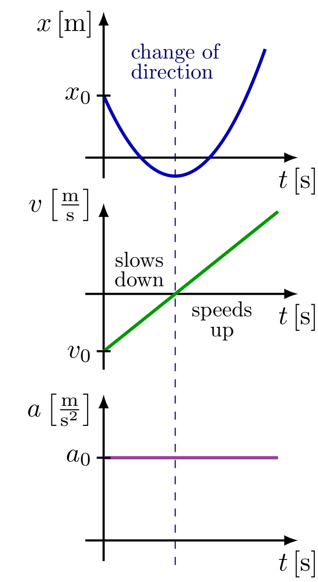
(a) Zero velocity and acceleration, i.e. standing still.



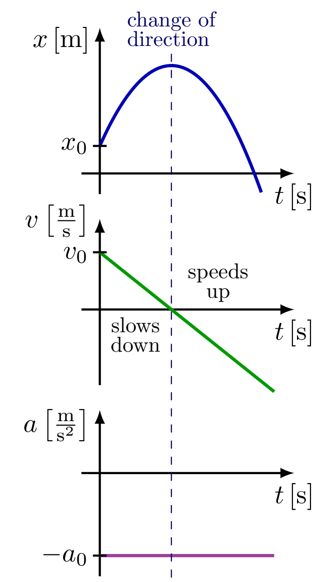
(b) Constant, positive velocity; zero acceleration.



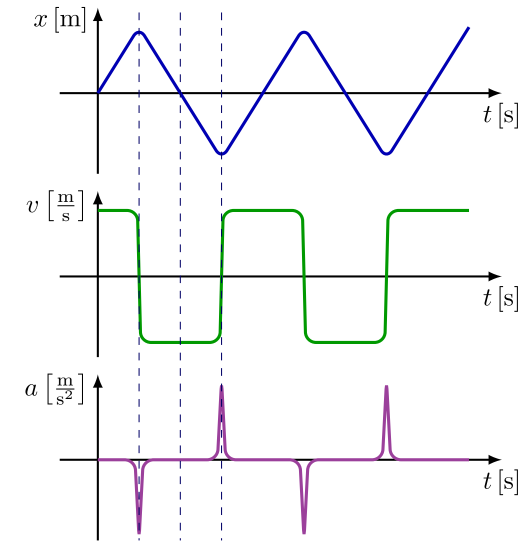
(c) Constant, negative velocity; zero acceleration.



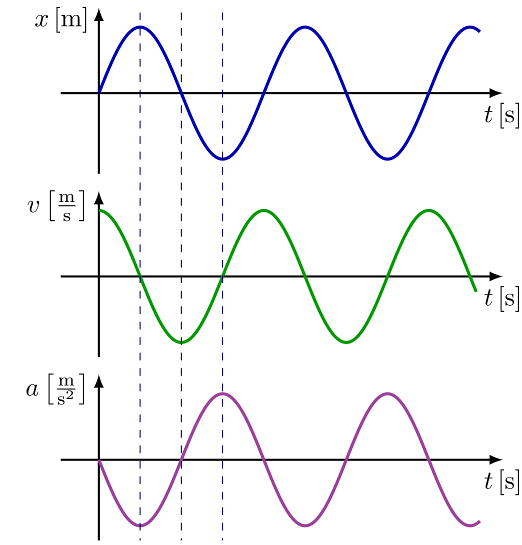
(d) Linear velocity; constant, positive acceleration.



(e) Linear velocity; constant, negative acceleration, like for a ball thrown upwards.



(f) Train bouncing back and forth with constant speed. The acceleration peaks when it quickly changed directions.



(g) Periodic one dimensional motion of a mass on a spring moving back and forth. Velocity is largest when $x = 0 = a$.

Figure 4.5: Several x-t curves of motion in one dimension, and its derivatives.

Chapter 5

Motion in Two Dimensions

In this chapter we will explore some of the basics of two dimensional motion. This includes a lot of nice real-life examples, like the trajectory of a ball you throw, or the orbit a satellite follows around the Earth.

In one dimension it makes less sense to use vectors, but in this chapter they will come to good use. We will use it for the position $\mathbf{r}$, but also the velocity $\mathbf{v}$ is a vector. In three dimensions:

$$\mathbf{v} = v_x\hat{\mathbf{x}} + v_y\hat{\mathbf{y}} + v_z\hat{\mathbf{z}}, \tag{5.1}$$

which has some direction and a length $v = |\mathbf{v}|$ (the speed). Each component has the dimensions of velocity and is a velocity in that respective direction. The beautifully useful thing about vectors is that you can treat their components independently if you choose the x and y axis wisely, and we will see some examples of this in the next sections.

The logic of deriving position to get the velocity also extends to vectors:

$$\mathbf{v}(t) = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}t}\hat{\mathbf{x}} + \frac{\mathrm{d}y}{\mathrm{d}t}\hat{\mathbf{y}} + \frac{\mathrm{d}z}{\mathrm{d}t}\hat{\mathbf{z}}, \tag{5.2}$$

as well as for deriving the velocity to get the acceleration:

$$\begin{aligned}\mathbf{a}(t) &= \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{\mathrm{d}v_x}{\mathrm{d}t}\hat{\mathbf{x}} + \frac{\mathrm{d}v_y}{\mathrm{d}t}\hat{\mathbf{y}} + \frac{\mathrm{d}v_z}{\mathrm{d}t}\hat{\mathbf{z}} \\ &= \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\end{aligned}$$

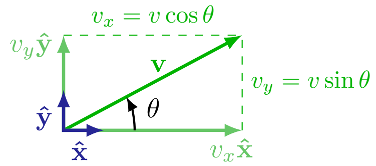
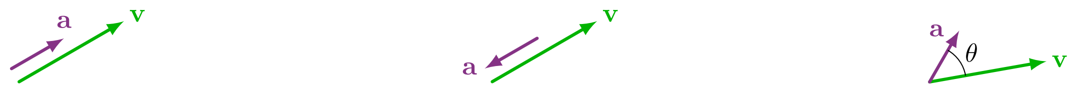


Figure 5.1: Breaking down a two-dimensional velocity vector into its two x and y components.



(a) Acceleration aligns: Speed increases, but velocity direction stays constant.

(b) Acceleration anti-aligns: Speed decreases, but velocity direction stays constant.

(c) Acceleration does not align: Direction changes. If $\mathbf{a}$ is not perpendicular the magnitude

Figure 5.2: Velocity $\mathbf{v}$ and acceleration vector $\mathbf{a}$.

Acceleration $\mathbf{a}$ codifies the change in velocity $\mathbf{v}$. It can change both the magnitude and direction, as shown in Fig. 5.2.

5.1 Parabolic motion

If you throw a rock through the air, its trajectory will describe a nice parabola. Why a parabola?

Say you throw the rock with some initial velocity v_0. If you throw at an angle θ between 0 and 90°, its velocity will have some x and y component, given by the vector

$$\mathbf{v}_0 = v_{0x}\hat{\mathbf{x}} + v_{0y}\hat{\mathbf{y}}, \tag{5.3}$$

or in terms of the total initial speed

$$|\mathbf{v}_0| = v_0 = \sqrt{v_x^2 + v_y^2}, \tag{5.4}$$

and angle θ:

$$\mathbf{v}_0 = v_0 \cos\theta\hat{\mathbf{x}} + v_0 \sin\theta\hat{\mathbf{y}}. \tag{5.5}$$

Now we will break down this vector into independent components to study the rock's trajectory. We see that it makes sense to choose the coordinates we have since gravity pulls in the vertical (y) direction, whereas it has no component in the x direction. For now we neglect air resistance, such that there is no acceleration (or deceleration) in the x direction, $a_x = 0$. In the y direction then, there is only the negative acceleration $a_y = -g$ due to gravity:

Projectile motion.

$$\begin{cases} x_\mathrm{d}(t) = x_0 + v_{0x}t \\ y_\mathrm{d}(t) = y_0 + v_{0y}t - \dfrac{gt^2}{2} \end{cases} \tag{5.6}$$

This set of equations will look like the parabola in Fig. 5.3. The velocity is given by

$$\begin{cases} v_x(t) = v_{0x} \\ v_y(t) = v_{0y} - gt \end{cases} \tag{5.7}$$

These are the basics of *ballistics*: the study of the *trajectory of projectiles*. They are important to understand if you want to hit your target when shooting bullets, missiles or rockets from a large distance, or less violently, throw a ball at your friend or in a hoop.

Notice that the y-x graph looks very similar to the y-t graph: Both are parabolic. Why is this? A simple way to see this is to realise that the y-t parabola is *parametrized* by time t. Time depends linearly on the position x, so we can substitute in an expression for t from $y_d(x)$ in Eq.5.6, and determine the formula for y vs. x, in which $y_d(x)$ will depend quadratically on x, taking the same form as $y_\mathrm{d}(t)$.

5.1.1 Example: Shooting a falling monkey

A monkey has been eating all the apples in our apple orchard! We want to shoot the monkey with a tranquilizer dart to capture it. We aim the dart gun straight at the monkey. We are ready to fire. But the monkey is smart: he is on to us. He knows that he can let gravity pull him out of the line of sight of the gun. He lets himself drop right when we pull the trigger. However, he forgot gravity is universal and will pull down the dart as well! Will the dart hit the monkey after all?

Let's fix the xy axis as in Fig. 5.4a; the origin is at the foot of the apple tree, with the y axis pointing vertically along the tree, and the x axis along the horizontal ground. (As mentioned in Section 3.7, this is an arbitrary choice. We could also have put the origin at the gun's nuzzle, or anywhere else.) Say the monkey hangs from a branch at a height h, and the dart gun's nozzle is at a horizontal distance d from the tree, so their positions are given by $(0, h)$ and $(d, 0)$, respectively.

Neglecting air resistance, the dart moves with constant velocity $v_x = v_{0x}$ in the x direction. Meanwhile gravity acts on the dart vertically, and accelerates it down in the y direction with acceleration $a_y = -g$. We can write the velocity components in Eq. (5.7) in terms of the total magnitude and smallest angle θ with the x axis

$$\begin{cases} v_x(t) = -v_0 \cos\theta \\ v_y(t) = v_0 \sin\theta - gt \end{cases} \tag{5.8}$$

The minus sign in the first line appears because the dart travels in the negative x direction. So starting from Eq. (5.6) and $(x_0, y_0) = (d, 0)$, the position is given by

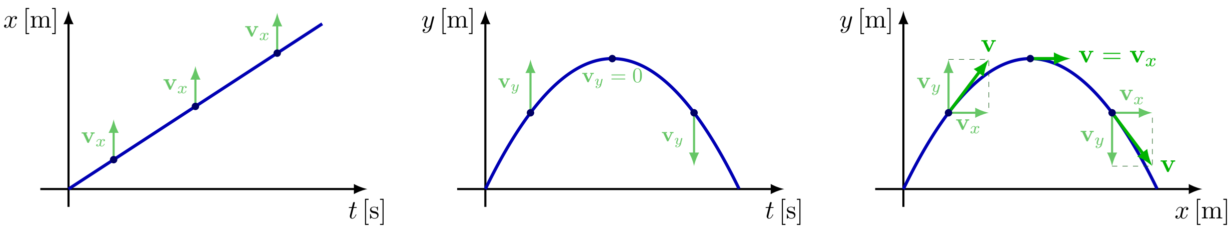
$$\begin{cases} x_\mathrm{d}(t) = d - v_0 \cos(\theta) t \\ y_\mathrm{d}(t) = v_0 \sin(\theta) t - \dfrac{gt^2}{2} \end{cases} \tag{5.9}$$

We assume the monkey lets himself drop from rest, i.e. the monkey has no initial velocity when it falls from height $y_\mathrm{m}(0) = h$:

$$\begin{cases} x_\mathrm{m}(t) = 0 \\ y_\mathrm{m}(t) = h - \dfrac{gt^2}{2} \end{cases} \tag{5.10}$$

Now we have all the equations set up. Will the dart hit the monkey? To answer this question, we need to know what the height of the monkey and dart are when the dart reaches the tree, $x = 0$, at time $t = t_1$. We can find time t_1 by setting the following condition:

$$x_\mathrm{d}(t_1) = 0 = d - v_0 \cos(\theta) t_1. \tag{5.11}$$



(a) x-t diagram. **(b)** y-t diagram. **(c)** y-x diagram.

Figure 5.3: Trajectory of a projectile. In the x direction (left), the projectile moves at constant velocity with respect to time. In the y direction (center), the projectile slows down due to gravity, stops, and then its velocity becomes negative. What we actually see, if we trace the movement of the projectile, is the y vs. x figure (right), where the blue curve is the trajectory, and the green arrows show the velocity components at different points along the trajectory.

This is graphically shown in Fig. 5.4b, where the two curves intersect. Then t_1 is given by

$$t_1 = \frac{d}{v_0 \cos\theta}. \tag{5.12}$$

At that same time, the height of the dart will be

$$y_\mathrm{d} = d\tan\theta - \frac{g}{2}\left(\frac{d}{v_0\cos\theta}\right)^2, \tag{5.13}$$

and that of the monkey:

$$y_\mathrm{m}(t_1) = h - \frac{g}{2}\left(\frac{d}{v_0\cos\theta}\right)^2. \tag{5.14}$$

You may recognize $d\tan\theta$ from trigonometry. Looking at the triangle in Fig. 5.4a,

$$h = d\tan\theta. \tag{5.15}$$

So $y_\mathrm{d} = y_\mathrm{m}$ at time t_1 as in Fig. 5.4c. The dart hits the monkey! The poor monkey is not as smart as it thinks.

What this experiment nicely shows is the you can "decouple" the motion into two perpendicular directions. You break down the velocity into two components, which you treat independently from each other. This is a very import idea that we will use in most problems we will see.

Notice that whether or not you hit the monkey is independent of the blow dart's initial velocity and angle, as long as it is large enough to reach the tree without hitting the ground first. The is the case as long as the dart is positive when reaching the tree, so if $y_\mathrm{m}(t_1) \geqslant 0$, or,

$$h \geqslant \frac{g}{2}\left(\frac{d}{v_0\cos\theta}\right)^2. \tag{5.16}$$

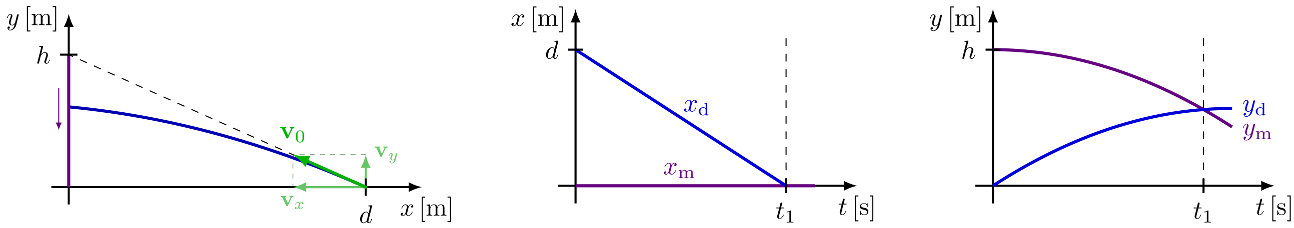
This is the condition on $v_{0x} = v_0\cos\theta$ for the dart to reach the tree before hitting the ground.

5.2 Interlude: Radians & polar coordinates

Before looking at circular motion, remember some of the basics of circles.

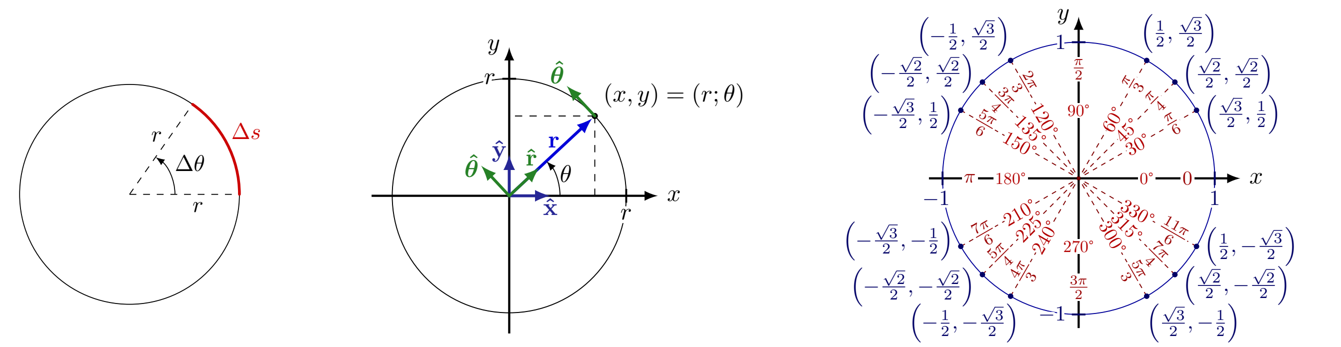
A *circular arc* is a line segment along a circle. Say it has some length Δs, then the angle $\Delta\theta$ subtended by the arc in units of *radians* is defined by

$$\Delta s = r\Delta\theta, \tag{5.17}$$



(a) x-y diagram. **(b)** x-t diagram. **(c)** y-t diagram.

Figure 5.4: The monkey drops itself from a branch at height h, when a dart is shot at a distance d from the tree, and with an initial velocity $\mathbf{v}_0$. The dart hits the monkey at time t_1.



(a) Definition of radians. **(b)** Polar coordinates and unit vectors. **(c)** Summary of angle, sine and cosine values.

Figure 5.5: Some circle basics.

where r is the radius of the arc like in Fig. 5.5a. This allows for a more natural units for angles. The *unit circle* has radius $r = 1$, such that its circumference is 2π, therefore in one full rotation,

$$360^\circ = 2\pi\,\text{rad} \approx 6.283\,\text{rad}. \tag{5.18}$$

So

$$1\,\text{rad} = \frac{180^\circ}{\pi} \approx 57.296^\circ, \tag{5.19}$$

The definition of arc length is consistent: The total circumference of a circle is simply the length of an arc spanning 360°: $\Delta s = 2\pi R$.

A different way to describe a point $P = (x, y)$ in a Cartesian coordinate system, is with *polar coordinates*, here indicated with a semicolon:

$$P = (x, y) = (r\cos\theta, r\sin\theta) = (r; \theta). \tag{5.20}$$

This is illustrated in Fig. 5.5b. The position vector in 2D points from the origin to P is

$$\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} \tag{5.21}$$

$$= r\cos\theta\,\hat{\mathbf{x}} + r\sin\theta\,\hat{\mathbf{y}}. \tag{5.22}$$

However, we can also choose any two other unit vectors that are not parallel (i.e. linearly independent) to each other. For this chapter, it's convenient to define the polar unit vectors, such that

$$\mathbf{r} = r_r\hat{\mathbf{r}} + r_\theta\hat{\boldsymbol{\theta}} = r\hat{\mathbf{r}}. \tag{5.23}$$

Here the direction of the unit vectors depend on the angle θ: $\hat{\mathbf{r}}$ points *radially* along the position vector, while $\hat{\boldsymbol{\theta}}$ points perpendicular to $\hat{\mathbf{r}}$ in the counterclockwise direction. In case of the position vector $\mathbf{r}$, only the radial component is nonzero $r_r = r$, while $r_\theta = 0$.

Table 5.1: Angles in radians and degrees.

θ [rad]	θ [°]	$\sin\theta$	$\cos\theta$	$\tan\theta$
0	0	0	1	—
1	$180/\pi$	0.841	0.540	1.557
$\pi/6$	30	$\sqrt{3}/2$	$1/2$	$\sqrt{3}$
$\pi/4$	45	$\sqrt{2}/2$	$\sqrt{2}/2$	1
$\pi/3$	60	$1/2$	$\sqrt{3}/2$	$\sqrt{3}/3$
$\pi/2$	90	1	0	—
π	180	0	−1	0
$3\pi/2$	270	−1	0	—
2π	360	0	1	0

Namely, θ is defined as the angle with the positive x axis in the counterclockwise direction. Written in terms of our trusty $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$:

$$\hat{\mathbf{r}} = \cos\theta\hat{\mathbf{x}} + \sin\theta\hat{\mathbf{y}} \tag{5.24}$$

$$\hat{\boldsymbol{\theta}} = -\sin\theta\hat{\mathbf{x}} + \cos\theta\hat{\mathbf{y}}, \tag{5.25}$$

such that they both indeed have length 1:

$$|\hat{\mathbf{r}}| = |\hat{\boldsymbol{\theta}}| = \sqrt{\cos^2\theta + \sin^2\theta} = 1, \tag{5.26}$$

and are perpendicular to each other:

$$\hat{\mathbf{r}} \cdot \hat{\boldsymbol{\theta}} = 0. \tag{5.27}$$

5.3 Uniform circular motion

Consider motion in a circle, like a satellite in orbit around the earth. Speed is the constant along its orbit, but the direction of its velocity is changing so it is always tangential to the orbit as in Fig. 5.6a. A change in velocity means there must be an acceleration. The acceleration that keeps the satellite in a circular orbit is called the *centripetal acceleration*. It is orthogonal to the velocity vector, so it does not affect its length: The change is only a "rotation" of the vector with respect to the center of the orbit.

If the velocity were constant in one direction, the satellite would move in a straight line. In that case, it would have travelled a distance $d = vt$ after some time t. From Fig. 5.6b, we see it has "fallen" a distance h instead, because of the centripetal acceleration. What is h? Notice from this figure that there is a right triangle, with sides about which Pythagoras tells us that

$$(vt)^2 + r^2 = (r+h)^2 \tag{5.28}$$

if the orbit is at a radius r. Rewriting,

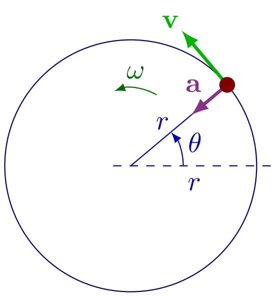
$$(vt)^2 = 2rh + h^2. \tag{5.29}$$

For a very small time t, the fall height h is very small as well. We can therefore make the approximation $h^2 \ll hr$, such that

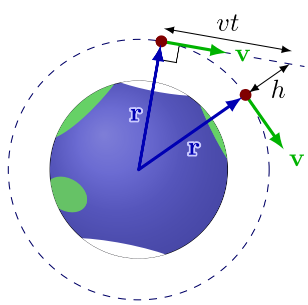
$$(vt)^2 = 2rh. \tag{5.30}$$

We can solve for h:

$$h \approx \frac{1}{2}\left(\frac{v^2}{r}\right)t^2. \tag{5.31}$$



(a) A constant centripetal acceleration is perpendicular to the velocity and changes it direction.



(b) A satellite orbiting the Earth due to gravity creating a centripetal force.

Figure 5.6: Uniform circular motion.

Compare this to our familiar equation for a falling object

$$x = \frac{1}{2}at^2, \tag{5.32}$$

and we recognize that

$$a = \frac{v^2}{r} \tag{5.33}$$

This is known as the centripetal acceleration. This equation says that if the satellite would move at twice the speed ($2v$), it would need four times the acceleration to stay in orbit at the same radius r. On the other hand, for the same velocity v, but half the radius ($r/2$), it would need double the acceleration.

The direction of the acceleration points from the satellite's position towards the center, so we can use the $-\hat{\mathbf{r}}$ unit vector, as in Fig. 5.5b:

Centripetal acceleration.

$$\mathbf{a}_\mathrm{c} = -\frac{v^2}{r}\hat{\mathbf{r}} = -r\omega^2\hat{\mathbf{r}}, \tag{5.34}$$

The satellite completes one full circle in one *period* T. The distance travelled is $2\pi r$, so it has a speed

$$v = |\mathbf{v}| = \frac{2\pi r}{T}. \tag{5.35}$$

The period then can be expressed as

Period for uniform circular motion.

$$T = \frac{2\pi r}{v}. \tag{5.36}$$

The number of times the satellite completes a full orbit per unit time is the *frequency* f:

Frequency for uniform circular motion.

$$f = \frac{1}{T} = \frac{v}{2\pi r}. \tag{5.37}$$

Remember the period has units of seconds, so frequency has units 1/s.

The *angular velocity*, or sometimes *angular frequency*, is the amount of radians, Δs covered per unit time. One full rotation is $\Delta s = 2\pi$, so

Angular velocity.

$$\omega = 2\pi f = \frac{2\pi}{T} = \frac{\mathrm{d}\theta}{\mathrm{d}t} \tag{5.38}$$

Notice it can also be understood as the time derivative of the angle, analogous to position x and velocity v. At a constant angular velocity,

$$\omega = \frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{\theta - \theta_0}{t - t_0}. \tag{5.39}$$

Assuming that $\theta(t_0) = \theta_0 = 0$ at $t_0 = 0$, we see that

Uniform rotation.

$$\theta(t) = \omega t. \tag{5.40}$$

This indeed look familiar to uniform linear motion in Eq. (4.5): $x = vt$.

For uniform circular motion, the angular frequency is

Angular velocity for uniform circular motion.

$$\omega = \frac{v}{r}. \tag{5.41}$$

This allows us to express the velocity also as

$$v = r\omega. \tag{5.42}$$

The centripetal acceleration then, can be rewritten as

$$a = r\omega^2 \tag{5.43}$$

We can derive Eq. (5.34) in a different way, starting from the position

Uniform circular motion.

$$\mathbf{r}(t) = r\cos(\omega t)\hat{\mathbf{x}} + r\sin(\omega t)\hat{\mathbf{y}}, \tag{5.44}$$

where $|\mathbf{r}| = r$ and ω are constant in time, and $\mathbf{r}(0) = r\hat{\mathbf{x}}$. The velocity they is given by

$$\mathbf{v}(t) = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = -r\omega\sin(\omega t)\hat{\mathbf{x}} + r\omega\cos(\omega t)\hat{\mathbf{y}}. \tag{5.45}$$

Similarly, the acceleration is

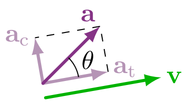
$$\begin{aligned}\mathbf{a}(t) &= \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} \\ &= -r\omega\cos(\omega t)\hat{\mathbf{x}} + -r\omega\sin(\omega t)\hat{\mathbf{y}}.\end{aligned}$$

So we find,

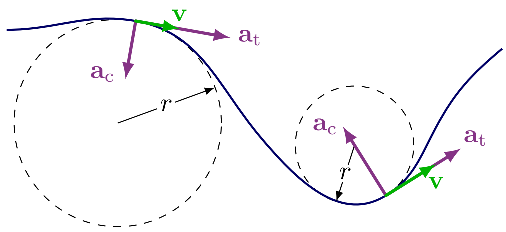
$$\mathbf{a}(t) = -\omega^2\mathbf{r}(t) = -\omega^2 r\hat{\mathbf{r}}(t), \tag{5.46}$$

such that the magnitude of the acceleration is

$$a = \omega^2 r \tag{5.47}$$



(a) The tangential acceleration $\mathbf{a}_\mathrm{t}$ points along $\mathbf{v}$, while the tangential acceleration $\mathbf{a}_\mathrm{c}$ is perpendicular.



(b) A random path with curvature. The centripetal acceleration changes the direction of the velocity, while the tangential acceleration changes its speed.

Figure 5.7: The acceleration vector $\mathbf{a}$ can be broken into centripetal acceleration $\mathbf{a}_\mathrm{c}$ and tangential acceleration $\mathbf{a}_\mathrm{t}$.

Example 5.1: A satellite moves around the earth, close to the surface, just 100 km above. How long does it take to go around? The radius of the earth is

$$r_\oplus = 6370\,\mathrm{km}. \tag{5.48}$$

The radius of the satellite's orbit therefore is $r = 6470\,\mathrm{km}$. There is an acceleration due to gravity, so the velocity to stay in orbit must be

$$v = \sqrt{rg} = 7.97\,\mathrm{km/s}. \tag{5.49}$$

Then the period is

$$T = \frac{2\pi r}{v} \approx 5100\,\mathrm{s} \tag{5.50}$$

Which corresponds to about 85 min . The actual International Space Station is at an average height of about 400 km and takes 92 minutes to fully orbit the Earth.

5.4 Motion along a general path

The acceleration vector along a random path can be broken up into two components as shown in Fig. 5.7b: the *tangential acceleration* a_t, and the *radial acceleration* a_r. Because the velocity vector is always tangential to the path, the tangential component appears when the speed $v = |\mathbf{v}|$ along the path changes. The radial component on the other hand, is always perpendicular to the velocity (and the tangential acceleration), and causes the velocity to change direction. This corresponds to the centripetal force in circular motion. One can always calculate the centripetal acceleration of a turn by measuring the radius of curvature, as shown in the dashed lines in Fig. 5.7b.

Chapter 6

Laws of Motion & Forces

6.1 Momentum

The measure of motion an object has depends on its mass and velocity, and is known as its *momentum*, which is defined as:

Linear momentum (one dimension).

$$p = mv = m\frac{\mathrm{d}x}{\mathrm{d}t} \tag{6.1}$$

If you have two object of different mass moving with the same velocity, the one with the largest mass will have more momentum. The dimensions clearly are mass times length divided by time, so one can use units kg m/s.

Just like velocity, it can be expressed as a vector with several components:

Linear momentum (vector).

$$\mathbf{p} = m\mathbf{v} = m\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}. \tag{6.2}$$

6.2 Newton's laws of motion

An unopposed *force* changes the direction of a mass, as prescribed by Newton's laws of motion. Forces have a direction and magnitude, and are therefore represented by vectors. The unit of force is named after Isaac Newton (1643–1727):

$$\mathrm{N} = \frac{\mathrm{kg\,m}}{\mathrm{s}^2}. \tag{6.3}$$

The *total force*, also *net* or *resultant force*, on some object is the sum of all forces acting on it:

$$\mathbf{F}_{\text{tot}} = \sum_i \mathbf{F}_i. \tag{6.4}$$

Now we are ready to look at Newton's laws of motions that form the foundation of classical mechanics. Newton formulated these laws of motion in the *Philosophiæ Naturalis Principia Mathematica* around 1687. These laws allows us to understand motion of masses and better define what forces actually *are*.

Newton's laws of motion.

1. ***Law of inertia**: An object either remains at rest, or continues to move in a straight line at a constant velocity, unless acted upon by a net force. If the net force is zero, then :*
$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = 0. \tag{6.5}$$
2. *A non-zero net force $\mathbf{F}_{\text{net}}$ will change the momentum of an object, according to*
$$\mathbf{F}_{\text{tot}} = \sum_i \mathbf{F}_i = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} \tag{6.6}$$
3. ***Law of action and reaction**: When one object exerts a force $\mathbf{F}_{12}$ on a second object, the second object simultaneously exerts a force $\mathbf{F}_{21}$ equal in magnitude but opposite in direction to the first object,*
$$\mathbf{F}_{12} = -\mathbf{F}_{21} \tag{6.7}$$

Notice that the way that we have written these laws, the first law is really just a specific case of the second law, in which there is no net force and therefore $\mathbf{a} = 0$. A particle with no net force acting on it is a *free particle*. The first law implies that the momentum of a free particle is constant in time. This law encapsulates the concept of *inertia*, which is the resistance of any object with a mass to any change in its velocity. Some typical examples are quickly pulling away a table cloth, without dragging everything on top of it with you. Or breaking off one sheet of toilet paper without having the whole roll unravel. It explains why objects in circular motion would fly off in a straight path, tangential to the circle, if the centripetal acceleration is suddenly removed. Finally, it is also the reason why you should never attach a trailer to your car with a rope.

If all the forces on an object cancel, we get an *equilibrium*:

Mechanical equilibrium.
$$\mathbf{F}_{tot} = \sum_i \mathbf{F}_i = 0. \tag{6.8}$$

At equilibrium, the first law tells us that that the object will stay at rest, or its velocity will be constant. A book that lies on a flat table is at equilibrium (see next section). A ladder leaning on a wall is at equilibrium. If you jump out of an airplane, very high up, the air resistance will grow with your speed until it *balances* the force of gravity, and you reach a *terminal velocity*, which is constant. At this point, the force of gravity and the force of wind resistance would cancel out, and you would be moving as a free particle with constant velocity.

We know that $\mathbf{p} = m\mathbf{v}$, such that in general

$$\mathbf{F}_{\text{tot}} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + \frac{\mathrm{d}m}{\mathrm{d}t}\mathbf{v}. \tag{6.9}$$

In most cases, m is constant, so the second term, which is the time derivative of the mass, vanishes, and we find the most well-known form of $F = ma$:

Newton's second law for constant mass.

$$\mathbf{F}_{\text{tot}} = m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = m\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = m\mathbf{a}. \tag{6.10}$$

The third law is all about *action and reaction*. When you push on a wall, the wall pushes back. If your force acting on the wall is $\mathbf{F}_{12}$, then the wall pushes on you with a force $\mathbf{F}_{21} = -\mathbf{F}_{12}$. Similarly, when you ride a skateboard, and you push the ground backwards with your foot, the ground pushes you forward in return. Another example is *recoil*: When you shoot a bullet, the gun will push you back as it repels the bullet and hot gas and metal go forward.

6.2.1 Interlude: Algorithm to solving force problems

Students who just learn about forces and Newton's laws often do not know where to start on a physics problem with forces. Typically you can use the following algorithm to systematically solve them:

1. Make a drawing and understand what is going on: What are the moving parts (*degrees of freedom* like positions and angles)? What are the forces on what object? Is there an equilibrium between the forces? Is there motion, acceleration? Use Newton's first law.
2. Write down for each mass the total force $\sum \mathbf{F}_i$ and apply Newton's second law. Choose a coordinate system or positive direction to simplify breaking down the force vector into components.
3. Note what is given and known, and what is unknown. Typically you need at least n independent equations to solve for n unknowns. Often you can use geometric equations to solve for angles or lengths.
4. Finally, sanity check and interpretation: Does the answer make sense? How does it depend on other variables? What is the physical interpretation of the result?

Similar steps can be used for a lot of other basic physics problems. Now let's look at some examples.

6.3 Gravitational force

The most common force in our lives is called *weight*, and results from the force of gravity,

Weight.

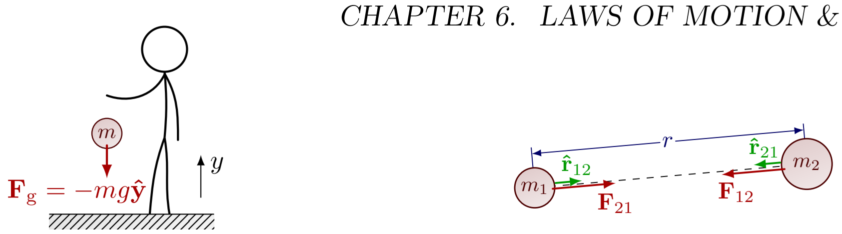
$$\textit{weight} = F_{\text{g}} = mg. \tag{6.11}$$

Here on Earth, the *gravitational acceleration* is around $g \approx 9.8\,\text{m/s}^2$. We can write it as a vector:

$$\mathbf{F}_{\text{g}} = m\mathbf{g}. \tag{6.12}$$

Where $\mathbf{F}$ and $\mathbf{g}$ both point downward to the center of the earth, like in Fig. 6.1a.

In general, however, two points of mass m and M attract each other with a force that acts along the line connecting them, and according to



(a) Weight $\mathbf{F}_\mathrm{g} = -mg\hat{\mathbf{y}}$ close to Earth's surface due to gravity.

(b) The attractive gravitational forces between masses m_1 and m_2.

Figure 6.1: Gravitational force.

Newton's law of universal gravitation.

$$F_\mathrm{g} = \frac{GmM}{r^2}, \tag{6.13}$$

where r is the distance between the mass points, and $G = 6.674 \times 10^{-11}\,\mathrm{m}^3\,\mathrm{kg}^{-1}\,\mathrm{s}^{-2}$ is the *gravitational constant*. The attractive force depends on the inverse of the distance squared: Gravity is four times weaker at double the distance, etc.

The vectorial form illustrated in Fig. 6.1b can be written as

$$\mathbf{F}_{21} = -\frac{Gm_1m_2}{r^2}\hat{\mathbf{r}}_{21}, \tag{6.14}$$

where $\mathbf{F}_{21}$ is the force of mass m_2 acting on m_1, and $\hat{\mathbf{r}}_{21}$ is the unit vector pointing from m_2 to m_1. The minus sign indicates attraction between the masses. Similarly, the force $\mathbf{F}_{12}$ is the force of mass m_1 acting on m_2, which by Newton's third law has equal magnitude, but points in the opposite direction: $\mathbf{F}_{12} = -\mathbf{F}_{21}$.

By comparing Eq. (6.11) to (6.11) to, we see that actually

$$g = \frac{MG}{r^2}. \tag{6.15}$$

So the size of g actually depends on the distance r to the center of the Earth, and the Earth's mass M. At Earth's surface, r is the Earth's radius, and it is easy to compute that $g \approx 9.8\,\mathrm{m/s}^2$. In practice, it also depends on the shape and mass density around you. Typically we assume that the Earth is spherically symmetric, which is good enough as a first-order approximation. It even allows us to approximate the Earth as a simple mass point with no radius at large scales.[1]

6.3.1 Force fields

Gravity is a nice example of a *force field*. A *field* in physics is a quantity (with units) that have some value in every point of space. A field can be a *scalar field* or a *vector field*. A scalar field $f(x, y, z) = f(\mathbf{r})$ takes on a single scalar value everywhere in space, with no direction, and we will see some examples in later chapters on potential energy. A vector field also has some direction in each point.

Close to Earth's surface, the magnitude of the gravitational force remain mostly constant and points in one direction only: downward. This is the simple case of a *uniform field*. The Earth's *gravitational field* can be expressed as

$$\mathbf{g} = -g\hat{\mathbf{z}}, \tag{6.16}$$

[1]In reality, the Earth is slightly "flattened" and bulges a bit at around the equator due to its rotation. Furthermore, land masses tend to be more dense and stick further out of the "ideal surface" of an oblate spheroid. We will learn later more about mass distributions of general bodies and centers of gravity later.

(a) Close to the Earth's surface, the field is uniform (constant magnitude and direction).

(b) At a large scale, the field depends on the distance r to the center of the Earth.

Figure 6.2: Gravitational force field $\mathbf{g}$.

where $\hat{\mathbf{z}}$ points down. The vector field $\mathbf{g}$ acts on masses. But this is only an approximation. At a larger scale, the earth creates a gravitational field

$$\mathbf{g}(\mathbf{r}) = -\frac{MG}{r^2}\hat{\mathbf{r}}, \tag{6.17}$$

which only depends on the distance from Earth's center.

6.4 Normal force

When an object with mass m rests on a table, something must counter the gravitational force to hold it in place. This is the *normal force* $\mathbf{F}_\mathrm{N}$ (or just $\mathbf{N}$ in some textbooks) the table's surface exerts on the mass. Because the objects stays put, all forces on the object must cancel:

$$\sum_i \mathbf{F}_i = \mathbf{F}_\mathrm{N} + \mathbf{F}_\mathrm{g} = 0. \tag{6.18}$$

It is called "normal" because it is orthogonal to the surface. In the simplest case of Fig. 6.3a, this is a one-dimensional problem, so fixing the y axis along the vertical, we have

$$\mathbf{F}_\mathrm{N} = F_\mathrm{N}\hat{\mathbf{y}} \tag{6.19}$$

$$\mathbf{F}_\mathrm{g} = -mg\hat{\mathbf{y}}. \tag{6.20}$$

So

$$F_\mathrm{N}\hat{\mathbf{y}} - mg\hat{\mathbf{y}} = 0, \tag{6.21}$$

or simply,

$$F_\mathrm{N} - mg = 0. \tag{6.22}$$

So the normal force is simply the weight!

$$F_\mathrm{N} = mg. \tag{6.23}$$

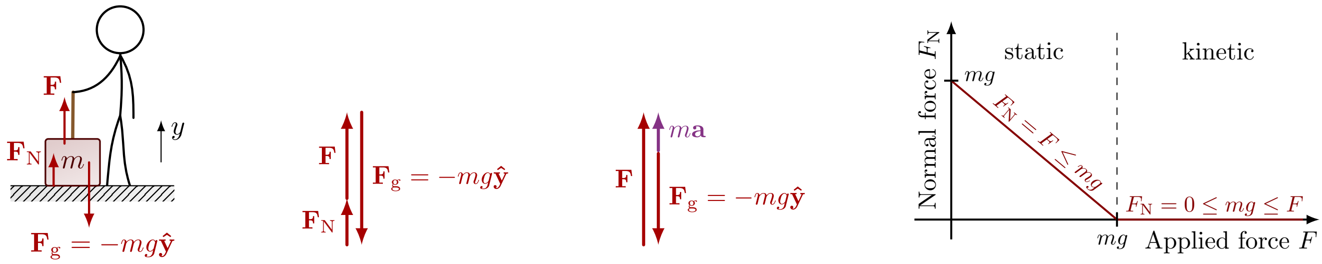
For example, if your mass is 80 kg, the normal force is about 784 N. On the moon, where $g_☾ = 1.625\,\mathrm{m/w^2} = 0.167g$, you would *weigh* 130 N, i.e. any scale calibrated to Earth's gravity would read 13 kg.

Remember that due to the third law, the book also pushes on the table, so the table experiences a force $-mg\hat{\mathbf{y}}$ from the book.

(a) The forces on a mass m at rest: The surface exerts a normal force $\mathbf{F}_\mathrm{N}$, and the Earth a gravitational one, $\mathbf{F}_\mathrm{g}$.

(b) The normal and gravitational force vectors create a balance.

Figure 6.3: Normal force.



(a) A person pulls up a mass with force **F**. **(b)** At equilibrium, the forces cancel. **(c)** Forces not canceling cause an acceleration. **(d)** The normal force becomes smaller as the applied **F** becomes larger.

Figure 6.4: Normal force.

Now, let's make it a bit more complicated. Imagine you slowly start pulling up the mass with a force F, as in Fig. 6.4a. You build up the force, and before applying enough to lift it up from the ground, you still maintain an equilibrium:

$$\sum_i \mathbf{F}_i = \mathbf{F}_\mathrm{N} + \mathbf{F}_\mathrm{g} + \mathbf{F} = 0, \tag{6.24}$$

or

$$F_\mathrm{N} - mg + F = 0, \tag{6.25}$$

where we assumed **F** is in the positive vertical directions. This time, the normal force is less than the weight mg:

$$F_\mathrm{N} = mg - F. \tag{6.26}$$

When you reach enough force to lift the mass from the ground, the normal force disappears. This is shown in Fig. 6.4d. At this points, the forces do not balance anymore, and you get an acceleration upward, given by

$$\sum_i \mathbf{F}_i = \mathbf{F}_\mathrm{g} + \mathbf{F} = m\mathbf{a}. \tag{6.27}$$

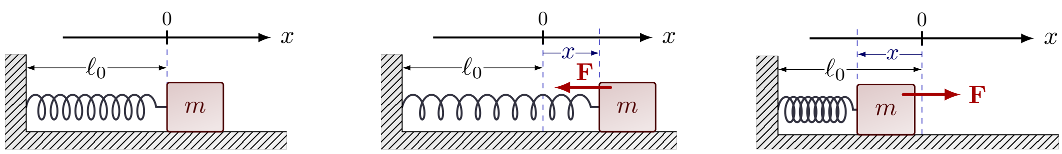
6.5 Springs & Hooke's law

A spring pulls when extended, and pushes when compressed. It always "wants" to return to its *rest length* ℓ_0. It does so with a force F, that is proportional to the change in length x shown in Fig. 6.5:

Hooke's law.

$$F = -kx. \tag{6.28}$$

The constant of proportionality k is the spring constant, which depends on the spring in question. This constant is a measure of the "stiffness" of the spring. The larger k, the stronger the spring will try to return to its rest length. The minus sign indicates that the spring's force points towards $x = 0$. Not all springs follow Hooke's law; those that do are called *Hookian*. Still, Hooke's law is often a very good approximation, and we will assume all springs in this course follow Hooke's law.



(a) Rest length ℓ_0. **(b)** Extension $x > 0$. **(c)** Compression $x < 0$.

Figure 6.5: Spring with a mass.

(a) Without a mass, the spring has length ℓ_0. **(b)** Elongation $\Delta\ell$ due to gravity.

Figure 6.6: Hanging spring with rest length ℓ_0.

6.5.1 Example 1: Vertical spring

Consider a mass hanging on a string as in Fig. 6.6. The weight from the mass will cause the spring to be extended a little bit to some length $\ell_0 + \Delta\ell$. At equilibrium, the spring force and weight balance

$$\sum_i \mathbf{F}_i = \mathbf{F} + m\mathbf{g} = 0, \tag{6.29}$$

or,

$$0 = k\Delta\ell - mg, \tag{6.30}$$

such that the extension is given by

$$\Delta\ell = \frac{mg}{k}. \tag{6.31}$$

6.5.2 Example 2: Double spring

Say you have a mass on a frictionless surface connected to two springs on either ends with different spring constants k_1 and k_2 as in Fig. 6.7. Assume the system is at rest, and the springs are not exactly at their own rest length ℓ_0. What then, are their relative changes in length?

First choose as convention the positive x direction to the right, as indicated in Fig. 6.7. This means that a positive x_1 corresponds to an extension of spring k_1, and negative x_1 actually means a compression, and vice versa with x_2 for spring k_2.

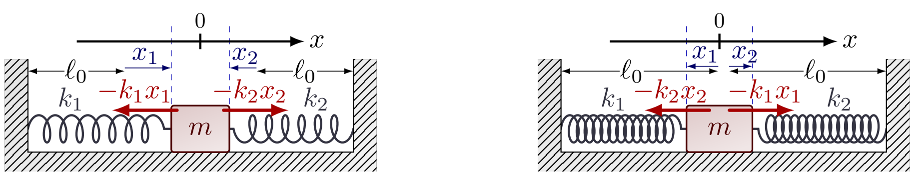
For the system to be at rest, their forces must balance,

$$\sum_i \mathbf{F}_i = -k_1 x_1 - k_2 x_2 = 0. \tag{6.32}$$

The rest position of the system is therefore given by

$$x_1 = -\frac{k_2}{k_1} x_2. \tag{6.33}$$

The minus sign indicates that the springs are either both extended as in Fig. 6.7a ($x_1 > 0$, $x_2 < 0$), both compressed as in Fig. 6.7b ($x_1 < 0$, $x_2 > 0$), or not extended at all ($x_1 = 0 = x_2$). As expected, the forces the springs exert on the mass are opposite. The spring with the largest constant, has the largest change in length.



(a) Both springs are extended. **(b)** Both springs are compressed.

Figure 6.7: Double spring at rest. Both springs have rest length ℓ_0.

(a) The forces acting on a suspended mass. **(b)** Piece of string, zoomed in: $\mathbf{T}_1 = -\mathbf{T}_2$.

Figure 6.8: Tension force on a mass suspended by a string.

6.6 Tension

Suppose a mass m hangs at the end of a string or rope, which is suspended from the ceiling like in Fig. 6.8a. The string pulls the mass with a force called the *tension*. Tension is very similar to the normal force, as it holds the mass in place:

$$\sum_i \mathbf{F}_i = \mathbf{T} + \mathbf{F}_\mathrm{g} = 0. \tag{6.34}$$

What is special about tension is that it it the same everywhere through the string. If you zoom into one piece of string, you see it experiences two forces on either end, as in Fig. 6.8b. If at rest:

$$\sum_i \mathbf{F}_i = \mathbf{T}_1 + \mathbf{T}_2 = 0. \tag{6.35}$$

So the tension has the same magnitude in both directions:

$$0 = T_1 - T_2. \tag{6.36}$$

Therefore we typically use one symbol $\mathbf{T}$ or T that is the same along a string. Furthermore, we typically assume that the mass of the string is negligible to simplify our problems.

This is where the fun of endless physics problems begins. Next, we will look at some examples of setups with pulleys, which redirect the tension in strings.

6.6.1 Example 1: Falling mass on a pulley

Consider a mass m_1 that can slide on a frictionless surface and is connected by a string to another mass m_2 hanging downward over a pulley as in Fig. 6.9a.

As mentioned above, the tension is the same everywhere in the string. The pulleys makes the string change direction by exerting a normal force on the piece of string it is in direct contact with. So the force $\mathbf{T}_1$ of the string pulling mass m_1 to the right, and the force $\mathbf{T}_2$ if it pulling mass m_2 up is of the same magnitude:

$$T = T_1 = T_2. \tag{6.37}$$

We do not care about the vertical forces on m_1 in this case; the interesting direction is that of motion. Assuming no friction, m_2 will fall down, pulling mass m_1 to the right. We will treat this "direction of motion" as one axis x as indicated in Fig. 6.9a, even though it changes "true" direction due to the pulley. We can write down Newton's second law for *each* mass in this direction of motion:

$$\begin{cases} m_1 a_1 = T_1 & \text{(mass 1)} \\ m_2 a_2 = m_2 g - T_2 & \text{(mass 2)} \end{cases} \tag{6.38}$$

Because the two masses are connected by the string they will have the same acceleration: If the m_1 moves by a distance Δx, m_2 moves by a distance Δx, if the m_1 moves at a speed

(a) Mass on a frictionless table is pulled by a hanging mass.

(b) Atwood machine with two pulleys: Two hanging masses are connecting by a string over two pulleys.

Figure 6.9: Setup of some simple pulley problems.

v, m_2 moves at a speed v, and the same goes for the acceleration. This works as long as we assume the string is not "stretchy". So $a = a_1 = a_2$. We then find that

$$\begin{cases} m_1 a = T \\ a = \dfrac{m_2}{m_1 + m_2} g \end{cases} \tag{6.39}$$

So in case $m_1 = m_2$, $a = g/2$.

6.6.2 Example 2: Two masses balancing over pulleys

An *Atwood machine*, shown in Fig. 6.9b, has two masses hanging by a string that connects them over two pulleys. This time, we write:

$$\begin{cases} m_1 a = T - m_1 g & \text{(mass 1)} \\ m_2 a = m_2 g - T & \text{(mass 2)} \end{cases} \tag{6.40}$$

Here we again have chosen the positive direction by setting an axis that follows the rope from m_1 to m_2. (Alternatively, we could have chosen the vertical y axis and set $a_1 = -a_2$, which means that if one mass accelerates in direction, the other accelerates in the opposite.) Given m_1, m_2 and g, we have two unknowns: T and a. But we have two equations, so we can solve it easily:

$$\begin{cases} a = \dfrac{T}{m_1} - g = g \\ T = \dfrac{2m_1 m_2}{m_1 + m_2} g \end{cases} \tag{6.41}$$

In case $m = m_1 = m_2$, $T = mg$ and $a = 0$, which makes sense.

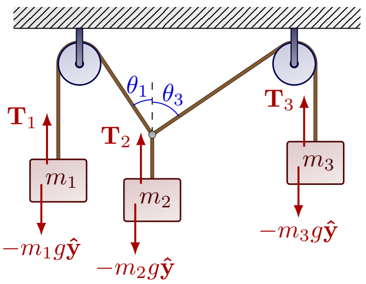
6.6.3 Example 3: Three unequal masses

Finally, look at Fig. 6.10a: Three masses are connected by strings, suspended by pulleys. We assume this system is in equilibrium. There are three tensions,

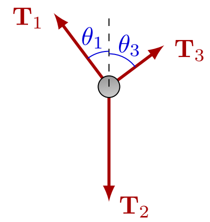
$$T_1 = m_1 g \tag{6.42}$$
$$T_2 = m_2 g \tag{6.43}$$
$$T_3 = m_3 g. \tag{6.44}$$

The interesting point is where the three strings meet, as in Fig. 6.10b:

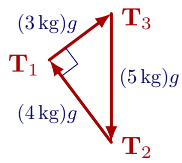
$$\mathbf{T}_1 = -T_1 \sin\theta_1 \hat{\mathbf{x}} + T_1 \cos\theta_1 \hat{\mathbf{y}} \tag{6.45}$$
$$\mathbf{T}_2 = -T_2 \hat{\mathbf{y}} \tag{6.46}$$
$$\mathbf{T}_3 = T_3 \sin\theta_3 \hat{\mathbf{x}} + T_3 \cos\theta_3 \hat{\mathbf{y}}. \tag{6.47}$$



(a) Setup of the masses, strings and pulleys, and the forces acting on each mass.



(b) Forces on the point where the three strings meet.



(c) The tensions form a right triangle for 5 : 4 : 3 ratios.

Figure 6.10: Third example of masses with strings and pulleys.

What are these angles θ_1 and θ_3? We can write down Newton's law to obtain two equation that can be solved for these two unknown angle if the masses are given:

$$\begin{cases} 0 = -T_1 \sin\theta_1 + T_3 \sin\theta_3 \\ 0 = T_1 \cos\theta_1 - T_2 + T_3 \cos\theta_3 \end{cases} \tag{6.48}$$

As we saw in class, a special combination is when $m_1 = 4\,\mathrm{kg}$, $m_2 = 5\,\mathrm{kg}$ and $m_2 = 3\,\mathrm{kg}$, in which case the two top strings will form a right angle, because the tension vectors will form a right triangle with sides of 5 : 4 : 3 ratios.

6.6.4 Example 4: Block & tackle

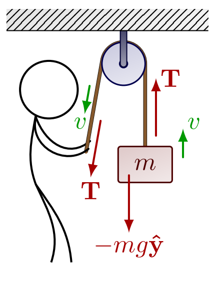
Pulleys are useful to lift heavy loads. When a person lifts up a mass from the ground, they also have to lift their own body weight to stand up. Consider the simple setup in Fig. 6.11a. Because the pulley on the ceiling redirects the tension in the rope downward, the person can now use their own body weight to their advantage to lift the mass. To lift the mass at a constant speed is, they have to pull with a force $T = mg$.

An even better setup is a "block and tackle" with two pulleys, shown in Fig. 6.11b. The pulley on the mass is sometimes called a "snatch block". This time, the force needed to lift the mass at a constant speed v is given by

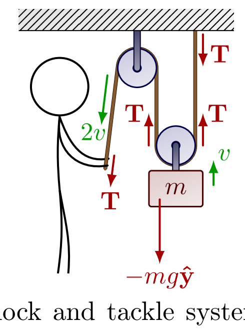
$$0 = T\hat{\mathbf{y}} + T\hat{\mathbf{y}} - mg\hat{\mathbf{y}}. \tag{6.49}$$

In other words, $T = mg/2$. It has halved with respect to the simple setup! Note that to lift the mass by a height h, the person has to pull the rope by a length $2h$.

Their are many variations one can make by winding the rope several times through the pulleys, or simply adding more pulleys.



(a) Lifting a mass with help from your body weight. The force is $T = mg$.



(b) A block and tackle system to halve the force, $T = mg/2$.

Figure 6.11: Using pulleys to lift masses with more ease.

6.7 Centripetal force

If you hold a bucket with water upside down, the water will pour out. But if you spin it around vertically and fast enough, the water will stay in the bucket. This is a consequence of the first law: The water has some velocity and because of its inertia, it wants to move in a straight line. However, there is a non-zero net force that "pushes" the water to change direction to stay in a circle path. This is the bottom of the bucket pushing the water. This is called the *centripetal force* $\mathbf{F}_\mathrm{c}$, which provides the centripetal acceleration we have seen in Section 5.3. To keep a mass m in a circle with radius r and tangential velocity $v = r\omega$, the net centripetal force according to Eq. (5.34) has to be

Centripetal force.

$$\mathbf{F}_\mathrm{c} = -m\frac{v^2}{r}\hat{\mathbf{r}} = -mr\omega^2\hat{\mathbf{r}}, \tag{6.50}$$

where the unit vector $\hat{\mathbf{r}}$ points radially from the center of the circle to the mass.

Other examples include the Earth orbiting the sun, where gravity acts as the centripetal force, or a mass m on a string you sling around very fast around your head. If it is the only force, the mass will move with a constant speed v in a circle with constant radius r. This is illustrated in Fig. 6.12.

You may have heard about the *centrifugal force*. This is the force the water "feels" when it is being swung around, and you might also experience it when a car takes a sharp turn, or you ride a roller coaster. This is however not a real force, but rather a *pseudoforce*. It is the consequence of being accelerated, while your inertial mass wants to move in a straight line. In the frame of reference of the water or car passenger, it seems like there is a force pulling it. We will talk more about inertial frame of references in Section 8.4.

6.7.1 Example: Mass on a string

Suppose you have a mass m hung on a string being slowly swung around in circles of radius r like in Fig. 6.13a. What is the angle of the string with the horizontal as a function of the string's length L and the angular frequency ω? There are two forces, the tension and gravitational force:

$$\sum \mathbf{F} = \mathbf{T} + m\mathbf{g} = m\mathbf{a}. \tag{6.51}$$

We know the velocity changes direction, because the centripetal force "pulls" the mass into a circular motion. So $\mathbf{a}$ is not zero. Choosing the coordinate system as in Fig. 6.13a, the forces can be decomposed into

$$\begin{cases} m\dfrac{v^2}{r} = T\sin\theta \\ 0 = T\cos\theta - mg \end{cases} \tag{6.52}$$

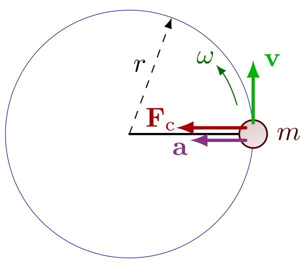
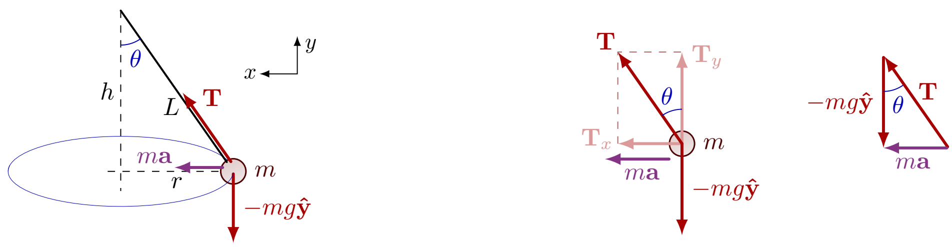


Figure 6.12: A constant centripetal force keeps a mass moving in a circle with constant radius r, angular velocity ω, and tangential velocity $v = r\omega$.



(a) A mass m on a string of length L swings in circles of radius r. There are two forces: the tension and weight.

(b) Breakdown of the forces on the mass: They do not balance and the tension causes a centripetal acceleration $\mathbf{a}$.

Figure 6.13: Example of a mass on a string.

Here we have used the fact that that the tension in the string will balance the gravitational force in the y direction to hold it at the same height (or equivalently, at the same angle θ), and simultaneously provide a centripetal force in the x direction:

$$\mathbf{a} = a_x\hat{\mathbf{x}} + a_y\hat{\mathbf{y}} \tag{6.53}$$

$$= \frac{v}{m}\hat{\mathbf{x}}. \tag{6.54}$$

By comparing the two formula in Eq. (6.52), we find that

$$\frac{v^2}{r\sin\theta} = \frac{g}{\cos\theta}. \tag{6.55}$$

Therefore,

$$\tan\theta = \frac{v^2}{rg} = \frac{r\omega^2}{g}, \tag{6.56}$$

where we have used $v = r\omega$. We can find yet another independent equation by looking at the right-angled triangle made by the string with the horizontal:

$$\tan\theta = \frac{r}{h}. \tag{6.57}$$

Comparing Eqs. (6.56) and (6.57), we get

$$\omega^2 = \frac{g}{h}. \tag{6.58}$$

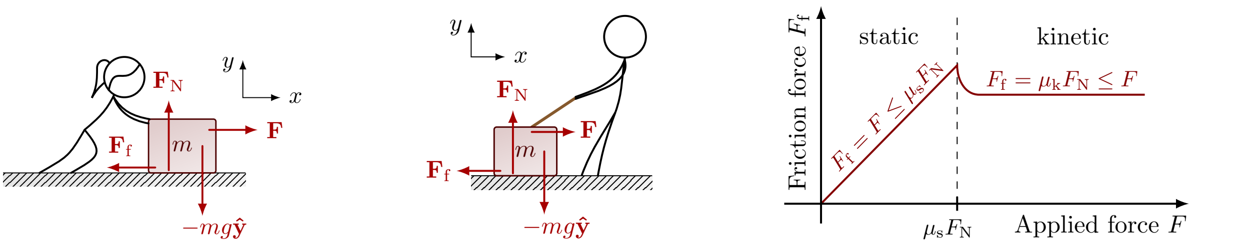
To find θ, we can substituting in $h = L\cos\theta$,

$$\omega^2 = \frac{g}{L\cos\theta}, \tag{6.59}$$

or

$$\theta = \mathrm{acos}\left(\frac{g}{\omega^2 L}\right). \tag{6.60}$$

The acos function is a monotone decreasing function, so this equation says the faster the mass swings around (larger ω, smaller $1/\omega^2$), the larger the angle θ, which makes intuitive sense.



(a) Pushing a mass.

(b) Pulling a mass.

(c) Two regimes of friction F_f, depending on the applied force F.

Figure 6.14: Frictional force $\mathbf{F}_\text{f}$ counteracts the applied force $\mathbf{F}$. Once $F > \mu_\text{s} F_\text{N}$, the friction cannot balance the applied force, and the mass starts moving with a constant friction $F_\text{f} = \mu_\text{k} F_\text{N}$.

6.8 Friction

Up until now we have assumed that our surfaces were frictionless. In reality, most surfaces are not infinitely smooth, so there will be a *frictional force* F_f (sometimes just f), pointing in the opposite direction of motion. Imagine you push or pull a mass across the ground with a force F, then friction will counteract the force of your push or pull. It turns out that friction is proportional to the normal force: If the block's mass is twice as large, the frictional force is twice as large as well. There will be a constant of proportionality μ that depends on the materials involved:

Friction during motion.

$$F_\text{f} = \mu F_\text{N}. \tag{6.61}$$

The direction will always be parallel to the surface and in the opposite direction of motion. The constant μ is the unitless *coefficient of friction*, and takes values between 0 and 1. $\mu = 0$ corresponds to no friction, and $\mu = 1$ to maximal friction. Each combination of two materials has two coefficients: the *static coefficient of friction* μ_s, and the *kinetic coefficient of friction* μ_k. Table 6.1 lists some concrete values of familiar materials. The static coefficient is only important if the object is pushed from rest to motion. Once in motion, the kinetic coefficient comes into play. Typically $\mu_\text{s} > \mu_\text{k}$: Once there is motion, the friction becomes a reduced a little bit. In case the force $F < \mu_\text{s} F_\text{N}$, the block stays at rest, and

Friction at rest.

$$F_\text{f} = F < \mu_\text{s} F_\text{N}. \tag{6.62}$$

This is plotted in Fig. 6.14c.

Table 6.1: Static and kinetic coefficient of friction for several combinations of materials.

Materials	μ_s	μ_k
Wood on wood	0.25–0.5	0.2
Steel on steel	0.74	0.57
Teflon on steel	0.04	0.04
Ice on ice	0.1	0.03
Synovial joint	0.01	0.003

(a) There are three forces on the mass: the normal force, friction and weight.

(b) The forces on the mass balance, as long as friction cancels the x component of the weight.

Figure 6.15: Frictional force on a mass on an inclined plane.

6.8.1 Example: Friction on an inclined plane

Now, let's look at an example where the table if inclined with some angle θ with respect to the horizontal. If the angle is small, friction will hold the block of mass in place, but if the angle is large enough, the block will start to slide down, due to gravity. What is the *critical angle* θ_c at which this happens? First we identify the three forces acting on the mass:

$$\sum \mathbf{F}_i = \mathbf{F}_\mathrm{g} + \mathbf{F}_\mathrm{f} + \mathbf{F}_\mathrm{N}. \tag{6.63}$$

This is clearly a two dimensional problem, so we break down the vectors into their components. It is convenient to use a coordinate system with the x axis pointing downward along the table's plane, and the y axis perpendicular to it, like in Fig. 6.15a. The y axis now makes an angle θ with the vertical, so gravity will have a x and y component:

$$\mathbf{F}_\mathrm{g} = F_{\mathrm{g}x}\hat{\mathbf{x}} + F_{\mathrm{g}y}\hat{\mathbf{y}} \tag{6.64}$$
$$= mg\sin\theta\hat{\mathbf{x}} - mg\cos\theta\hat{\mathbf{y}}. \tag{6.65}$$

Notice that if $\theta = 0$, the x component disappears, just as expected. Assume θ is small enough for the block to stay at rest. Then

$$\begin{cases} 0 = F_\mathrm{g} - mg\cos\theta \\ 0 = F_\mathrm{N} + mg\sin\theta - F_\mathrm{f} \end{cases} \tag{6.66}$$

are the conditions for equilibrium. Because the normal force F_N is pointing upward and perpendicular to the table, and the friction F_g point opposes the x component of gravity that would make the object slide. To stay at rest, we have the condition

$$mg\cos\theta = F_\mathrm{g} < \mu_\mathrm{s} F_\mathrm{N}. \tag{6.67}$$

This defines the critical point at which the block starts sliding:

$$mg\cos\theta_\mathrm{c} = F_\mathrm{g} = \mu_\mathrm{s} F_\mathrm{N}. \tag{6.68}$$

Substituting this into Eq. (6.66), we find

$$\mu = \tan\theta_\mathrm{c}. \tag{6.69}$$

This result allows us to measure μ_s for any combination of materials by measuring the critical angle θ_c when the block starts sliding. The larger μ_s, the larger the critical angle θ_c at which the mass starts to move.

6.8.2 Drag

Drag is the frictional force due to air resistance.

Consider an airplane flying at a constant velocity. Simplified, it experiences four forces:

- Gravity F_{g} pulling the plane downward, as usual.
- Thrust F_{thrust} by the engines propelling the airplane forward.
- Drag F_{drag} opposing the thrust due to air resistance.
- Lift F_{lift} by the wings, pushing the airplane upward.

What is special about the drag force is that it depends on the velocity. In simple cases, it is proportional to the cross-sectional area of the plane and the velocity squared:

$$F_{\text{drag}} \propto A v^2. \tag{6.70}$$

Take for example two planes with the same engine (so the same thrust), but one with an area two times as large, $A_1 = 2A_2$. Flying at constant speeds, what will their velocities be relative to each other? Since the thrust are the same, the drag forces have to be the same. We can therefore find that

$$\begin{aligned} F_{\text{drag},1} &= F_{\text{drag},2} \\ CA_1 v_1^2 &= CA_2 v_2^2 \\ A_1 v_1^2 &= \frac{1}{2} A_1 v_2^2 \end{aligned}$$

where C is some constant that is the same between both planes. So $v_2 = \sqrt{2} v_1$. If the "wider" plane is moving with a speed $v_1 \sim 800\,\text{km}h$, the smaller plane with the same thrust will fly at $v_2 \sim 1130\,\text{km}h$.

Chapter 7

Work & Energy

7.1 Work

In this chapter, we are going to discuss the relationship between force and energy. Recall the units of force are

$$\mathrm{N} = \frac{\mathrm{kg} \cdot \mathrm{m}}{\mathrm{s}^2}, \tag{7.1}$$

which is easy to remember using $F = ma$. Now, the units of energy are *Joules*:

$$\mathrm{J} = \frac{\mathrm{kg} \cdot \mathrm{m}^2}{\mathrm{s}^2} = \mathrm{N} \cdot \mathrm{m} \tag{7.2}$$

In fact, the amount of *energy* that is transferred to an object by a force F by moving it over a distance Δx can be defined as

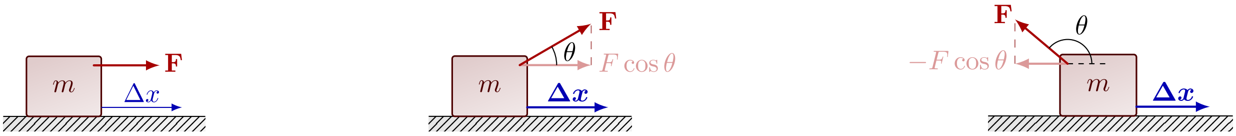
$$W = F\Delta x. \tag{7.3}$$

This is the *work* performed by the force F, see Fig. 7.1a. And indeed, *work* allows us to define the amount of energy that objects can possess. Next semester we will also see another type of energy transfer, namely by *heat* in *thermodynamics*.

Equation (7.3) is a simple case where the displacement is parallel to the force. Work can be positive or negative: It is positive if the force,F, points in the same direction as the displacement, and negative if they are opposite. More generally, the force vector $\mathbf{F}$ and displacement vector $\boldsymbol{\Delta x}$ may not be parallel, as in Fig. 7.1b:

> **Work.** *The work done by a force F to move another object by a displacement $\boldsymbol{\Delta x}$ is*
>
> $$W = \mathbf{F} \cdot \boldsymbol{\Delta x} = F\Delta x \cos\theta, \tag{7.4}$$
>
> *where θ is the angle between $\mathbf{F}$ and $\boldsymbol{\Delta x}$.*

Notice that $F\cos\theta$ can be interpreted as the component of $\mathbf{F}$ that is along the $\boldsymbol{\Delta x}$ direction, which is why the scalar product is so useful here. We see that if $0 < \theta < \pi/2$, the work will be positive because $\cos\theta > 0$ (Fig. 7.1b), while if $\pi/2 < \theta < 3\pi/2$, the work will be negative ($\cos\theta < 0$, see Fig. 7.1c). If the force and displacement are perpendicular to each other, $W = \mathbf{F} \cdot \boldsymbol{\Delta x} = 0$, the work will be zero: The force $\mathbf{F}$ cannot not cause any displacement in the $\boldsymbol{\Delta x}$ direction.



(a) The force and displacement are aligned.

(b) The work is positive if the force and displacement are in the same direction.

(c) The work is negative if the force and displacement are in the opposite direction.

Figure 7.1: A force $\mathbf{F}$ can transfer energy to an object by moving it by a displacement $\boldsymbol{\Delta x}$. In that case it does work $W = \mathbf{F} \cdot \boldsymbol{\Delta x}$. We are interested in the component of the force parallel to the displacement.

7.2 Kinetic energy

Newton's second law says that an unbalanced force or sum of forces causes an acceleration: $\sum \mathbf{F} = m\mathbf{a}$. Remember Torelli's equation (4.22) to find the final velocity from the initial velocity v_0, acceleration a, and displacement Δx without caring about the time it took:

$$v^2 = v_0^2 + 2a\Delta x. \tag{7.5}$$

So we can express the acceleration as

$$a = \frac{v^2 - v_0^2}{2\Delta x}. \tag{7.6}$$

Therefore, in one dimension:

$$F\Delta x = \left(m\frac{v^2 - v_0^2}{2\Delta x} \right) \Delta x \tag{7.7}$$

$$= \frac{mv^2}{2} - \frac{mv_0^2}{2}. \tag{7.8}$$

We see that there are two quantities *before* and *after*, which we define as the *kinetic energy* K:

Kinetic energy.

$$K = \frac{mv^2}{2}. \tag{7.9}$$

This is the amount of energy due to the movement of an object. It has no direction; it is a scalar. It only depends on the *total* velocity squared and its mass. We also find another important result,

Work-energy theorem. *The total work done on an object causes a change in kinetic energy:*

$$W = \Delta K = K_\mathrm{f} - K_\mathrm{i} \tag{7.10}$$

$$= \frac{m{v_\mathrm{f}}^2}{2} - \frac{m{v_\mathrm{i}}^2}{2}. \tag{7.11}$$

Here the subscript "f" refers to "final", and "i" to "initial". Total means that we need to consider the sum of all forces: The work-energy theorem only works for the total force, since we used Newton's second law in the derivation. It is best to look at an example.

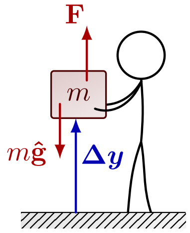


Figure 7.2: Arnold lifts up a weight of mass m to a height h with a constant force F. The displacement vector is $\mathbf{\Delta y} = h\hat{\mathbf{y}}$.

7.2.1 Example: Lifting a weight

Consider Arnold lifting a block mass $m = 5\,\mathrm{kg}$ by a height $h = 2\,\mathrm{m}$ with a force $F = 500\,\mathrm{N}$. There are two forces: the one by Arnold, and the weight due to gravity, see Fig. 7.2. Several questions come up:

1. What is the work done by Arnold?
2. What is the work done by gravity?
3. What is the final velocity of the block?

The answers are simple enough:

1. Arnold's force is parallel to the displacement, $\theta = 0$, so $W_\mathrm{A} = Fh\cos\theta = 1000\,\mathrm{J}$.
2. Gravity opposes the displacement, $\theta = 180°$, so $W_\mathrm{g} = mgh\cos\theta = -100\,\mathrm{J}$.
3. The total work on the block is $W_\mathrm{tot} = 900\,\mathrm{N} = \Delta K$, so the final velocity upwards is

$$v_\mathrm{f} = \sqrt{\frac{2W_\mathrm{tot}}{m}} = 19\frac{\mathrm{m}}{\mathrm{s}}. \tag{7.12}$$

If Arnold moved the block with a constant force F, then this would be the final velocity v_f at height h.

7.2.2 Work integral over a path

So far we have only looked at a constant force F. In this case the work, $W = F\Delta x$, can be thought of as the blue area in Fig. 7.3a. If the force $F = F(x)$ varies as a function of displacement x, we can use an integral

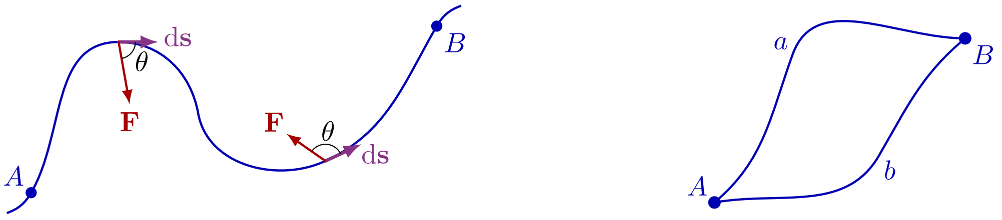
$$W = \int_{x_1}^{x_2} F\mathrm{d}x. \tag{7.13}$$

Even more generally, the moved object can follow a three-dimensional path where the direction of the displacement and force varies as illustrated in Fig. 7.4a. Let's call s the

(a) A constant force F.

(b) A general force F.

Figure 7.3: Work is the integral of $\mathbf{F} \cdot \hat{\mathbf{x}}$ over $\mathrm{d}x$.



(a) A general path. **(b)** A closed path.

Figure 7.4: Integral.

distance along the path, and the vector $\mathrm{d}\mathbf{s}$ a small displacement that is tangential to the path.[1] In three dimensions,

$$\mathrm{d}\mathbf{s} = \mathrm{d}x\hat{\mathbf{x}} + \mathrm{d}y\hat{\mathbf{y}} + \mathrm{d}z\hat{\mathbf{z}}. \tag{7.14}$$

The work done by the vector $\mathbf{F}(\mathbf{s})$ is the integral along the path

Work along a path.

$$W = \int_{s_1}^{s_2} \mathbf{F} \cdot \mathrm{d}\mathbf{s}\,, \tag{7.15}$$

with $\mathrm{d}\mathbf{s}$ as the differential element. This is a *line integral.* Notice that we are actually only interested in the component of the force $\mathbf{F}$ that is parallel to the displacement $\mathrm{d}\mathbf{s}$,

$$\mathbf{F} \cdot \mathrm{d}\mathbf{s} = F_\mathrm{t}\,\mathrm{d}s\,, \tag{7.16}$$

where F_t is the component of $\mathbf{F}$ that is tangential to the path. And of course, $\mathrm{d}\mathbf{s}$ is also parallel to the path, with $\mathrm{d}s$ being the length of $\mathrm{d}\mathbf{s}$.

Let's demonstrate the work-energy theorem for this general path. We want to arrive at Eq. (7.11), so we need to retrieve the velocity from somewhere. We can use Newton's second law:

$$\mathbf{F}(\mathbf{s}) = m\mathbf{a} = m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}, \tag{7.17}$$

where $\mathbf{v}$ is the velocity along the path, such that

$$\mathbf{v} = \frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t}. \tag{7.18}$$

Remember from Section 5.4 that only the tangential acceleration changes the magnitude of the velocity vector. The tangential acceleration a_t is caused by the tangential component F_t, so:

$$F_\mathrm{t}(\mathbf{s}) = ma_\mathrm{t} = m\frac{\mathrm{d}v}{\mathrm{d}t}. \tag{7.19}$$

We can now solve the integral (7.15) by substituting $\mathrm{d}s$ with $\mathrm{d}v$ using the chain rule

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d}v}{\mathrm{d}s}v. \tag{7.20}$$

So

$$F_\mathrm{t}(\mathbf{s})\,\mathrm{d}s = mv\,\mathrm{d}v\,. \tag{7.21}$$

and we solve the integral

$$W = \int_{v_\mathrm{i}}^{v_\mathrm{f}} mv\,\mathrm{d}v \tag{7.22}$$

$$= \frac{m{v_\mathrm{f}}^2}{2} - \frac{m{v_\mathrm{i}}^2}{2}, \tag{7.23}$$

[1]Not to be confused with the $r(t)$, the radial distance to the origin, and $\mathbf{r}(t)$, the accompanying position vector, which points from the origin to a point along the path.

where we integrate over v and set the limits to $v_\mathrm{i} = v(s_1)$ and $v_\mathrm{f} = v(s_2)$. We have shown that the work-energy theorem also holds for our path integral:

Work-energy theorem for a general path.

$$W = \int_{s_1}^{s_2} \mathbf{F} \cdot \mathrm{d}\mathbf{s} = \frac{m{v_\mathrm{f}}^2}{2} - \frac{m{v_\mathrm{i}}^2}{2}. \tag{7.24}$$

Again, because we used Newton's second law in the derivation, W has to be the total work done by the total force.

Sometimes it might be useful to write this integral in a time dependent way. Using Eq. (7.18),

Work time-integral.

$$W = \int_{t_1}^{t_2} \mathbf{F} \cdot \mathbf{v}\, \mathrm{d}t. \tag{7.25}$$

This can be used if $F(t)$ and $v(t)$ are known as a function of time.

7.3 Conservative & non-conservative forces

Suppose we want to move a mass from point A to B. There might be several paths we can take, as illustrated in Fig. 7.4b. Ask yourself: Is the total work different on different paths? The answer depends on the type of force: A force is called *conservative* if the work is the same for any path between two points. Otherwise, if the work is different for different paths, the force is called *non-conservative*.

So for a conservative force, the work W_a and W_b along paths a and b in Fig. 7.4b, respectively, is the same:

$$W_a = W_b. \tag{7.26}$$

This implies we can move back in a loop, such that the total work is zero!

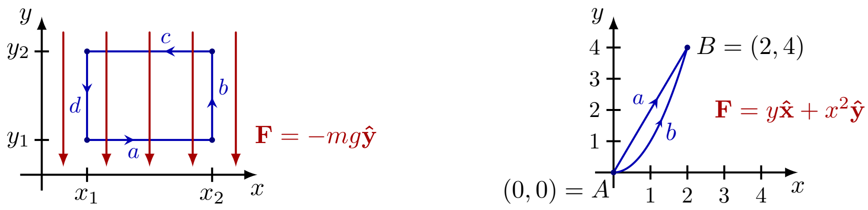
$$W_\mathrm{tot} = W_a - W_b = 0 \tag{7.27}$$

For any path:

Work of conservative force on a closed path.

$$W = \oint \mathbf{F} \cdot \mathrm{d}\mathbf{s} = 0. \tag{7.28}$$

The circle on the integral symbol indicates the path is closed. Let's make it a bit more concrete by looking at some examples.



(a) A closed path in a gravitational force field.

(b) A closed path in a non-conservative field.

Figure 7.5: Paths in force fields.

7.3.1 Example 1: Gravity

As discussed in Section 6.3, gravity provides a uniform vector field for a mass m:

$$\mathbf{F}_{\mathrm{g}} = -mg\hat{\mathbf{y}}. \tag{7.29}$$

It is constant in size and direction everywhere. Suppose you follow a closed path as in Fig. 7.5a. It has four segments: moving left (a), up (b), right (c) and down (d). It is easy to compute the work performed by gravity in this rectangular loop. We can break up the loop integral into its four segments:

$$W = \oint \mathbf{F} \cdot \mathrm{d}\mathbf{s} \tag{7.30}$$

$$= \int_{x_1}^{x_2} \mathbf{F} \cdot \hat{\mathbf{x}}\, \mathrm{d}x + \int_{y_1}^{y_2} \mathbf{F} \cdot \hat{\mathbf{y}}\, \mathrm{d}y + \int_{x_2}^{x_1} \mathbf{F} \cdot \hat{\mathbf{x}}\, \mathrm{d}x + \int_{y_2}^{y_1} \mathbf{F} \cdot \hat{\mathbf{y}}\, \mathrm{d}y\,. \tag{7.31}$$

The horizontal segments, a and c, are perpendicular to gravity, so do not contribute. Only the vertical ones remain, but cancel each other:

$$W = -mg(y_2 - y_2) - mg(y_1 - y_2) = 0. \tag{7.32}$$

So the total work done on a mass making a loop is zero. Therefore, gravity is a conservative force.

7.3.2 Example 2: $F = y\hat{\mathbf{x}} + x^2\hat{\mathbf{x}}$

Now consider the vector field

$$F = y\hat{\mathbf{x}} + x^2\hat{\mathbf{y}}. \tag{7.33}$$

Is this force conservative? Let's look at an arbitrary loop in the xy plane. Consider path a following $y = 2x$, and path b following $y = x^2$. They both start at $A = (0, 0)$ and meet in $B = (4, 2)$. To compute the work, one can integrate over each component separately, using the fact that $\mathbf{ds} = \mathrm{d}x\hat{\mathbf{x}} + \mathrm{d}y\hat{\mathbf{y}}$, so we get two terms for the work in the x direction and in the y direction:

$$W = \int_0^2 y\, \mathrm{d}x + \int_0^4 x^2\, \mathrm{d}y$$

First compute the work along path a by rewriting the integral using the constraint $y = 2x$ or $x = y/2$:

$$W_a = \int_0^2 2x\, \mathrm{d}x + \int_0^4 \left(\frac{y}{2}\right)^2 \mathrm{d}y = \frac{28}{3}.$$

Compare to the work along path b with relation $y = x^2$:

$$W_b = \int_0^2 x^2 \,\mathrm{d}x + \int_0^4 y \,\mathrm{d}y = \frac{32}{3}.$$

They are not the same! The work done depends on the path that was taken. We have proven by counterexample that this force is not conservative.

7.4 Potential energy

When defining work, we mentioned that energy is transferred to the moved object. In fact, for conservative forces, you can "store" energy in an object, which can be reused later. Notice that potential energy is negative of the work done by the force. This type of energy is called *potential energy*.

Potential energy.

$$\Delta U = U_\mathrm{f} - U_\mathrm{i} \tag{7.34}$$

$$= -W = -\int_{s_1}^{s_2} \mathbf{F} \cdot \mathrm{d}\mathbf{s} \tag{7.35}$$

7.4.1 Gravitational potential energy

Gravity is the classic example. When Arnold lifts up the mass in Fig. 7.2, gravity as a conservative fore will do negative work, so it will *add* positive gravitational potential energy to the mass. Arnold lifted the mass from a lower to a higher potential. The difference, from the formula for potential energy, is

$$\Delta U = -\int_0^h (-mg)\mathrm{d}y = mgh. \tag{7.36}$$

The minus sign in the integrand indicates that the gravity opposes the vertical displacement. Like the example in Section 7.3.1, only the height difference h is important.

Gravitational potential energy difference.

$$\Delta U = mgy_2 - mgy_1 = mgh, \tag{7.37}$$

where $h = y_2 - y_1$. This is also clear by performing an indefinite integral;

$$U(y) = \int mg\mathrm{d}y = mgy + U_0, \tag{7.38}$$

where U_0 is an integration constant. In a nice box:

Gravitational potential energy.

$$U(y) = mgy + U_0, \tag{7.39}$$

This constant U_0 is somewhat arbitrary and can be set to anything you like. In the problem with Arnold, it is natural to choose the ground for setting

$$U_0 = U(0) = 0. \tag{7.40}$$

Here, $y = 0$ is our chosen *reference level* where we set $U(0) = 0$. It is clear that $U(y) > 0$ is positive as long as $y > 0$, and negative if $y < 0$.

We can calculate the potential energy more generally for a gravitational force. Remember the force between two masses is

$$\mathbf{F}_\mathrm{g}(r) = -G\frac{mM}{r^2}\hat{\mathbf{r}}. \tag{7.41}$$

The only interesting direction is radial, and clearly

$$\hat{\mathbf{r}} \cdot \mathrm{d}\mathbf{s} = \mathrm{d}r. \tag{7.42}$$

The work a mass does to attract another mass from distance r_1 to r_2 by means of gravity is positive if $r_2 < r_1$ because the force and displacement are aligned:

$$W = \int_{r_1}^{r_2} G\frac{mM}{r^2}\mathrm{d}r = G\frac{mM}{r_1} - G\frac{mM}{r_2}. \tag{7.43}$$

Note this result also holds for $r_2 > r_1$, when the work becomes negative. It is very convenient to set a reference point r_1 at infinity, such that the first term disappears:

$$\lim_{r\to\infty} G\frac{mM}{r} = 0. \tag{7.44}$$

So for this common choice,

Gravitational potential energy.

$$U(r) = -G\frac{mM}{r}. \tag{7.45}$$

Notice that this form of the potential energy is always negative, because it is always *lower* than the reference level at infinity.

7.4.2 Spring energy

As discussed in Section 6.5, the force by a spring is

$$\mathbf{F} = -kx\hat{\mathbf{x}}. \tag{7.46}$$

The potential energy to move a mass from the rest length $x = 0$ to an extension x is

$$U = -\int_0^x (-kx)\,\mathrm{d}x = \frac{1}{2}kx^2. \tag{7.47}$$

We choose a reference level $U(0) = 0$ for when the spring has zero extension or compression, $x = 0$, which is a natural choice.

Potential energy of a spring.

$$U(x) = \frac{1}{2}kx^2, \tag{7.48}$$

7.5 Energy conservation

If we consider all sources of energy, then the sum of energies is conserved before and after any situation.

Law of conservation of energy. *Energy can neither be created nor destroyed, only converted from one form of energy to another:*

$$E_{before} = E_{after}. \tag{7.49}$$

If there are only conservative forces in play, we only have to consider *mechanical energy*, the sum of kinetic and potential energy:

Conservation of mechanical energy (for conservative forces).

$$(K + U)_{before} = (K + U)_{after}. \tag{7.50}$$

7.5.1 Example 1: Ramp

Consider dropping a mass from rest at a height H on a frictionless ramp. The mass will gain speed and jump off at a height $h < H$, just like in Fig. 7.6. What is the final velocity $\mathbf{v}$ at the jump-off point? First write down the energy *before* and *after*:

$$K_\mathrm{i} + U_\mathrm{i} = K_\mathrm{f} + U_\mathrm{f} \tag{7.51}$$

$$0 + mgH = \frac{mv^2}{2} + mgh. \tag{7.52}$$

The initial kinetic energy is zero if it starts at rest. Notice the final kinetic energy equals the difference in potential energy,

$$K_\mathrm{f} = \frac{mv^2}{2} = mg(H - h). \tag{7.53}$$

The final velocity does not depend on the mass, which cancels out:

$$v = \sqrt{2g(H - h)}. \tag{7.54}$$

This is the velocity's magnitude. The angle of the velocity depends on the shape of the ramp.

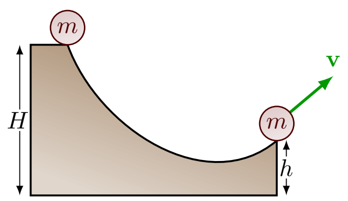
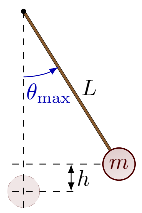
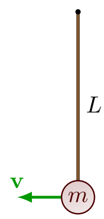


Figure 7.6: A closed path in a gravitational force field.



(a) Pendulum released from rest at an angle θ_{max}.



(b) Pendulum has the maximum speed at $\theta = 0$.

Figure 7.7: Pendulum.

7.5.2 Example 2: Pendulum

Pendulums are a favorite example for physicists. Suppose a pendulum with mass m is released from rest at an angle $\theta = \theta_{\text{max}}$. What is its speed at $\theta = 0$? Solving this problem with forces and our kinematic equations is cumbersome, since the acceleration is not constant. The important component is the tangential one due to gravity, which depends on the angle θ. Instead, energy conservation offers a quick solution. Comparing before at $\theta = \theta_{\text{max}}$ and after at $\theta = 0$:

$$K_{\text{i}} + U_{\text{i}} = K_{\text{f}} + U_{\text{f}} \tag{7.55}$$

$$0 + mgh = \frac{mv^2}{2} + 0, \tag{7.56}$$

From the triangle in Fig. 7.7a, we see that

$$h = L - L\cos\theta_{\text{max}}, \tag{7.57}$$

so the velocity at $\theta = 0$ is

$$v = \sqrt{2gL(1 - \cos\theta_{\text{max}})}. \tag{7.58}$$

Since all the potential energy is converted into kinetic energy at $\theta = 0$, this will be where the maximum speed is reached.

7.6 Energy loss due to friction

Friction is not a conservative force. When present, it always opposes the movement of an object, so it always performs negative work. This causes a loss in energy E_{loss}. Energy conservation still holds in the universe as a whole, but the energy is *dissipated* into a new, "unusable" form of energy, like heat or sound.

For example, if the ramp in the last section had some friction, we would rewrite the law of energy conservation as

Energy conservation with friction.

$$(K + U)_{before} = (K + U)_{after} + E_{loss}. \tag{7.59}$$

So the mass-ramp system has more mechanical energy before than after:

$$(K + U)_{\text{before}} > (K + U)_{\text{after}}. \tag{7.60}$$

If we only look at the mass-ramp system, mechanical energy is not conserved, but if we factor in the energy that was dissipated to the surroundings in other forms, energy conservation does still hold.

In a simple case where a mass moves a distance Δx over a surface, the work done by friction is

$$W = -\mu_{\mathrm{k}} F_{\mathrm{N}} \Delta x. \tag{7.61}$$

So the energy loss is

Energy loss due to friction.

$$E_{loss} = \mu_{\mathrm{k}} F_{\mathrm{N}} \Delta x. \tag{7.62}$$

Recall the mass on an inclined plane in Section 6.8.1. If it is let go from rest at a height h and there is no friction, the energy before equals the energy after:

$$mgh = \frac{mv^2}{2}, \tag{7.63}$$

and the final velocity is similar to that of the mass from the ramp:

$$v = \sqrt{2gh}. \tag{7.64}$$

But if the surface has some friction, there will be an energy loss:

$$E_{\mathrm{loss}} = \mu_{\mathrm{k}} \left(mg\cos\theta\right)\left(\frac{h}{\sin\theta}\right) = \mu_{\mathrm{k}} mgh \cot\theta, \tag{7.65}$$

with cotangent $\cot = \cos\theta / \sin\theta$. The new condition for energy conservation is

$$mgh = \frac{mv^2}{2} + \mu_{\mathrm{k}} mgh \cot\theta. \tag{7.66}$$

The final velocity is

$$v = \sqrt{2(1 - \mu_{\mathrm{k}} \cot\theta) gh}. \tag{7.67}$$

7.7 Power

When Arnold lifts up the mass, does the amount of work he does depend on how fast he does it? No. From the integral, we see that the total work only depends on the path taken, as long as the force is constant. So what is different?

We could define the integrand of work as

$$\mathrm{d}W = \mathbf{F} \cdot \mathrm{d}\mathbf{s}\,. \tag{7.68}$$

But remember from Eq. (7.18) that velocity is the time-derivative of $\mathbf{s}$, so we can write the displacement infinitesimal as

$$\mathrm{d}\mathbf{s} = \mathbf{v}\mathrm{d}t, \tag{7.69}$$

so as we did to derive the time-integral Eq. (7.25),

$$\mathrm{d}W = \mathbf{F} \cdot \mathbf{v}\mathrm{d}t. \tag{7.70}$$

This leads to a time derivative of work

Power.

$$P = \frac{\mathrm{d}W}{\mathrm{d}t} = \mathbf{F} \cdot \mathbf{v}, \tag{7.71}$$

called *power* P, which has units of *Watts*,

$$\mathrm{W} = \frac{\mathrm{J}}{\mathrm{s}} = \frac{\mathrm{kg} \cdot \mathrm{m}^2}{\mathrm{s}^3}. \tag{7.72}$$

Power is useful to quantify how much energy you spend per unit of time to perform some work. It can also be used to quantify how much energy you *lose*, for example via friction.

For example, if Arnold hoists up a mass of 1 kg to a height of 10 m in 10 s with a constant force exactly opposing gravity $F = mg$, his power output depends on the product of force and velocity :

$$P \approx 10\,\mathrm{W}. \tag{7.73}$$

The faster Arnold picks up the mass, the higher the power.

The typical incandescent light bulb uses 60 W of electrical power to produce light, most of which was "wasted" in the form of heat. Only about 5% is turned into visible light. Nowadays, energy-saving light bulbs need only about 15 W for the same light output, and the equivalent LED light would need only about 10 W. Table 7.1 lists some other typical power values of object and natural phenomena.

Table 7.1: Some typical power values of some objects or phenomena. Note that some of these numbers of objects include energy that is put out as heat, not just "useful" electrical or mechanical energy that can do work.

Object of phenomena	Power P [W]
Supernova at peak	5×10^{37}
The sun	4×10^{26}
Nuclear power plant	3×10^{9}
Car ($\sim$ 100 hp)	8×10^{4}
Clothes dryer	4000
Hair dryer	1200
Horsepower (hp)	735.5
Refrigerator	100–400
Desktop computer and monitor	200-400
Heat from person at rest	100
Incandescent light bulb	30–100
Human heart	10
Phone charger	2–6
Phone	0.1–0.5

Chapter 8

Conservation of Momentum

Suppose you have two trains colliding on a rail. Newton's third law states that their forces are equal, but opposite, so

$$F_{12} = -F_{21}, \tag{8.1}$$

or in terms of momentum, using Newton's second law,

$$\frac{\mathrm{d}p_1}{\mathrm{d}t} = -\frac{\mathrm{d}p_2}{\mathrm{d}t}. \tag{8.2}$$

Rewriting this, it's clear that the sum of momentum is constant with time,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(p_1 + p_2\right) = 0. \tag{8.3}$$

So the total momentum of the two trains is the same before and after the collision,

$$\left(p_1 + p_2\right)_{\text{before}} = \left(p_1 + p_2\right)_{\text{after}}. \tag{8.4}$$

or, assuming the trains remain in one piece,

$$\left(m_1 v_1 + m_2 v_2\right)_{\text{before}} = \left(m_1 v_1 + m_2 v_2\right)_{\text{after}}, \tag{8.5}$$

This is *conservation of (linear) momentum.* It holds no matter how complicated the forces are between them.

It is easy to generalize this to a bunch of masses interacting in three-dimensions. Say the total momentum of all interacting masses is $\mathbf{p}_{\text{tot}} = \sum \mathbf{p}_i$. Due to Newton's third law, all internal forces will cancel each other when we consider these masses as one *system.* So we only need to consider the total external force $\mathbf{F}$. Then Newton's second law says that the change of momentum in time is given by

$$\mathbf{F} = \frac{\mathrm{d}\mathbf{p}_{\text{tot}}}{\mathrm{d}t}, \tag{8.6}$$

or as an integral over some time period

$$\Delta\mathbf{p}_{\text{tot}} = \int_0^{\Delta t} \mathbf{F}\,\mathrm{d}t. \tag{8.7}$$

As long as there are no external forces, $\mathbf{F} = 0$, we call this an *isolated system*, and we see that the change in total momentum is zero:

$$\Delta\mathbf{p}_{\text{tot}} = 0. \tag{8.8}$$

So total momentum is conserved in an isolated system.

Law of conservation of linear momentum.

$$\left(\sum \mathbf{p}_i\right)_{\mathit{before}} = \left(\sum \mathbf{p}_i\right)_{\mathit{after}}. \tag{8.9}$$

This holds assuming the ensemble of masses is isolated: There is no interaction with some external force, and none of the masses leave the system, and no new masses enter.

If there are multiple external forces acting on different masses inside the system, but they still add up to zero in total, $\mathbf{F} = 0$, then momentum will still be conserved, but there could be an overall change in "rotation". We will see this in more detail in the next chapter on torque and angular momentum (as opposed to linear momentum).

Notice that the momentum is a vector, so momentum is conserved in each of the three spatial direction independently, so we could have three equations for momentum conservation. Most examples of collisions we will see involve only two masses colliding, and so we can often reduce the problem to two dimensions.

Now, sometimes things break or merge after a collision, so the number of masses before and after can change. However, in an isolated system the total mass before and after will be the same, as matter can not be created nor destroyed,[1]

$$\left(\sum m_i\right)_{\text{before}} = \left(\sum m_i\right)_{\text{after}}. \tag{8.10}$$

So for a isolated system with constant mass, we can rewrite Eq. (8.9) in terms of the masses and their respective velocities:

Conservation of linear momentum of a group of masses.

$$\left(\sum m_i \mathbf{v}_i\right)_{\mathit{before}} = \left(\sum m_i \mathbf{v}_i\right)_{\mathit{after}}. \tag{8.11}$$

8.1 Elastic & inelastic collisions

In elastic collisions, both the total mechanical energy and momentum are conserved:

Elastic collision.

$$\begin{cases} (K+U)_{\mathit{before}} = (K+U)_{\mathit{after}} & (8.12) \\ \left(\sum \mathbf{p}_i\right)_{\mathit{before}} = \left(\sum \mathbf{p}_i\right)_{\mathit{after}} & (8.13) \end{cases}$$

Notice that in three dimensions, this will give you four equations. For a two dimensional problem, it will just be three. If the total potential energy is unchanged before and after the collision, then

$$U_{\text{i}} = U_{\text{f}}, \tag{8.14}$$

and the total kinetic energy must also be unchanged before and after,

$$K_{\text{i}} = K_{\text{f}}. \tag{8.15}$$

[1] *Conservation of mass* is assumed in classical physics, but since Einstein's theory of special relativity we know this is no longer the case. Mass can be converted into energy and vice versa. However, momentum is still conserved, but simply in a different form: Massless particles like light particles (photons) have momentum. This fact is relevant in nuclear and particles physics.

We will see a lot of examples of this case.

In inelastic collisions, some energy gets lost in the form of heat, sound and deformation. In this case, momentum is still conserved, but mechanical energy is not.

Inelastic collision.

$$\begin{cases} (K+U)_{before} \neq (K+U)_{after} & (8.16) \\ \left(\sum \mathbf{p}_i\right)_{before} = \left(\sum \mathbf{p}_i\right)_{after} & (8.17) \end{cases}$$

8.1.1 Example 1: Two masses colliding elastically in 1D

Consider the masses about to collide in Fig. 8.1a. If the collision is elastic and there is no potential energy to be considered, then their total kinetic energy and linear momentum are conserved:

$$\begin{cases} \frac{m_1 v_1^2}{2} + \frac{m_2 v_2^2}{2} = \frac{m_1 v_1'^2}{2} + \frac{m_2 v_2'^2}{2} & (8.18) \\ m_1 v_1 + m_2 v_2 = m_1 v_1' + m_2 v_2' & (8.19) \end{cases}$$

where prime $'$ indicates the velocities after the collision. Given the masses m_1 and m_2, and initial velocities v_1 and v_2, we have two unknowns, the final velocities v_1' and v_2', which can be solved for because there are two independent equations.

Notice we have not included negative signs to indicate some direction before and after the collision. In Fig. 8.1a, the initial velocity of mass 1 is negative, $v_1 < 0$, and that of mass 2 is positive, $v_2 > 0$, if we define the positive x direction to the left. After the collision in Fig. 8.1b, $v_1' > 0$ and $v_2' < 0$. It is important to know that it does not matter if you assume the velocity v_i is negative, or instead put a negative sign in the equation for momentum equation, as long as you are consistent.

One of the simplest case of this problem is when $m_1 = m_2$ and $v_1 = -v_2$. Then all we need is conservation of momentum,

$$0 = mv_1 + mv_2 = mv_1' + mv_2', \tag{8.20}$$

to find that the final velocities are $v_2' = -v_1'$.

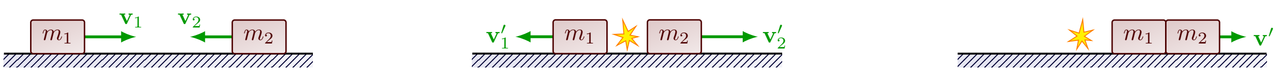
But what if $m_2 = 2m_1$ and $v_2 = -v_1/2$, such that the total momentum is zero?

$$0 = mv_1 + 2mv_2 = mv_1' + 2mv_2', \tag{8.21}$$

and

$$v_2' = -\frac{v_1'}{2}. \tag{8.22}$$

Using conservation of kinetic energy, we can also find v_1' given m and one of the intitial velocities.



(a) Two masses approach before head-on collision.

(b) After an elastic collision, mechanical energy and momentum are conserved.

(c) After a fully inelastic collision, the masses stick together.

Figure 8.1: One-dimensional collision of two masses.

8.1.2 Example 2: Two masses colliding inelastically in 1D

Let's look at the same problem, but now assume that the collision is *fully inelastic*. Fully would mean that after the collision the masses stick together as in Fig. 8.1c. Luckily, momentum is still conserved, so we can write

$$m_1 v_1 + m_2 v_2 = (m_1 + m_2) v', \tag{8.23}$$

where v' is the final velocity of the merged mass $m_1 + m_2$. In this simple case, there is only one unknown, v', so we will not miss our energy equation. The solution for v' given the initial masses and velocities is

$$v' = \frac{m_1 v_1 + m_2 v_2}{m_1 + m_2}. \tag{8.24}$$

8.1.3 Example 3: Walking on ice

Conservation of momentum does not always involve collisions. Suppose Brian is standing on a plank on frictionless ice. At first, both he and the plank are at rest. When he starts walking, he pushes the the plank backwards. Without outside forces like friction, momentum is conserved. At the beginning, the total momentum is zero, so once Brian starts walking with some velocity v_B, the plank has to compensate:

$$0 = m_\mathrm{p} v_\mathrm{p} + m_\mathrm{B} v_\mathrm{B}. \tag{8.25}$$

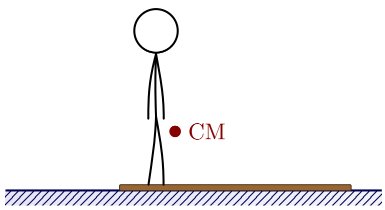
So,

$$v_\mathrm{p} = \frac{m_\mathrm{B}}{m_\mathrm{p}} v_\mathrm{B}. \tag{8.26}$$

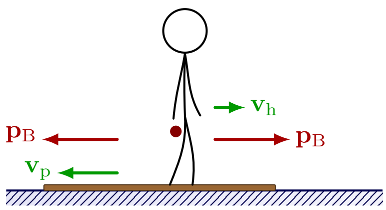
8.1.4 Example 4: Two masses colliding inelastically in 2D

Let's look at a collision in two dimensions. The simplest case is a fully inelastic one. Consider to cars colliding at an intersection. After the collision the get stuck together, and ignoring friction completely for now, they keep moving with some constant velocity $\mathbf{v}'$, as in Fig. 8.3. With this choice of x and y axes, the initial momenta before and after are

$$\begin{cases} \mathbf{p} = m_1 v_1 \hat{\mathbf{x}} + m_2 v_2 \hat{\mathbf{y}} & (8.27) \\ \mathbf{p}' = (m_1 + m_2) v'_x \hat{\mathbf{x}} + (m_1 + m_2) v'_y \hat{\mathbf{y}} & (8.28) \end{cases}$$



(a) Before: Brian and the plank are at rest. The total momentum is zero.



(b) After: Brian starts walking, and the plank "recoils". The total momentum is still zero.

Figure 8.2: Brian stands on a wooden plank on frictionless ice. The center of mass (Section 8.3) is closest to Brian's center, who is much heavier than the plank. As Brian walks, the center of mass stays constant in space.

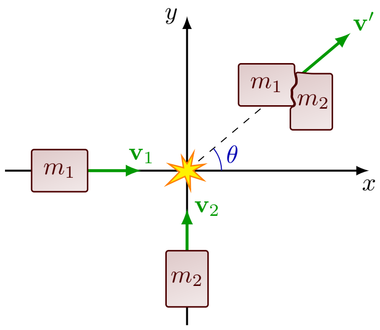


Figure 8.3: Two cars collide at an intersection and get stuck together.

So invoking momentum of conservation for each component,

$$\begin{cases} m_1 v_1 = (m_1 + m_2)v'_x & (8.29) \\ m_2 v_2 = (m_1 + m_2)v'_y & (8.30) \end{cases}$$

So the final velocity if given by

$$\mathbf{v}' = \frac{m_1}{m_1 + m_2} v_1 \hat{\mathbf{x}} + \frac{m_2}{m_1 + m_2} v_2 \hat{\mathbf{y}}. \tag{8.31}$$

8.1.5 Example 5: Ballistic pendulum

Another example of a fully example we have seen in class. We shoot a bullet at high velocity into a block suspended from the ceiling by a string. There are two parts to this problem: conservation of momentum in the inelastic collision, followed by conservation of energy when the block-plus-bullet swing upward. In the first part, the total momentum before and after is given by the momentum of the bullet.

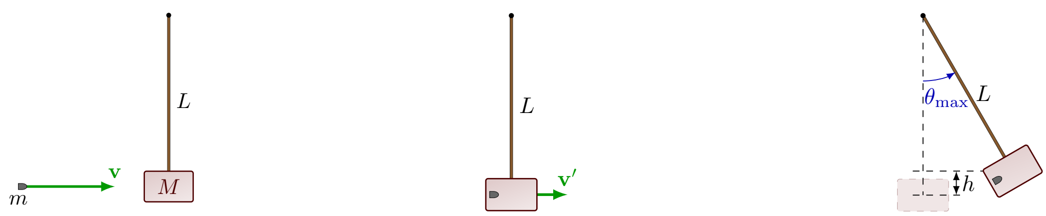
$$mv = (m + M)v'. \tag{8.32}$$

The bullet may be very light, but has a very high velocity, so it has a considerable momentum. This gives us the final velocity

$$v' = \frac{m}{m + M} v. \tag{8.33}$$

In the next part, energy is conserved, and we can follow our derivation in Section 7.5.2. The total mechanical energy before the block moves and after the block reaches its highest point is

$$\frac{(m + M)v'^2}{2} = (m + M)gh. \tag{8.34}$$



(a) Before: Bullet is shot at a pendulum at rest.

(b) After: Bullet gets stuck into the block.

(c) Later: Pendulum reaches is highest point $\theta = \theta_{\max}$.

Figure 8.4: Ballistic pendulum: A bullet is shot into a block suspended from the ceiling. This is an inelastic collision.

So, the maximum height is given by

$$h = \frac{v'^2}{2g} = \left(\frac{m}{m+M}\right)^2 \frac{v^2}{2g}. \tag{8.35}$$

8.2 Impulse

Newton's second law says that a force acting on a mass changes its momentum:

$$\mathbf{F} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}. \tag{8.36}$$

The force causes an acceleration. The change in momentum depends on how long the force acts on the mass. If the force is constant,

$$\Delta\mathbf{p} = \mathbf{F}\mathrm{d}t. \tag{8.37}$$

However, the forces involved in a collision are often more complicated than just a constant force. Typically, the force changes continuously with time. In that case, the change of momentum is given by

Impulse.

$$I = \Delta\mathbf{p} = \int_{t_1}^{t_2} \mathbf{F}(t)\mathrm{d}t. \tag{8.38}$$

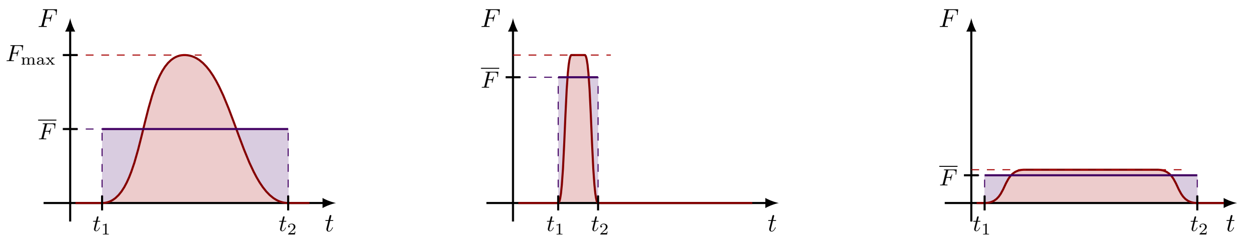
which is called an *impulse* I.

The impulse is the area under the F-t graph, as in Fig. 8.5a. The time-averaged force $\overline{F}$ gives the same impulse as a constant force, which is useful for simplifying problems.

$$I = \overline{F}\Delta t. \tag{8.39}$$

If a material is soft, it can deform under a force. This typically takes a bit of time, so inelastic materials like pillows and bouncy balls tend to "soften" the blow of a hard force.

This is why cars have airbags. Without an airbag, your face with slam very hard into the steering wheel in a very short time (Fig. 8.5b). An airbag will quickly inflate to catch your head before it hits the steering wheel, and then deflate to apply a smaller force over a relatively longer period of time in order to minimize your injury (Fig. 8.5c). However, airbags also have a large surface area, so in addition, it will spread out the force over your whole body, and not just you head.



(a) Force changes with time. **(b)** Short, hard hit, like a car collision. **(c)** Long, soft hit, like an airbag.

Figure 8.5: Impulse is the integral of force versus time. It can be approximated with a constant force, that is the time-averaged force $\overline{F}$.

8.3 Center of mass

Consider a bomb at rest as in Fig. 8.6b. When it explodes, the fragments go in all directions like in Fig. 8.6c, but assuming no external forces, the total momentum, which is zero, is conserved:

$$0 = \sum p_i = \sum m_i \frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t}, \tag{8.40}$$

where m_i are the bomb fragments after the explosion. Assuming the masses are constant,

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum m_i \mathbf{r}_i\right). \tag{8.41}$$

Here, $m_i\mathbf{r}_i$ is called the *moment* of mass m_i. So the sum of moments is constant in time. This leads to the definition of the *center of mass* (CM), which is this sum normalized by the total mass $M = \sum m_i$.

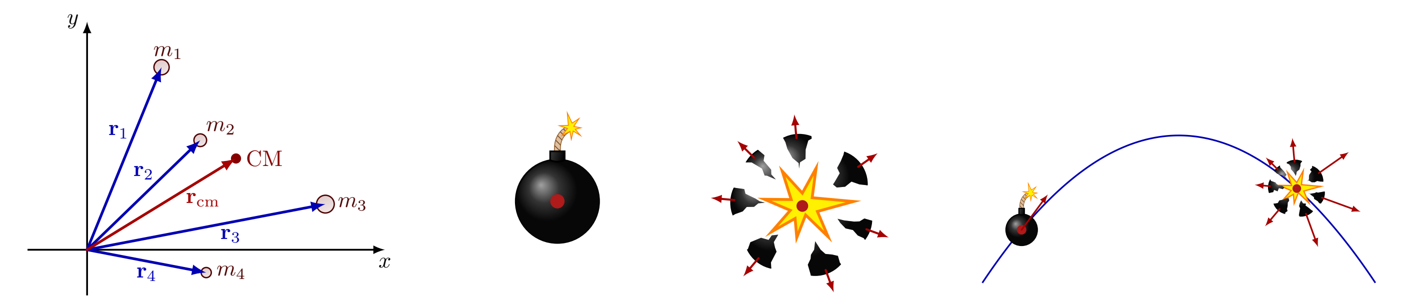
Center of mass.

$$\mathbf{r}_{\mathrm{cm}} = \frac{\sum m_i \mathbf{r}_i}{\sum m_i}. \tag{8.42}$$

The center of mass is like the "averaged position" of mass.

If there are no external forces, such that momentum is conserved, then the center of mass will always be in the exact same point in space. Notice that the point is constant, but the vector $\mathbf{r}_{\mathrm{cm}}$ does depend on the choice of coordinate system. In case of Brian walking on the plank in Section 8.1.4 and Fig. 8.2, the center of mass is also constant. The center of mass of a human standing straight is typically just under the navel, but once they start moving their limbs, the center of gravity will shift. This is an important element in the physics of sports like martial arts and dance.

Similarly, we can define a center-of-mass velocity and momentum



(a) A system of masses in two dimensions.

(b) A bomb at rest an its CM.

(c) The fragments have the same CM.

(d) The center of mass of the fragments follows the same trajectory.

Figure 8.6: Center of mass (red dot) is like the mass-averaged position of a group of masses, or continuous body. If momentum is conserved, the center of mass is constant in space. If there is a non-zero external force, the center of mass will move as if all the mass is concentrated in it and the external force act only on it.

Center-of-mass velocity and momentum.

$$\mathbf{v}_{\text{cm}} = \frac{\mathrm{d}\mathbf{r}_{\text{cm}}}{\mathrm{d}t} = \frac{\sum m_i \mathbf{v}_i}{\sum m_i} \tag{8.43}$$

$$\mathbf{p}_{\text{cm}} = \sum m_i \mathbf{v}_{\text{cm}} = \sum m_i \mathbf{v}_i. \tag{8.44}$$

But what if you throw the bomb in a gravity field instead (Fig. 8.6d)? Momentum is not conserved anymore. However, it turns out that the center of mass will still follow a parabola, as it would if the bomb never exploded! To see this, notice that treating the fragments as one system, the total force is

$$\mathbf{F} = \frac{\mathrm{d}}{\mathrm{d}t} \sum \mathbf{p}_i = \frac{\mathrm{d}\mathbf{p}_{\text{cm}}}{\mathrm{d}t}, \tag{8.45}$$

so there will be an acceleration on the center of mass

$$\mathbf{F} = \frac{\mathrm{d}}{\mathrm{d}t} \sum m_i \mathbf{a}_i = M \frac{\mathrm{d}\mathbf{a}_{\text{cm}}}{\mathrm{d}t}, \tag{8.46}$$

defined in a similar way as the center-of-mass velocity. So to summarize:

The center of mass of a system of masses m_i moves in the same way as a single point with mass $M = \sum m_i$ moves under a force $\mathbf{F} = \sum \mathbf{F}_i$, where $\mathbf{F}_i$ is the total force on mass m_i.

The center of mass of a spherical planet like the Earth is in its center by geometrical symmetry. This is why in a lot of problems, we can simplify the Earth as a mass point.

The definition of the center of mass can be extended for a continuous body employing integrals, which is the limit of the sum over infinitesimal small masses.

Center of mass of a continuous body.

$$\mathbf{r}_{\text{cm}} = \frac{\int \mathbf{r}\,\mathrm{d}m}{\int \mathrm{d}m}, \tag{8.47}$$

where the denominator is the total mass of the body. To solve this integral, the infinitesimal $\mathrm{d}m$ often can be rewritten in terms of the mass density $\rho(\mathbf{r})$ and Cartesian (x, y, z) or spherical coordinates (r, θ, ϕ) in three dimensions, or polar (r, θ) in two dimensions. We will see some examples of such integrals over $\mathrm{d}m$ in the next chapter.

8.3.1 Example: Two mass in 1D

Consider the two masses m_1 and m_2 in Fig. 8.7. What is the center of mass? First set the origin $x = 0$ in the middle between the masses. Then the center of mass is given by

$$x_{\text{cm}} = -\frac{d}{2} m_1 + \frac{d}{2} m_2 = \frac{d}{2}(m_2 - m_1). \tag{8.48}$$

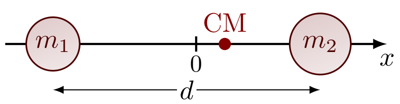


Figure 8.7: Center of mass (red dot) of two masses in one dimension.

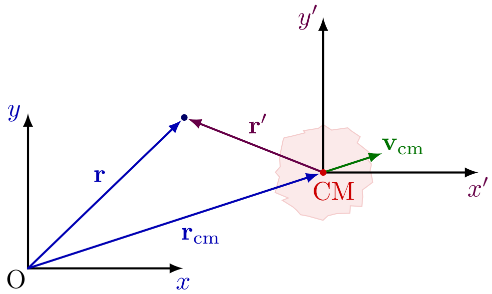


Figure 8.8: The center-of-mass frame moves with a constant velocity $\mathbf{v}_{\text{cm}}$ in the lab frame. A point in the center of mass frame has a different position vector $\mathbf{r}'$ than in the lab frame, $\mathbf{r}$).

8.4 Inertial frames of reference

In most problems we have seen so far, we always chose a coordinate system that was not moving, a frame of reference that is "fixed" with respect to its surroundings. This is called the *laboratory frame of reference*, because the "laboratory" where the experiment is performed is at rest.

However, sometimes it is convenient to choose our coordinate system such that the center of mass in this frame is fixed to zero:

Center-of-mass frame of reference.

$$\mathbf{r}'_{\text{cm}} = 0. \tag{8.49}$$

Using this trick, the total momentum will be zero in this frame of reference for an isolated system:

$$\mathbf{p}'_{\text{cm}} = \sum m_i v_i = 0. \tag{8.50}$$

A coordinate system with the above property is called the *center-of-mass frame of reference*. From Fig. 8.8, we can see that the new position vector in the center-of-mass frame can be written as

$$\mathbf{r}' = \mathbf{r} - \mathbf{r}_{\text{cm}}, \tag{8.51}$$

where the position vector $\mathbf{r}$ and center-of-mass $\mathbf{r}_{\text{cm}}$ are defined with respect to the origin of the lab frame. The prime symbol $'$ indicates that the $\mathbf{r}'$ is defined in a frame other than the lab frame, pointing from the origin of the center-of-mass frame (which is the center of mass $\mathbf{r}'_{\text{cm}} = 0$). It is not to be confused with the first derivative.

Often, the center-of-mass frame is moving with respect of the lab frame. If the center-of-mass frame started at the lab frame's origin $\mathbf{r}_{\text{cm}} = 0$ at time $t = 0$ and is moving at a constant velocity $\mathbf{v}_{\text{cm}}$, then its origin is given by $\mathbf{r}_{\text{cm}} = \mathbf{v}_{\text{cm}} t$ in the lab frame. The transformation of the position vector going from the center-of-mass frame to lab frame is given by:

Galilean transformation (to a center-of-mass frame).

$$\mathbf{r}' = \mathbf{r} - \mathbf{v}_{\text{cm}} t. \tag{8.52}$$

If the center-of-mass velocity is only in the x direction for simplicity, $\mathbf{v}_{\text{cm}} = v_{\text{cm}}\hat{\mathbf{x}}$, the coordinates of the position vector in the center-of-mass frame are given by

$$\begin{cases} x' = x - v_{\text{cm}}t & (8.53) \\ y' = y & (8.54) \\ z' = z & (8.55) \end{cases}$$

This is a called a *Galilean transformation*, and can be used to relate the coordinates between any two frames of reference that are moving relative to each other with some constant velocity $\mathbf{v}$. In this section, we focused on the center-of-mass frame with $\mathbf{v} = \mathbf{v}_{\text{cm}}$.

We will see in Section 10 what happens if a frame of reference is accelerated.

Chapter 9

Torque & Angular Momentum

9.1 Torque

To get a wheel rotating around its axis, you need to apply a force somewhere on the wheel. To describe how it starts rotating, we need to know the force $\mathbf{F}$, but also where on the wheel it acts. This is encapsulated by the concept of *torque*

Torque.

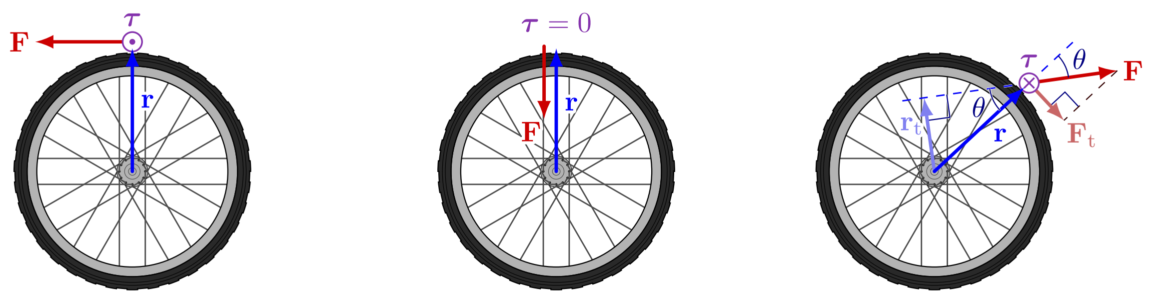
$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}, \tag{9.1}$$

where $\mathbf{r}$ points from the origin to the point where $\mathbf{F}$ acts, as in Fig. 9.1. Because this is a product involving a distance to a reference point, it is also sometimes called *moment of force*. From Section 3.6, we know the length of the cross product is

$$\tau = rF\sin\theta, \tag{9.2}$$

where θ is the smallest angle between the $\mathbf{F}$ and $\mathbf{r}$ vectors. The torque vector $\boldsymbol{\tau}$ is perpendicular to both the $\mathbf{F}$ and $\mathbf{r}$, and its direction of the torque is given by the right-hand rule (Fig. 3.4).

One important thing to note here, is that position vector $\mathbf{r}$ appears in Eq. (9.1). Clearly, the definition of torque depends on the chosen origin. In two reference frames with different origins, you will have a different torque. However, there is again a natural choice, which is putting the origin on the axis of rotation. In the case of a flat wheel, this choice is



(a) Force and position vector are perpendicular. The torque $\boldsymbol{\tau}$ points out the paper.

(b) Force and position vector are parallel. The torque $\boldsymbol{\tau}$ is zero, because $\sin\theta = 0$

(c) Force and position vector are make an angle θ. The torque $\boldsymbol{\tau}$ points into the paper.

Figure 9.1: Torque $\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$ on a bicycle wheel by a force acting on the tyre. The axis of rotation is fixed to the wheel's own axis.

putting the origin in the wheel's center that is assumed to be fixed. If you apply a force perpendicular to the position vector and in the plane of rotation (i.e. perpendicular to the axis or rotation as well), as in Fig. 9.1a, you will apply maximum torque for some force ($\sin\theta = 1$), and cause maximum rotation. If you instead just push on the wheel directly towards the axis of rotation, there is no torque ($\sin\theta = 0$), and nothing will happen, as in Fig. 9.1b. More generally, the force can make some angle pushing or pulling the wheel at some point, as in Fig. 9.1c. In this case, only the tangential component $\mathbf{F}_\mathrm{t}$ that is perpendicular to the axis of rotation and position vector is important. Its magnitude is exactly $F_\mathrm{t} = F\sin\theta$. Another way to consider torque is to consider the component of the position vector that is perpendicular to the force vector. This $\mathbf{r}_\mathrm{t}$ is called the *lever arm*. The size of the torque can thus be written in either way:

$$\tau = r_\mathrm{t}F = rF_\mathrm{t} \tag{9.3}$$

Torque has units of N m, which is the same units as energy. But do not think of torque as an energy, think of it as the rotational analogue of force.

9.2 Angular acceleration

Let's look again at the circular motion of a mass as in Fig. 5.6. In Section 5.3 we considered uniform circular motion, where the angular velocity ω and radius r was constant:

$$\theta(t) = \theta_0 + \omega t. \tag{9.4}$$

What if the angle is being accelerated, due to a torque? Just like the angular velocity is the change of the angle in time,

$$\omega = \frac{\mathrm{d}\theta}{\mathrm{d}t}, \tag{9.5}$$

the *angular acceleration* α is the change of angular velocity in time,

$$\alpha = \frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{\mathrm{d}^2\theta}{\mathrm{d}t^2}. \tag{9.6}$$

If the acceleration and radius are constant, then we have a similar formula to linear motion with uniform acceleration (Eq. (4.12)):

Circular motion with uniform acceleration.

$$\begin{cases} \theta(t) = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2 & (9.7) \\ \omega(t) = \omega_0 + \alpha t. & (9.8) \end{cases}$$

The angular velocity is the same for every point on a rotating object, but the velocity depends on the how far the point is from the center of rotation,

$$v = r\omega. \tag{9.9}$$

Also, the distance a point on the edge of a rotating object travels depends on the angle it rotates, but also its distance from the center of rotation (see Section 5.2),

$$s = r\theta. \tag{9.10}$$

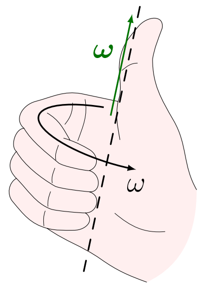
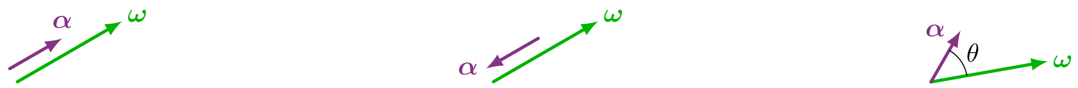


Figure 9.2: Right-hand rule for the direction of $\boldsymbol{\omega}$: Curl the fingers of your right hand around the axis of rotation, pointing in the direction of motion. Your right-hand thumb will point along $\boldsymbol{\omega}$.



(a) If $\boldsymbol{\alpha}$ aligns with $\boldsymbol{\omega}$, then $|\boldsymbol{\omega}|$ increases, but its direction stays constant. If $|\boldsymbol{\omega}|$ is positive, so is $|\boldsymbol{\alpha}|$.

(b) If $\boldsymbol{\alpha}$ is opposite to $\boldsymbol{\omega}$, $|\boldsymbol{\omega}|$ decreases, and $\boldsymbol{\omega}$ would eventually point in the opposite direction. At the start, if $|\boldsymbol{\omega}|$ is positive, $|\boldsymbol{\alpha}|$ is negative.

(c) $\boldsymbol{\alpha}$ does not align with $\boldsymbol{\omega}$, so the object would change the direction of its rotation if it is free to do so.

Figure 9.3: Angular velocity $\boldsymbol{\omega}$ and angular acceleration vector $\boldsymbol{\alpha}$. Compare this to linear velocity and acceleration Fig. 5.2.

Likewise, the acceleration of a point on the object relates to the angular acceleration *alpha* by

$$a = r\alpha. \tag{9.11}$$

It is useful to define the vectors $\boldsymbol{\omega}$ and $\boldsymbol{\alpha}$ for the angular velocity and angular acceleration, respectively. The angular velocity vector is conventionally defined to point along the axis of rotation, with the direction given by the right-hand rule in Fig. 9.2. The angular acceleration vector $\boldsymbol{\alpha}$ in turn, indicates a change in size and/or direction of $\boldsymbol{\omega}$, as illustrated in Fig. 9.3c: If $\boldsymbol{\omega}$ and $\boldsymbol{\alpha}$ align in the same direction, the angular velocity will increase, and decrease if $\boldsymbol{\omega}$ and $\boldsymbol{\alpha}$ anti-align.

9.3 Rotational equilibrium & moment of inertia

Let's look at a simple case where a force $\mathbf{F}$ is applied to a single point mass m that is constrained to move in a circle of constant radius r, as in Fig. 9.4. If the force is unbalanced, Newton's second law and Eq. (9.11) tell us that

$$F_{\mathrm{t}} = ma = mr\alpha. \tag{9.12}$$

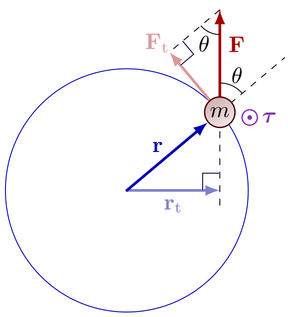


Figure 9.4: A force $\mathbf{F}$ acting on a single mass point m moving in a circle of constant radius r creates a torque $\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$.

Multiplying both sides by r, we retrieve the torque:

$$\tau = rF_t = mr^2\alpha, \tag{9.13}$$

or

$$\tau = I\alpha, \tag{9.14}$$

with the *moment of inertia* I of mass m at a distance r from the axis of rotation.

Moment of inertia of a single mass point.

$$I = mr^2 \tag{9.15}$$

The direction of the angular acceleration $\boldsymbol{\alpha}$ is given by the direction of the torque $\boldsymbol{\tau}$:

Newton's second law for rotation.

$$\boldsymbol{\tau} = I\boldsymbol{\alpha} \tag{9.16}$$

Notice the similarity to $\mathbf{F} = m\mathbf{a}$: Torque $\boldsymbol{\tau}$ acts as a force, moment of inertia I acts as inertial mass, and $\boldsymbol{\alpha}$ acts as the acceleration.

Just like with forces, several torques can act on a body at the same time. They may want to rotate the body in different directions. If we want to know the actual rotation that will happen in that case, we are interested in the *total torque*,

$$\sum \boldsymbol{\tau}_i = \sum (\mathbf{r}_i \times \mathbf{F}_i) = I\boldsymbol{\alpha}. \tag{9.17}$$

The torque may act on different parts of the body, as given by the position vector $\mathbf{r}_i$ So if all torques balance, there will be no rotation.

Rotational equilibrium.

$$\boldsymbol{\tau}_{tot} = \sum \boldsymbol{\tau}_i = 0. \tag{9.18}$$

Like with linear motion, you need to define an origin and a positive direction. Clockwise (cw) and counterclockwise (ccw) torques have opposite sign, so at equilibrium, they must balance:

$$\sum \boldsymbol{\tau}_i^{\text{cw}} = \sum \boldsymbol{\tau}_i^{\text{ccw}}. \tag{9.19}$$

Equation (9.15) is the moment of inertia I of a single mass point. What about more complicated massive bodies? Moment of inertia is actually an additive quantity, so for N mass points,

Moment of inertia of a group of points.

$$I = \sum_i^N m_i r_i^2. \tag{9.20}$$

Just like the center of mass, we can extend this to a continuous body,

Moment of inertia of a continuous body.

$$I = \int r^2 \mathrm{d}m. \tag{9.21}$$

Typically, the infinitesimal $\mathrm{d}m$ can be expressed in terms of some *mass volume density* $\rho(\mathbf{r})$. For a three-dimensional integral,

$$\mathrm{d}m = \rho(\mathbf{r})\mathrm{d}V, \tag{9.22}$$

where $\mathrm{d}V$ is an infinitesimal element of volume. The density ρ has units $\mathrm{kg\,m^{-3}}$. If you have a thin surface with some mass, we would compute a two-dimensional integral. We can use the area mass density $\sigma(\mathbf{r})$ with units $\mathrm{kg\,m^{-2}}$ instead,

$$\mathrm{d}m = \sigma(\mathbf{r})\mathrm{d}A, \tag{9.23}$$

where $\mathrm{d}A$ is is the infinitesimal element of surface area. Finally, in one dimension, we use the linear mass density $\lambda(\mathbf{r})$ with units $\mathrm{kg\,m^{-1}}$,

$$\mathrm{d}m = \lambda(\mathbf{r})\mathrm{d}x. \tag{9.24}$$

If the body is *homogeneous*, or *uniformly distributed*, then these density distributions are constant in space. For volume, this means that the volume mass density is the total mass divided by the total volume, $\rho = M/V$. Similarly, for a surface with total area A and total homogeneous mass M, $\sigma = M/A$, and for a line of total length L, $\lambda = M/L$.

Notice that moment of inertia is defined as a sum or integral. One interesting property that follows immediately, is that it is additive. If you combine two objects with moment of inertia I_1 and I_2, defined with respect to the same rotation axis, then the total moment of inertia is simply

$$I = I_1 + I_2. \tag{9.25}$$

9.3.1 Example 1: Balancing two masses on a seesaw

Consider the seesaw in Fig. 9.5, where two masses m_1 and m_2 are balanced on a plank. There are three forces: normal force $\mathbf{F}_\mathrm{N}$ of the pivot on the plank, and the gravitational forces $m_1\mathbf{g}$ and $m_2\mathbf{g}$ on the masses. The pivot point is the point of rotation, so the total torque is

$$\sum \boldsymbol{\tau}_i = \tau_1 - \tau_2 = r_1 m_1 g - r_2 m_2 g, \tag{9.26}$$

where we have defined the positive direction of the angle θ counterclockwise. In terms of Eq. (9.19),

$$r_1 m_1 g = r_2 m_2 g. \tag{9.27}$$

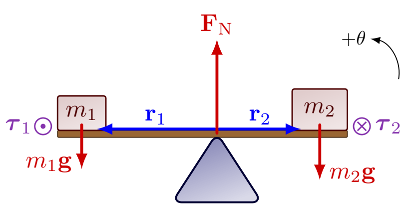
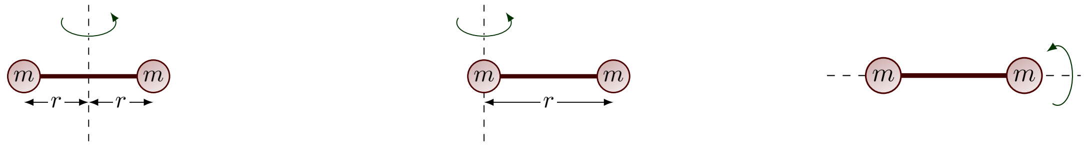


Figure 9.5: A seesaw: The torques on the masses on either end counteract. If they balance, there will be no rotation (if they start from rest). By convention, the positive direction of rotation is counterclockwise.



(a) Axis though the middle point between the masses.

(b) Axis though one of the masses.

(c) Axis though both masses.

Figure 9.6: Moment of inertia of two heavy masses m connected by a rod of negligible mass. The moment of inertia depends on the axis of rotation.

The normal force creates no torque with respect to the pivot point, as $\mathbf{r} = 0$. To balance, the torques must cancel, so the condition for equilibrium is

$$\frac{r_1}{r_2} = \frac{m_1}{m_2}. \tag{9.28}$$

This makes sense: If $m_1 = m_2$, then the distance of each mass to the pivot must be equal, $r_1 = r_2$. If one mass is twice as large, take $m_2 = 2m_1$, then it must be twice as close to the pivot to balance the other mass, $r_2 = r_2/2$.

We recognize the center of mass from Eq. (9.26):

$$r_{\text{cm}} = r_1 m_1 - r_2 m_2 = 0, \tag{9.29}$$

which is zero in case of rotational equilibrium. This is an example of an unstable equilibrium, which we will study in more detail later. Even though the forces balance if the center of mass is at the pivot, $r_{\text{cm}} = 0$, a small push will cause the seesaw to tip.

9.3.2 Example 2: Moment of inertia of two masses

But what if the torques do not balance, and the total torque is non-zero? Then there will be an angular acceleration $\boldsymbol{\alpha}$, and thus a rotation. To know $\boldsymbol{\alpha}$, we need to compute the moment of inertia of the seesaw.

Let's approximate the seesaw as a massless rod connecting two mass points m, as shown in Fig. 9.6a. Say the distance from the center to either masses is r, then the moment of inertia is simply the sum,

$$I = I_1 + I_2 \tag{9.30}$$
$$= mr^2 + mr^2 = 2mr^2. \tag{9.31}$$

The angular velocity then is given by

$$\alpha = \frac{\tau}{I}. \tag{9.32}$$

Suppose we move the masses closer to the middle, at a new distance $r' = r/2$, then the moment of inertia changes to

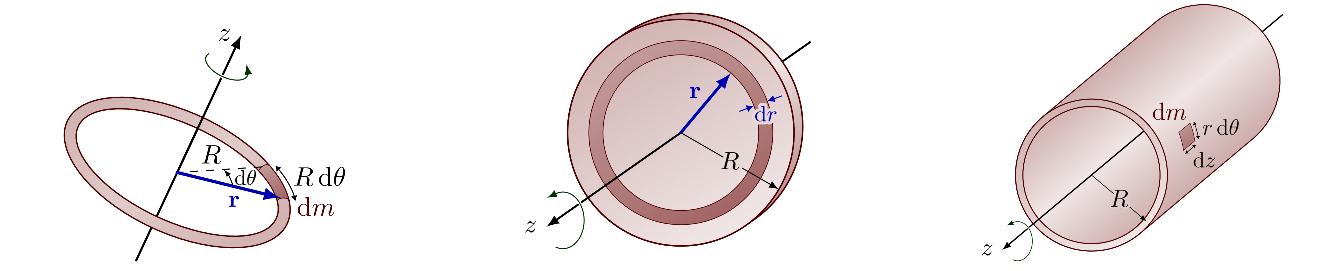
$$I' = mr'^2 = \frac{I}{4}. \tag{9.33}$$

So for a given torque τ, the angular acceleration of these two configurations will be different:

$$\alpha' = \frac{\tau}{I'} = 4\alpha_1. \tag{9.34}$$

This means that for the same torque, the angular acceleration will be larger the closer the masses are to the point of rotation.

This result is very similar to a ice skater pulling her arms to her body to spin faster. With arms spread out, the ice skater's moment of inertia is larger, and her speed of rotation is less, than when they are close to her body.



(a) Homogeneous ring with radius R, mass M.

(b) Homogeneous hollow cylinder with radius R, length L and mass M.

(c) Homogeneous disk of radius R and mass M.

Figure 9.7: Computing the moment of inertia with respect to the axis of radial symmetry by integration over $\mathrm{d}m$.

9.3.3 Example 3: Moment of inertia of a ring

Let's look at a ring of radius R with its mass M distributed homogeneously along its length (Fig. 9.7a). We neglect its thickness. The ring has a linear density of

$$\lambda = \frac{M}{2\pi R}, \tag{9.35}$$

based on its circumference $s = 2\pi R$. An infinitesimal segment on the ring, subtended by an angle $\mathrm{d}\theta$, has an arc length $R\,\mathrm{d}\theta$. This small segment therefore has a mass $\mathrm{d}m = \lambda R\,\mathrm{d}\theta$. The full integral (9.21) over the ring can thus be written as

$$I = \int_0^{2\pi} R^2 \left(\frac{M}{2\pi R}\right)(R\,\mathrm{d}\theta) = MR^2. \tag{9.36}$$

9.3.4 Example 4: Moment of inertia of a hollow cylinder

Now consider a hollow cylinder of radius R, length L and uniform mass M, as in Fig. 9.7b. The surface density is the mass divided by the total surface $A = 2\pi RL$,

$$\sigma = \frac{M}{2\pi RL}. \tag{9.37}$$

A small piece of the surface can be expressed in cylindrical coordinates (r, θ, z), in which case it has sides $r\,\mathrm{d}\theta$ and $\mathrm{d}z$, such that its area is

$$\mathrm{d}A = R\,\mathrm{d}\theta\,\mathrm{d}z. \tag{9.38}$$

So the moment of inertia with respect to the cylinder's axis as rotation axis becomes

$$I = \int_0^L \int_0^{2\pi} R^2 \left(\frac{M}{2\pi RL}\right)(R\,\mathrm{d}\theta\,\mathrm{d}z) = MR^2. \tag{9.39}$$

This is the same result as for a simple mass point and a ring! The length L does not matter. The simple reason for this is that each mass point on the cylinder's surface is at a constant distance R from the axis of rotation.

9.3.5 Example 5: Moment of inertia of a solid cylinder

Take again a cylinder rotating about its axis, but assume it is solid. If it is homogeneous, it has volume mass density

$$\rho = \frac{M}{\pi R^2 L}. \tag{9.40}$$

A small block of volume can also be expressed in cylindrical coordinates. It will have a top area $\mathrm{d}A = r\,\mathrm{d}\theta\,\mathrm{d}z$ and a thickness $\mathrm{d}r$,

$$\mathrm{d}V = \mathrm{d}A\,\mathrm{d}r = r\,\mathrm{d}\theta\,\mathrm{d}r\,\mathrm{d}z. \tag{9.41}$$

So the integral becomes

$$I = \int_0^L \int_0^R \int_0^{2\pi} r^2 \left(\frac{M}{\pi R^2 L}\right) (r\,\mathrm{d}\theta\,\mathrm{d}r\,\mathrm{d}z) = \frac{MR^2}{2}. \tag{9.42}$$

9.3.6 Example 6: Moment of inertia of a disk

Now let's look at a homogeneous disk. It has a surface mass density

$$\sigma = \frac{M}{\pi R^2}. \tag{9.43}$$

distributed over the area πR^2. One method to solving the integral is using a small piece at radius $r < R$, $\mathrm{d}A = r\,\mathrm{d}\theta\mathrm{d}r$. However, since the disk is assumed homogeneous, it is easier to divide the disk into concentric rings of thickness $\mathrm{d}r$, such that each ring has an area $\mathrm{d}A = 2\pi r\mathrm{d}r$[1], which we get from taking the integral of the area. The integral for I becomes

$$I = \int_0^R r^2 \left(\frac{M}{\pi R^2}\right) (2\pi r\mathrm{d}r) = \frac{MR^2}{2}. \tag{9.44}$$

This is the same as for a cylinder!

9.3.7 Example 7: Moment of inertia of a hollow sphere

Let's look at a hollow sphere. If the mass M is homogeneously distributed over the spherical surface with area $A = 4\pi R^2$, it has area mass density

$$\sigma = \frac{M}{4\pi R^2}. \tag{9.45}$$

A spherical area element of sides $R\sin\theta\,\mathrm{d}\phi$ and $R\,\mathrm{d}\theta$ can be expressed in spherical coordinates (Fig. 9.8)

$$\mathrm{d}A = R^2 \sin\theta \mathrm{d}\theta\,\mathrm{d}\phi\,. \tag{9.46}$$

The moment of inertia is

$$I = \int_0^{2\pi} \int_0^{\pi} R^2 \left(\frac{M}{4\pi R^2}\right) (R^2 \sin\theta\mathrm{d}\theta\ \mathrm{d}\phi) = \frac{2MR^2}{3}. \tag{9.47}$$

[1]Because the integrand does not depend on θ, this is actually the same as integrating over $\mathrm{d}\theta$ between 0 and 2π.

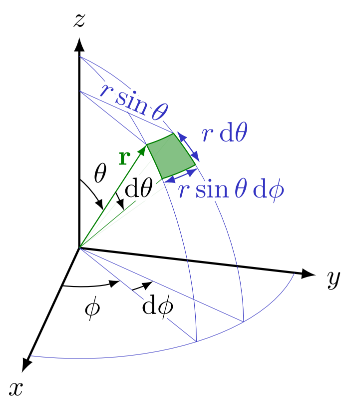


Figure 9.8: Spherical coordinates (r, θ, ϕ) and an infinitesimal area element with sides $r \sin\theta \, \mathrm{d}phi$ and $r \, \mathrm{d}\theta$. Radius r is always positive, running form 0 to $+\infty$, while the polar angle θ runs from 0 to π, and azimuthal angle from 0 to 2π.

9.3.8 Example 8: Moment of inertia of a solid sphere

Now consider a homogeneous, solid sphere with volume $V = 4\pi R^3/3$ and volume mass density

$$\rho = \frac{M}{4\pi R^3/3}. \tag{9.48}$$

A small block of volume at radius r can again be expressed in spherical coordinates.

$$\mathrm{d}V = \mathrm{d}A \, \mathrm{d}r = r^2 \sin\theta \mathrm{d}\theta \; \mathrm{d}\phi \; \mathrm{d}r. \tag{9.49}$$

So the integral becomes

$$I = \int_0^R \int_0^{2\pi} \int_0^{\pi} r^2 \left(\frac{M}{4\pi R^3/3} \right) (r^2 \sin\theta \mathrm{d}\theta \; \mathrm{d}\phi \; \mathrm{d}r) = \frac{2MR^2}{5}. \tag{9.50}$$

9.3.9 Example 9: Large wheel/disk

Suppose you have a mass m suspended by a string that goes over a pulley and then is winded around a large, wheel of radius R and mass M, as in Fig. 9.9. The mass is let go from rest, and falls down, such that the wheel starts spinning counterclockwise. For the mass m, the second law is

$$\sum_i F = mg - T = ma. \tag{9.51}$$

The wheel will stay at the same origin, but will start to rotate as the string starts pulling. Therefore, we will use Newton's second law for torques:

$$\sum_i \tau = TR = I\alpha. \tag{9.52}$$

The wheel has a mass M and radius R, with the mass distributed evenly across the disk, so the moment of inertia is $\frac{1}{2}MR^2$. Each point on the string has an acceleration a, therefore each point at the edge of the disk has an angular acceleration of

$$\alpha = \frac{a}{R}. \tag{9.53}$$

Therefore, the right side of Eq. (9.52) becomes

$$TR = \left(\frac{1}{2}MR^2\right)\left(\frac{a}{R}\right) = \frac{1}{2}MRa, \tag{9.54}$$

such that the tension is $T = \frac{1}{2}Ma$. Using Eq. (9.51),

$$a = \frac{m}{m + \frac{1}{2}M}g. \tag{9.55}$$

So the larger M, the smaller the acceleration α or a. This makes sense because a disk with more mass, means more moment of inertia.

9.4 Kinetic energy of rotation

Say we have an object rotating at some angular velocity ω. What is its kinetic energy? First, we assume the object can be divided into many small masses m_i that are each moving around the axis of rotation at the same angular velocity ω. In the familiar linear motion, each mass m_i with velocity v_i has a kinetic energy

$$K_i = \frac{1}{2}mv_i^2. \tag{9.56}$$

If a part of the mass m_i is at a distance r_i from the axis of rotation, we can calculate the total kinetic energy of a rotating body using the fact that $v_i = r_i\omega$:

$$K = \sum_i \frac{1}{2}m_i(r_i\omega)^2 = \frac{1}{2}\left(\sum_i m_i r_i^2\right)\omega^2, \tag{9.57}$$

where we have regrouped the terms. We recognize the moment of inertia in the parentheses, so

Kinetic energy of rotation.

$$K = \frac{1}{2}I\omega^2. \tag{9.58}$$

So what is the work to move the mass m_i over some arc length $\mathrm{d}s_i$? Say this work is done by the force F_{it} that is tangential to the circle, then

$$\mathrm{d}W_i = F_{i\mathrm{T}}\mathrm{d}s_i, \tag{9.59}$$

or because $\mathrm{d}s_i = r_i\mathrm{d}\theta$ and Eq. (9.3),

$$\mathrm{d}W_i = \tau\,\mathrm{d}\theta_i\,. \tag{9.60}$$

So in general,

Work delivered by torque.

$$W = \int_{\theta_1}^{\theta_2} \tau\mathrm{d}\theta. \tag{9.61}$$

The power, which is the time derivative of W, is then given by

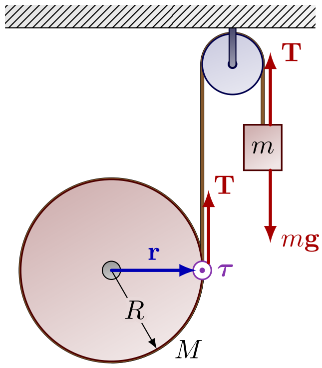


Figure 9.9: A hanging mass m is connected by a string to a large, disk of radius R and mass M. The string is wound several times around the disk, such that it is free to spin due to a torque from the tension.

Power by torque.

$$P = \tau \frac{\mathrm{d}\theta}{\mathrm{d}t} = \tau\omega. \tag{9.62}$$

This looks indeed similar to $P = Fv$, with the analogies of force with torque, and velocity with angular velocity.

9.4.1 Example: Large wheel connected to a suspended mass

Look back at Example 4 and Fig. 9.9. What is the angular velocity ω of the disk after the mass m falls some height h from rest? This can easily be solved in terms of energy, assuming there is no friction. In the beginning, the wheel and mass are at rest, so there is only some potential energy stored in mass m. Once it has fallen over a height h, it will have lost that potential energy, which will be converted into kinetic energy of its own motion and of the rotation of the disk:

$$mgh = \frac{1}{2}\left(\frac{1}{2}MR^2\right)\omega^2 + \frac{1}{2}mv^2, \tag{9.63}$$

where v is the velocity of the falling mass, which has to be $v = R\omega$, as this is also the velocity of any point on the edge of the disk, and the mass is connected to the disk by the string. We can now solve for ω:

$$\omega = \sqrt{\frac{2mgh}{R^2(\frac{1}{2}M + m)}}. \tag{9.64}$$

9.4.2 Kinetic energy of a system of particles

Say you have a group of particles, each with a mass m_i and velocity $\mathbf{v}_i$, then the total kinetic energy of this system is simply the sum

$$K = \sum_i \frac{1}{2} m_i \mathbf{v}_i^2, \tag{9.65}$$

representing the magnitude squared as $v^2 = \mathbf{v} \cdot \mathbf{v} = \mathbf{v}^2$ (see Eq. (3.15)). Let's rewrite this a bit, using the particles' velocity with respect to the center of mass

$$\mathbf{v}_i = \mathbf{v}_{\mathrm{cm}} + \mathbf{v}_{i,\mathrm{cm}}, \tag{9.66}$$

where $\mathbf{v}_{\rm cm}$ is the center-of-mass velocity, and $\mathbf{v}_{i,\rm cm}$ is the velocity of particle i with respect of the center of mass, or equivalently, in the center-of-mass frame, see Fig. 8.8. The total kinetic energy is therefore

$$K = \sum_i \frac{1}{2} m_i \left(\mathbf{v}_{i,\rm cm} + \mathbf{v}_{\rm cm}\right)^2 . \tag{9.67}$$

Using the fact that the scalar productive is distributive with vector sum (see Eq. (3.14)),

$$K = \sum_i \frac{1}{2} m_i v_{\rm cm}^2 + \sum_i \frac{1}{2} m_i v_{i,\rm cm}^2 + 2\mathbf{v}_{\rm cm} \cdot \sum_i m_i \mathbf{v}_{i,\rm cm} \tag{9.68}$$

Now remember from Section 8.3 that in the center-of-mass frame, all momenta adds up to zero, so

$$0 = \sum_i \mathbf{p}_i = \sum_i m_i \mathbf{v}_{i,\rm cm}. \tag{9.69}$$

In other words, we can write the total kinetic energy of these mass particles as

Kinetic energy of a system of particles.

$$K = \frac{1}{2} M v_{\rm cm}^2 + \sum_i \frac{1}{2} m v_{i,\rm cm}^2, \tag{9.70}$$

where $M = \sum_i m_i$ is the total mass. The first term is the total kinetic energy due to the linear motion of the center of mass. The second term

$$K_{\rm rel} = \sum_i \frac{1}{2} m v_{i,\rm cm}^2 \tag{9.71}$$

is the sum of kinetic energy due to the motion $\mathbf{v}_{i,\rm cm}$ of the particle relative to the center of mass.

If all particles are rotating around the center of mass with some angular velocity ω, then $v_{i,\rm cm} = r_{i,\rm cm}\omega$, and we can rewrite Eq. (9.70) in terms of moment of inertia I with respect to the center of mass:

Total kinetic energy of a rotating rigid body.

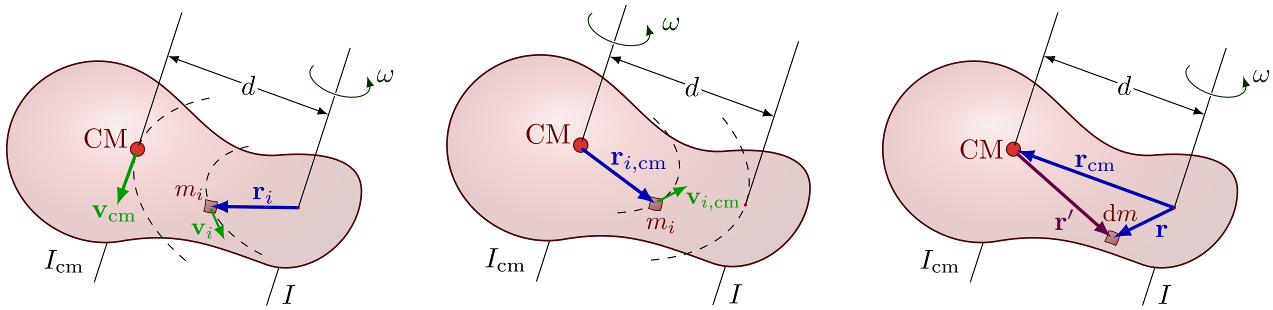
$$K = \frac{1}{2} M v_{\rm cm}^2 + \frac{1}{2} I \omega^2 \tag{9.72}$$

$$= K_{\rm trans} + K_{\rm rot}. \tag{9.73}$$

So the total kinetic energy of this system is the *translational* plus the *rotational kinetic energy*. We can effectively "decouple" the two types motions and treat them separately.

9.5 Parallel axis theorem

It is important to remember that the moment of inertia is defined with respect to a some rotation axis. It can therefore depend not only on the direction of the axis, but also the position with respect to the body of mass. In most examples we have see before, the axis goes through the center of mass of a highly symmetric object, like a ring, disk, cylinder or



(a) A rigid body of mass. **(b)** Center-of-mass frame of (a). **(c)** With position vectors.

Figure 9.10: Parallel axis theorem: The moment of inertia with respect to an axis parallel to an axis through the center of mass is $I = I_{\rm cm} + Md^2$, where d is the distance between the axes.

sphere, and we can therefore exploit the symmetry to simplify our calculations. Another trick we can deploy is the so-called *parallel axis* theorem, which we will derive now.

Consider some rigid body with mass M that is rotating around some axis that does not go through the center of mass, but is at a distance d from this point, as illustrated in Fig. 9.10a. If the body rotates with an angular velocity ω, then the center of mass has a velocity $v_{\rm cm} = d\omega$. In the center-of-mass frame as in Fig. 9.10b, each piece of mass in the body rotates around the center of mass with a velocity $v_{i,\rm cm} = r_{i,\rm cm}\omega$, so Eq. (9.70) becomes

$$K_{\rm cm} = \frac{1}{2}M(d\omega)^2 + \frac{1}{2}\sum_i m_i(r_{i,\rm cm}\omega)^2. \tag{9.74}$$

Now we can rewrite this in terms of the moment of inertia $I_{\rm cm} = \sum_i m_i r_{i,\rm cm}^2$ with respect to the center of mass:

$$K = \frac{1}{2}\left(I_{\rm cm} + Md^2\right)\omega^2. \tag{9.75}$$

We have a useful result:

Parallel axis theorem (Steiner's theorem).

$$I = I_{\rm cm} + Md^2 \tag{9.76}$$

This theorem is very useful to calculate the moment of inertia. If we know the moment of inertia with respect to an axis of rotation through the center of mass, $I_{\rm cm}$, we can always find the moment of inertia I with respect to any other axis that is parallel; all we have to do is add the term Md^2, where d is the distance between the parallel axes.

Notice the parallel axis theorem implies that the moment of inertia is minimal if it is with respect to a rotation axis through the center of mass for a given direction. It can only be larger if a parallel axis does not go through the center of mass, because Md^2 is always a positive quantity.

9.5.1 Extra: Alternative proof

Most sources give an alternative proof with integrals over an infinitesimal mass dm in a continuous body. Let's look at the position vectors in Fig. 9.10c. Like Eq. (8.51) in Section 8.4,

$$\mathbf{r} = \mathbf{r}' + \mathbf{r}_{\rm cm}. \tag{9.77}$$

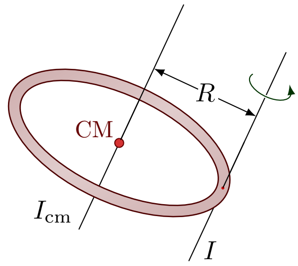


Figure 9.11: Applying the parallel axis theorem to a massive ring with $d = R$.

then

$$I = \int \mathbf{r}^2 \mathrm{d}m = \int (\mathbf{r}' + \mathbf{r}_{\mathrm{cm}})^2 \mathrm{d}m \tag{9.78}$$

which, thanks to the distributive property Eq. (3.14), becomes

$$I = \int \mathbf{r}'^2 \mathrm{d}m + \int \mathbf{r}_{\mathrm{cm}}{}^2 \mathrm{d}m + 2\mathbf{r}_{\mathrm{cm}} \cdot \int \mathbf{r}' \mathrm{d}m. \tag{9.79}$$

The first integral is exactly the moment of inertia defined with respect to the parallel axis through the center of mass, I_{cm}. Because $\mathbf{r}_{\mathrm{cm}}^2 = d^2$ is constant, the second integral reduces to Md^2 with the total mass $M = \int \mathrm{d}m$. Finally, the last integral is the center of mass in the center-of-mass frame, which is by definition zero according to Eq. (8.49). Again, we find

$$I = I_{\mathrm{cm}} + Md^2. \tag{9.80}$$

9.5.2 Example 1: Moment of inertia of two masses (revisited)

Remember that for the moment of inertia of two masses rotating around an axis through their center of mass shown in Fig. 9.6b was $I_{\mathrm{cm}} = 2mr^2$. What is the moment of inertia I' if the axis is through one of the masses as in Fig. 9.6a? Applying the parallel axis theorem with $d = r$,

$$I = 2mr^2 + 2mr^2 = 4mr^2. \tag{9.81}$$

So it the total rod has length $L = 2r$, $I = mL^2$. This is consistent with our definition (9.20) for the moment of inertia of a single mass point m.

9.5.3 Example 2: Moment of inertia of a ring (revisited)

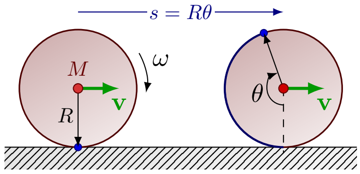
As an example: What is the moment of inertia I for a rotating ring through its edge, as in Fig. 9.11? We know that the moment of inertia of a ring with respect to a rotation axis through its center and perpendicular to the plane the ring lays in, is $I_{\mathrm{cm}} = MR^2$ Therefore, if the new axis of rotation is at the ring's edge, $h = R$, and the new moment of inertia is

$$I = 2MR^2. \tag{9.82}$$

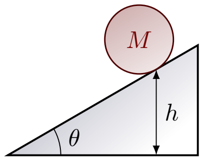
9.6 Rolling

Consider a cylinder rolling over a flat surface as in Fig. 9.12a. Say it traveled some distance s. If the cylinder was rolling without slipping, it simultaneously rotated by an angle θ that is related to the distance as

$$s = R\theta, \tag{9.83}$$



(a) Condition for rolling without slipping is that the linear distanced s equals the arc length (thick dark blue) of the rotated angle θ.



(b) A cylinder is released from rest off a ramp. To compute its final velocity, we simply need energy conservation.

Figure 9.12: Cylinder rolling without slipping.

because it should be the same as the arc length. The velocity of the center of mass then has to be the time-derivative

$$v = \frac{\mathrm{d}s}{\mathrm{d}t} = R\frac{\mathrm{d}\theta}{\mathrm{d}t}. \tag{9.84}$$

This gives us the condition

Condition of rolling without slipping.

$$v = r\omega \tag{9.85}$$

So without slipping, any rolling object with mass M and moment of inertia I has a total kinetic energy

Kinetic energy of a rolling object.

$$K = \frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2, \tag{9.86}$$

with $\omega = v/r$. The first term is due to its linear motion, and the second due to its rotation, as in Eq. (9.72).

9.6.1 Example: Rolling off a ramp

Consider a cylinder with radius R and mass M that is let go from a ramp. How fast will it move at the bottom of a ramp with inclination θ, if it starts from rest at a height h? We can now easily solve this with energy conservation. At the top, the cylinder has a potential energy mgh, and at the bottom it only has kinetic energy,

$$Mgh = \frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2. \tag{9.87}$$

We know that $I = MR^2/2$ for a solid cylinder and $\omega = v/r$ for rolling without slipping, so the last equation becomes

$$Mgh = \frac{1}{2}Mv^2 + \frac{M}{4}v^2. \tag{9.88}$$

and we find

$$v = \sqrt{\frac{4}{3}gh}. \tag{9.89}$$

If we instead roll a hollow cylinder of the same mass, we need to plug $I = MR^2$ into Eq. (9.87), and find

$$v = \sqrt{gh}. \tag{9.90}$$

So the hollow cylinder will roll down slower, which is to be expected as it has larger moment of inertia.

9.7 Angular momentum

In linear motion, we have $\mathbf{p} = m\mathbf{v}$ and

$$\sum_i \mathbf{F} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}. \tag{9.91}$$

If there is no net external force, $\sum_i F_i = 0$, then

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = 0, \tag{9.92}$$

and we say momentum is conserved. This means that in an isolated system, the initial and final momentum are always equal, $\mathbf{p}_\mathrm{i} = \mathbf{p}_\mathrm{f}$.

In rotational motion we can define something similar, starting from the definition of torque

$$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F} = \mathbf{r} \times \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}. \tag{9.93}$$

Notice that the time derivative of the cross product of $\mathbf{r} \times \mathbf{p}$ can be found using the fact that the cross product respects the product rule for derivatives

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r} \times \mathbf{p}) = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} \times \mathbf{p} + \mathbf{r} \times \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}. \tag{9.94}$$

The first term on the right-hand side is zero because the cross product of two parallel vectors is zero, $\mathbf{v} \times \mathbf{p} = 0$. The second term on the right is the torque from Eq. (9.93). We obtain then

$$\boldsymbol{\tau} = \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{r} \times \mathbf{p}). \tag{9.95}$$

The quantity in the parentheses is the rotational equivalent of momentum, and is appropriately called the *angular momentum*,

Angular momentum of a point particle.

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}. \tag{9.96}$$

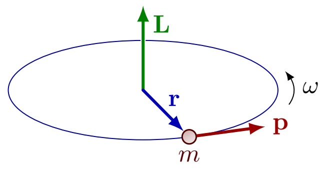
For a single mass point, the angular momentum with respect to some given origin is

$$\mathbf{L} = \mathbf{r} \times m\mathbf{v}. \tag{9.97}$$

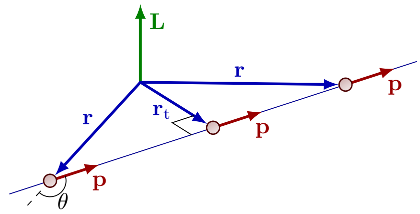
From the derivation above, we also learn that just like a net force causes a change in momentum by Newton's second law, a net torque causes a change in angular momentum.

Newton's second law for rotation.

$$\boldsymbol{\tau} = \frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}. \tag{9.98}$$



(a) Uniform circular motion.



(b) Uniform linear motion.

Figure 9.13: Angular momentum of a single point particle.

Of course, we may have multiple torques, therefore it is important to sum the torques to determine the motion. Imagine a ball at the end of a rope being swung around in circles as in Fig. 9.13a. It has some angular momentum $\mathbf{L} = \mathbf{r} \times m\mathbf{v}$. In a flat circle, $\mathbf{v} \perp \mathbf{r}$ such that

$$|\mathbf{r} \times \mathbf{v}| = rv. \tag{9.99}$$

So the magnitude of angular momentum of a mass m moving in a circle is

$$L = mrv. \tag{9.100}$$

Because the tangential velocity is $v = r\omega$,

$$L = mr^2\omega. \tag{9.101}$$

We can plug in the moment of inertia $I = mr^2$, and find

$$L = I\omega. \tag{9.102}$$

More generally, for a rotation we can write it as a vector

Angular momentum (circular motion).

$$\mathbf{L} = I\boldsymbol{\omega}. \tag{9.103}$$

The $\boldsymbol{\omega}$ and thus $\mathbf{L}$ follow the right-hand rule in Fig. 9.2. Notice that this equation looks very similar to $\mathbf{p} = m\mathbf{v}$. This is consistent with our earlier result,

$$\boldsymbol{\tau} = \frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = I\boldsymbol{\alpha}. \tag{9.104}$$

9.7.1 Conservation of angular momentum

If there are no external torques, then there is no angular acceleration, and angular momentum is conserved:

$$0 = \frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = I\boldsymbol{\alpha}, \tag{9.105}$$

, similar to the case of $\mathbf{p}$, and

Law of conservation of angular momentum.

$$(\mathbf{L})_{before} = (\mathbf{L})_{after}. \tag{9.106}$$

A well-known example is a figure skater who spins faster after pulling her arms in. With the arms closer to the body, her moment of inertia I is smaller, and the angular velocity ω has to increase to compensate and keep the angular momentum L stay constant,

$$(I\omega)_{\text{before}} = (I\omega)_{\text{after}}. \tag{9.107}$$

9.7.2 Example 1: Neutron star

Another cool example are neutron stars, which are known for being extremely dense, and spinning incredibly fast. When a supergiant star of about 10 to 25 solar masses $M_\odot$ reaches its end, it explodes in a supernova and only leaves behind its core which collapses under its own gravity. The neutron star is this collapsed core with a typical mass of $1.4M_\odot$, crunched into amazingly small radius of $R_{\rm ns} = 10\,{\rm km}$. Say the core initially had a radius of $R_{\rm core} = 8000\,{\rm km}$, by what factor will the rotation frequency increase when the core is compressed into a neutron star of radius $R_{\rm ns} = 10\,{\rm km}$ and the same mass? For simplicity, suppose that the core and neutron star are homogeneous, solid spheres, then we can use Eq. (9.50), such that their moment of inertia $I = 2MR^2/5$ depends on R^2. The frequency of rotation of the neutron star will be

$$f_{\rm ns} = \frac{I_{\rm core}}{I_{\rm ns}} f_{\rm core} \tag{9.108}$$

$$= \frac{R_{\rm core}^2}{R_{\rm ns}^2} f_{\rm core} = 1 \times 10^6 f_{\rm core}. \tag{9.109}$$

So if the core was originally spinning with a period of $T_{\rm core} = 1000\,{\rm s}$, then the neutron star will spin with a period of $T_{\rm ns} = 1.6\,{\rm ms}$, or 640 full revolutions every second! This is indeed a typical period of a neutron star.

9.7.3 Example 2: Particle moving in a straight line

A particle moving uniformly in a straight line has an angular momentum relative to some point not on the line, even though it does not “rotate” around that point. In this case, the position vector $\mathbf{r}$ would be changing, so

$$L = rmv\sin\theta. \tag{9.110}$$

Clearly, $\mathbf{L} = 0$ is everywhere with respect to a point on the line because the angle $\theta = 0$ or π. The point with respect to which $\mathbf{L}$ is defined, is called the *point of closest approach*. The lever arm $\mathbf{r}_{\rm t}$ points to this point and is perpendicular to the particle's trajectory. Looking at the left triangle in Fig. 9.13b, we see that $r = r_{\rm t}/\sin\theta$, and so in any point $\mathbf{r}$ on the path,

$$L = r_{\rm t}mv. \tag{9.111}$$

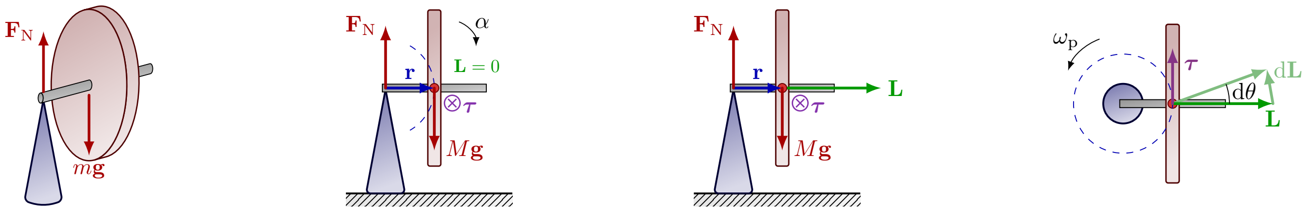
This is clearly constant everywhere on the path, which is consistent with conservation of angular momentum.

9.8 Precession

Suppose you place a disk on a pivot by one of its handles as in Fig. 9.14. If the wheel does not spin, the wheel will simply fall over because there is an unbalanced torque of $\tau = rmg$. But if we spin the wheel with an angular velocity ω, there is still the same torque, however this time, there is a nonzero angular momentum $\mathbf{L}$ along the axis of the disks rotation. $\mathbf{L}$ changes in the direction of $\boldsymbol{\tau}$.

$$\mathrm{d}\mathbf{L} = \boldsymbol{\tau}\mathrm{d}t = rMg\,\mathrm{d}t\,\hat{\boldsymbol{\tau}}. \tag{9.112}$$

Because $\mathbf{L}$ and $\boldsymbol{\tau}$ are perpendicular, the magnitude of $\mathbf{L}$ stays constant, and only it's direction changes. The wheel does not fall down counter to intuition, but rotates around the pivot. This is called *precession*.



(a) The handle allows the disk to spin around its axis and around the pivot.

(b) The disk does not spin, and it will fall down due to an unbalanced torque $\boldsymbol{\tau}$.

(c) The disk spins, creating an angular momentum $\mathbf{L}$. Torque $\boldsymbol{\tau}$ will cause a precession.

(d) Torque $\boldsymbol{\tau}$ perpendicular to angular momentum $\mathbf{L}$, will only change its direction.

Figure 9.14: A disk with mass M is placed on a pivot with one of its handles. If the disk spins with angular velocity ω, then it will have an angular momentum $\mathbf{L}$. A torque will cause the angular momentum to change, leading to the phenomenon known as precession. Precession occurs with angular velocity ω_p, and the wheel does not fall down.

If we look from above as in Fig. 9.14d, we can see that as the axis of rotation changes by some angle $\mathrm{d}\theta$, so does the $\mathbf{L}$ vector change by the vector $\mathrm{d}\mathbf{L}$.

$$\mathrm{d}L = L\mathrm{d}\theta. \tag{9.113}$$

We find

$$\mathrm{d}\theta = \frac{\mathrm{d}L}{L} = \frac{rMg\mathrm{d}t}{L}, \tag{9.114}$$

or,

$$\omega_\mathrm{p} \equiv \frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{rMg}{L}, \tag{9.115}$$

which is how fast $\mathbf{L}$ goes around the pivot. If the wheel is spinning with an angular velocity ω, then the angular momentum can be expressed as $L = I\omega$. Therefore, we can express the frequency of precession as a function of some constants to do with the mass, radius, gravity, and rotational inertia, as well as an angular velocity ω, which is variable.

Precession.

$$\omega_\mathrm{p} = \frac{rMg}{I\omega}. \tag{9.116}$$

From this, we see that as ω gets smaller, ω_p gets bigger.

Precession appears in toy spinning tops and gyroscopes, but Earth also describes a precession. The Earth rotates about once around its own axis per day, and fully precesses once every 25 772 years.

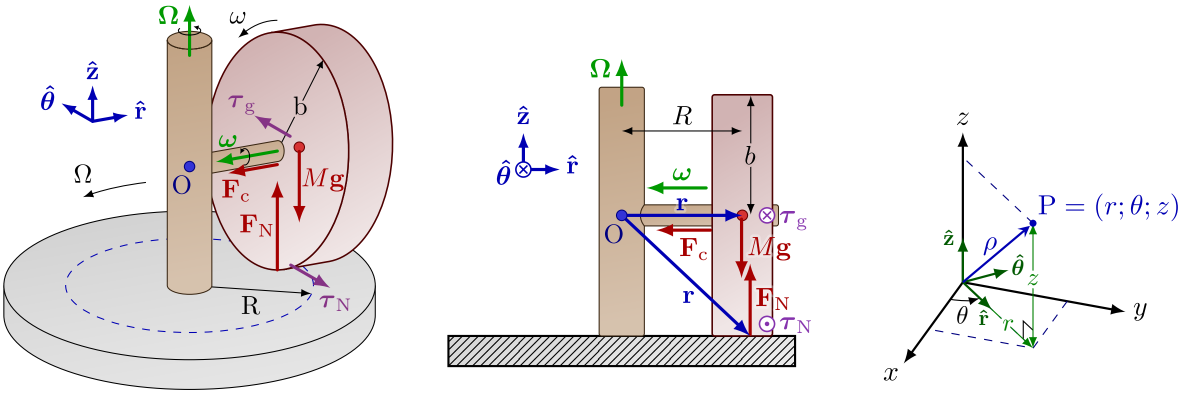
9.9 Application: Millstone

We will now look at a more complicated example of applying our knowledge of angular momentum, torques, and the rolling condition.

In a millstone like illustrated in Fig. 9.15, a heavy "wheel", called a *runner stone*, is driven by a wooden shaft to roll around a flat *bed stone*. This type of mill is sometimes called a *kollergang* or *edge mill*. Our goal is to study the angular moment and determine with what normal force $\mathbf{F}$, the runner stone grinds grain.

First, say that the millstone is not rotating and is simply standing still. Then, there will only be two forces on the runner stone, which balance;

$$0 = \mathbf{F}_\mathrm{N} + M\mathbf{g}, \tag{9.117}$$



(a) 3D view. **(b)** Side view with position vectors. **(c)** Cylindrical coordinates.

Figure 9.15: A millstone: A runner stone (red) of radius b and mass M rolls around a bed stone (grey) driven by a wooden shaft (vertical axle). As it rolls with an angular velocity ω around its own axis, it traces a radius R with an angular velocity Ω.

such that $F_\mathrm{N} = Mg$.

Now assume the runner stone of radius b is rolling about its own axle with an angular velocity ω, such that it goes around a circle of radius R with an angular velocity Ω. Because a massive body is rotating, there has to be some angular momentum. Furthermore, the rotating runner stone goes around a circle with Ω, so its angular momentum must be constantly changing, implying there is a net torque that is nonzero.

Let's break down all the vectors we know of. The most convenient coordinate system is the cylindrical one shown in Fig. 9.15c with the origin at point O. The unit vectors $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}$ rotate with the angle $\theta = \Omega t$, like the the 2D polar coordinates defined in Section 5.2. We can immediately identify three forces: the gravitational, normal and centripetal force,

$$\mathbf{F}_\mathrm{g} = M\mathbf{g} = -Mg\hat{\mathbf{z}} \tag{9.118}$$
$$\mathbf{F}_\mathrm{N} = F_\mathrm{N}\hat{\mathbf{z}} \tag{9.119}$$
$$\mathbf{F}_\mathrm{c} = -M\Omega\hat{\mathbf{r}}, \tag{9.120}$$

respectively. In our analysis below, we will see that we are still missing one force from the shaft to explain the change in angular momentum. About point O, we have two nonzero torques $\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$,

$$\boldsymbol{\tau}_\mathrm{g} = RMg\hat{\boldsymbol{\theta}} \tag{9.121}$$
$$\boldsymbol{\tau}_\mathrm{N} = -RF_\mathrm{N}\hat{\boldsymbol{\theta}} \tag{9.122}$$
$$\boldsymbol{\tau}_\mathrm{c} = 0, \tag{9.123}$$

where we used the geometry in Fig. 9.15b.

There are two simultaneous rotations, given by:

$$\boldsymbol{\omega} = -\omega\hat{\mathbf{r}} \tag{9.124}$$
$$\boldsymbol{\Omega} = \Omega\hat{\mathbf{z}}. \tag{9.125}$$

Because the runner stone is rolling, we have the rolling condition of Eq. (9.85),

$$v = b\omega, \tag{9.126}$$

while it also goes around the circle with the same tangential velocity,

$$v = R\Omega. \tag{9.127}$$

The angular velocities therefore relate as

$$\omega = \frac{R\Omega}{b}. \tag{9.128}$$

Because the z and r axes are perpendicular to each other are along the disk's axes of symmetry[2], we can separate two components in the total angular momentum,

$$\mathbf{L} = I_z\boldsymbol{\omega} + I_r\boldsymbol{\Omega}, \tag{9.129}$$

where the moments of inertia for disks are

$$I_z = \frac{1}{2}Mb^2 \tag{9.130}$$

$$I_r = \frac{1}{4}Mb^2 + MR^2 \tag{9.131}$$

with respect to the z and r axis, respectively. In the last equation, we simply looked up the moment of inertia of a disk about an axis through the disk's plane elsewhere, and added MR^2 according to the parallel axis theorem with $d = R$.

Now, the change in angular momentum is given by the net torque,

$$\sum \boldsymbol{\tau} = \boldsymbol{\tau}_{\mathrm{g}} + \boldsymbol{\tau}_{\mathrm{N}} = \frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t}. \tag{9.132}$$

When we write the angular momentum in terms of unit vectors,

$$\mathbf{L} = I_r\Omega\hat{\mathbf{z}} - I_z\omega\hat{\mathbf{r}}, \tag{9.133}$$

the only time-dependent piece is the unit vector $\hat{\mathbf{r}}$, while $\hat{\mathbf{z}}$, I_z and I_r are constant. This is because the direction of $\boldsymbol{\omega}$ along $\hat{\mathbf{r}}$ keeps changing with angular velocity

$$\Omega = \frac{\mathrm{d}\theta}{\mathrm{d}t}. \tag{9.134}$$

Therefore, we can write $\hat{\mathbf{r}} = \cos(\Omega t)\hat{\mathbf{x}} + \sin(\Omega t)\hat{\mathbf{y}}$ like in Eq. (5.24) with $\theta = \Omega t$, and

$$\frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t} = -\Omega\sin(\Omega t)\hat{\mathbf{x}} + \Omega\cos(\Omega t)\hat{\mathbf{y}} = \Omega\hat{\boldsymbol{\theta}}. \tag{9.135}$$

The only component of Eq. (9.133) that survives derivative with respect to time is

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = -I_z\omega\frac{\mathrm{d}\hat{\mathbf{r}}}{\mathrm{d}t} \tag{9.136}$$

$$= -\frac{bRM\Omega^2}{2}\hat{\boldsymbol{\theta}}. \tag{9.137}$$

Put everything together in Eq. (9.132):

$$RMg\hat{\boldsymbol{\theta}} - RF_{\mathrm{N}}\hat{\boldsymbol{\theta}} = -\frac{bRM\Omega^2}{2}\hat{\boldsymbol{\theta}}, \tag{9.138}$$

and we see that when the mill is moving, the normal force is larger than the runner stone's weight

$$F_{\mathrm{N}} = M\left(g + \frac{b\Omega^2}{2}\right) > Mg. \tag{9.139}$$

The center of mass of the runner stone does not move vertically, so for all forces in the z direction to balance, there has to be an extra downward force coming from the shaft, that has a zero torque about O,

$$\mathbf{F}_{\mathrm{shaft}} = -M\Omega\hat{\mathbf{r}} - \frac{b\Omega^2}{2}\hat{\mathbf{z}}. \tag{9.140}$$

[2] To study general rotations in 3D, one needs to construct a 3×3 matrix called the *inertia tensor*. For a well chosen set of axes perpendicular to each other (often along axes of symmetry), this matrix is diagonal, which allows us to "decouple" the rotations around these different axes in the angular momentum.

9.10 Stability

Mechanical equilibrium (6.8) is when all forces balance. Consider for example a ball than sits on top of a round hill. The normal force $\mathbf{F}_\mathrm{N}$ balances the ball's weight $m\mathbf{g}$, but the smallest sideways push can cause it to roll down the hill. This is therefore called an *unstable equilibrium*. On the other hand, if the ball sits in a valley or trough, a small push will always cause it to be pulled back to its point equilibrium. This is called a *stable equilibrium*. In physics this is often thought of in terms of energy potentials, as in Fig. 9.16a. The "valley" is called a *potential well*, and returns in the study of planetary attraction, but also nuclear attraction.

Some potentials might have another local minimum elsewhere (like a lake on a mountain), where the ball can sit stably. However, if the push is large enough, it can cause the ball to roll over a local maximum, and drop into an even lower minimum. This is called a *metastable equilibrium*. The amount of energy to get it over the "barrier" is called *activation energy*, which is a useful concept in chemistry. Namely, you have to pay some small activation energy to let the ball release a big amount of, often useful, energy. In case of a ball in a local minimum of a hill, the activation energy is $E = mg\Delta y$, where Δy is the height difference with the barrier.

Lastly, if a ball sits on a completely flat surface, it is in a *neutral equilibrium*. The ball does not restore itself to its original position if you give it a small push (like in a stable equilibrium), but it also does not spontaneously gain energy (unstable equilibrium).

But we also have seen rotational equilibrium in Section 9.3. In this case there will be no angular acceleration, because all torques cancel. If both the sum of forces and the sum of torques are zero, you have a *static equilibrium*.

Static equilibrium.

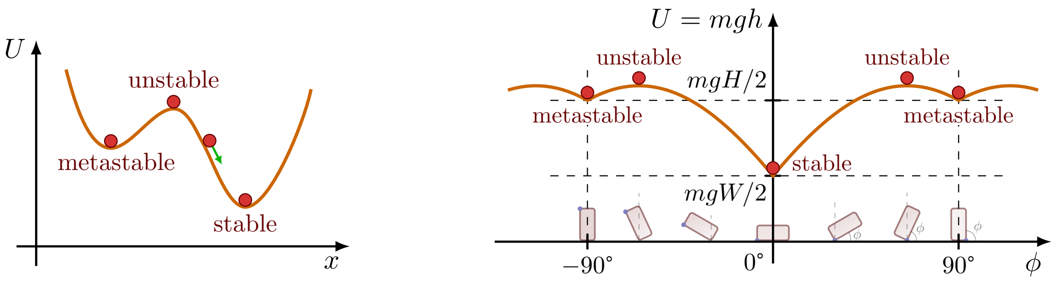
$$\begin{cases} \sum \mathbf{F}_i = 0 & (9.141) \\ \sum \boldsymbol{\tau}_i = 0 & (9.142) \end{cases}$$

If the body starts from rest, the first equation implies that the center of mass will not have a linear motion, and the second equation says that there will not be rotation.

We have already seen one example in the seesaw Fig. 9.5 and in precession of the disk Fig. 9.14, let's look at some other real life ones.

9.10.1 Example 1: Ladder

Ladders can be deceivingly dangerous. You can not only fall off them, but the ladder can also slide down if you are not careful. Let's look at the ladder leaning against a wall in Fig. 9.17a. There are four forces: the ladder's weight $M\mathbf{g}$, the ground's friction $\mathbf{F}_\mathrm{f}$, the



(a) A ball on a hill. **(b)** A rectangular block with width W and height H (Fig. 9.19).

Figure 9.16: Potential wells with stable, metastable and unstable equilibria.

normal force $\mathbf{F}_\mathrm{N}$ from the ground, and the normal force from the wall $\mathbf{F}_\mathrm{W}$. For simplicity, we assume the wall has no friction. What is the condition for the ladder staying in place? What should the angle θ and friction coefficient be μ_s? The ladder is stable as long as it is in a static equilibrium:

$$\begin{cases} 0 = \mathbf{F}_\mathrm{N} + \mathbf{F}_\mathrm{f} + M\mathbf{g} + \mathbf{F}_\mathrm{W} & (9.143) \\ 0 = \boldsymbol{\tau}_\mathrm{N} + \boldsymbol{\tau}_\mathrm{f} + \boldsymbol{\tau}_\mathrm{g} + \boldsymbol{\tau}_\mathrm{W} & (9.144) \end{cases}$$

Using the xy axes as in Fig. 9.17a, choosing clockwise as the positive direction of rotation, and defining the torque with respect to ladder's lowest point,

$$\begin{cases} 0 = F_\mathrm{N} - Mg & (9.145) \\ 0 = F_\mathrm{f} - F_\mathrm{W} & (9.146) \\ 0 = \left(\frac{L}{2}\cos\theta\right)(Mg) - (L\sin\theta)(F_\mathrm{W}) & (9.147) \end{cases}$$

The friction depends on how much the wall pushes the ladder (remember Eq. (6.62) and Fig. 6.14c). The lower the ladder leans against the wall, the larger the friction has to be to balance. We are interested in the extreme case, where the angle θ is just small enough to be stable without the ladder slipping down. The maximum static friction is $F_\mathrm{f} = \mu_\mathrm{s} F_\mathrm{N}$, so these equations become

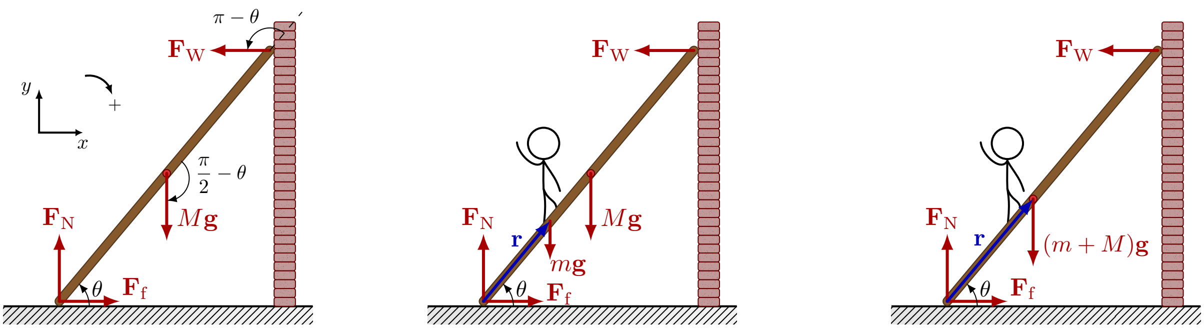
$$\begin{cases} F_\mathrm{N} = Mg & (9.148) \\ F_\mathrm{W} = \mu_\mathrm{s} Mg & (9.149) \\ F_\mathrm{W} = \dfrac{Mg}{2\tan\theta_\mathrm{min}} & (9.150) \end{cases}$$

This allows us to relate the angle to the static friction coefficient μ_s:

$$\tan\theta_\mathrm{min} = \frac{1}{2\mu_\mathrm{s}}. \qquad (9.151)$$

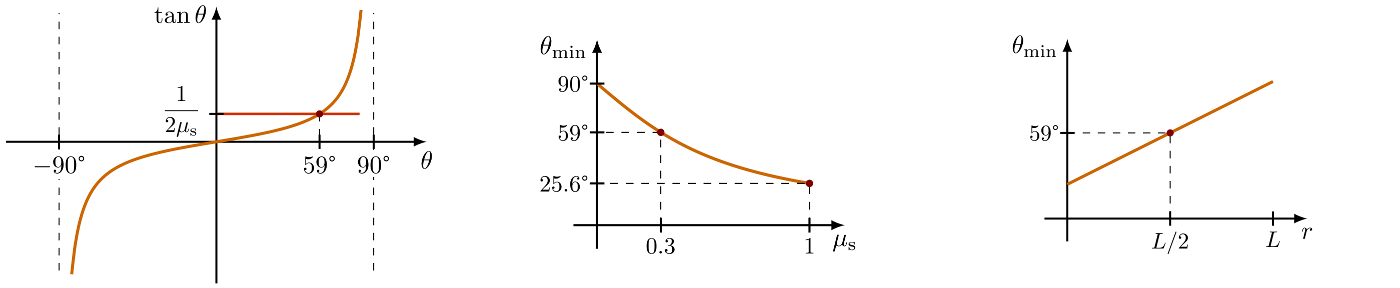
So the angle does not depend on the mass or gravity. If $\mu_\mathrm{s} = 0.3$, then the minimum angle is about $\theta_\mathrm{min} = 59°$. This is a numerical solution one finds with a calculator. Graphically it is the intersection point shown in Fig. 9.18a. Fig. 9.18b shows the angle θ_min as a function of θ_min.

But what if a person of mass m starts climbing the ladder? The new condition for



(a) Without a person. **(b)** With a person of mass m, there is an extra weight. **(c)** Alternative to (b), take the ladder-person as one system, shifting the CM.

Figure 9.17: Ladder of length L leaning against a wall. There is the weight, friction, normal force from the ground, and normal force from the wall. We assume no friction on the wall.



(a) Numerical solution for $\mu_\mathrm{s} = 0.3$: Finding the intersection with $\tan\theta$.

(b) Minimum angle θ_min as a function of μ_s (without an person).

(c) Minimum angle θ_min as a function of the person's position with $\mu_\mathrm{s} = 0.3$.

Figure 9.18: Plots of relation between μ_s and angle θ in Fig. 9.17a.

static equilibrium is

$$\begin{cases} 0 = F_\mathrm{N} - Mg - mg & (9.152) \\ 0 = F_\mathrm{f} - F_\mathrm{W} & (9.153) \\ 0 = (r\cos\theta_\mathrm{min})\,(mg) + \left(\dfrac{L}{2}\cos\theta_\mathrm{min}\right)(Mg) - (L\sin\theta_\mathrm{min})(F_\mathrm{W}) & (9.154) \end{cases}$$

where r is the position of the person on the ladder, $r = 0$ being on the bottom, $r = L$ on top. This time, the solution is given by

$$\tan\theta_\mathrm{min} = \frac{rm + LM/2}{\mu_\mathrm{s}(m+M)L}. \tag{9.155}$$

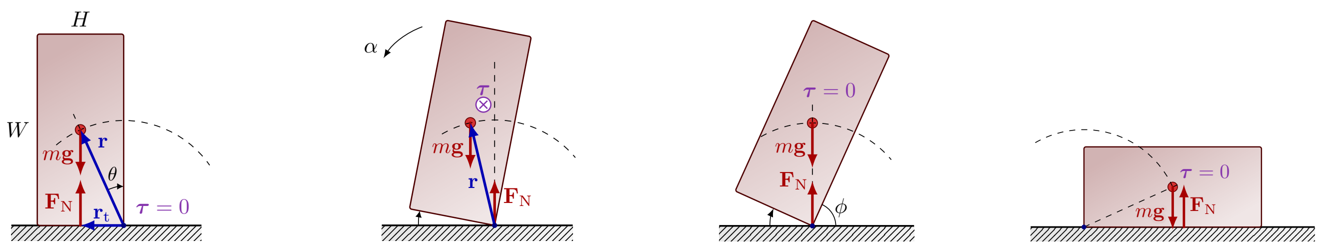
A special case is when the person stands in the middle of the ladder, $r = L/2$, and we again find that $\tan\theta_\mathrm{min} = 1/2\mu_\mathrm{s}$. This means that in this case, it is as if there is no person. What if the person stand lower or higher? Because Eq. (9.155) increases with r, as shown in Fig. 9.18c, and because tan is a monotonously increasing function, it follows that the minimum angle θ_min is larger if $x > L/2$ ($\tan\theta_\mathrm{min} > 1/2\mu_\mathrm{s}$), and smaller if $x < L/2$ ($\tan\theta_\mathrm{min} < 1/2\mu_\mathrm{s}$). So, if the ladder were set at the minimum angle without a person, and then a person walks up the ladder, after they pass the mid-point, the ladder would become unstable and fall !

9.10.2 Example 2: Rectangular block

Consider the rectangular block with width W, height H and homogeneous mass m in Fig. 9.19. Intuitively, we know that the block is more stable if it rests on its long side than its short one. Why?

Say the block rests on its short side as in Fig. 9.19a. Then all torques will balance,

$$(r\sin\theta)mg - r_\mathrm{t}F_\mathrm{N} = 0, \tag{9.156}$$



(a) Metastable equilibrium: Standing on short side. A small push can topple it.

(b) No equilibrium: there is an unbalanced nonzero torque.

(c) Unstable equilibrium: The center of mass is right above the pivot point.

(d) Stable equilibrium: Lying on it long side. It takes more energy to pivot it.

Figure 9.19: Stability of a rectangular block of width W, height H and mass m.

because $F_{\rm N} = mg$ and $r_{\rm t} = r\sin\theta$. This is a metastable state, because a small push will manage to not topple it and cause the block to return to this equilibrium position. However, if the push is large enough it will fall over to a more stable position on its long side.

Say you pivot the block just a little bit as in Fig. 9.19b over its right corner, but with the center of mass on the left side of the pivot point. If you let it go from rest, there will be an unbalanced, restoring torque due to gravity, $\boldsymbol{\tau} = \mathbf{r} \times m\mathbf{g}$.

But what if you balance it *just* right, such that the center of mass is exactly above the pivot point? Now, the torque due to gravity vanishes, because it is antiparallel to the position vector, and thus $\mathbf{r} \times m\mathbf{g} = 0$. This is an unstable equilibrium, because the smallest pushes in will cause it to fall. This is why it is hard to balance a pencil on its tip.

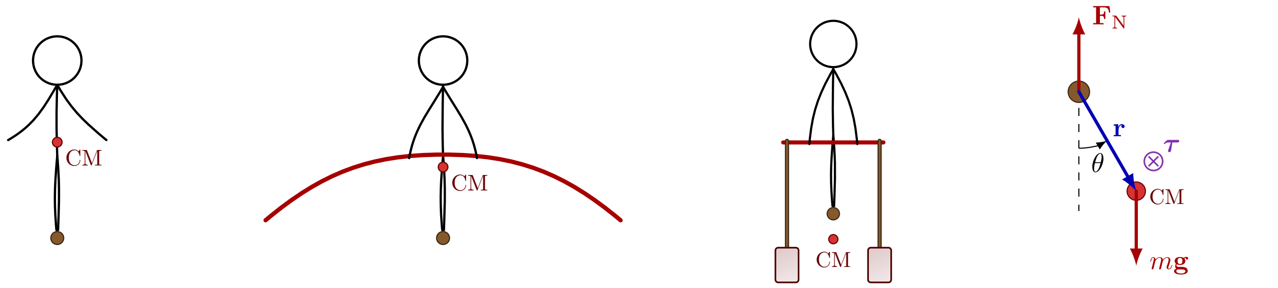
This problem can be expressed in terms of potential wells, by realizing that the block has a maximum gravitational energy $U = mgh$ (with $h > H/2$) when it is balance on one of its corners. When it rests on its short side, it has a lower potential energy ($U = mgH/2$), because the center of mass is lower, and when it lays on its long side even lower ($U = mgW/2$). This is plotted in Fig. 9.16b.

9.10.3 Example 3: Tightrope artist

Walking on a tightrope is a classic example of a balancing act. Without any help, the center of mass of a tightrope artist is above the pivot point, meaning that any small push can cause her to tip over. By extending the arms, the artist can increases her moment of inertia and make small adjustments. By carrying a rod as in Fig. 9.20b, she increases the moment of inertia event more, and has more control over it. A bent rod will also lower the center of mass, making her more stable. But to become fully stable, one can “cheat” by carrying heavy weights that lower the center of mass to *below* the rope, as shown in Fig. 9.20c. Any small rotation will be counteracted by a restoring torque pulling the center of mass back to below the pivot point, Fig. 9.20d. This is how self-balancing toys work.

9.11 Summary

We have seen many formulas for rotational motion in this chapter. Luckily they are closely related to linear motion, and there is a lot of concepts and formulas that are similar. They are summarized and compared in Table 9.1. Make sure you understand in what case these formulas can be applied and in which cases they cannot.



(a) Balancing with arms, to increase the moment of inertia.

(b) Balancing with long rod to increase moment of inertia even more, and slightly lower the center of mass.

(c) Stabilize by lowering the center of mass below the rope.

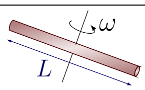
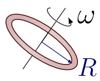
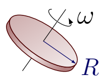
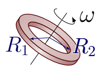
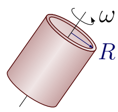
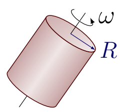
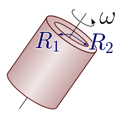
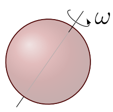
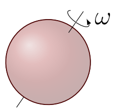
(d) Restoring torque when the center of mass is below the pivot point.

Figure 9.20: Tightrope artist walks on a rope.

Table 9.1: Comparison of translational (or linear) and rotational (or angular) formulas.

Name	Translational	Rotational
Position	x	θ
Velocity	$v = \dfrac{\mathrm{d}x}{\mathrm{d}t}$	$\omega = \dfrac{\mathrm{d}\theta}{\mathrm{d}t}$
Acceleration	$a = \dfrac{\mathrm{d}^2x}{\mathrm{d}t^2}$	$\alpha = \dfrac{\mathrm{d}^2\theta}{\mathrm{d}t^2}$
Position, constant a or α	$x = x_0 + v_0t + \frac{1}{2}at^2$	$\theta = \theta_0 + \omega_0t + \frac{1}{2}\alpha t^2$
Velocity, constant a or α	$v = v_0 + at$	$\omega = \omega_0 + \alpha t$
Inertia	m	$I = \int r^2\mathrm{d}m$
Force (torque)	$\mathbf{F}$	$\boldsymbol{\tau} = \mathbf{r} \times \mathbf{F}$
Momentum	$\mathbf{p} = m\mathbf{v}$	$\mathbf{L} = \mathbf{r} \times \mathbf{p} = I\boldsymbol{\omega}$
Newton's second law	$\sum \mathbf{F} = \dfrac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = m\mathbf{a}$	$\sum \boldsymbol{\tau} = \dfrac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = I\boldsymbol{\alpha}$
Work	$W = \mathbf{F} \cdot \boldsymbol{\Delta x}$	$W = \tau\Delta\theta$
Kinetic energy	$K = \frac{1}{2}mv^2$	$K = \frac{1}{2}I\omega^2$
Power	$P = \mathbf{F} \cdot \mathbf{v}$	$P = \tau\omega$
Momentum conservation	$\sum \mathbf{p}_\mathrm{i} = \sum \mathbf{p}_\mathrm{f}$	$\sum \mathbf{L}_\mathrm{i} = \sum \mathbf{L}_\mathrm{f}$
Equilibrium	$\sum \mathbf{F} = 0$	$\sum \boldsymbol{\tau} = 0$

Table 9.2: Summary of moments of inertia of several common shapes about the main axis.

Shape		Moment of inertia
Rod		$I = \dfrac{1}{12}ML^2$
Ring or loop		$I = MR^2$
Solid disk		$I = \dfrac{1}{2}MR^2$
Solid disk with hole		$I = \dfrac{1}{2}M(R_1^2 + R_2^2)$
Hollow cylinder		$I = MR^2$
Solid cylinder		$I = \dfrac{1}{2}MR^2$
Solid cylinder with hole		$I = \dfrac{1}{2}M(R_1^2 + R_2^2)$
Hollow sphere		$I = \dfrac{2}{3}MR^2$
Solid sphere		$I = \dfrac{2}{5}MR^2$

Chapter 10

Non-Inertial Reference Frames & Pseudo Forces

10.1 Inertial reference frames

In Section 8.4 we saw the Galilean coordinate transformation Eq. (8.53) between two reference frames that move with a uniform velocity with respect to each other. Two observers in each of these reference frames will see different things happen, but the physical prediction stays the same. All you need to do is transform the coordinates.

Take for example, Alice in frame S, who is standing at rest with respect to the ground, and Bob in frame S′, who is in a car moving with constant velocity v with respect to Alice and frame S. Alice sees Bob moving away with a velocity v, and Bob, who thinks he is at rest, sees Alice moving away from him with a velocity v. When Alice throws up a ball, it will follow a one-dimensional vertically path according to

$$\begin{cases} x(t) = 0 & (10.1) \\ y(t) = v_{0y}t - \dfrac{gt^2}{2} & (10.2) \end{cases}$$

When Bob throws up his own ball with the same initial velocity v_{0y}, he will see it follow a one-dimensional path in his own frame as well. He will describe it with

$$\begin{cases} x'(t) = 0 & (10.3) \\ y'(t) = v_{0y}t - \dfrac{gt^2}{2} & (10.4) \end{cases}$$

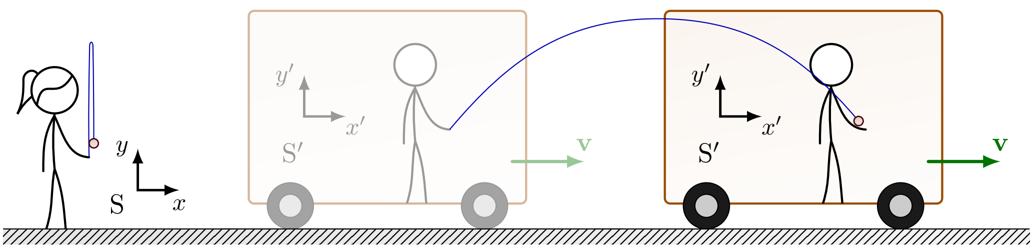


Figure 10.1: Throwing a ball in two inertial reference frames S and S′ moving at a constant relative velocity v. Alice in S sees Bob's ball moving in a parabolic trajectory, while Bob sees his ball moving only vertically.

However, Alice in S will see Bob's ball form a parabolic path as in Fig. 10.1. Alice can use Galilean transformation to move Eqs. (10.3) and (10.4) to her own coordinate system:

$$\begin{cases} x(t) = x'(t) + vt = vt & (10.5) \\ y(t) = y'(t) = v_{0y}t - \dfrac{gt^2}{2} & (10.6) \end{cases}$$

which is indeed a parabola. From Bob's perspective in S′, Alice's ball will also follow a parabola, but in the negative x' direction. Even though they arrive at different descriptions, the physics (i.e. Newton's laws and gravity) are the same, and their results are related by a simple coordinate transformation.

10.2 Non-inertial reference frames

But what if a frame is accelerated with respect to another one? Consider again Alice in frame S and Bob in frame S′, but this time Bob's car is sped up with a uniform acceleration a. Say there is a mass m suspended from the car's ceiling by a wire, which was initially at rest. The mass m has some inertia, so once the car starts moving, the wire pulls the mass along with a tension $\mathbf{T}$. This causes the mass and wire to deflect with an angle θ as shown in Fig. 10.2a. Newton's second law says

$$m\mathbf{a} = m\mathbf{g} + \mathbf{T}, \tag{10.7}$$

where the horizontal component of the tension causes an acceleration $\mathbf{a}$, as illustrated in Fig. 10.2b. But this is in Alice's frame S, which is at rest. Bob, however, thinks he and the mass are the ones at rest and sees a mysterious force $\mathbf{F}'$ pulling the mass:

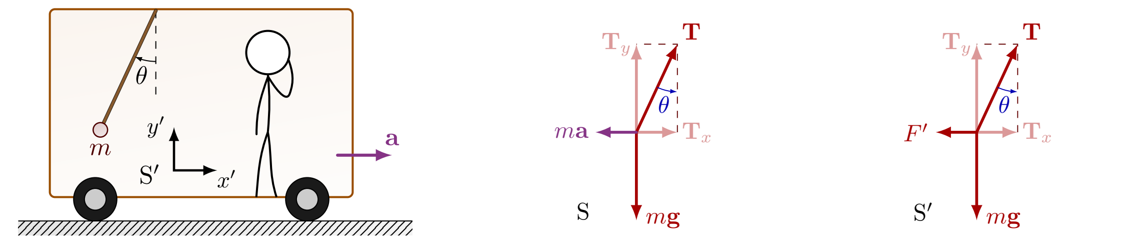
$$0 = m\mathbf{g} + \mathbf{T} + \mathbf{F}'. \tag{10.8}$$

The mass is at equilibrium in S′. So clearly, this new force is given by the acceleration,

$$\mathbf{F}' = -m\mathbf{a}. \tag{10.9}$$

In fact, Bob will also *feel* this force $\mathbf{F}'$ and might fall over if he is not careful. This is a so-called *pseudo force*, or also a *fictitious* or *inertial force*. Because masses have inertia, pseudo forces like $\mathbf{F}'$ or the centrifugal force mentioned in Section 6.7 will appear in accelerated frames like S'. A frame that does not have pseudo forces is called an *inertial frame of reference*.

Inertial frame of reference. *In an inertial frame of reference, all bodies with a zero net force acting upon them do not accelerate.*



(a) Bob in non-inertial frame S′ is confused about the mass deflecting.

(b) Force balance in inertial frame S.

(c) Force balance in Bob's non-inertial frame S′.

Figure 10.2: When Bob's car accelerates, the mass suspended from the ceiling seems to experience a (pseudo) force $\mathbf{F}'$ in accelerated frame S′.

This is actually one of the reasons why Newton wisely included the first law, which might seem redundant as a special case of the second law ($\mathbf{F} = 0$ such that $\mathbf{a} = 0$). By including the first law in addition to the second law, one can postulate that only (real) forces in an inertial reference frame can accelerate masses, otherwise a mass will stay at rest or move in a straight line with constant speed. Equivalently, you can define inertial frames of reference as those reference frames in which Newton laws hold true.

Note that any frame moving with a constant $\mathbf{v}_{\mathrm{cm}}$ with respect to an inertial lab frame, is also an inertial frame of reference.

A *non-inertial reference frame* then, is a(n accelerated) frame where pseudo forces arise and Newton's law are violated. Some typical examples are a passenger in a car that is going around the bend (centripetal acceleration), an astronaut being launched in a rocket, or a skydiver in free fall (before hitting terminal velocity).

The perceived force due to acceleration is often expressed in units of *gravitational force equivalent*, or *g-force*. For example, the driver in a fast race car that accelerates by $a = g$ will feel a g-force of "$1\,g$", which is the force the seat pushes back on them. A Space Shuttle launch or reentry is typicall $3\,g$, and the highest g-force on a roller coaster (in 2020) is up to a whopping $6.3\,g$ on the Tower of Terror in at Gold Reef City in Gauteng, South Africa.

We typically assume a lab frame at rest on the earth's surface is an inertial. However, this is not strictly true due to the Earth's rotation. A section below will explore the Coriolis effect among others, which cause a ball that is dropped to the ground to not follow an exactly straight path, seemingly violating Newton's first law until the rotations are considered.

10.2.1 Coordinate transformation

The physical prediction between two frames that are accelerated with respect to each other is different, because an observer in one frame will claim there is an acceleration, while the other will claim there is a force. The coordinate transformation becomes

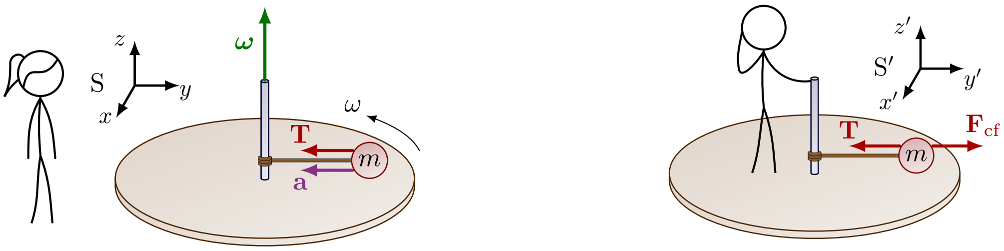
Coordinate transformation to an accelerated frame.

$$\mathbf{r}' = \mathbf{r} - \frac{\mathbf{a}_{\mathrm{frame}} t^2}{2}, \tag{10.10}$$

where frame S$'$ is moving with a uniform acceleration $\mathbf{v}_{\mathrm{frame}}$ with respect to frame S, and has no initial velocity. For simplicity, take the initial velocity and acceleration between the frames to be only in the x direction, in which case the coordinates of the position vector $\mathbf{r}'$ in the accelerated S$'$ frame are given by

$$\begin{cases} x' = x - \dfrac{a_{\mathrm{frame}} t^2}{2} & (10.11) \\ y' = y & (10.12) \\ z' = z & (10.13) \end{cases}$$

Notice that from Bob's perspective in S$'$, Alice is actually the one that is accelerated by $-\mathbf{a}_{\mathrm{frame}}$.



(a) Alice in inertial frame S sees that the mass experiences a centripetal force from tension $\mathbf{T}$.

(b) Bob in non-inertial frame S′ sees that the mass experiences a centrifugal force creating a tension $\mathbf{T}$.

Figure 10.3: A rotational frame gives rise to the fictious centrigual force. Mass m rotates with a merry-go-around (red disk) and is radially held in place by a wire with tension $\mathbf{T}$.

10.3 Rotating reference frames

10.3.1 Centrifugal force

Now consider a uniformly rotating frame as in Fig. 10.3. Alice looks from the side in inertial frame S, while Bob is on the rotating merry-go-around in rotating frame S′. A mass is suspended to the middle by a wire and rotates in circles with the merry-go-around. To hold the mass radially in place, the wire provides a centripetal force with its tension $\mathbf{T}$,

$$m\mathbf{a} = \mathbf{T}. \tag{10.14}$$

Meanwhile, Bob scratches his head because he notices there is a mysterious centrifugal force creating a tension in the wire:

$$0 = \mathbf{T} + \mathbf{F}_{\text{cf}}. \tag{10.15}$$

Again, the mass is at equilibrium in S′ and the forces balance. Comparing the latter two equations,

Centrifugal force. *The centrifugal force is opposite to the centripetal acceleration*

$$\mathbf{F}_{\text{cf}} = -m\mathbf{a} = \frac{mv^2}{r}\hat{\mathbf{r}}. \tag{10.16}$$

where we assumed the centripetal acceleration in frame S points to the origin on rotation axis, just as in Eq. (5.34).

10.3.2 Extra: Coordinate transformation of rotation

This section is extra for the interested reader, and is not part of the main curriculum. Some knowledge of matrices and their properties is assumed.

So what is the coordinate transformation between S and S′? Suppose the coordinate systems of the two frame have the same origin, and S′ rotates counterclockwise with a uniform angular velocity ω, as in Fig. 10.4a. A point P that has coordinates (x, y) in S, will have coordinates

$$\begin{cases} x' = x\cos(\omega t) + y\sin(\omega t) & (10.17) \\ y' = -x\sin(\omega t) + y\cos(\omega t) & (10.18) \end{cases}$$

(a) A point or position vector **r** has different coordinates in the S and S′ frame.

(b) The coordinate transformation $(x, y) \mapsto (x', y')$ can be derived by finding the right triangles and angles.

Figure 10.4: Frame S' is rotated counterclockwise with a time-dependent angle $\theta = \omega t$ with respect to frame S.

in frame S′, as can be gleaned from Fig. 10.4b. In terms of linear algebra, this corresponds to the linear transformation (see Section 3.8)

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos(\omega t) & \sin(\omega t) \\ -\sin(\omega t) & \cos(\omega t) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{10.19}$$

where the matrix $R(\omega t)$ is the *clockwise-rotation matrix* with rotation angle $\theta = \omega t$

$$R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{10.20}$$

This matrix has the nice property that its transpose is its inverse:

$$R^{-1}(\theta) = R^{\mathrm{T}}(\theta) = R(-\theta), \tag{10.21}$$

which is a rotation in the opposite direction $(-\theta)$. Equivalently,

$$RR^{\mathrm{T}} = R^{\mathrm{T}}R = \mathbb{1}, \tag{10.22}$$

where $\mathbb{1}$ is the 3×3 identity matrix. In linear algebra, matrices with this property are called *orthogonal*. It implies that the transformation preserves the length in the scalar product, and therefore any length:

$$\mathbf{r}' \cdot \mathbf{r}' = (R\mathbf{r}) \cdot (R\mathbf{r}) = (R\mathbf{r})^{\mathrm{T}}(R\mathbf{r}) = \mathbf{r}^{\mathrm{T}} \underbrace{(R^{\mathrm{T}}R)}_{\mathbb{1}} \mathbf{r} = \mathbf{r} \cdot \mathbf{r}, \tag{10.23}$$

where $\mathbf{r} = (x, y)$ and $\mathbf{r}' = R\mathbf{r} = (x', y')$, and we wrote the scalar product as a transpose (see Eq. (3.17)). This is exactly what we want for rotations; changing the directions of all vectors without changing their lengths.

Notice that with the above choice of coordinate systems, the position vectors $\mathbf{r}' = \mathbf{r}$ are identical because the origins coincide,

$$\begin{aligned} \mathbf{r}' &= x'\hat{\mathbf{x}}' + y'\hat{\mathbf{y}}' + z'\hat{\mathbf{z}}' \qquad (10.24) \\ &= x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} = \mathbf{r}, \qquad (10.25) \end{aligned}$$

even though their coordinates are different. The unit vectors in the rotating S′ frame are given by

$$\begin{cases} \hat{\mathbf{x}}' = \cos(\omega t)\hat{\mathbf{x}} + \sin(\omega t)\hat{\mathbf{y}} & (10.26) \\ \hat{\mathbf{y}}' = -\sin(\omega t)\hat{\mathbf{x}} + \cos(\omega t)\hat{\mathbf{y}} & (10.27) \\ \hat{\mathbf{z}}' = \hat{\mathbf{z}} & (10.28) \end{cases}$$

The unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ are constant with time in S, but the ones from S$'$ change with time in S as

$$\begin{cases} \dfrac{\mathrm{d}\hat{\mathbf{x}}'}{\mathrm{d}t} = -\omega\sin(\omega t)\hat{\mathbf{x}} + \omega\cos(\omega t)\hat{\mathbf{y}} & (10.29) \\ \dfrac{\mathrm{d}\hat{\mathbf{y}}'}{\mathrm{d}t} = -\omega\cos(\omega t)\hat{\mathbf{x}} - \omega\sin(\omega t)\hat{\mathbf{y}} & (10.30) \\ \dfrac{\mathrm{d}\hat{\mathbf{z}}'}{\mathrm{d}t} = 0 & (10.31) \end{cases}$$

Because the rotation assumed to be counterclockwise in Fig. 10.4a, the constant angular frequency vector is $\boldsymbol{\omega} = \omega\hat{\mathbf{z}}$. So we can rewrite the above time-derivatives as

$$\frac{\mathrm{d}\hat{\mathbf{x}}'}{\mathrm{d}t} = \boldsymbol{\omega}\times\hat{\mathbf{x}}' \tag{10.32}$$

$$\frac{\mathrm{d}\hat{\mathbf{y}}'}{\mathrm{d}t} = \boldsymbol{\omega}\times\hat{\mathbf{y}}', \tag{10.33}$$

such that the derivatives are orthogonal to $\boldsymbol{\omega}$ and the corresponding unit vector. It turns out that this holds more generally for any choice of direction of $\boldsymbol{\omega}$,

$$\frac{\mathrm{d}\hat{\mathbf{x}}'}{\mathrm{d}t} = \boldsymbol{\omega}\times\hat{\mathbf{x}}', \quad \frac{\mathrm{d}\hat{\mathbf{y}}'}{\mathrm{d}t} = \boldsymbol{\omega}\times\hat{\mathbf{y}}', \quad \frac{\mathrm{d}\hat{\mathbf{z}}'}{\mathrm{d}t} = \boldsymbol{\omega}\times\hat{\mathbf{z}}'. \tag{10.34}$$

So the change in direction of these unit vector is always perpendicular to the rotation axis and to the given unit vector itself, so its size is constant, as expected. In other words, the change is tangential. Clearly, if a unit vector is parallel to the rotation axis $\boldsymbol{\omega}$, it will not change direction (e.g. $\hat{\mathbf{x}}' \parallel \boldsymbol{\omega} \Rightarrow \boldsymbol{\omega}\times\hat{\mathbf{x}}' = 0$). On the other hand, the change of direction will be maximal if the unit vector is perpendicular to $\boldsymbol{\omega}$ (e.g. $\hat{\mathbf{x}}' \perp \boldsymbol{\omega} \Rightarrow |\boldsymbol{\omega}\times\hat{\mathbf{x}}'| = \omega$).

10.3.3 Extra: Time derivation of a vector function

This section provides some mathematical formulas to help derive the pseudo forces observed in a rotating frame.

Take any time-dependent vector function $\mathbf{A}$ in inertial frame S,

$$\mathbf{A} = A_x\hat{\mathbf{x}} + A_y\hat{\mathbf{y}} + A_z\hat{\mathbf{z}}. \tag{10.35}$$

This vector can be anything, including the position $\mathbf{r}$. The time derivative of $\mathbf{A}$ is given by the derivative of its coordinates in S

$$\left.\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\right|_{\mathrm{S}} = \frac{\mathrm{d}A_x}{\mathrm{d}t}\hat{\mathbf{x}} + \frac{\mathrm{d}A_y}{\mathrm{d}t}\hat{\mathbf{y}} + \frac{\mathrm{d}A_z}{\mathrm{d}t}\hat{\mathbf{z}}, \tag{10.36}$$

because the unit vectors are constant in time in frame S. Now look at the *same* vector, but with expressed in coordinates of a rotating frame S$'$,

$$\mathbf{A} = A'_x\hat{\mathbf{x}}' + A'_y\hat{\mathbf{y}}' + A'_z\hat{\mathbf{z}}', \tag{10.37}$$

where the coordinates and unit vectors are defined in S$'$. The time derivative of $\mathbf{A}$ in S, with S$'$ coordinate system is

$$\left.\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\right|_{\mathrm{S}} = \left(\frac{\mathrm{d}A'_x}{\mathrm{d}t}\hat{\mathbf{x}}' + \frac{\mathrm{d}A'_y}{\mathrm{d}t}\hat{\mathbf{y}}' + \frac{\mathrm{d}A'_z}{\mathrm{d}t}\hat{\mathbf{z}}'\right) + \left(A'_x\frac{\mathrm{d}\hat{\mathbf{x}}'}{\mathrm{d}t} + A'_y\frac{\mathrm{d}\hat{\mathbf{y}}'}{\mathrm{d}t} + A'_z\frac{\mathrm{d}\hat{\mathbf{z}}'}{\mathrm{d}t}\right) \tag{10.38}$$

due to the product rule. The terms between the first parenthesis in Eq. (10.38) is the change of A in non-inertial frame S′, wherein the unit vectors $\hat{\mathbf{x}}'$, $\hat{\mathbf{y}}'$ and $\hat{\mathbf{z}}'$ are constant. Therefore, Eq. (10.38) becomes

$$\left.\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\right|_{\mathrm{S}} = \left.\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\right|_{\mathrm{S}'} + \boldsymbol{\omega} \times \left(A'_x\hat{\mathbf{x}}' + A'_y\hat{\mathbf{y}}' + A'_z\hat{\mathbf{z}}'\right), \tag{10.39}$$

where we also substituted Eq. (10.34). So using Eq. (10.37), we find the following useful result.

Kinematic transport theorem.

$$\left.\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\right|_{\mathrm{S}} = \left.\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\right|_{\mathrm{S}'} + \boldsymbol{\omega} \times \mathbf{A}. \tag{10.40}$$

So the change of $\mathbf{A}$ in frame S is related to its change relative to the rotating frame S′ by simply adding the term $\boldsymbol{\omega} \times \mathbf{A}$ tangentially to $\mathbf{A}$.

10.3.4 Pseudo forces in a rotating system

The centrifugal force is not the only pseudo force that appears in a rotating system. We will now formally derive the other two that arise starting from the transport theorem (10.40) and deriving a relation between the acceleration in inertial reference frame S and rotating reference frame S′ to see if extra terms show up that indicate the presence of pseudo forces.

In Eq. (10.40), $\mathbf{A}$ can be any vector. For example, the position vector $\mathbf{A} = \mathbf{r} = \mathbf{r}'$,

$$\left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{S}} = \left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{S}'} + \boldsymbol{\omega} \times \mathbf{r}, \tag{10.41}$$

These are the velocities $\mathbf{v}$ in the S frame, and $\mathbf{v}'$ in the S′ frame

$$\mathbf{v} = \mathbf{v}' + \boldsymbol{\omega} \times \mathbf{r}, \tag{10.42}$$

where $\boldsymbol{\omega} \times \mathbf{r}$ encodes the tangential velocity from the rotation $\boldsymbol{\omega}$. By deriving expression (10.41) with respect to time, we find the acceleration $\mathbf{a}$ in frame S,

$$\mathbf{a} = \left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{S}} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{S}'} + \boldsymbol{\omega} \times \mathbf{r}\right). \tag{10.43}$$

Following the product rule,

$$\mathbf{a} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{S}'}\right) + \frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \mathbf{r} + \boldsymbol{\omega} \times \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} \tag{10.44}$$

We can apply the transport theorem (10.40) to $\mathbf{r}$ in the last term, which is Eq. (10.41), and to the vector between the parentheses in the first term, which is

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{S}'}\right) = \left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{S}'} + \boldsymbol{\omega} \times \left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{S}'}. \tag{10.45}$$

Putting everything together,

$$\mathbf{a} = \left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{S}'} + 2\boldsymbol{\omega} \times \left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{S}'} + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) + \frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \mathbf{r}. \tag{10.46}$$

Because the position vector $\mathbf{r} = \mathbf{r}'$ in case of rotation in the same origin, we can define the velocity and acceleration measured in frame S′ as

$$\mathbf{v}' = \left.\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|_{\mathrm{S}'}, \quad \mathbf{a}' = \left.\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|_{\mathrm{S}'}, \tag{10.47}$$

and write the result as

Acceleration in rotating frame.

$$\mathbf{a} = \mathbf{a}' + 2\boldsymbol{\omega} \times \mathbf{v}' + \boldsymbol{\omega}(\times\boldsymbol{\omega} \times \mathbf{r}') + \frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \mathbf{r}'. \tag{10.48}$$

We see that indeed extra terms pop up in the acceleration $\mathbf{a}'$ in the S′ frame. This should give us a hint of new pseudo forces! This is easy to see if you first invoke Newton's second law in the inertial frame S,

$$m\mathbf{a} = \sum \mathbf{F}, \tag{10.49}$$

where $\sum \mathbf{F}$ is the total force observed in inertial frame S. After substituting Eq. (10.48), we see that in the rotating frame S′

Newton's second law in a rotating system.

$$m\mathbf{a}' = \underbrace{\sum \mathbf{F}}_{\textit{real}} \underbrace{- 2m\boldsymbol{\omega} \times \mathbf{v}'}_{\textit{Coriolis}} \underbrace{- m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}')}_{\textit{centrifugal}} \underbrace{- m\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \mathbf{r}'}_{\textit{Euler}}. \tag{10.50}$$

These are three extra pseudo forces observed in the rotating system S′.

Pseudo forces in a rotating system.

- *The Coriolis force* $\mathbf{F}_{\mathrm{Cor}} = -2m\boldsymbol{\omega} \times \mathbf{v}'$ *appears when the particle is moving in the S′* ($\mathbf{v}' \neq 0$).
- *The centrifugal force* $\mathbf{F}_{\mathrm{cf}} = -m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}')$ *appears if the particle is at a nonzero distance from the rotational axis* ($\mathbf{r}' \neq 0$).
- *The Euler force* $\mathbf{F}_{\mathrm{Eul}} = -m\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \times \mathbf{r}'$ *appears when the reference frame experiences an angular acceleration* ($\mathrm{d}\boldsymbol{\omega}/\mathrm{d}t \neq 0$).

Notice that if $\boldsymbol{\omega} \perp \mathbf{r}'$, that the centrifugal force points again in the radial direction,

$$\mathbf{F}_{\mathrm{cf}} = -m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}') = m\omega^2 r\hat{\mathbf{r}}', \tag{10.51}$$

because $\hat{\boldsymbol{\omega}} \times (\hat{\boldsymbol{\omega}} \times \hat{\mathbf{r}}) = -\hat{\mathbf{r}}$.

10.3.5 Coriolis force

Let's focus on the Coriolis force

$$\mathbf{F}_{\text{Cor}} = -2m\boldsymbol{\omega} \times \mathbf{v}'. \tag{10.52}$$

This force is perpendicular to the rotation $\boldsymbol{\omega}$ and the velocity $\mathbf{v}'$ in the rotating frame S′. It is also proportional to the angular frequency ω and speed v'.

Suppose that the velocity moves in the rotation plane, i.e. $\mathbf{v}' \perp \boldsymbol{\omega}$ and take $\boldsymbol{\omega} = \omega\hat{\mathbf{z}}$, then the Coriolis force for a velocity $\mathbf{v}' = v'_x\hat{\mathbf{x}}' + v'_y\hat{\mathbf{y}}'$ is

$$\mathbf{F}_{\text{Cor}} = 2m\omega(v'_y\hat{\mathbf{x}}' - v'_x\hat{\mathbf{y}}'). \tag{10.53}$$

This is a bit hard to interpret, so let's break down $\mathbf{v}'$ into polar coordinates in the $x'y'$ plane of the inertial frame S′,

$$\mathbf{v}' = v'_r\hat{\mathbf{r}} + v'_\theta\hat{\boldsymbol{\theta}}, \tag{10.54}$$

where $\hat{\mathbf{r}}' = \hat{\mathbf{r}}$ and $\hat{\boldsymbol{\theta}}' = \hat{\boldsymbol{\theta}}$ between S and S′. In this case, the Coriolis force is given by

$$\mathbf{F}_{\text{Cor}} = -2m\boldsymbol{\omega} \times (v'_r\hat{\mathbf{r}} + v'_\theta\hat{\boldsymbol{\theta}}). \tag{10.55}$$

Because $\hat{\mathbf{z}} \times \hat{\mathbf{r}} = \hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{z}} \times \hat{\boldsymbol{\theta}} = -\hat{\mathbf{r}}$,

$$\mathbf{F}_{\text{Cor}} = 2m\omega(v'_\theta\hat{\mathbf{r}} - v'_r\hat{\boldsymbol{\theta}}). \tag{10.56}$$

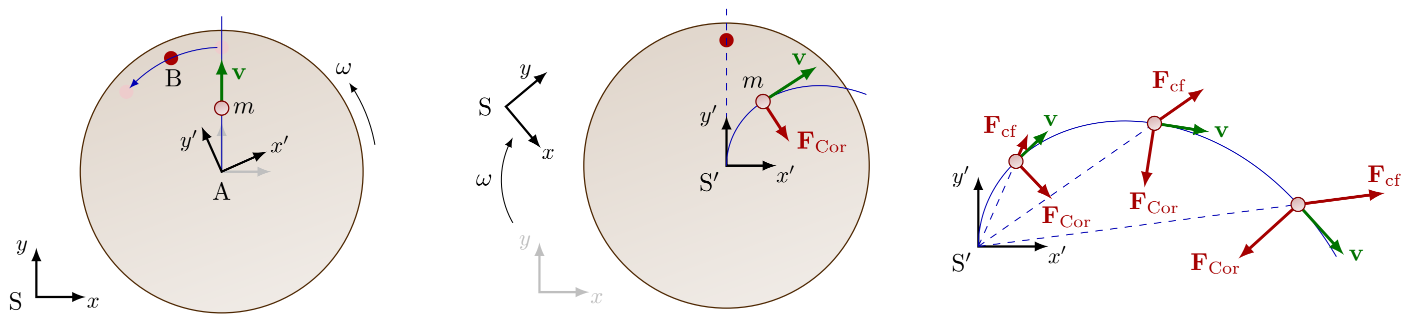
10.3.6 Example 1: Throwing a ball on rotating disk

Suppose Alice stands in the middle of a large rotating disk and throws a ball to Bob standing at the disk's edge. When she throws the ball at $t = 0$, it will follow a straight path in the inertial frame S to where Bob was at $t = 0$, if the ball moves along the y axis,

$$\mathbf{r}(t) = vt\hat{\mathbf{y}}. \tag{10.57}$$

By the time the ball reaches the edge, Bob has already rotated away as in Fig. 10.5a. From the perspective of both Alice and Bob in the rotating frame S′, the ball follows a curved path as in Fig. 10.5b. In S′, the ball seems to experience a mysterious force, namely the Coriolis force. The coordinates of $\mathbf{r}(t)$ in S′ are given by

$$\mathbf{r}'(t) = vt\sin(\omega t)\hat{\mathbf{x}}' + vt\cos(\omega t)\hat{\mathbf{y}}', \tag{10.58}$$



(a) Mass m moves in a straight line in inertial frame S.

(b) Mass m moves in a spiral in rotating frame S′.

(c) The spiral motion in S′ is caused by the centrifugal and Coriolis force.

Figure 10.5: Alice (A) throws a ball of mass m from the center of a disk to Bob (B) at the edge of the disk. Because its velocity is nonzero in the rotating frame S′, a Coriolis force will cause a deflection in the rotating S′.

or in polar coordinates,

$$\mathbf{r}'(t) = vt\hat{\mathbf{r}}' - \omega t\hat{\boldsymbol{\theta}}', \tag{10.59}$$

which describes a clockwise spiral in S′, i.e. a uniform circular motion $\theta = \omega t$, except that the radius $r = vt$ increases with time. The deflection is caused by the Coriolis force $\mathbf{F}_{\text{Cor}}$, which is perpendicular the $\mathbf{v}$. At the same time there is the centrifugal force $\mathbf{F}_{\text{cf}}$ which seems to radially increase the velocity, as shown in Fig. 10.5c.

10.3.7 Example 2: A ball at rest

What if the ball simply stays at rest at some radial distance r from the origin in the inertial frame S? Without loss of generality, say its coordinates are given by

$$\mathbf{r}(t) = r\hat{\mathbf{x}} = r\hat{\mathbf{r}} \tag{10.60}$$

and $\mathbf{v} = 0$ in S, and

$$\mathbf{r}'(t) = r\cos(\omega t)\hat{\mathbf{x}}' + r\sin(\omega t)\hat{\mathbf{y}}' = r\hat{\mathbf{r}}' \tag{10.61}$$

in the rotating frame S′, meaning the ball appears to rotate clockwise around the origin. The velocity in S′ is given by

$$\mathbf{v}'(t) = -r\omega\sin(\omega t)\hat{\mathbf{x}}' + r\omega\cos(\omega t)\hat{\mathbf{y}}' \tag{10.62}$$

$$= -r\omega\hat{\boldsymbol{\theta}}', \tag{10.63}$$

such that it is tangential to the position $\mathbf{r}$, and has size $v = r\omega$, as one would expect for a uniform circular motion. Note this is also consistent with the transport theorem Eq. (10.42) for $\mathbf{r}$. The velocity $\mathbf{v}'$ is nonzero in S′, so there is a Coriolis force (10.56). In this case it is in the negative radial direction

$$\mathbf{F}_{\text{Cor}} = -2m\omega^2 r\hat{\mathbf{r}}', \tag{10.64}$$

while the centrifugal force is

$$\mathbf{F}_{\text{cf}} = m\omega^2 r\hat{\mathbf{r}}'. \tag{10.65}$$

Putting this together in the second law in S′,

$$m\mathbf{a} = \mathbf{F}_{\text{Cor}} + \mathbf{F}_{\text{cf}} = -m\omega^2 r\hat{\mathbf{r}}', \tag{10.66}$$

which is consistent with the centripetal force (5.43) for uniform circular motion.

10.3.8 Example 3: Centrigufal force on Earth

The Earth spins, and is therefore not an non-inertial reference frame. Assuming the rotation is constant over time, there will be a Coriolis and centrifugal force, but no Euler force. Let's compare the former two depending where you are on Earth. Because of symmetry, the only important piece of information is your latitude θ. Anywhere at the equator, $\phi = 0$, while at the North Pole it is $\phi = 90°$ and at the South Pole it is $\phi = -90°$. The centrifugal force only depends on the radial distance $r = R\cos\phi$ from the rotation axis, where $R \approx 6370\,\text{km}$ is Earth radius. So the centrifugal force is

$$F_{\text{cf}} = m\omega^2 r = m\omega^2 R\cos\phi, \tag{10.67}$$

where $\omega \approx 2\pi/1\,\text{day} = 7.27 \times 10^{-5}\,\text{s}$. The centrifugal force is zero at the poles, and biggest at the equator, where for 1 kg

$$F_{\text{cf}} = m\omega^2 R = 0.034\,\text{N}.$$

The centrifugal acceleration that opposes the gravitational field g is

$$a_{\text{cf}} = \omega^2 R = 0.034\,\frac{\text{m}}{\text{s}}.$$

In other words, if at the poles weight is measured as simple mg, then the difference in measured weight at the equator is

$$-\frac{F_{\text{cf}}}{mg} = -\frac{\omega^2 R}{g} = -0.35\%$$

In reality the difference is closer to 0.5%, as the Earth is not a perfect sphere. Due to the centrifugal force, the Earth is "flattened" a bit, and looks more like oblate spheroid. As a consequence, the gravity is stronger at the poles, which are closer to Earth's center (about 6378 km) than at the equator (about 6357 km).

10.3.9 Example 4: Coriolis force on Earth

How does the centrifugal force compare tot the Coriolis force? This force is maximum when the velocity is perpendicular to the rotation axis, $\mathbf{v}' \perp \boldsymbol{\omega}$, when

$$F_{\text{Cor}} = 2m\omega v \tag{10.68}$$

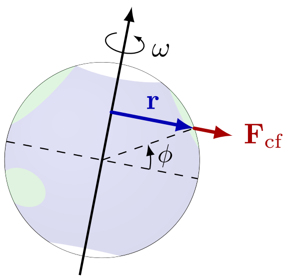
So for a 1 kg object moving at an appreciable speed of 100 km/h,

$$F_{\text{Cor}} = 0.0040\,\text{N},$$

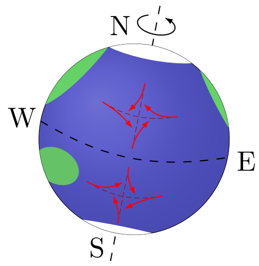
almost ten times smaller than the centrifugal force for the same mass.

Still, the Coriolis effect has a large impact on the directions of things like winds and ocean currents. In the Northern Hemisphere, a mass moving from west to east will experience a southward deflection, and vice verse, a mass moving from east to west will notice a northward deflection. Similar for longitudinal north-south travel. This will cause winds on the Northern Hemisphere to spiral counterclockwise as in Fig. 10.6b. On the Southern hemisphere it will be clockwise instead. Despite what you may have seen on the Simpsons, this effect is not strong enough to determine the sense of a vortex in a kitchen drain of toilet, as other initial condition can influence the direction.

Interestingly, this effect is also noticeable on the railroad tracks for trains moving longitudinally: A train moving from north to south in the Northern Hemisphere will put more force on the west track (right in facing the direction of motion), while a train moving



(a) The centrigufal force on Earth's surface depends on the latitude ϕ.



(b) The Coriolis force depends on which direction you move.

Figure 10.6: Earth rotates, so pseudo forces arise.

from south to north will put more force on the east track (again right). This means that right tracks in the Northern Hemisphere will tend to wear out faster than the left one. A similar effect is seen in the shape of river beds.

Another phenomena that the Coriolis effect influences are jet streams. As hot air moves from the high to low pressure zones, it tends to be deflected eastward in both hemispheres, creating jet stream between 9 and 16 km. These jet streams have a large impact on flight times. Eastward planes pick up a tail wind and arrive quicker than without, while westbound planes take longer. For example, planes flying from New York to London take roughly seven hours, while the return takes about eight.

Chapter 11

Stress & Strain

Up until now, we have implicitly assumed that solid objects are not deformed under force. In reality, even hard solids can be deformed and have some elastic and plastic properties.

A force or pressure that deforms an object, is called *stress* σ. Under stress, an object will deform, which is measured by *strain* ϵ.

Figure 11.1 illustrates several types of strains. For example, when you push on either side of a beam, you will cause it to compress. If you instead twist it with torques, you create torsion.

Each type of stress and strain will have its own measure of resistance to deformation, which typically follows the formula

Generalized Hooke's law.

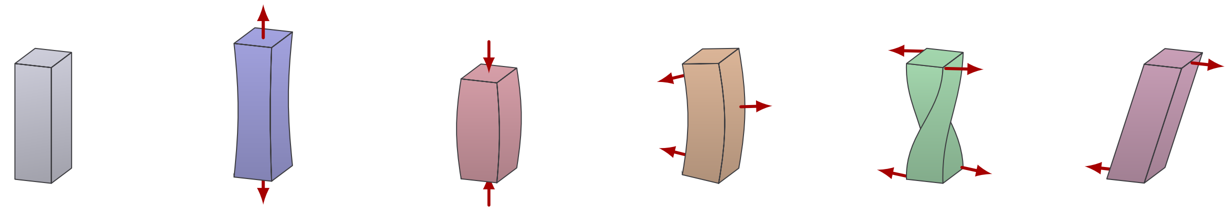
$$\textit{elastic modulus} = \frac{\textit{stress}}{\textit{strain}} = \frac{\sigma}{\epsilon}. \tag{11.1}$$

Note that form this similar to Hooke's law

$$k = \frac{F}{\Delta x}, \tag{11.2}$$

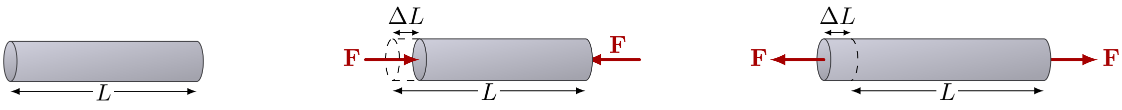
where the spring constant k is the resistance a spring offers to change in length Δx (strain) under a force F (stress). In this course we see

- *Young's modulus*: solid's resistance to change in length;
- *Shear modulus*: solid's resistance to shearing;
- *Bulk modulus*: solid or fluid's resistance to change in volume (Sections 13.5.2 and 16.3).



(a) No stress **(b)** Tension **(c)** Compression **(d)** Bending **(e)** Torsion **(f)** Shearing

Figure 11.1: Illustration of different types of strain on a solid under stress.



(a) No stress **(b)** Compression **(c)** Tension

Figure 11.2: Closer look at the compression and tension of a solid beam of rest length L.

11.1 Young's modulus

Imagine you have a rod of length L. If we apply a force on the ends with areas A, how much does the length change? As a measure of a given solid's compressibility along its length, we introduce the Young's modulus Υ as

$$\frac{F}{A} = \Upsilon \frac{\Delta L}{L}, \quad (11.3)$$

where F is the force applied, A is the area and ΔL is the change of length. (Υ is the Greek symbol for upsilon.) In other words,

Young's modulus.

$$\Upsilon = \frac{F/A}{\Delta L/L} = \frac{stress}{strain}. \quad (11.4)$$

It has units N/m^2 or Pascal like pressure (see Chapter 16), and varies per material. Some values are listed by Table 11.1. Note that Young's modulus also measures the *tensile strength*, i.e. the material's resistance to pulling by a tension force.

A lot of materials behave similar qualitative behavior under stress that is illustrated in the stress-versus-strain plot shown in Fig. 11.3. We see that for small strain and stresses, the relationship is linear, following Hooke's law Eq. (11.1). This regime is called the *elastic region*: Any deformation will be undone by removing the stress. After the *elastic limit*, the material the strain does not increases linearly anymore, but keeps increasing until its ultimate strength. This is the *plastic region*, where some of the deformation will be permanent, even after the stress is removed. Increasing the stress beyond this point will eventually lead to breaking the object.

Table 11.1: Young's modulus for several materials .

Material	Young's modulus Υ [GN/m^2 or GPa]
Brass	90
Bone	9
Stainless steel	180

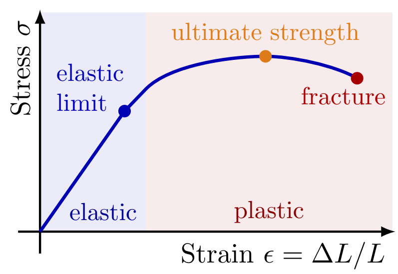


Figure 11.3: Stress-strain curve showing the qualitative behavior of many materials under stress. In the elastic region, Hooke's linear law is followed. After the elastic limit, in the plastic region, deformation caused by stress can become permanent.

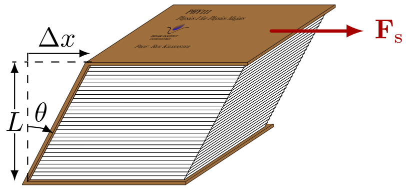


Figure 11.4: A book is sheared by a shearing force $\mathbf{F}_\mathrm{s}$.

11.2 Shear modulus

Let's say you put a book flat on a table top and push on top of it with a lot of pressure, as shown in Fig. 11.4. Now slowly add a horizontal force, such that the book starts to deform. This tangential force on an object is called a *shear force* $\mathbf{F}_\mathrm{s}$. It produces a *shear stress* to the object,

$$\sigma = \frac{F_\mathrm{s}}{A}, \tag{11.5}$$

and a shear strain

$$\epsilon = \frac{\Delta x}{L} = \tan\theta, \tag{11.6}$$

where Δx is displacement on the sheared top. The *shear modulus* then is defined as the resistance to shearing,

Shear modulus.

$$M_\mathrm{s} = \frac{F_\mathrm{s}/A}{\Delta x/L} = \frac{F_\mathrm{s}/A}{\tan\theta}. \tag{11.7}$$

Is is also known as the *torsion modulus*: If you apply a torque to an object, it will produce a twisting angle as in Fig. 11.1e.

Part II

Oscillations and Waves

Chapter 12

Harmonic Oscillations

12.1 Interlude: Taylor expansion

Physicists like to model the physical world by starting with simplifications and approximations. One mathematical tool indispensable to physicists is Taylor approximation. From analysis, we remember that

Taylor series. *The Taylor series of a function f that is infinitely differentiable in a is the power series*

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + ... \tag{12.1}$$

$$= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{12.2}$$

Accordingly, the *nth Taylor polynomial* is the polynomial of degree n,

$$T_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i. \tag{12.3}$$

T_1 is the tangent line to $f(x)$ at $x = a$ with slope $f'(a)$. Typically, the more terms you add, the better you approximate f around $x = a$, $f(x) \approx T_n(x)$. Also the closer you stay to $x = a$ (so small $|x - a|$), the smaller the difference $|T_n(x) - f(x)|$.

A special case of a Taylor series is when $a = 0$.

Maclaurin series.

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + ... \tag{12.4}$$

$$= \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}x^n \tag{12.5}$$

12.1.1 Example 1: Cubic function

As a simple example, consider a cubic function

$$f(x) = x^3 - x^2 - x + 1. \tag{12.6}$$

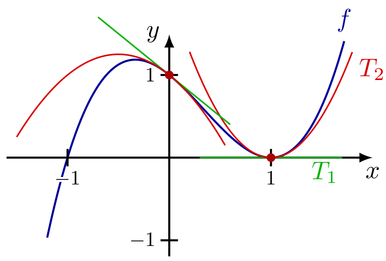


Figure 12.1: First (green) and second (red) degree Taylor polynomials that approximate $f(x) = x^3 - x^2 - x + 1$. T_1 is a straight line in $x = 0$ and $x = 1$, while T_2 is a parabola.

The first, second and third Taylor polynomials in $x = 0$ are

$$T_1(x) = 1 - x \tag{12.7}$$

$$T_2(x) = 1 - x - x^2 \tag{12.8}$$

$$T_3(x) = 1 - x - x^2 + x^3 \tag{12.9}$$

Notice that with $T_3(x)$ we retrieve the original function $f(x)$, because they are both third-degree polynomial. Similarly, in a different point, $x = 1$, they are

$$T_1(x) = 0 \tag{12.10}$$

$$T_2(x) = 2(x-1)^2 \tag{12.11}$$

$$T_3(x) = 2(x-1)^2 + (x-1)^3 \tag{12.12}$$

Notice that the last result $T_3(x)$ is again the exact same as $f(x)$ and Eq. (12.9).

12.1.2 Example 2: Sine

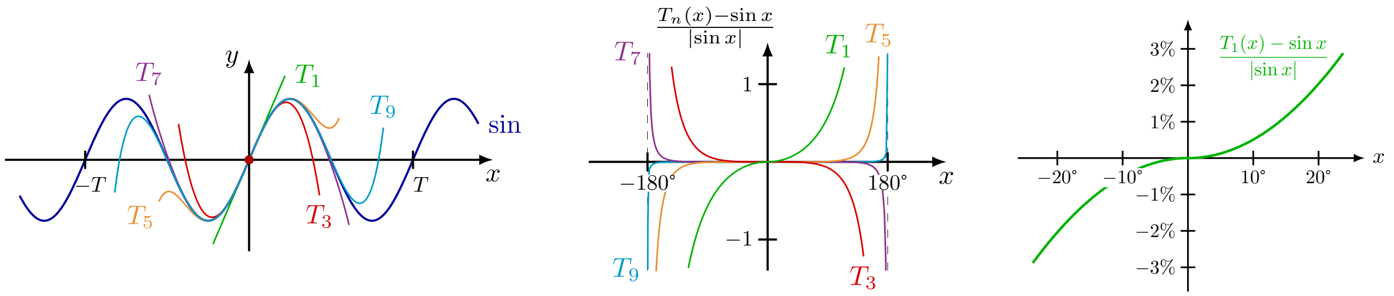
For the purposes of this course, it is important to know the Taylor expansion of sine:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} + \ldots \tag{12.13}$$

$$= \sum_n^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}. \tag{12.14}$$

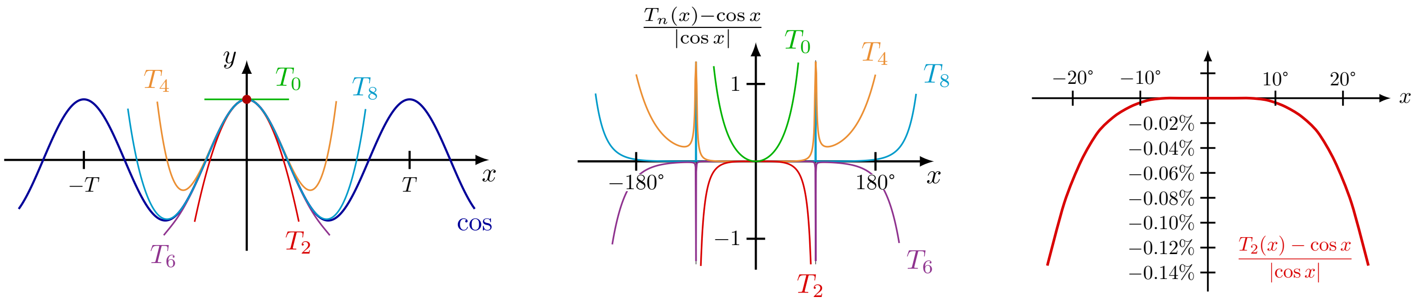
The series only has terms of odd degree. Several Taylor polynomials are shown in Fig. 12.2a. The first-order approximation of a sine function is a linear one,

$$\sin x \approx T_1(x) = x. \tag{12.15}$$



(a) Taylor polynomials T_n, see Eq. (12.13). **(b)** Relative difference with T_n. **(c)** Relative difference with T_1.

Figure 12.2: Taylor expansion approximating sine $f(x) = \sin x$. The first-order approximation $T_1(x) = x$ (green) is a straight line. All Taylor polynomials of an odd function like $\sin x$ have an odd degree.



(a) Taylor polynomials T_n, see Eq. (12.17). **(b)** Relative difference with T_n. **(c)** Relative difference with T_2.

Figure 12.3: Taylor expansion approximating $f(x) = \cos x$. The first-order approximation $T_2(x) = 1 - x^2/2$ (red) is a parabola. All the Taylor polynomials of $\cos x$ have an even degree.

They are shown in Fig. 12.1. To see how each terms improve the approximation, one can use as a measure the *relative difference* defined as

$$\frac{T_n(x) - \sin x}{|\sin x|}. \tag{12.16}$$

This is plotted in Fig. 12.2b. The closer to $x = 0$, the smaller the difference, the better the approximation; and the more terms, the larger the range around $x = 0$ where there is a good approximation. Clearly, the fifth degree polynomial T_5 is a better approximation to sin than a parabola T_3, which is in turn better than a straight line T_1. In physics, we often consider smaller values of x, and the first-order approximation T_1 would be enough. We will see the case of small-angle approximation below, but it is also to setup differential equations by expanding a function in terms of some infinitesimal value $\mathrm{d}x$.

12.1.3 Example 3: Cosine

Cosine is very similar, except now, only terms of even degree survive;

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} + ... \tag{12.17}$$

$$= \sum_{n}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}. \tag{12.18}$$

The lowest-degree polynomials are shown in Fig. 12.3a. So the first-order approximation of a cosine function is a quadratic function, and not a linear one!

12.2 Simple harmonic oscillator

Harmonic oscillators correspond to cases where acceleration is proportional to the displacement, and in the opposite direction. They appear in many places in physics, all the way from physical springs and pendulums, down to molecules and atoms.

Reconsider the mass on a frictionless surface and fixed to a wall by a spring discussed in Section 6.5 and shown in Fig. 6.5. If you pull or push the mass, the length of the spring changes, and the spring will try to bring the mass back to position $x = 0$. If you push or pull it from of its equilibrium position, and let it go, it will accelerate, gaining velocity that causes it to overshoot its rest position. In absence of other forces like friction, the spring will keep trying to return the mass to $x = 0$, but overshooting it each time, *oscillating* back and forth. By Hooke's law (6.28), the spring force is $F = -kx$, so

$$ma = -kx, \tag{12.19}$$

where x is the displacement of the mass with respect to the rest position, or equivalently, the extension of the spring, while k is the spring constant. Because $F = -kx$ is the only force, this is called a *simple harmonic oscillator*. We can rewrite this in terms of x as

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -\frac{k}{m}x, \tag{12.20}$$

or

Equation of simple harmonic motion.

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \omega^2 x = 0, \tag{12.21}$$

with "some" constant $\omega^2 = k/m$ for a spring-mass system. Remember the form of this equation, because we will see equations of this form appearing in many different problems in physics. In the Newton's dot notation (Section 4.4):

$$\ddot{x} + \omega^2 x = 0. \tag{12.22}$$

As the mass oscillates back and forth, the position $x = x(t)$ depends on time t. Equation 12.21 is a differential equation, which is the so-called *equation of motion* for x. So what is the solution? The solution needs to be a function $x(t)$ whose second derivative is itself times some negative constant $-\omega^2 = -k/m$. We know two such (real) functions! Cosine and sine. Let's use the ansatz (i.e. an educated guess)

$$x(t) = A\cos(\omega t), \tag{12.23}$$

where A is a constant called the *amplitude*, and ω is the *angular frequency* with units rad/s. Remember from Section 5.3 that $\omega = 2\pi f = 2\pi/T$ for *frequency* f and *period* T. The first derivative is the velocity,

$$v(t) = \frac{\mathrm{d}x}{\mathrm{d}t} = -A\omega\sin(\omega t), \tag{12.24}$$

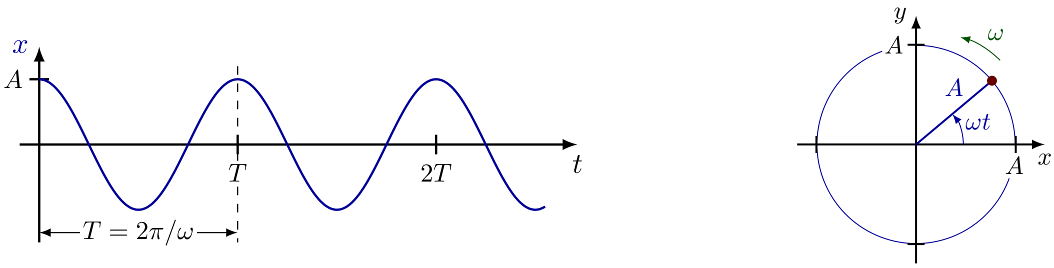
and the second one is the acceleration

$$a(t) = \frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -A\omega^2\cos(\omega t) = -\omega^2 x(t). \tag{12.25}$$

We see that our ansatz solves Eq. 12.21. Comparing Eq. 12.21 to Eq. 12.20, we see that the constant ω is actually the angular frequency, which we choose to be positive, and which relates to the spring constant and mass as in,

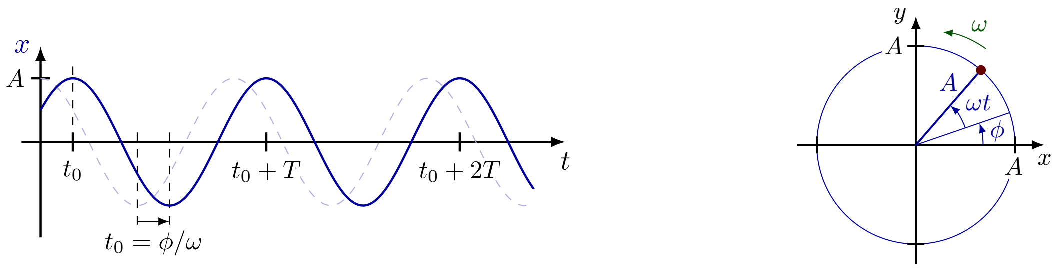
Angular frequency of a harmonic oscillator.

$$\omega = \sqrt{\frac{k}{m}}. \tag{12.26}$$



(a) $x(t) = A\cos(\omega t)$. **(b)** A point moving uniformly in a circle.

Figure 12.4: Simple harmonic oscillator with amplitude A, uniform angular frequency ω and period $T = 2\pi/\omega$.



(a) $x(t) = A\cos(\omega t - \phi)$. **(b)** A point moving uniformly in a circle.

Figure 12.5: Simple harmonic oscillator with a non-zero phase shift.

We can also convert this result to the period of oscillation with units of seconds,

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{k}{m}}. \tag{12.27}$$

This is shown in Fig. 12.4a. Note that our solution Eq. (12.23) assumes that $x(0) = A$, i.e. the oscillator starts at maximum displacement when it is released from rest. This is our *initial condition* that determines this arbitrary constant. Another solution would be to choose a different starting time. So a more general solution includes some phase ϕ:

Solution of a simple harmonic oscillator.

$$x(t) = A\cos(\omega t - \phi). \tag{12.28}$$

This phase is like a shift $t_0 = \phi/\omega$ in time, shown in Fig. 12.5a,

$$x(t) = A\cos\big[\omega(t - t_0)\big]. \tag{12.29}$$

12.2.1 Initial conditions

So at $t = 0$, the oscillation in Eq. (12.23) starts with

$$\begin{cases} x(0) = A\cos\phi & (12.30) \\ v(0) = A\omega\sin\phi & (12.31) \end{cases}$$

Equation (12.28) has two arbitrary constants, A and ϕ, while $\omega = \sqrt{k/m}$ is determined by the properties of the spring-mass system. The two unknown constants are determined by two independent *initial conditions*. Typically this is a constraint of the form $x(t_0) = x_0$ and $v(t_0) = v_0$ with constant t_0, x_0 and v_0. A trivial example of a set of intitial conditions is given by

$$\begin{cases} x(0) = x_0 & (12.32) \\ v(0) = 0 & (12.33) \end{cases}$$

in which case $A = x_0$ and $\phi = 0$, and we find Eq. (12.23) again. A different simple example is

$$\begin{cases} x(0) = 0 & (12.34) \\ v(0) = v_0 & (12.35) \end{cases}$$

such that $A = v_0/\omega$ and $\phi = \pi/2$.

12.2.2 General solution

It is immediately obvious from Eq. (12.28) that sine is also a valid solution because $\sin(x) = \cos(x + \pi/2)$. Therefore, one way to think of a harmonic oscillator is as the vertical and horizontal projections of a point moving uniformly in a circle, as shown in Figs. 12.4b and 12.5b. Mathematical analysis says that if you have two linearly independent solutions of a second-order differential equation, like $A\cos(\omega t)$ and $B\sin(\omega t)$ of Eq. (12.21), then the most general solution is the linear combination of these two equations:

General solution of a simple harmonic oscillator.

$$x(t) = A\cos(\omega t) + B\sin(\omega t). \tag{12.36}$$

However, if the amplitudes A and B, and ω are constant, you can always rewrite this linear combination as just one cosine or sine with some phase ϕ by using trigonometric identities like $\cos(x - y) = \cos x \cos y + \sin x \sin y$, setting $A = A_0 \cos\phi$ and $B = A_0 \sin\phi$,

$$A\cos(\omega t) + B\sin(\omega t) = (A_0\cos\phi)\cos(\omega t) + (A_0\sin\phi)\sin(\omega t) \tag{12.37}$$

$$= A_0\cos(\omega t - \phi) \tag{12.38}$$

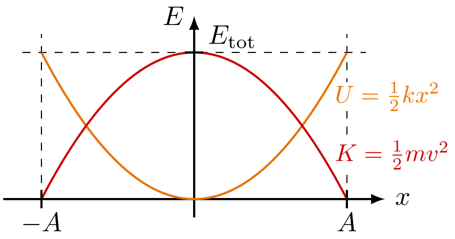
The maximum displacements are at $x = \pm A$. In our derivation above, we saw that the velocity and acceleration also have oscillatory behavior, according to sine or cosine. A comparison of the phase of the position, velocity and acceleration is shown earlier in Fig. 4.5g. The maximum velocity is $\pm A\omega$, while the maximum acceleration is $\pm A\omega^2$. The velocity is $\pi/2$ out of phase with the position. On the other hand, the acceleration is a full π out of phase with the position, which means that it always points to the opposite direction than the position vector, just like the force exerted by the spring.

Section 12.6 will show how the complete set of solutions and initial conditions can be visualized in a phase diagram.

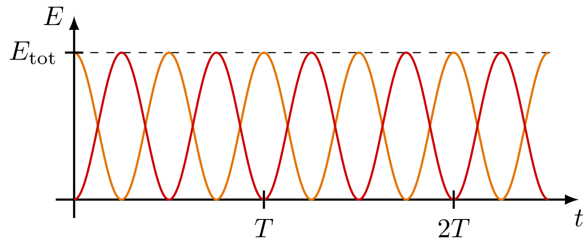
12.2.3 Energy of a harmonic oscillator

From Section 7.4.2, the potential energy is

$$U(t) = \frac{kx^2}{2} = \frac{kA^2}{2}\cos^2(\omega t), \tag{12.39}$$



(a) Versus the degree of freedom, position x.



(b) Versus time.

Figure 12.6: Energy $E = U + K$ of a harmonic oscillator with period T.

if the position $x(t) = A\cos(\omega t)$. The total energy of the mass-spring system has to stay constant if we neglect non-conservative forces like friction, so the potential and kinetic energy both oscillate in time and are out of phase with respect to each other:

$$E_{\text{tot}}(t) = \frac{kx^2}{2} + \frac{mv^2}{2} \tag{12.40}$$

$$= \frac{kA^2}{2}\cos^2(\omega t) + \frac{m\omega^2 A^2}{2}\sin^2(\omega t). \tag{12.41}$$

Notice that the total energy is indeed constant,

$$E_{\text{tot}} = \frac{kA^2}{2} = \frac{m\omega^2 A^2}{2} \tag{12.42}$$

and is proportional to the amplitude squared, $E_{\text{tot}} \propto A^2$. The maximum kinetic energy is when $x = 0$ and $v = \pm\omega A$, as expected. This is shown in Fig. 12.6.

12.2.4 Vertical harmonic oscillator

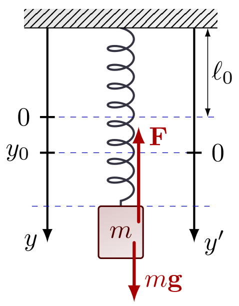
Now remember the mass hanging on a spring from the ceiling from Section 6.5 and Fig. 6.6. This is shown in Fig. 12.7. If the weight and spring force balance,

$$0 = -mg + ky \tag{12.43}$$

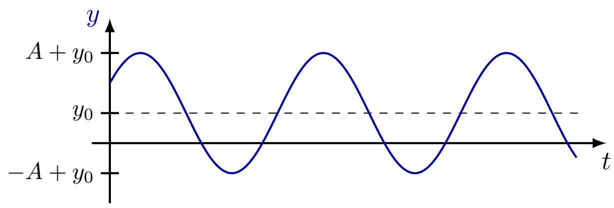
where y is the extension of the spring, with respect to its normal rest length of ℓ_0. Because gravity pulls down the mass, the spring will be extended by a constant

$$y_0 = \frac{mg}{k}. \tag{12.44}$$

So at equilibrium and rest, $y = y_0$, and the new rest length of the spring is $\ell_0 + y_0$. But if we extend the spring even more by some distance y', the total extension is $y = y_0 + y'$.



(a) Vertical spring.



(b) Oscillation around offset y_0: $y(t) = y_0 + A\cos(\omega t)$.

Figure 12.7: Spring with rest length ℓ_0 is hung vertically. Because of the weight, the new rest length is $\ell_0 + y_0$.

When we let go the mass from rest, Newton's second law becomes

$$m\frac{\mathrm{d}^2y}{\mathrm{d}t} = -mg + ky, \tag{12.45}$$

or in terms of $y' = y - y_0$,

$$m\frac{\mathrm{d}^2y'}{\mathrm{d}t} = -mg + k(y' + y_0) = -ky'. \tag{12.46}$$

So we again find a simple harmonic oscillation, but this time around $y = y_0$, instead of around $y = 0$. The solution is for the form

$$y(t) = y_0 + y'(t) = y_0 + A\cos(\omega t - \phi). \tag{12.47}$$

12.2.5 Double spring

Consider a mass on a frictionless surface connected to two springs on either ends as in Fig. 12.8. Assume the springs have spring constants k_1 and k_2, respectively, and at equilibrium they are at their rest length ℓ_0 (i.e. neither compressed nor extended). Any displacement x by the mass from its rest position $x = 0$ causes a force imbalance given by

$$ma = -k_1x - k_2x. \tag{12.48}$$

If we move the mass to the right by $x > 0$, spring k_1 gets extended, while k_2 is compressed, and vice versa. Their forces, therefore, always point in the same x direction. Notice that the effective spring constant of the whole system is $k_{\text{eff}} = k_1 + k_2$. The equation of motion are

$$\frac{\mathrm{d}^2x}{\mathrm{d}t} + \frac{k_1 + k_2}{m}x = 0. \tag{12.49}$$

So this acts as a simple harmonic oscillator with angular frequency

$$\omega = \sqrt{\frac{k_1 + k_2}{m}}. \tag{12.50}$$

What if the springs are not at their own rest length when the system is at its rest position? Remember the example in Section 6.5.2 where we found the rest position of the system is given by $x_1 = -(k_2/k_1)x_2$. This time, if the mass is displaced by x from its rest position $x = 0$,

$$ma = -k_1(x + x_1) - k_2(x + x_2). \tag{12.51}$$

Because of Eq. (6.32), this reduces to Eq. (12.48), so the system has the same dynamics as before.

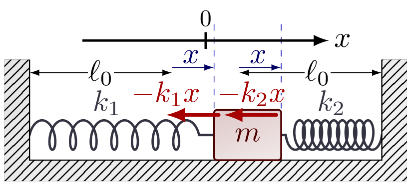
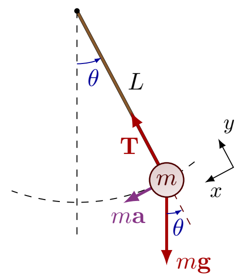
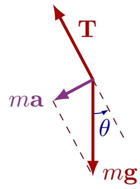


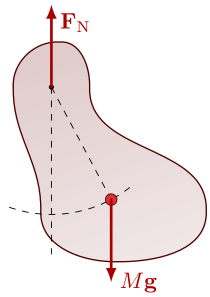
Figure 12.8: Double oscillator: Two springs with constants k_1 and k_2.



(a) Simple pendulum with mass on a string.



(b) Forces and acceleration vectors.



(c) Physical pendulum with mass m and moment of inertia I.

Figure 12.9: A pendulum is a harmonic oscillator.

12.3 Pendulum

Another physicist's favorite example of a harmonic oscillator is a pendulum. A mass is suspended by a string of fixed length L to the ceiling. At equilibrium, the mass just hangs vertically at rest. The string makes an angle $\theta = 0$ with the vertical. When we pull or push the mass from its rest position, gravity will pull the mass back towards its rest position, and the mass will oscillate back and forth.

Let's decompose the forces on the mass. It is easier to consider forces in the radial and tangential directions, instead of vertical and horizontal ones, as in Fig. 12.9a. The total force is

$$\sum \mathbf{F} = \mathbf{T} - m\mathbf{g}. \tag{12.52}$$

The mass will only swing in the tangential direction, so the forces must balance radially:

$$\begin{cases} 0 = T - mg\cos\theta & (12.53) \\ ma = -mg\sin\theta & (12.54) \end{cases}$$

The degree of freedom is the arc length $s = L\theta$ from the rest position $\theta = 0$, so

$$a = \frac{\mathrm{d}^2 s}{\mathrm{d}t} = -g\sin\theta. \tag{12.55}$$

This is a *nonlinear differential equation* due to the sine function, which is not easy to solve... We therefore use a common trick, called the *small-angle approximation*: We assume that the angle θ is small, such that we can use the Taylor expansion Eq. (12.15),

> **Small-angle approximation.**
> $$\sin\theta \approx \theta. \tag{12.56}$$

From Fig. 12.2c, we see that this is a very good approximation. The relative error is about

$$\frac{10^\circ - \sin 10^\circ}{|\sin 10^\circ|} \approx 0.51\%, \tag{12.57}$$

where $10^\circ \approx 0.1754$. The error is about 1.0% for 14°, and 2.1% for 20°.

Now we can rewrite Eq. (12.55), using $s = L\theta$ and assuming small θ we find

Linearized equation of motion for a pendulum.

$$\frac{\mathrm{d}^2 s}{\mathrm{d}t} + \frac{g}{L}s = 0. \tag{12.58}$$

We recognize this equation as the one for simple harmonic oscillators, Eq. (12.21)! So for small angles, a pendulum behaves as a simple harmonic oscillator. The solution must be of the form

$$s(t) = s_0 \cos(\omega t - \phi), \tag{12.59}$$

with the amplitude $s(0) = s_0$, and an angular frequency

Angular frequency of a pendulum.

$$\omega = \sqrt{\frac{g}{L}}. \tag{12.60}$$

Therefore, the period is

$$T = 2\pi\sqrt{\frac{L}{g}}. \tag{12.61}$$

Notice that this time, the period does not depend on the mass m, but only on the length L of the string, and the gravitational acceleration g.

Perhaps surprisingly, the period does not depend on the initial conditions; the amplitude s_0 which is how high you let it go. However, it is clear that the larger the amplitude s_0, the larger the maximum velocity $v_{\mathrm{max}} = \pm\omega s_0$ at $s = 0$ (or $\theta = 0$). The property that the period of a pendulum (for small angles) is independent of its amplitude is called *isochronism*, and was discovered by Galileo Galilei around 1602. This makes pendulums particularly useful as timekeepers. The first pendulum clock was designed by Christiaan Huygens in 1656, and this was one of the most accurate timekeeping technology all the way until the 1930s.

Notice that the angle $\theta = s/L$ also oscillates harmonically,

$$\theta(t) = \theta_0 \cos(\omega t - \phi), \tag{12.62}$$

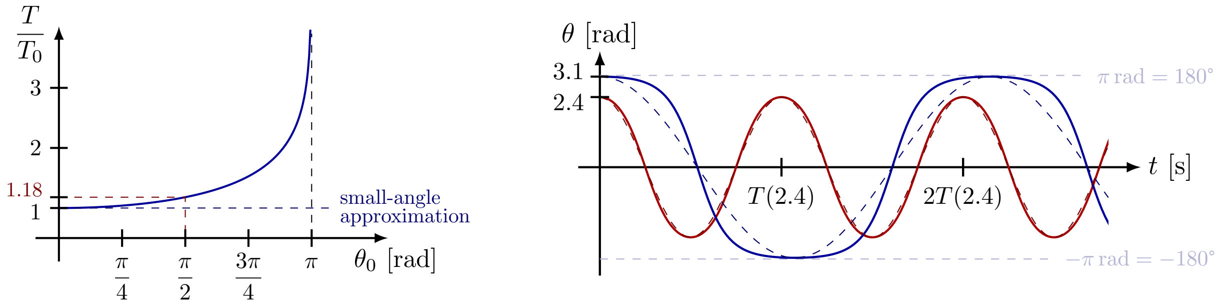
where the pendulum was let go from rest at $\theta(0) = \theta_0$ at $t = 0$.

Similarly, the tension also has a harmonic oscillation below the minimum tension $T_0 = mg$ at $\theta = 0$. By using the small angle approximation for Eq. (12.53),

$$T(t) = mg\left(1 - \frac{\theta_0^2}{2}\cos^2(\omega t - \phi)\right). \tag{12.63}$$

12.3.1 Extra: Exact solution

The exact solution of the original nonlinear pendulum equation (12.55) falls outside the scope of this course. It is worth noting however, that it is not isochronous, meaning the frequency does depend on the amplitude θ_0: As a pendulum clock winds down due to loss of energy, the frequency decreases ever so slightly. However, it turns out that the harmonic solutions using the small angle approximation hold surprisingly well. Figure 12.10a shows



(a) Ratio of real period $T(\theta_0)$ and approximate, constant period $T_0 = \sqrt{L/g}$. At $\theta_0 = \pi/2\,\mathrm{rad} = 45°$ the difference is 18%.

(b) Comparison of exact (solid line) and a cosine with the same period (dashed line). At large angles, the solution becomes more "square" than cosine, and has a longer period.

Figure 12.10: The exact solution of pendulum with period $T(\theta_0)$ depends on the amplitude θ_0, while the solution of the small-angle approximation has a constant period $T_0 = \sqrt{L/g}$.

that even for an angle $\theta_0 < \pi/6rad = 30°$, the difference between the real period $T = T(\theta_0)$ and the period $T_0 = \sqrt{L/g}$ according to the small-angle approximation is less than 1.8%. At the most extreme angle amplitude of $\theta_0 = \pi\,\mathrm{rad} = 180°$, the difference is almost fourfold! But of course, small differences add up over time. So the smaller the initial amplitude of a pendulum clock, the smaller this effect, and thus the more accurate the time.

Besides the period, the shape of the exact solution still looks quite like a cosine, even for an angle amplitude up to $\theta_0 = 2.4\,\mathrm{rad} \approx 138°$ as shown in Fig. 12.10b. For larger amplitudes, besides much longer periods, the curve becomes more "square".

12.3.2 Physical pendulum

What about the case of a body of arbitrary shape? Say a body is fixed in some point and allowed to swing like a pendulum, Fig. 12.9c. There is a torque on the center-of-mass due to its weight $F = mg\sin\theta$:

$$I\alpha = L(mg\sin\theta). \tag{12.64}$$

So using the small-angle approximation, the equation of motion becomes

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t} + \frac{Lmg}{I}\theta = 0. \tag{12.65}$$

This is once again a simple harmonic oscillator with an angular frequency

Angular frequency of a pendulum.

$$\omega = \sqrt{\frac{Lmg}{I}}. \tag{12.66}$$

In the trivial example of a single point mass m at the end of the pendulum, the moment of inertia is $I = L^2 m$, and we retrieve Eq. (12.60).

12.4 Damped harmonic oscillators

The previous sections discussed some ideal simple harmonic oscillators that have a constant energy, but in reality there frictional forces that cause the oscillation to "die out". The simplest example is again a mass-spring system, but where this time there is a drag force F_d that is opposite to the velocity. It is typically of the form

$$F_\mathrm{d} = -bv, \tag{12.67}$$

where it is opposite and proportional to the velocity, and there is some *drag constant* b with units N s/m = kg/s. Newton's second law is opposite to the velocity. It is typically of the form

$$ma = -bv - kx, \tag{12.68}$$

so the new equation of motion is slightly modified to

Equation of damped harmonic motion.

$$\frac{\mathrm{d}^2x}{\mathrm{d}t} + \frac{b}{m}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{k}{m}x = 0. \tag{12.69}$$

Or in dot notation,

$$\ddot{x} + \frac{b}{m}\dot{x} + \frac{k}{m}x = 0. \tag{12.70}$$

If $b = 0$, this reduces to a simple harmonic oscillation Eq. (12.21).

So what is the solution of this equation? Qualitatively, we expect some solution that oscillates, but with an amplitude that is damped over time. It should still oscillates with some frequency ω similar to the undamped frequency $\omega_0 = \sqrt{k/m}$ called the *natural angular frequency* or *resonant frequency*,

$$\omega \sim \omega_0 = \sqrt{\frac{k}{m}}. \tag{12.71}$$

12.4.1 Energy loss

Because the amplitude decreases, the energy must decrease with time. The loss of energy is due to the drag force,

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \mathbf{F}_\mathrm{d} \cdot \mathbf{v} = -bv^2 < 0. \tag{12.72}$$

We haven't yet shown how $v(t)$ depends with time, but we can instead look at the average amount of energy lost in one cycle (one period T). In a given cycle, the average kinetic energy is given by some average speed,

$$\overline{K} = \left\langle \frac{mv^2}{2} \right\rangle = \frac{m}{2}\left\langle v^2 \right\rangle. \tag{12.73}$$

Due to conservation of energy, the total energy E is the sum of the average potential energy $\overline{U}$ and average kinetic energy $\overline{K}$.

$$E = \overline{U} + \overline{K} \tag{12.74}$$

Because the potential and kinetic energies oscillate out of phase of each other, we can safely assume that their averages are about the equally same,

$$E = \frac{1}{2}m\left\langle v^2 \right\rangle + \frac{1}{2}m\left\langle v^2 \right\rangle = m\left\langle v^2 \right\rangle. \tag{12.75}$$

So approximating $v^2 \approx \langle v^2 \rangle$, we write Eq. (12.72) as

$$\frac{\mathrm{d}E}{\mathrm{d}t} \approx -b\langle v^2 \rangle = -b\frac{E}{m}. \tag{12.76}$$

This is a differential equation, where the derivative of the solution $E(t)$ is again itself, but with an extra coefficient $-b/m$. This is like an exponential! If we integrate

$$\int \frac{\mathrm{d}E}{E} = \int -\frac{b}{m}\,\mathrm{d}t, \tag{12.77}$$

we find that the solution is indeed an exponential

Energy of a damped harmonic oscillation.

$$E(t) = E_0 e^{-\frac{b}{m}t}, \tag{12.78}$$

where E_0 is some integration constant given by the initial condition $E(0) = E_0$. Because the exponential's argument is negative, there is an exponential decay with time constant $\tau = m/b$. This is the time τ needed to decrease the energy by a factor of $1/e$. If the damping is small enough,

$$\frac{\Delta E}{E} = -\frac{b}{m}T, \tag{12.79}$$

where ΔE is the amount of energy lost in one period T. And $\Delta E/E$ is the fractional energy lost per period.

12.4.2 Quality factor

Damping is often quantified by a dimensionless quantity Q called the *quality factor* or *Q factor*

Quality factor.

$$Q = \frac{2\pi E}{|\Delta E|}, \tag{12.80}$$

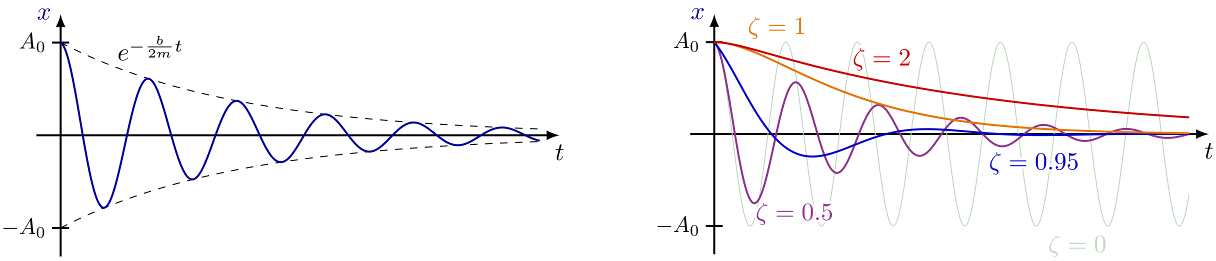
such that the fractional energy lost is

$$\frac{|\Delta E|}{E} = \frac{2\pi}{Q}. \tag{12.81}$$

Comparing this to Eq. (12.79) for a damping force $F_\mathrm{d} = -bv$,

Quality factor for damping drag force.

$$Q = \frac{2\pi m}{bT} = \frac{2\pi\tau}{T}. \tag{12.82}$$



(a) Underdamped oscillation with $\zeta < 1$ and envelope $A(t) = A_0 e^{-bt/2m}$.

(b) Critically damped $\zeta = 1$, overdamped $\zeta > 1$.

Figure 12.11: Damped oscillation with damping ratio $\zeta = b/b_c$.

12.4.3 Underdamping

Remember from Eq. (12.42), that the energy for a simple harmonic oscillator is proportional to the amplitude squared. So at $t = 0$, $E(0) = E_0 \propto A_0^2$, while at some later time t, $E(t) \propto A(t)^2$, where $A(0) = A_0$. Therefore,

$$\frac{E}{E_0} = \frac{A^2}{A_0^2} = e^{-\frac{b}{m}t}, \tag{12.83}$$

or

$$A(t) = A_0 e^{-\frac{b}{2m}t}. \tag{12.84}$$

This tells us that the amplitude of the harmonic oscillation is damped by an exponential decay, namely

Underdamped harmonic oscillation.

$$x(t) = A_0 e^{-\frac{b}{2m}t} \cos(\omega t - \phi). \tag{12.85}$$

The exponential term with time constant $\tau = 2m/b$ is called an envelope, as it modulates the amplitude of the oscillation, as shown in Fig. 12.11a.

The above derivation might seem a bit hand-wavy. Let's use Eq. (12.85) as an ansatz, and show that it indeed solves the equation of motion (12.69):

$$\begin{aligned} 0 =& A_0 e^{-\frac{b}{2m}t} \left(\frac{b^2}{4m^2} \cos(\omega t - \phi) + \frac{b}{m} \omega \sin(\omega t - \phi) - \omega^2 \cos(\omega t - \phi) \right) \\ &+ \frac{b}{m} A_0 e^{-\frac{b}{2m}t} \left(-\frac{b}{2m} \cos(\omega t - \phi) - \omega \sin(\omega t - \phi) \right) \\ &+ \frac{k}{m} \left(A_0 e^{-\frac{b}{2m}t} \cos(\omega t - \phi) \right). \end{aligned} \tag{12.86}$$

Simplyfing a bit,

$$0 = \left(\frac{k}{m} - \frac{b^2}{4m^2} - \omega^2 \right) \cos(\omega t - \phi). \tag{12.87}$$

And in fact, Eq. (12.85) is a valid solution if the sum between these parentheses vanishes. This condition provides us the formula for the angular frequency:

Angular frequency of an underdamped harmonic oscillation.

$$\omega = \sqrt{\frac{k}{m} - \frac{b^2}{4m^2}} = \sqrt{\omega_0^2 - \frac{1}{\tau^2}}, \tag{12.88}$$

where $\tau = 2m/b$ the time constant τ for the oscillation. Clearly, if the damping is small, $b \to 0$, then $\omega \to \omega_0$. Conversely, the larger b, the larger $\omega > \omega_0$. So in addition to an exponential decay, the oscillation is slowed down by a drag force opposing its motion, as expected.

Like before, another linearly independent solution is found by replacing cosine with sine in Eq. (12.85).

12.4.4 Critical damping and overdamping

But what if the damping is very large? At some point, Eq. (12.88) is not well defined anymore, namely when the argument of the square root becomes negative. The critical point where this is about to happen is given by the *critical damping coefficient*

Critical damping coefficient.

$$b_\mathrm{c} = 2m\omega_0 = 2\sqrt{km}, \tag{12.89}$$

for which there is no oscillation anymore, $\omega = 0$. The ratio $\zeta = b/b_\mathrm{c}$, given by the Greek letter "zeta", is the called the *damping ratio*. There are four different cases, or *regimes*, where you will have different solutions to Eq. (12.69):

- $b = 0$ or $\zeta = 0$: no damping;
- $0 < b < b_\mathrm{c}$ or $0 < \zeta < 1$: *underdamping*;
- $b = b_\mathrm{c}$ or $\zeta = 1$: *critical damping*;
- $b > b_\mathrm{c}$ or $\zeta > 1$: *overdamping*.

So far we have mostly assumed underdamping, where the damping is not too large such that there is still an oscillation given by Eq. (12.85) and Eq. (12.88) before the energy completely decays away. The regimes with larger damping, critical damping, and overdamping have different solutions $x(t)$ that will not be discussed here. The main thing we need to understand is that starting from critical damping, there is no oscillation anymore at all, and the system will just return to its rest position $x = 0$ at $t \to \infty$. These different cases are compared in the plot 12.11b. Notice that for overdamping it takes longer to return to $x = 0$, because the drag force really opposes any motion. In the extreme case where $b \to \infty$, or equivalently $\zeta \to \infty$, the system will stay at its initial position $x(0)$.

We will revisit damped oscillations in Section 14.3, where we will use complex numbers to solve differential equations.

12.4.5 Energy (revisited)

For an underdamped oscillation Eq. (12.85), it is clear that the energy is in fact given by Eq. (12.78):

$$E(t) = \frac{kx^2}{2} + \frac{mv^2}{2}, \tag{12.90}$$

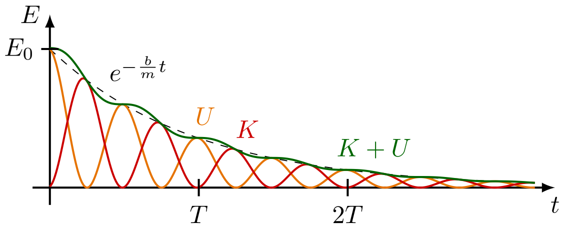


Figure 12.12: The potential (orange) and kinetic energy of an underdamped oscillation (red) have an envelope $E_0 e^{-bt/m}$, while the total energy (green) oscillates around the average $\overline{E}(t) = E_0 e^{-bt/m}$. Compare to Fig. 12.6.

which is not constant anymore as in Eq. (12.42) and in Fig. 12.12. The initial energy E_0 is given by the maximum extension $x = A_0$

$$E_0 = \frac{kA_0^2}{2}. \tag{12.91}$$

The total energy decays exponentially, but not quite. Since loss of energy depends on the velocity which oscillates,

$$v(t) = -A_0 e^{-\frac{b}{2m}t}\left(\frac{b}{2m}\sin(\omega t - \phi) + \omega\cos(\omega t - \phi)\right), \tag{12.92}$$

the loss of energy also oscillates: When the velocity is at its maximum in the oscillation, the loss of energy is highest, and when $v \approx 0$, the energy loss is small to zero, and the energy roughly constant. However, the average energy in each period will still follow an exponential decay, as shown by the green line in Fig. 12.12. So Eq. (12.78) does not hold exactly, but the average does.

12.5 Driven oscillation & resonance

In a damped oscillation, energy is lost. But we can "pump" energy into the system to counteract this energy loss. We consider a driving force of in the form of

$$F(t) = F_0 \cos(\omega t) \tag{12.93}$$

that oscillates harmonically with some fixed ω, that can be different than the natural angular frequency ω_0. The relation between ω_0 and ω is measured by the *Q factor*

Q factor for driven oscillation.

$$Q = \frac{\omega_0}{|\omega_0 - \omega|} = \frac{\omega_0}{|\Delta\omega|}. \tag{12.94}$$

The smaller $\Delta\omega$, the larger Q, and the more power $P = \mathrm{d}E/\mathrm{d}t > 0$ is give to the oscillator.

We continue with our example of a simple mass-spring system with a drag force $F_\mathrm{d} = -bv$, and add the applied force F:

$$ma = -bv - kx + F_0 \cos(\omega t). \tag{12.95}$$

The equation of motion is

Equation of driven harmonic motion.

$$\frac{\mathrm{d}^2x}{\mathrm{d}t} + \frac{b}{m}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{k}{m}x - F_0\cos(\omega t) = 0, \tag{12.96}$$

or,

$$\ddot{x} + \frac{b}{m}\dot{x} + \frac{k}{m}x - F_0\cos(\omega t) = 0. \tag{12.97}$$

This is called a *driven* or *forced oscillation*. There are special methods that can be used to solve this differential equation, but this falls outside the scope of this course. However, the result is important for understanding resonant behavior. The solution is a simple harmonic oscillator for the form

Driven harmonic oscillation.

$$x(t) = A\cos(\omega t - \phi) \tag{12.98}$$

with constant amplitude

$$A = \frac{F_0}{\sqrt{m^2(\omega_0^2 - \omega^2)^2 + b^2\omega^2}}, \tag{12.99}$$

and phase

$$\tan\phi = \frac{b\omega}{m(\omega_0^2 - \omega^2)}, \tag{12.100}$$

When an external force starts driving an oscillation, it typically takes some short time before the system stabilizes. But once it is stabilized, its harmonic oscillation will be described by Eq. (12.98). The amplitude A as a function of the driving ω has a special feature shown in Fig. 12.13. It peaks at the natural frequency ω_0. This is called a *resonance*. We can learn several things from this:

- The larger the driving F_0, the larger the amplitude A, $F_0 \propto A_0$.
- The amplitude A tends to be larger around $\omega \sim \omega_0$.
- For a fixed F_0 and m, the amplitude A becomes the same for all values of b, when $\omega \to 0$.
- The smaller b, the larger the amplitude A, and the narrower and larger the resonance around $\omega = \omega_0$. In particular, if $b \to 0$, the drag force disappears and the resonance becomes arbitrary large. This can be understood from the fact that the driving force will keep pumping energy into the system with no limit.

12.5.1 Real-life examples

There are many other examples of resonant phenomena in physics, all the way from orbital mechanics to electronics, to atomic physics. One typical real-life example of a resonance is when someone is pushing you on a swing. A swing is like a pendulum, and when your playmate pushes you at the right time, they can make you swing higher and higher by (i.e.

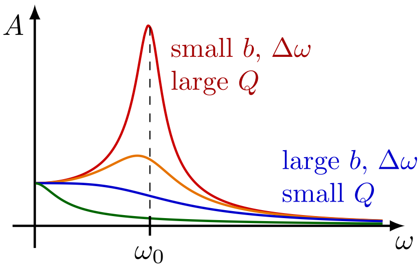


Figure 12.13: Resonance with Q factor.

a higher amplitude). If you have ever pushed someone on the swing before, you know the right moment is at the highest point, just when the swing is about to come down again. By pushing at this exact moment each time, you add energy to the system every period $T = 2\pi/\omega_0$ and in phase, building up its amplitude.

Most objects, even rigid ones, have some natural frequencies ω_0 of $f_0 = \omega_0/2\pi$. Finding the natural frequency of an object can make it vibrate harder than it would otherwise with some other random frequency. Sometimes this reaches the point where the object breaks. Objects can have multiple natural frequencies, which depend on their shape, mass and structure, but also in which direction the driving force is applied.

For example, you can shatter a wine glass if you play an acoustic sound at the natural frequency of the glass and loud enough. The natural frequencies vary between glasses but they typically have one or two natural frequencies in the 400–2500 Hz range. This is why a wine glass "sings" when you rub around its rim with a moist finger, which causes the glass to vibrate, with just the right frequency. You can lower the natural frequency by pouring water in the glass. If you have multiple identical wine glasses, but filled with different amounts of water, you now have different notes to play a song, like some musicians do.

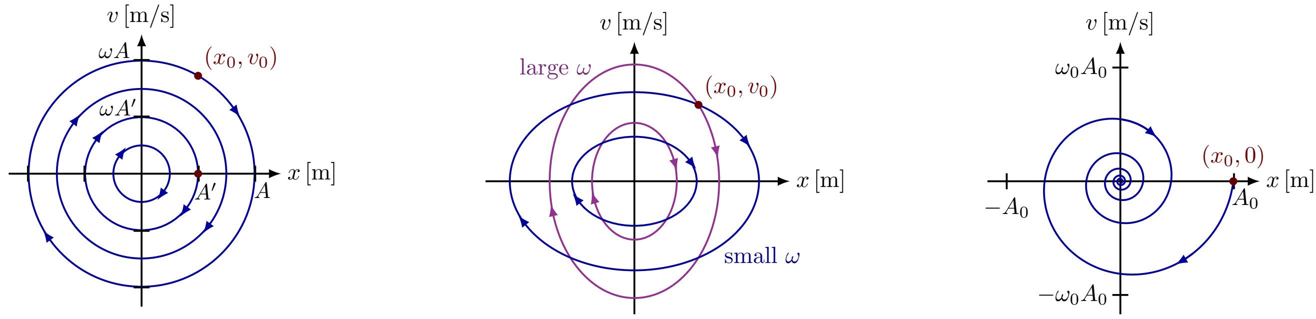
The pillars of offshore oil platforms standing in sea water are constantly hit by sea waves. Engineers have to take into account the natural frequency of the whole oil platform and come up with solutions to prevent the waves from causing the platform to shake uncontrollably when the waves hit its resonance.

The dramatic collapse of the Tacoma Narrows Bridge in 1940 due to strong winds is an oft-cited example of resonance, but this can be better explained by complicated aerodynamical effects.

12.6 Extra: Phase diagrams

One useful tool to analyze the behavior of systems are so-called *phase diagrams*. These are planes with as axes some variables of the system, for example y vs. x, or v vs. x. Any point in this plane is a possible state of the system. As time passes, the state of the system can change, an so the point will move. The evolution of a system will therefore form a line, or *trajectory*, in the phase plane.

This is why it is often used in context of differential equations. A differential equation describes the change of a system, and each solutions will describe unique trajectories. One can conveniently display and summarize all the possible solutions of a differential equation in the phase plane. This is called a *phase portrait*. An initial condition is some point, or *initial state*, where the line "starts". Any other point on the line can be taken as an initial condition with the same solution, but with some time offset. Different trajectories can never cross. What is more, trajectories have some direction, going from "early" to "later" time, past to future. Differential equation are therefore sometimes visualized as vectors in the phase plane.



(a) The solutions of simple harmonic oscillator form ellipses. The radius depends on the initial condition.

(b) The trajectory shape depends on $\omega = \sqrt{k/m}$. Two systems have different trajectories for the same initial condition.

(c) An underdamped harmonic oscillator spirals to $(x, v) = (0, 0)$, as it loses energy.

Figure 12.14: Phase portraits for (x, v) of a simple and damped oscillator. The trajectories (solid lines) depends on the system parameters and initial condition (red points).

Take for example the simple harmonic oscillator of a spring. Figure 12.14a illustrates that each solution is an ellipse where the system moves in the counterclockwise direction. This is easy to understand when you realize that if $x \propto \cos(\omega t)$, then $v \propto \sin(\omega t)$. The radii of the ellipse depends on the amplitude A and angular frequency ω. The amplitude A in turn, is determined by the initial conditions (x_0, v_0), while $\omega = \sqrt{k/m}$ is a parameter of the system. For a system with a given ω, the trajectories will look different, as shown in Fig. 12.14b. These are examples of *closed trajectories*.

Because the kinetic energy $K \propto v^2$, and the potential energy $U \propto x^2$, we can also identify the vertical direction in the xv plane as the kinetic energy, and the horizontal as the potential one. The larger the radius, the larger the total energy.

Next, consider the underdamped oscillator. As time passes, the amplitude of both x and v will exponentially decrease. This looks like a spiral, as shown in Fig. 12.14c.

Lastly, the pendulum looks similar to a simple harmonic oscillator with closed trajectories for angle amplitudes θ_0 below π, or, 180°. However, if the pendulum has enough velocity, or equivalently enough kinetic energy, it can spin all the way around. As long as it does not lose energy, it will keep spinning without changing direction. These form *open trajectories* in the phase plane, as illustrated in Fig. 12.15. The border between these two cases where the angle amplitude is $\theta_0 = \pi\,\text{rad} = 180°$, is called the *separatrix*. This is the case when the velocity of the mass m is just about enough to reach the highest point

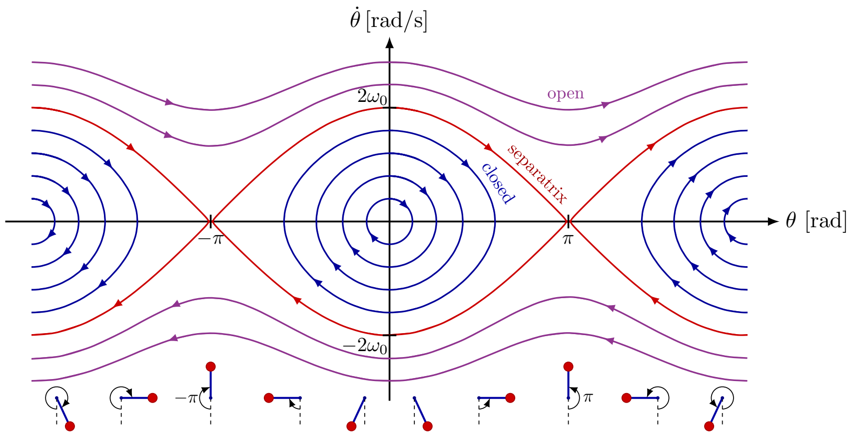


Figure 12.15: The phase portrait of a pendulum with closed (blue) and open trajectories (purple). The edge case where the amplitude is $\theta_0 = \pi\,\text{rad}$ and the maximal $\dot{\theta}$ is $2\omega_0$, forms a separatrix (red). Note that because the angle θ is cyclic; the horizontal axis “wraps“ around itself, e.g. any angle θ is the same point in space as $\theta \pm 2\pi\,\text{rad}$, and therefore corresponds to the same state.

$h = 2L$. The maximal kinetic energy is therefore

$$\frac{mv_{\max}^2}{2} = 2mgL. \tag{12.101}$$

Because $v = L\,\mathrm{d}\theta/\mathrm{d}t = L\dot{\theta}$,

$$\dot{\theta}_{\max} = 2\sqrt{\frac{g}{L}} = 2\omega_0. \tag{12.102}$$

Anything smaller, and the pendulum will swing back and forth, anything larger, and the pendulum will just rotate in one direction. Its maximal velocity will be at the bottom, and it minimum at the top. Clockwise or counterclockwise depends on the initial v_0.

12.7 Application of simple harmonic oscillator

The phenomenon of a simple harmonic oscillator will returns frequently in physics. The reason why it is so ubiquitous can be understood with Taylor approximations. In most of these cases, a mass oscillates around some point of stable equilibrium. The simple harmonic oscillator is often a good first-order approximation because the total force F depends on a displacement x, and can be expanded by Taylor expansion,

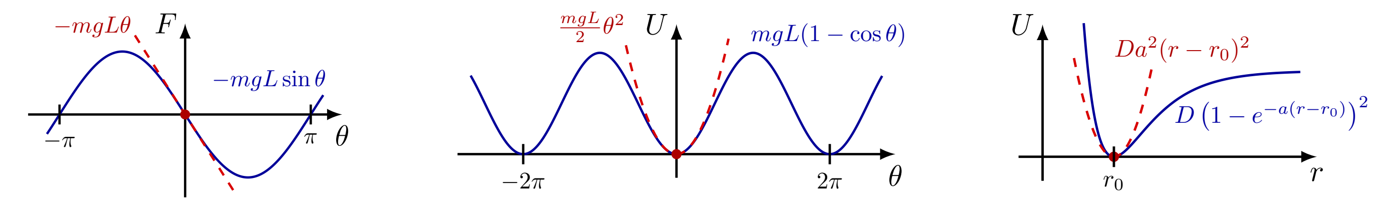
$$F(x) \approx F'(0)x + F''(0)\frac{x^2}{2} + ... \tag{12.103}$$

where F' is the first derivative of the force, which always pulls the mass back to $x = 0$, such that $F'(0) < 0$. $F(0)$ must be zero, otherwise $x = 0$ would not correspond to an equilibrium. If the second derivative $F''(0)$ and higher derivatives are relatively small around $x = 0$, then this form approximates Hooke's law $F(x) \propto x$, leading to a harmonic oscillator as in Eq. (12.19). One such example is a pendulum with $F(\theta) = -mg\sin\theta \approx -mg\theta$ for small θ.

An alternative way to formulate this argument is in terms of potential energy for some conservative force. A stable equilibrium means that there is some local minimum in the potential energy around $x = 0$ (see Section 9.10). In that case, the potential forms a well that can be approximated with a parabola in $x = 0$;

$$U(x) \approx U(0) + U''(0)\frac{x^2}{2} + U'''(0)\frac{x^3}{3!} + ... \tag{12.104}$$

Ignoring the arbitrary constant $U(0)$, assuming $U''(0) > 0$ for a stable point, and that higher order derivatives are relatively small in $x = 0$, we recognize the potential energy for a simple harmonic oscillator, $U(x) \propto x^2$. One example where the minimum of a potential is well approximated by a parabola, is the vibrations of diatomic molecules. It is modeled by the Morse potential, shown in Fig. 12.16c, which is a function of the atom separation r. This potential is indeed behaves like a simple harmonic oscillator at first-order approximation.



(a) Force of a pendulum is $F(\theta) = -mgL\sin\theta \approx -mgL\theta$.

(b) Potential energy of a pendulum is $U(\theta) = mgL(1-\cos\theta) \approx mgL\theta^2/2$.

(c) Molecular vibrations can be approximated by harmonic oscillator.

Figure 12.16: Many systems with a stable equilibrium can be approximated by a simple harmonic oscillator.

Chapter 13

Waves

Waves, very broadly, are disturbances that travel through a medium. It is a phenomenon that appears in many different places in physics. The medium can be one-dimensional, like a string or a metal bar, also two-dimensional membrane, like the skin of a drum, or even three-dimensional, like a bulk of fluid. The disturbance can be *transverse* (perpendicular to the direction of propagation), or *longitudinal* (in the direction of propagation). In particular, will have a closer look at *traveling* and *standing waves* in strings and air (sound), as well as *interference*, the *Doppler effect*, and how waves carry energy.

Next semester, PHY121 will cover Huygens's principle, refraction, diffraction, interference patterns in more detail in the context of optics.

13.1 Transverse waves

A *transverse wave* is wave that causes a disturbance in the perpendicular *direction of the propagation*. Let's take a simple wave that travels in one direction, like a disturbance on a string after yanking it. We break down the motion in two directions: x direction for propagation, and y direction for the disturbance. At $t = 0$, the disturbance has a shape within the medium that is given by some function, say $y(x,t) = f(x)$ as in Fig. 13.1a. f can be any function, although we will focus mostly on sine waves.

The wave travels in the x direction with some constant velocity v, but the wave's shape stays the same. In the reference frame S' that moves at the same velocity as the wave, we see that at any time t,

$$y'(x,t) = f(x). \tag{13.1}$$

Going back to the lab frame S, we use the Galilean transformation Eq. (8.53) to find our

(a) Shape of wave at time $t = 0$ can be some function f in space variable x.

(b) The shape moved by a distance $v\Delta t$ at time $t = \Delta t$.

Figure 13.1: A transverse wave is a traveling wave that distorts a medium in the direction perpendicular to the direction of propagation.

coordinates

$$\begin{cases} y = y' & (13.2) \\ x = x' + vt. & (13.3) \end{cases}$$

Therefore, in the S frame, the wave can be described as

Traveling wave.

$$y(x,t) = f(x - vt). \tag{13.4}$$

If $v > 0$, the expression above describes a wave moving in the positive x direction, i.e. to the right in Fig. 13.1b. For a wave moving in the negative x direction, we have instead

$$y(x,t) = f(x + vt). \tag{13.5}$$

13.1.1 Sinusoidal waves

If we keep moving a string up and down, we can generate a periodic wave which has some wavelength λ and period T. The *wavelength* is the minimum distance λ at any given time between two points where the wave repeats itself. Similarly, the *period* is the minimum time interval T at a given point in space, when the wave repeats.

One example of such a periodic wave, is a *sine wave*, which is a simple harmonic oscillator. A sine wave can be described by

$$y(x,t) = A\sin(kx - \omega t) \tag{13.6}$$

with *amplitude* A, *angular velocity* ω and *wavenumber* k. It is illustrated in Fig. 13.2. More generally, a *phase* ϕ can appear.

Sine wave.

$$y(x,t) = A\sin(kx - \omega t - \phi) \tag{13.7}$$

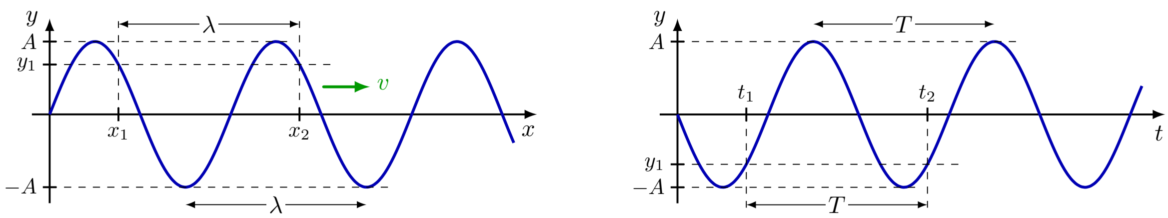
But what exactly is k? Without loss of generality, take time $t = 0$, where $y(x,0) = A\sin(kx)$. Take any two points that are separated by one wavelength λ, say x_1 and $x_2 = x_1 + \lambda$, such that the corresponding y values are the same, $y(x_1,0) = y(x_2,0)$, as in Fig. 13.2a. Therefore,

$$A\sin(kx_1) = A\sin\big(k(x_1 + \lambda)\big), \tag{13.8}$$

and so we must have

$$k(x_1 + \lambda) = kx_1 + 2\pi n \tag{13.9}$$

for integer $n = 0, \pm1, \pm2, ...$ Since the wavelength λ is the *minimum* distance between two *different* points, $n = 1$, we find that k must be:



(a) Whole wave in space at time $t = 0$, given by $y(x,0) = A\sin(kx)$.

(b) Local disturbance at position $x = 0$, given by $y(0,t) = -A\sin(\omega t)$.

Figure 13.2: A space and time slice of a travelling sine wave $y(x,t) = A\sin(kx - \omega t)$.

Wavenumber.

$$k = \frac{2\pi}{\lambda}. \tag{13.10}$$

Therefore, k is a measure of how many wavelengths λ "fit" in 2π, the period of a sine wave. Clearly, k has dimension of inverse length with units rad/m.

In a similar way, we find that for a space slice $y(0,t) = -A\sin(\omega t)$, two moments in time t_1 and $t_2 = t_1 + T$ must have

$$\omega(t_1 + T) = \omega t_1 + 2\pi, \tag{13.11}$$

such that we find

$$\omega = \frac{2\pi}{T}, \tag{13.12}$$

as previously in Sections 5.3 and 12.2. Note the equivalence between k for space and ω for time. A comparison of all these different parameters and their relations is given by Table 13.1. A sine wave Eq. (13.7) can be expressed as

$$y(x,t) = A\sin\left[2\pi\left(\frac{x}{\lambda} - \frac{t}{T}\right)\right]. \tag{13.13}$$

One can immediately read off that if $x_2 = x_1 + \lambda$ at a fixed time t, $y(x_1,t) = y(x_2,t)$. Similarly, when $t_2 = t_1 + T$, $y(x,t_1) = y(x,t_2)$ for any x.

Furthermore, any two space points x_1 and $x_2 = x_1 + n\lambda$ that differ only in an integer multiple $n = 0, \pm1, \pm2, ...$ of the wavelength λ at a fixed time, are said to be *in phase*. Similarly, in a fixed space point, the wave has the same phase at any two times t_1 and $t_2 = t_1 + nT$ that only differ by an integer multiple n of the period T.

Notice that we can also rewrite Eq. (13.7) as a traveling wave Eq. (13.4),

$$y(x,t) = A\sin\big(k(x - vt)\big). \tag{13.14}$$

So what is the velocity v? Comparing the this form to Eq. (13.7), it is simply

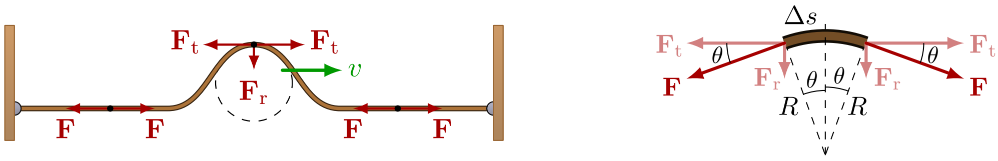
Velocity of a sine wave.

$$v = \frac{\omega}{k}. \tag{13.15}$$

Comparing to Eq. (13.13), we have more generally for a traveling, periodic wave,

Velocity of a traveling, periodic wave.

$$v = \frac{\lambda}{T} = \lambda f. \tag{13.16}$$



(a) Forces on a string. All across the string, there is a constant tension **F**.

(b) Small segment of length Δs experiences a tension **F** on either side.

Figure 13.3: The tension in a string is increased due to a disturbance.

13.1.2 Speed of a wave on a string

The propagation speed of a wave depends on the medium. It does not depend on how fast the disturbance is caused. Consider a wave on a string with a length L and of uniform mass M as in Eq. (13.3a). The mass density per unit length is therefore

$$\mu = \frac{M}{L}. \tag{13.17}$$

The string has some tension F along its length. Let's focus on a small segment at the peak of the disturbance, and assume it can be approximated by a short segment of a circle with some radius R and that subtends an angle 2θ. Zooming into this small segment, the tension on either side can be decomposed into a tangential and radial part (Fig. 13.3b). The tangential components tend to cancel out over the disturbance, while the radial components tend to add up. The segment has an arc length $\Delta s = 2R\theta$. This means it has a mass of

$$m = 2\mu R\theta. \tag{13.18}$$

The sum of the radial forces is given by

$$\sum F_\mathrm{r} = 2F\sin\theta. \tag{13.19}$$

If θ is very small, we can use the small-angle approximation (12.56), such that

$$\sum F_r \approx 2F\theta. \tag{13.20}$$

This force points radially toward the center of the circle. It is convenient to consider a reference frame S$'$ that moves along with the disturbance at a velocity v. In this frame, the wave shape is stationary, but the string segment on the top moves with a tangential velocity v to the left, which is caused by the centripetal force

$$2F\theta = m\frac{v^2}{R} = (2\mu R\theta)\frac{v^2}{R}, \tag{13.21}$$

where we used the centripetal acceleration Eq. (5.34). We see that R and θ, the variables related to this particular segment, cancel out. We find an interesting relation

Table 13.1: Summary of parameters of a traveling sine wave. Their units are given in terms of seconds x = s, meters x = m, and/or radians.

Units	Spatial	Temporal
x	Wavelength λ	Period T
$\frac{1}{\mathrm{x}}$	–	Frequency $f = \frac{1}{T}$
$\frac{\mathrm{rad}}{\mathrm{x}}$	Wavenumber $k = \frac{2\pi}{\lambda}$	Angular frequency $\omega = \frac{2\pi}{T}$
$\frac{\mathrm{m}}{\mathrm{s}}$	Velocity $v = \frac{\lambda}{T} = \lambda f = \frac{\omega}{k}$	

Speed of wave on string.

$$v = \sqrt{\frac{F}{\mu}}. \tag{13.22}$$

This means that the wave velocity only depends on the mass density μ and tension F of the string. In fact, the formula for the velocity of mechanical waves in other media, very often has form

$$v = \sqrt{\frac{\text{elastic property}}{\text{inertial property}}}, \tag{13.23}$$

where the "plastic property" can refer to the tension, stiffness or compressibility of the medium, and "inertial property" refers to for example the mass density.

13.2 Wave equation

These wave functions can generally be obtained by solving a differential equation with the appropriate boundary equations. Such a differential equation is called the wave equation, and its solutions will be of the form $y(x,t) = f(x - vt)$ and $y(x,t) = f(x + vt)$.

To find the wave equation, consider a small segment of a vibrating string. Figure 13.4 shows the segment is pulled by two different tensions $\mathbf{F}_1$ and $\mathbf{F}_2$ in opposite directions, but not exactly parallel. They make different angles with the horizontal. The small segment has dimension Δx and Δy, so the slope s of the small string tension is given by

$$s = \frac{\Delta y}{\Delta x} = \tan\theta, \tag{13.24}$$

where θ is the angle of the tangent on the string segment. Taking the infinitesimal limit,

$$s = \frac{\partial y}{\partial x} = \tan\theta, \tag{13.25}$$

where we take a partial differential w.r.t. to x, as there are several variables like time t in play, that we want to keep constant.

Meanwhile, if we assume the wave is transverse, we only care about the vertical components of the tension, and assume the horizontal ones cancel. The total vertical force is

$$\sum F_y = F\sin\theta_1 - F\sin\theta_2, \tag{13.26}$$

where we assume the magnitude of the tensions, F, is the same along the whole string. We can now use our small-angle approximation (12.56) trick again to say that $\sin\theta \sim \theta \sim \tan\theta$. Substituting,

$$\sum F_y = F\tan\theta_1 - F\tan\theta_2 = F(s_1 - s_2) = F\Delta s. \tag{13.27}$$

Let's apply Newton's second law for the y direction,

$$F\Delta s = ma_y = (\mu\Delta x)\frac{\partial^2 y}{\partial t}, \tag{13.28}$$

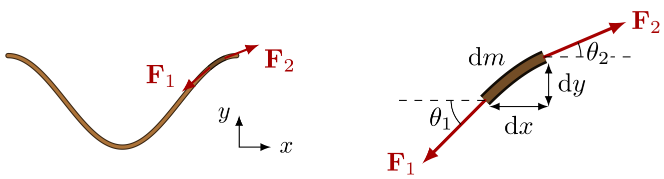


Figure 13.4: The tension in a small segment of a string.

with linear mass density μ. Taking the ratio

$$\frac{\Delta s}{\Delta x} = \frac{\mu}{F}\frac{\partial^2 y}{\partial t^2}, \tag{13.29}$$

where we recognize the velocity $v^2 = F/\mu$ from the last section. Taking the infinitesimal limit again,

$$\lim_{\Delta x \to 0} \frac{\Delta s}{\Delta x} = \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2} \tag{13.30}$$

Comparing this to Eq. (13.25), which leads to

$$\frac{\partial s}{\partial x} = \frac{\partial^2 y}{\partial x^2}, \tag{13.31}$$

we find the desired wave equation in 1D.

Wave equation (1D).

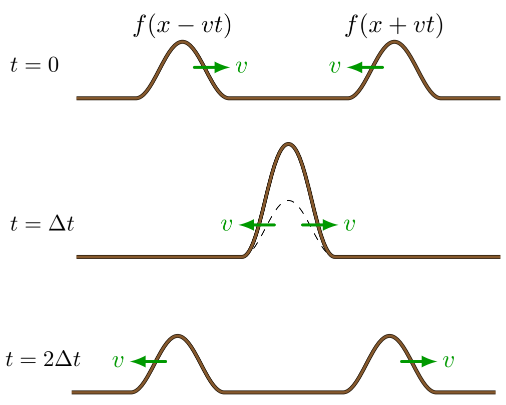
$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 y}{\partial t^2}. \tag{13.32}$$

13.3 Superposition & interference

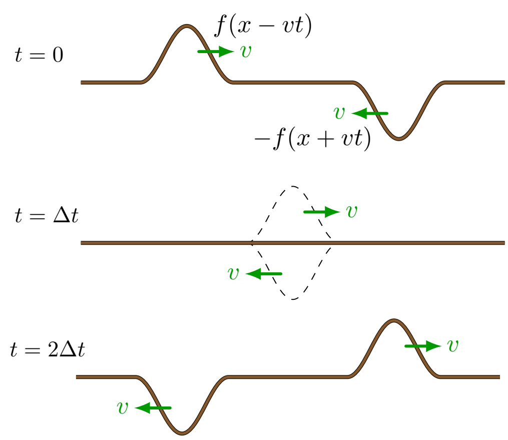
If we have two waves traveling along the same path, we can often simply add them linearly. This is called a *superposition*. For example, two waves traveling in the opposite direction can be added as

$$y(x,t) = f_1(x - v_1 t) + f_2(x + v_2 t). \tag{13.33}$$

If $f_1(x) = -f_2(x)$ for any x, then at the time the waves meet, the will completely cancel, as in Fig. 13.5b.



(a) Constructive interference happens when two oppositely waves meet on a string.



(b) Destructive interference. If the waves are the same but for a sign, they cancel completely.

Figure 13.5: Superposition between two oppositely travelling waves in the same medium is a simple linear sum.

13.3.1 Phase difference

What happens when two sine waves have the same amplitude, wave number and angular frequency, but a constant difference in phase? Consider

$$y_1(x,t) = A\sin(kx - \omega t) \tag{13.34}$$
$$y_2(x,t) = A\sin(kx - \omega t - \phi). \tag{13.35}$$

Adding them together, there are several interesting cases. If $\phi = 2\pi n$ for some integer n, the superposition $y_1 + y_2$ will have *constructive interference*, doubling their individual sizes,

$$y_1(x,t) + y_2(x,t) = 2A\sin(kx - \omega t). \tag{13.36}$$

If $\phi = n\pi$ for an uneven integer $n = \pm1, \pm3, \pm5, ...$, they cancel in all time and space point,

$$y_1(x,t) + y_2(x,t) = 0, \tag{13.37}$$

which is *destructive interference.*

In general, for some phase ϕ,

$$y_1(x,t) + y_2(x,t) = A\sin(kx - \omega t) + A\sin(kx - \omega t - \phi), \tag{13.38}$$

we can use the sum rule

$$\sin\theta_1 + \sin\theta_2 = 2\cos\left(\frac{\theta_1 - \theta_2}{2}\right)\sin\left(\frac{\theta_1 + \theta_2}{2}\right). \tag{13.39}$$

So the interference of two waves,

$$y_1(x,t) + y_2(x,t) = 2A\cos\left(\frac{\phi}{2}\right)\sin\left(kx - \omega t - \frac{\phi}{2}\right). \tag{13.40}$$

Two separate sources that create waves with the same shape, wavelength, period, but only differ in a constant phase, are said to be *coherent.*

13.3.2 Interference patterns in space

Say you have two coherent sources, S_1 and S_2, like in Fig. 13.7, that spherically emit waves, i.e. in all radial directions equally. Assuming they are coherent sine waves,

$$y_1(r,y) = A\sin(kr - \omega t) \tag{13.41}$$
$$y_2(r,y) = A\sin(kr - \omega t), \tag{13.42}$$

with the radial distance r. However, the sources are separated in space, and therefore in each point, there will be a phase difference, which depends on the path difference, namely

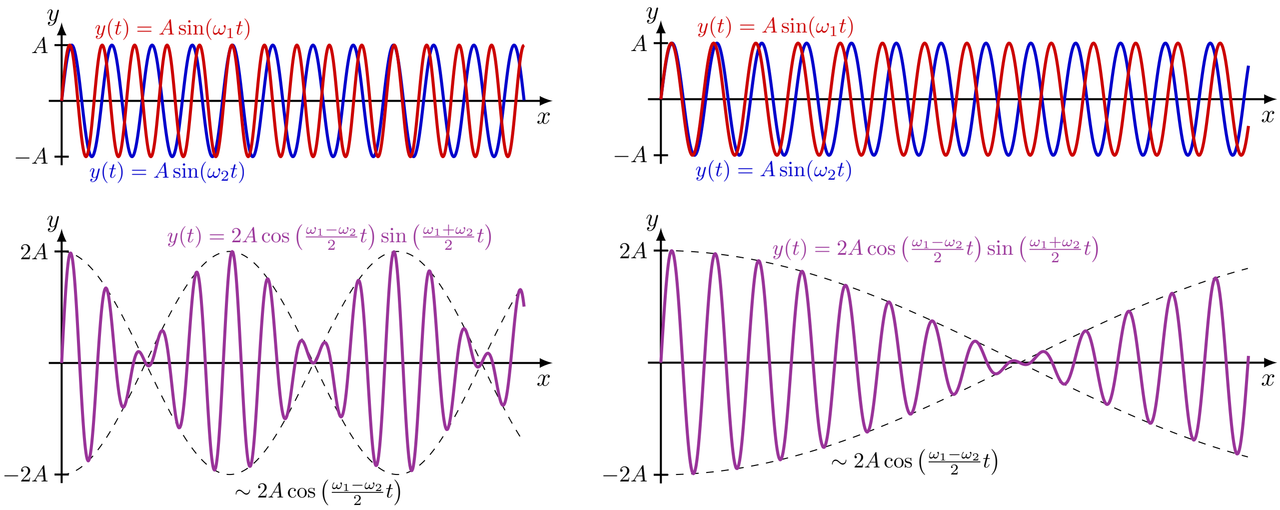
$$\Delta\phi = (kr_2 - \omega t) - (kr_1 - \omega t) = k\Delta r, \tag{13.43}$$

where $\Delta r = r_2 - r_1$. So

$$\Delta\phi = 2\pi\frac{\Delta r}{\lambda}. \tag{13.44}$$

Some special points are where $\Delta\phi = 2\pi n$ or $\Delta r = n\lambda$ for some integer $n = 0, \pm1, \pm2, ...$, such that the waves add constructively. Destructive interference happens when $\Delta r = n\lambda/2$ for odd integers $n = \pm1, \pm3, \pm5, ...$

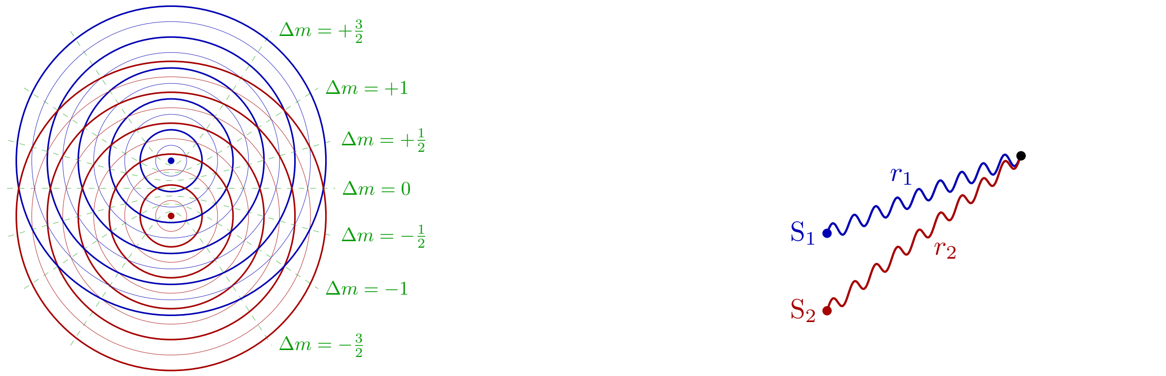
Next semester, PHY121 will cover interference patterns in more detail.



(a) Large difference $\omega_2 - \omega_1$.

(b) Small difference $\omega_2 - \omega_1$.

Figure 13.6: Beats: Interference pattern between two oscillations with a different frequencies $f_i = \omega/2\pi$.



(a) Thick lines are maxima, thin lines are minima. The green lines connect points with destructive and constructive interference.

(b) Path difference $\Delta r = r_2 - r_1$ will determine the type of interference in a point.

Figure 13.7: Two point sources create an interference pattern. There is destructive interference if the path difference $r_2 - r_1 = \Delta m\lambda$ where Δm is half-integer, constructive if Δm is an integer.

(a) Fixed on a rigid support.

(b) Loose end on a free support.

Figure 13.8: Reflection.

(a) From lighter string to heavier string.

(b) Loose end.

Figure 13.9: Partial transmission and reflections between strings of different mass densities.

13.3.3 Frequency difference & beats

Previously, in Eq. (13.34), we assumed the amplitude and frequency were the same. What happens when the frequencies are *not* quite the same ? Say the sine waves of two sources meet in some point in space, causing the oscillation

$$y_1(t) = A\sin(\omega_1 t) \tag{13.45}$$

$$y_1(t) = A\sin(\omega_2 t), \tag{13.46}$$

where $\omega_1 \neq \omega_2$, and we ignore any phase difference for simplicity. Using again the sum rule Eq. (13.39), we find

$$y_1(t) + y_2(t) = 2A\cos\left(\frac{\omega_1 - \omega_2}{2}t\right)\sin\left(\frac{\omega_1 + \omega_2}{2}t\right). \tag{13.47}$$

The combination is represented as two oscillations going on at the same time, first a fast oscillation with frequency $(f_1 + f_2)/2$, and secondly a slower oscillation with frequency $|f_1 - f_2|/2$, where $f_i = \omega_i/2\pi$. The slower frequency can be thought of as the modulation of the amplitude, creating the dashed envelopes shown in Fig. 13.6. Notice that the smaller the difference in frequency, the longer the modulation. Furthermore, if $f_1 \sim f_2$, then the larger frequency will almost be double the original ones $\sim 2f_1$.

In acoustics, this can happen with sound waves, and it is called a *beat.* Because the amplitude varies up and down twice per cycle, $f_{\text{beat}} = |f_1 - f_2|$ is called the *beat frequency.*

13.4 Reflection & transmission

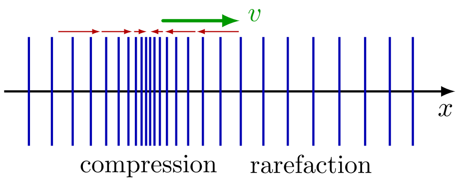
The classic picture of a wave is a shape wave traveling along a string. It is interesting to look at what happens when the wave reaches the end of the string. There are several cases.

If the string is fixed on one rigid support, a wave reaching the end will be *reflected*, but the direction of the disturbance will be "flipped" vertically as in Fig. 13.8a. If the support is free, such that the string's end can freely move up and down, the wave will still be reflected, but not inverted.

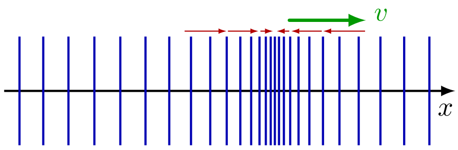
Another case, is where the string is connects to another string with a different mass density μ. In case the other string is heavier, only part of the wave will be transmitted at a lower velocity, while part of the wave will be flipped and reflected on the first string with a smaller amplitude to conserve energy. Conversely, a wave on a heavy string will transmit most of the wave to a lighter one, and part will be reflected, without being flipped.

13.5 Longitudinal waves

So far we discussed mostly transverse waves, where the disturbance is in the y-axis direction, perpendicular to the direction of propagation (x axis). If the disturbance instead is along the direction of propagation, we are dealing with a longitudinal wave. The classic example is sound, where the air molecules will move back and forth. In this case, the disturbance is given by a longitudinal displacement s along x of the air molecules. Other forms of longitudinal waves can be expressed with density or pressure variations along the propagation.



(a) Time $t = 0$.



(b) Time $t = \Delta t$.

Figure 13.10: A traveling longitudinal wave is when the distortion happens along the direction of propagation, here shown as a local displacement.

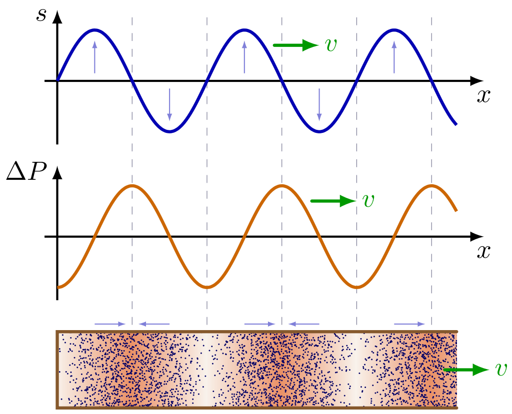


Figure 13.11: Sound wave traveling in a tube of air, shown as a local, average displacement s of air molecules in the longitudinal (x) direction (blue), and a local pressure variation ΔP (orange), 90° out of phase with s.

13.5.1 Sound waves

A hand clap in midair will create a disturbance in the air that travels in all directions. This disturbance is what hits our ear drum and our brain interprets as sound (after it is converted to electric stimuli). Looking more closely, air molecules will locally vibrate back and forth, which can be expressed as an average displacement s along x. Because there will be a *compression* in some places, where more air molecules bunch up, and a *rarefaction* in others, where they are depleted, an equivalent way of describing this is as a change in pressure, ΔP, as in Fig. 13.11. Although we mostly think of sound as traveling through air, it can also travel through any gas, liquid or solid medium.

For a sinusoidal sound wave, the pressure wave ΔP is always 90° ahead of the displacement wave s, as shown in Fig. 13.11. We can therefore write them as

$$s(x,t) = s_0 \sin(kx - \omega t) \tag{13.48}$$
$$\Delta P(x,t) = \Delta P_0 \sin(kx - \omega t - \pi/2). \tag{13.49}$$

The relation between their respective amplitudes is given by

Soundwave amplitude relationship.

$$\Delta P_0 = \rho \omega v s_0, \tag{13.50}$$

with the mass density ρ.

13.5.2 Speed of sound in fluids

What is this velocity of sound in a given fluid? Consider a tube filled with a fluid of density ρ. We move a piston at one end of the tube with cross sectional area A to suddenly increase the pressure. However, the fluid will resist compression which is given by the *bulk modulus*

$$B = -\Delta P \frac{\Delta V}{V}, \tag{13.51}$$

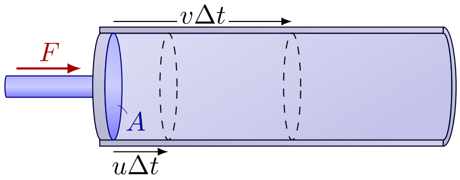


Figure 13.12: A piston with area A quickly moves some fluid in a pipe with some force F and velocity u, creating a pressure wave with the speed of sound v.

discussed in more detail in Section 16.3. First we quickly compress the piston in some short time interval Δt and some velocity u. It moves a distance $\Delta x_{\text{piston}} = u\Delta t$, and increases the fluid's pressure by

$$\Delta P = \frac{F}{A}, \tag{13.52}$$

where F is the force exerted by the piston, and A is the tube's area. This creates a pulse of pressure, a sound wave, traveling with some velocity v through the tube. The pressure wave moves a distance $\Delta x = v\Delta t$ through the air in the same time interval Δt. Over this distance, a volume $V = A\Delta x$ of fluid with total mass

$$m = \rho A v \Delta t \tag{13.53}$$

is moved. We want to know the velocity v of the wave in the fluid, we can derive with conservation of momentum. The impulse (Section 8.2) created by the piston,

$$F\Delta t = (A\Delta P)\Delta t, \tag{13.54}$$

causes a change in momentum of the fluid,

$$A\Delta P\Delta t = \Delta p_{\text{fluid}}. \tag{13.55}$$

The change in momentum can also be expressed as mass times velocity

$$\Delta p_{\text{fluid}} = (\rho A v \Delta t)u. \tag{13.56}$$

Comparing Eq. (13.55) and (13.56), Δt and area A drops out,

$$\Delta P = \rho v u. \tag{13.57}$$

From the definition of the bulk modulus Eq. (13.51),

$$-B\frac{\Delta V}{V} = \rho v u. \tag{13.58}$$

Here, $V = Av\Delta t$ is the volume of the fluid, and $\Delta V = Au\Delta t$ is the volume of fluid moved by the piston, so

$$\frac{\Delta V}{V} = \frac{u}{v}. \tag{13.59}$$

Plugging this into Eq. (13.58), we see that u cancels out, which leads us to our result

Speed of sound in fluids.

$$v = \sqrt{\frac{B}{\rho}}. \tag{13.60}$$

In words, the speed of sound in some fluid is determined by

$$v = \sqrt{\frac{\text{compressibility}}{\text{mass density}}}. \tag{13.61}$$

For air, $B = 142\,\text{kPa}$, $\rho = 1.2\,\text{kg/m}^3$. So we can calculate how fast air travels in over some distance Δx. While the above calculation is a good first approximation, the speed of sound also lightly depends on the temperature of the air. As the temperature increases, the air molecules move faster, allowing a pressure wave to pass through more quickly.

13.5.3 Speed of sound in solids

In fluids, the speed of sound was given by the bulk modulus B, So if we want to study sound traveling through a solid rod, we need to ask how compressible a solid is? Now, the compressibility of our medium is expressed by Young's modulus, and it turns out that indeed

Speed of sound in solids.

$$v_{solid} = \sqrt{\frac{\Upsilon}{\rho}}. \tag{13.62}$$

As an example, brass has a density of $\rho = 8600\,\text{kg/m}^3$ and Young's modulus $\Upsilon = 90\,\text{GN/m}^2$, so the speed of sound is about ten times faster in brass than in air:

$$v_{\text{brass}} = 3234\,\text{m/s}. \tag{13.63}$$

13.6 Standing waves

The previous sections covered traveling waves. Another very interesting case with many applications are a standing waves. They are important to understand how one can produce different notes on music instruments, for example. In some cases, the superposition of to oppositely traveling waves will give rise to a standing wave.

Let's see how. Consider two sine waves meeting on a string, one moving to the right and one moving to the left,

$$y_{\text{R}}(x,t) = A\sin(kx - \omega t) \tag{13.64}$$

$$y_{\text{L}}(x,t) = A\sin(kx + \omega t), \tag{13.65}$$

respectively. Their superposition can rewritten with sum rule Eq. (13.39) as

$$y_{\text{R}}(x,t) + y_{\text{L}}(x,t) = 2A\cos(\omega t)\sin(kx). \tag{13.66}$$

Figure 13.13 shows that the superposition does not seem to be traveling anymore in any direction. Instead, there is a standing wave with shape $\sin(kx)$, whose amplitude $2A\cos(\omega t)$ oscillates in time with frequency $f = \omega/2\pi$.

There are space points where there never is a disturbance. They are given by

$$2A\cos(\omega t)\sin(kx) = 0, \tag{13.67}$$

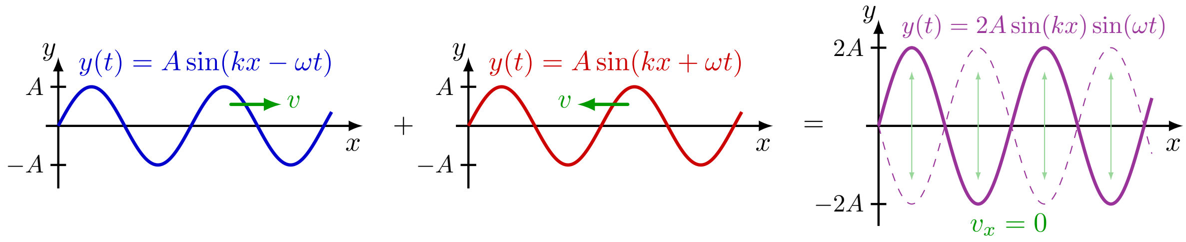


Figure 13.13: Two oppositely traveling sine waves with the same wave length and velocity form a standing wave with amplitude $2A$ and period $T = 2\pi/\omega$.

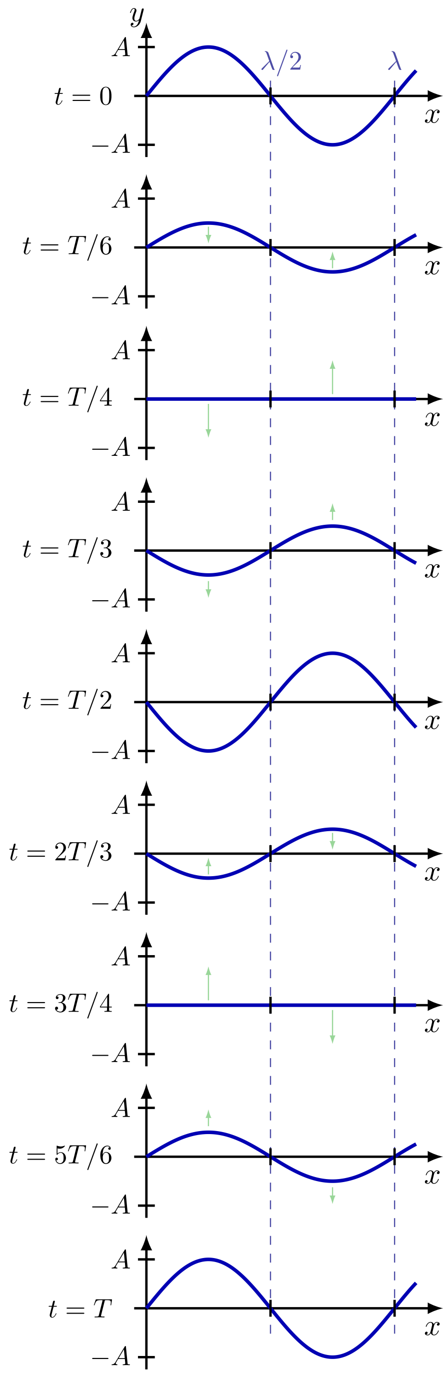
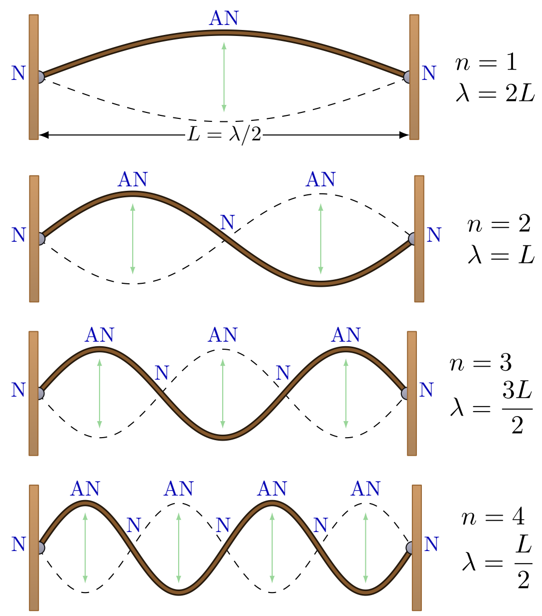
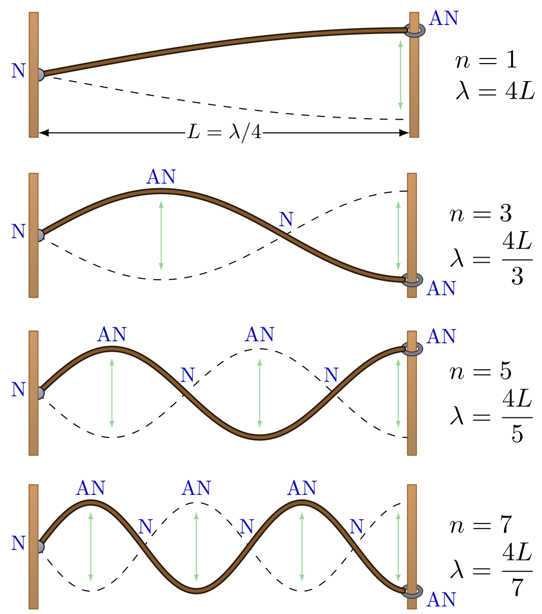


Figure 13.14: Standing wave at different times.



(a) Fixed on both ends.



(b) Loose on one end.

Figure 13.15: Standing waves on a string with different boundary conditions.

meaning that they must satisfy the condition

$$x = \frac{n\pi}{k} = \frac{n\lambda}{2}, \tag{13.68}$$

for integers $n = 0, \pm 1, \pm 2, ...$ These are so-called *nodes*. Similarly, there are *anti-nodes*, where the wave reaches its maximum variation at given times. They are given by

$$\sin(kx) = 1, \tag{13.69}$$

or,

$$x = \frac{n\pi}{2k} = \frac{n\lambda}{4}, \tag{13.70}$$

for odd integers $n = \pm 1, \pm 3, \pm 5, ...$

13.6.1 On a string

As discussed in Section 13.4, when waves are confined by a boundary, they will reflect. If the reflected waves "hit" other incoming waves, they can combine into standing waves. Let's study this in more detail for sine waves on a string of length L.

Fixed ends

First consider that both ends of the strings are fixed by rigid supports, as in Fig. 13.14a. Now, notice that the standing wave must have nodes $y(x, t) = 0$ at $x = 0$ and $x = L$, where string cannot move. A standing wave must be of the form

$$y(x, t) = 2A\cos(\omega_n t)\sin(k_n x) \tag{13.71}$$

with boundary condition

$$\sin(k_n L) = 0. \tag{13.72}$$

This happens when

Standing waves on a string fixed on both ends.

$$\lambda_n = \frac{2L}{n}, \qquad k_n = \frac{n\pi}{L} \tag{13.73}$$

for non-zero integers $n = 1, 2, 3, ...$

The first case $n = 1$ corresponds to the standing wave which has only two nodes, on either end, and one anti-node. For $n = 2$, there are two anti-nodes, and a third node shows up in the middle. So for any n, the respective standing waves has $n + 1$ nodes and n anti-nodes.

One fixed, one loose end

The same calculation can be done for a boundary condition where one end is left loose, as in Fig. 13.15. In this case, the open end is an anti-node, because the standing wave reaches its maximum here. This provides the boundary condition

$$\sin(kL) = 1. \tag{13.74}$$

We find that

Standing waves on a string fixed on only one ends.

$$\lambda_n = \frac{4L}{n}, \qquad k_n = \frac{n\pi}{2L} \tag{13.75}$$

for odd integers $n = 1, 3, 5, \ldots$.

For n, the corresponding standing waves has $(n+1)/2$ nodes and anti-nodes.

13.6.2 Resonant frequencies

Remember that for a wave, the wavelength λ and frequency f are linked by the wave velocity $v = \lambda f$, which in turn is determined by the properties of the medium, e.g. $v = \sqrt{F/\mu}$ for a string. Therefore, for a given medium, there are certain frequencies, called *resonant frequencies*, at which the medium will spontaneously form standing waves.

Take for example a string with length L, fixed on both ends. In the last section, we saw that the standing waves can be classified with some integer n. Such a standing wave is called the n^{th} *harmonic* and a member of a discrete set of allowed standing waves, called the *harmonic series*. The frequency of the n^{th} harmonic is

Frequency of the n^{th} harmonic.

$$f_n = \frac{v}{\lambda_n}. \tag{13.76}$$

For a string, we found Eq. (13.22) and Eq. (13.73), so

Frequency of the n^{th} harmonic on a string with fixed ends.

$$f_n = \frac{n}{2L}\frac{F}{\mu}. \tag{13.77}$$

The first case $n = 1$ corresponds to the *first* or *fundamental harmonic*. Notice that the frequency of any harmonic,

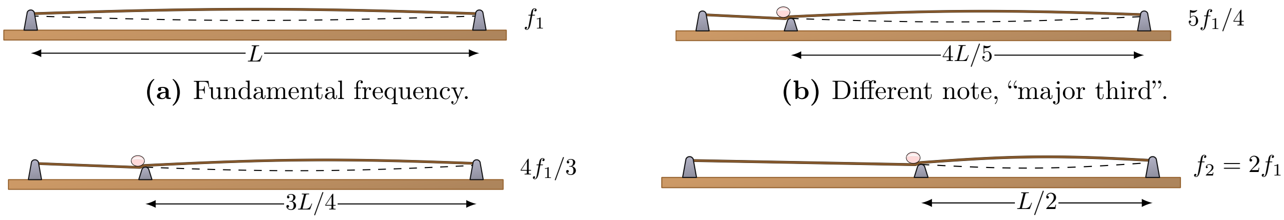
$$f_n = n f_1, \tag{13.78}$$

can be expressed as the multiple of the *fundamental frequency*

$$f_1 = \frac{v}{2L}, \tag{13.79}$$

which is the lowest possible frequency to form a standing wave.

You can experiment with this at home. All you need is to tie a string to some rigid support like a door knob, and smoothly pull it up and down. As you increase the frequency, you will first find the fundamental f_1. To find the next harmonic, you need to go twice as fast, $f_2 = 2f_1$.



(a) Fundamental frequency.

(b) Different note, "major third".

(c) Different note, "perfect fourth".

(d) Same note as (a), but one octave up.

Figure 13.16: By changing the length of a guitar string, different notes can be played.

13.6.3 String instruments

Many musical instrument, including guitars, pianos and violins, use strings to produce sound. Humans experience the sound a vibrating string makes as a *note* with a certain *pitch*, or frequency.

The harmonic frequency only depend on three properties of the string: the length L, mass density μ, and tension F. This is why a guitar string can be tuned by adjusting the tension with the knobs, and why each guitar string with a different mass density, can reach a different frequency. Heavier strings will produce lower notes.

At the same time, if you put your finger on the middle of the guitar string as in Fig. 13.16, you double the frequency, producing the same note, but at a higher pitch. In music, this jump between the same notes is called an *octave*.

13.6.4 Musical notes

Very broadly speaking, musical notes are a set of frequencies that to humans sound "good" or "harmonious" when played together. Pythagoras had discovered that this happens when the ratio of the frequencies is (approximately) a rational number, i.e. the ratio of two whole numbers, like 5/4, 4/3 or 3/2. In the western music, the frequency interval of each full octave is traditionally divided into twelve notes, or *semi-tones*, that follow some set of ratios. To play a different note on a string, you have to put your finger in just the right spot to get the note that corresponds to a frequency that has the correct ratio with respect to the other notes. Figures 13.16b and 13.16c show some examples.

Conventionally, the musical note "A" (also known as "la") is tuned to 440 Hz and all other notes can be derived from that. This choice is largely arbitrary, but by fixing the frequency of one note, different instruments can be adjusted to be in tune when playing together.

13.6.5 In a pipe of air

Standing waves can also form in a tube of air. The derivation of the expressions is very similar to strings, but one needs to consider different boundary conditions.

Closed tube

First consider a closed tube of length L filled with air as in Fig. 13.17. Say you move a piston or hit a membrane on one end to produce sinusoidal pressure waves. When the pressure wave hits the other closed end at $x = L$, it will reflect the wave and interfere with the incoming wave. Under the right conditions, a standing wave will form.

Because the air molecules cannot move at the closed ends, the longitudinal displacement s has to be zero at all times. Therefore, there are displacement nodes in the standing s

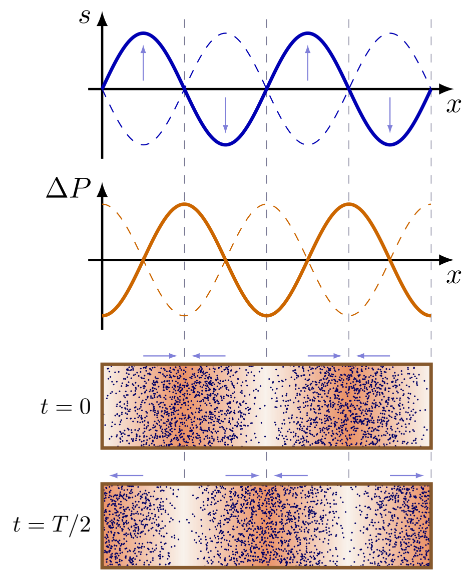
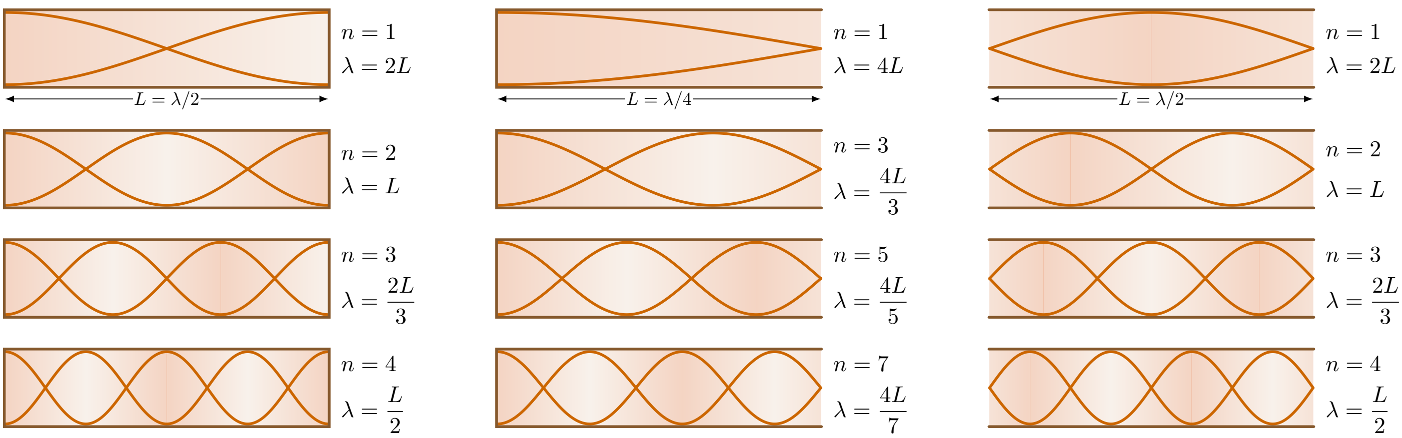


Figure 13.17: Standing wave of harmonic $n = 4$ in a closed tube filled with air. Top plot shows the local, average displacement s of air molecules in the longitudinal (x) direction (blue). The second plot shows the pressure variation ΔP (orange), which is 90° out of phase with s. The solid line is the wave at $t = 0$, while the dashed line is at $t = T/2$, when the wave is reversed.



(a) Closed on both ends.

(b) Half-open tube. Only odd harmonics appear.

(c) Open on both ends.

Figure 13.18: Harmonics of standing pressure waves in tubes of air. Closed ends have maximal pressure differences and coincides with a pressure anti-node (displacement node). Open ends, where the air is in contact with the atmospheric pressure, the pressure variation is zero, and corresponds to a pressure node (displacement anti-node).

wave at either ends. This also means that at either ends, the air molecules will bunch up the most, creating the peak pressure variations ΔP, or pressure anti-nodes. The pressure wave will therefore have the form

$$\Delta P(x,t) = 2A\cos(\omega_n t)\cos(k_n x) \tag{13.80}$$

with boundary condition

$$\cos(k_n L) = 1. \tag{13.81}$$

Therefore,

Standing waves in a closed or open tube.

$$\lambda_n = \frac{2L}{n}, \qquad k_n = \frac{n\pi}{L} \tag{13.82}$$

for non-zero integers $n = 1, 2, 3, ...$

Notice that this formula works if the two ends are both closed or both open, since in both cases the same number of waves fit in the tube. Several harmonics are shown in Fig. 13.18a.

Half-open tube

Now consider the case one end is open as in Fig. 13.18b. At the open end, the air molecules are fully free to move back and forth, while the pressure must be more or less fixed, due to the exposure to atmospheric pressure. Therefore, the new boundary condition for Eq. (13.80) is

$$\cos(k_n L) = 0, \tag{13.83}$$

so that

Standing waves in a half-open tube.

$$\lambda_n = \frac{4L}{n}, \qquad k_n = \frac{n\pi}{2L} \tag{13.84}$$

for odd integers $n = 1, 3, 5, ...$

Open tube

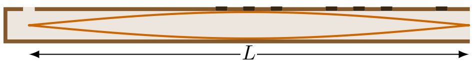
Standing waves in a tube open on both ends must have nodes on those ends, and therefore have the form

$$\Delta P(x,t) = 2A\cos(\omega_n t)\sin(k_n x), \tag{13.85}$$

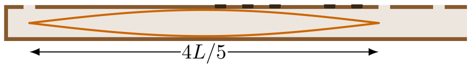
and has same solutions as in Eq. (13.82).

Wind instruments

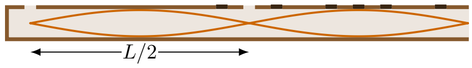
Wind instruments use standing waves in a pipe to produce notes. They come in many shapes, cylindrical like flutes and organs, or conical like clarinets, trumpets or saxophones. Typically, by controlling the openings of holes and valves, a player can change which harmonic is excited by changing which holes are in contact with the atmosphere and form



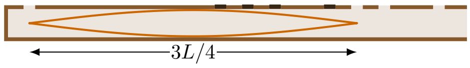
(a) Fundamental frequency.



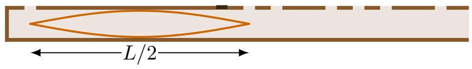
(b) Different note, "major third".



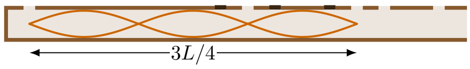
(c) Second harmonic of (a), same note, one octave up.



(d) Different note, "perfect fourth".



(e) Same note as (a), one octave up.



(f) Third harmonic of the "perfect fourth".

Figure 13.19: Acoustics of a flute. By controlling the air holes and air flow, different notes can be played. Opening the lowest holes makes the flute effectively shorter. By opening a hole in the middle, a node can be created to excite a higher harmonic of a given note.

nodes. In the case of a trombone, the length is increased by moving the slide. Organs, trumpets, saxophones and clarinets are examples of half-open pipes, while flutes are open pipes.

13.7 Energy transmission in a wave

Waves can carry energy and do work. For example, as a wave travels from left to right on a string under tension, it can lift some weight that hangs on the string. The pressure waves of sound can make other objects like ear drums or glasses vibrate by transferring energy through pressure.

13.7.1 Power of a wave on a string

Consider a sine wave with amplitude A and angular frequency ω on a string. Let's look at a small segment like we did before in Fig. 13.4. The segment has some length Δs and mass $\Delta m = \mu \Delta s$. Because we are considering a sine wave, our piece of string is a simple harmonic oscillator moving up and down. Last chapter, we figured out that the total energy (potential plus kinetic) of an oscillator is

$$E = \frac{1}{2} k A^2, \tag{13.86}$$

where $k = m\omega^2$ is the spring constant, not to be confused with the wave number. For our small segment with mass $\Delta m = \mu \Delta s$, we therefore have

$$\Delta E = \frac{1}{2} (\mu \Delta s) \omega^2 A^2. \tag{13.87}$$

The power of energy that the wave carries across the string then is

$$P = \frac{\Delta E}{\Delta t} = \frac{1}{2} \mu \omega^2 v A^2, \tag{13.88}$$

where $v = \Delta s / \Delta t$ is the wave's velocity. (Do not confuse P with pressure.) Notice that the power is proportional to A^2, ω^2, μ and v.

13.7.2 Exciting harmonics

Vibrating systems generally have multiple standing waves that are superimposed. In general, you can think of a wave as the sum of several harmonics with varying amplitudes A_n:

$$y(x,t) = \sum_n A_n \cos(\omega_n t + \phi_n) \sin(k_n x). \tag{13.89}$$

$\omega_n^2 A_n^2$ is the relative energy of the n^{th} harmonic w.r.t. the other harmonics. The phase ϕ_n depends on the initial conditions.

The location of the initial impact on the string that creates the standing wave, determines the values of each harmonic's amplitude A_n, and therefore their relative energies. Like a guitar player, we can "pluck" or hit a string to create a standing wave. If we pull exactly in the middle, at $x = L/2$, just the fundamental harmonic $n = 1$ with one anti-node in the middle is excited. However, if we pick at any other random place, we can excite more than one harmonic, and get some set of amplitudes A_n, called a *spectrum*. The higher-order harmonics are called the *overtones* to the fundamental. The resulting standing wave will be some superposition of harmonics. Often, most energy will still go into the fundamental.

We have seen that a wave on a string depends on the length L of the string. The same is true for waves on two- or three-dimensional objects. As standing waves strongly depend on the boundary conditions and the location of the (anti-)nodes, differently shaped objects of the same material have different wave functions. And again, plucking or hitting the object at different locations will create a different set of A_n. An example of standing waves in 2D, is the so-called Chlandi plate, which are simple square metal plates that, when vibrated, create beautifully regular patterns in a thin layer of sand, lying on top. In music, drum head and cymbals also form standing waves when hit by the drummer.

Equation (13.89) and what is discussed above will be put in more context in in Chapter 15 on Fourier analysis.

13.7.3 Energy density & intensity

Consider again a source that emits waves spherically in three dimensions, Fig. 13.20. As each wavefront spreads out, its energy is distributed over a larger area. Therefore, it is useful to define

Intensity.

$$I = \frac{P}{A}, \tag{13.90}$$

where P is the power, and A is the area of the wavefront. In case of spherically emitted waves, the surface area is $A = 4\pi r^2$. So as the wave moves outward, the area increased with r^2, while the intensity decreases with $1/r^2$.

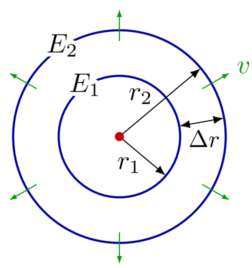


Figure 13.20: Cross section of two spherical wavefronts that carry some energy. The wave speed is v.

Let's look at the relationship between intensity I and the energy density defined as the energy per unit volume,

Energy density.

$$\eta = \frac{\Delta E}{\Delta V}. \tag{13.91}$$

Take a spherical shell of radial thickness Δr as in Fig. 13.20. If the wave travels with velocity v, it will take a wavefront $\Delta t = \Delta r/v$ to go through the shell. The additional energy that is contained in this spherical shell is given by the energy density

$$\Delta E = \eta \Delta V. \tag{13.92}$$

A very thin spherical shell has volume $\Delta V = A\Delta r$, so

$$\Delta E = \eta A(v\Delta t). \tag{13.93}$$

The power then, is given by

$$P = \eta v A \tag{13.94}$$

Intensity of a wave is therefore the energy density times velocity,

Intensity of a wave.

$$I = \eta v. \tag{13.95}$$

Sound wave

But what is η? Let's consider a sinusoidal sound wave. We remember for a simple harmonic oscillator that $E = m\omega^2 A^2/2$. In case of sound, the air molecules vibrate back and forth with an amplitude s_0, so the energy in a small volume ΔV with mass $m = \rho\Delta V$ is given by

$$\Delta E = \frac{1}{2}(\rho\Delta V)\omega^2 s_0^2. \tag{13.96}$$

The energy density therefore is

$$\eta = \frac{1}{2}\rho\omega^2 s_0^2, \tag{13.97}$$

or in term of pressure variations, using Eq. (13.50),

$$\eta = \frac{\Delta P_0^2}{2\rho v^2}. \tag{13.98}$$

Therefore, the intensity of a sound wave is given

Intensity of a sound wave.

$$I = \frac{\Delta P_0^2}{2\rho v} \tag{13.99}$$

13.7.4 Decibel scale

The human ear can hear a range intensities from $I = 10 \times 10^{-12}\,\mathrm{W/m^2}$ (breathing) up to $1\,\mathrm{W/m^2}$ (rock concert speakers at 2 m). Because this range spans over twelve orders of magnitude, intensity of sound, i.e. "loudness", is measured with the logarithmic *decibel scale*, defined as

Decibel scale for intensity.

$$\beta = 10 \log \left(\frac{I}{I_0} \right) \,\mathrm{dB}, \tag{13.100}$$

where the reference intensity is $I_0 = 10^{-12}\,\mathrm{W/m^2}$ as per convention. Although β is dimensionless, it is counted in units of decibels (dB). The logarithm is understood to be in base 10, so $\log(10^x) = x$. Therefore, if $I = 10^{-12}\,\mathrm{W/m^2}$, $\beta = 0\,\mathrm{dB}$, while for $I = 1\,\mathrm{W/m^2}$, $\beta = 120\,\mathrm{dB}$.

Logarithms convert multiplication to addition, which is useful to span large orders of magnitudes. Say you are playing music at an intensity I, and you increase the volume of your speaker by $\Delta\beta = 10\,\mathrm{dB}$, then the new intensity is given by

$$10 \log \left(\frac{I'}{I_0} \right) \,\mathrm{dB} = 10 \log \left(\frac{I}{I_0} \right) \,\mathrm{dB} + \Delta\beta, \tag{13.101}$$

which leads to

$$I' = 10^{\frac{\Delta\beta}{10\,\mathrm{dB}}} I. \tag{13.102}$$

So a sound that is $\Delta\beta = 10\,\mathrm{dB}$ louder actually has an intensity that is larger by a factor ten. An increase of $\Delta\beta = 20\,\mathrm{dB}$ is an increase of factor 100, $\Delta\beta = 3\,\mathrm{dB}$ is approximately a factor 2 and so on. This is depicted in Fig. 13.21. Remember that the intensity depends on distance. Additionally, how loud humans perceive a sound, also strongly depends on the frequency: Very low or very high frequencies do not seem as loud at the same pressure as the middle range to which our ears are most sensitive.

Because $I \propto \Delta P^2$, as per Eq. (13.99), a slightly different decibel scale is defined in terms of a pressure wave:

Decibel scale for pressure amplitude.

$$\beta_P = 20 \log \left(\frac{\Delta P}{\Delta P_0} \right) \,\mathrm{dB}, \tag{13.103}$$

with reference pressure $\Delta P_0 = 20\,\mu\mathrm{Pa} = 2 \times 10^{-5}\,\mathrm{Pa}$ in air. This very roughly corresponds to I_0. This time, an increase of $\Delta\beta_P = 20\,\mathrm{dB}$ is a factor $10^{\frac{\Delta\beta_P}{20\,\mathrm{dB}}} = 10$, etc.

The decibel scale returns in other fields like electronics as well.

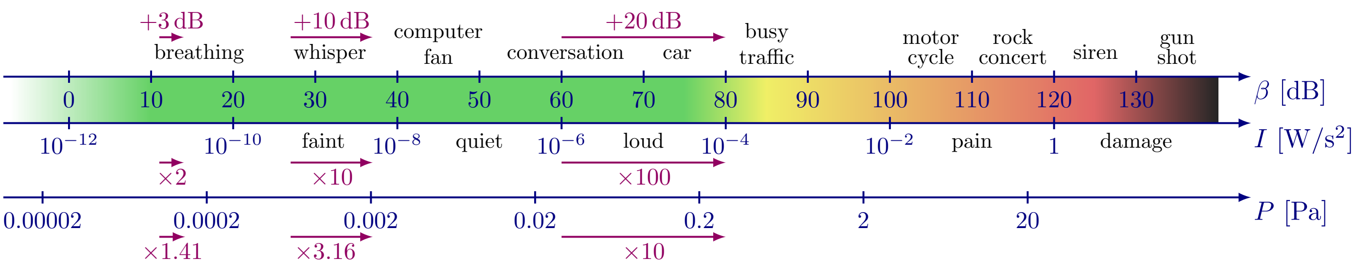
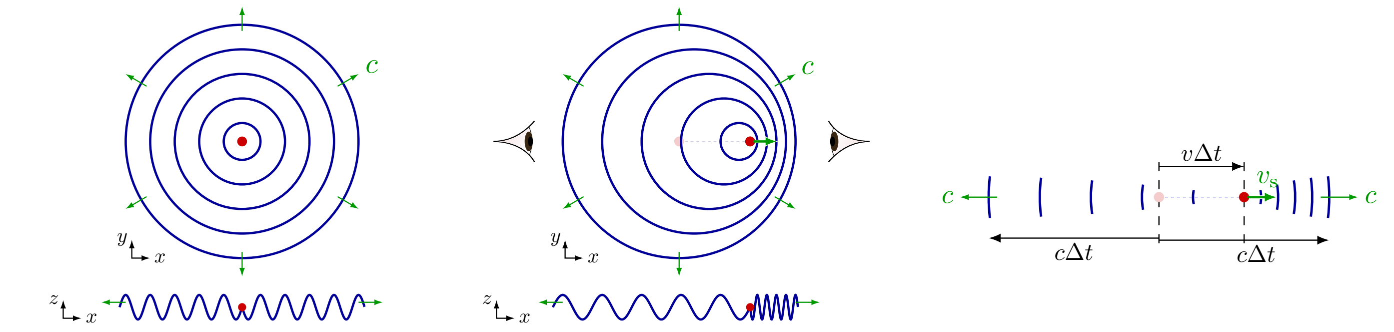


Figure 13.21: Decibel scale for sound pressure with some typical examples.



(a) Stationary source (red point) emitting waves with speed c.

(b) Source moving with velocity v. Wavefronts get compressed in front.

(c) Wavefronts in front and back traveled the same distance after Δt.

Figure 13.22: Doppler effect in a wavefront diagram: When a source emitting wave moves, the distance between wavefronts (maxima) will vary in each direction.

13.8 Doppler effect

When a source emitting waves moves relative to a receiver the the observed frequency f' is not the same as the frequency f at which the source emits the waves. This is called the *Doppler effect*. We experience it in our daily lives, when a fast car with a loud engine or an ambulance with its sirens on zooms past us. The same effect, however, can also be observed with light in astronomy, such as the light emitted by stars moving relative to us.

13.8.1 Moving source

To understand, first consider the stationary source in Fig. 13.22a that emits waves in all spherical directions with constant frequency f and wave speed c. Each wavefront (the wave's maximum) forms a radially expanding circle with the source as center. The distance between two wavefronts is always λ. But something interesting happens when the source starts moving in some direction with constant velocity $v_\mathrm{s} < c$, as in Fig. 13.22b. Say that the source emits N waves in some time interval Δt,

$$N = f\Delta t. \tag{13.104}$$

After Δt, the first wavefront has traveled a distance $c\Delta t$ in each direction. The source, however, has also and moves a distance $v_\mathrm{s}\Delta t$. So in the direction of the source's motion, N wavefronts are compressed into a distance $c\Delta t - v_\mathrm{s}\Delta t$, while to the opposite direction they are stretched over $c\Delta t + v_\mathrm{s}\Delta t$.

So what does an observer see who stand in the front of the moving source? To them, the wavelength appears shorter:

$$\lambda_\text{front} = \frac{(c - v_\mathrm{s})\Delta t}{N} = \frac{c - v_\mathrm{s}}{f}, \tag{13.105}$$

because they count N wavefront over a distance $(c - v_\mathrm{s})\Delta t$. Similarly, for an observer standing in the back, and the source is moving away from them,

$$\lambda_\text{back} = \frac{c + v_\mathrm{s}}{f}. \tag{13.106}$$

This can be converted back with Eq. (13.16) to the frequency measured by the observer;

Doppler effect for a moving source.

$$f'_{front} = \left(\frac{c}{c - v_{\mathrm{s}}}\right) f \tag{13.107}$$

$$f'_{back} = \left(\frac{c}{c + v_{\mathrm{s}}}\right) f. \tag{13.108}$$

13.8.2 Moving observer

If source is stationary, but an observer is moving at a speed v_{r}, we see the same effect. The observer moving toward the source counts more wavefront in time Δt,

$$N = \frac{(c + v_{\mathrm{r}})\Delta t}{\lambda}. \tag{13.109}$$

In terms of frequency,

$$f'_{\mathrm{toward}} = \frac{N}{\Delta t} = \frac{c + v_{\mathrm{r}}}{\lambda}. = \left(\frac{c + v_{\mathrm{r}}}{c}\right) f. \tag{13.110}$$

Same for an observer moving away from the source

$$f'_{\mathrm{away}} = \left(\frac{c - v_{\mathrm{s}}}{c}\right) f. \tag{13.111}$$

13.8.3 General formula

In general, we can combine the frequencies into one formula

Doppler effect.

$$f' = \left(\frac{c + v_{\mathrm{r}}}{c - v_{\mathrm{s}}}\right) f, \tag{13.112}$$

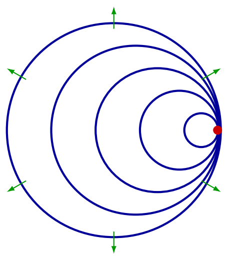
where $v_{\mathrm{s}} > 0$ if the source is moving toward the observer, $v_{\mathrm{s}} < 0$ if the source is moving away from the observer, while $v_{\mathrm{r}} > 0$ if the observer is moving toward the source, and $v_{\mathrm{r}} < 0$ if the observer is moving away.

Note that above about 10% of the speed of light, some corrections are needed to the equations above.

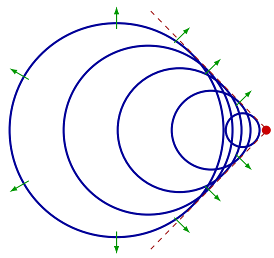
13.8.4 Sonic boom

If the source accelerates, it will at some point reach the speed of the wave, $v_{\mathrm{s}} = c$. In case of sound, the wavefronts of the pressure waves will coincide and add up to create a huge pressure shock, known as a *sonic boom.* If the source source like a supersonic plane moves faster than sound, $v_{\mathrm{s}} > c$, wave fronts will still line up sideways in the shape of a growing cone as in Fig. 13.23b, leaving behind a large shock wave. This is similar to the triangular wake of a boat sailing on water, when the boat is quicker than the waves on the water surface.

Besides supersonic planes, the crack of a whip or bullet are also examples of sonic booms.



(a) When the source moves at the speed of sound, $v_s = c$, wavefronts align in the front.



(b) If the source moves faster than the speed of sound, $v_s > c$, the wave fronts align sideways (red line).

Figure 13.23: Breaking the sound barrier: A supersonic source causes a shockwave, called a sonic boom.

Chapter 14

Complex Numbers

Even though everything we observe and measure in physics has a real value, complex numbers still are useful and have many applications; from oscillators and waves to electronics and quantum mechanics. As we will see in this chapter, they are particularly handy for solving differential equations like those for harmonic oscillators.

14.1 Basics

The *imaginary number* is defined as

$$i = \sqrt{-1}. \tag{14.1}$$

Notice that it follows that

$$i^2 = -1, \quad i^4 = +1, \quad \frac{1}{i} = -i. \tag{14.2}$$

Complex number have the general form

$$z = x + yi \tag{14.3}$$

with a *real part* $\mathrm{Re}[z] = x$ and an *imaginary part* $\mathrm{Im}[z] = y$, They can behave like vectors in the *complex plane* with coordinates (x, y). This plane is spanned by the *real axis* (values of x) and the *imaginary axis* (values of y), which are perpendicular to each other as shown in Fig. 14.1. Complex number are also useful to write down in terms of polar angles $(r; \theta)$:

$$z = r\cos\theta + ir\sin\theta, \tag{14.4}$$

where the radius $|z| = r$ is the *modulus* and the polar angle $\mathrm{Arg}[z] = \theta$ is the *argument*, given by

$$r = \sqrt{x^2 + y^2}, \quad \tan\theta = \frac{y}{x}. \tag{14.5}$$

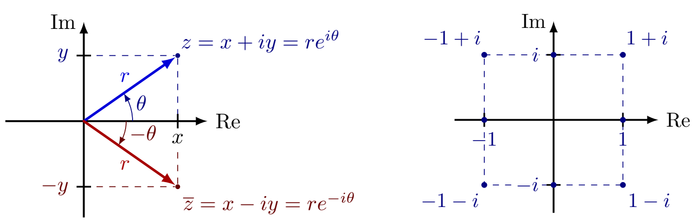


Figure 14.1: Complex plane with the real and imaginary axis. A complex number $z = x + iy$ is given by the coordinates $x = \mathrm{Re}[z]$ and $y = \mathrm{Im}[z]$. The complex conjugate $\overline{z} = x - iy$ is the reflection of z with respect to the real axis.

14.1.1 Euler's formula

It turns out that any complex number z can be written as

$$z = re^{i\theta} = r\exp(i\theta) \tag{14.6}$$

There are many different proofs of this formula, but it can be shown most readily with a Taylor series. For the exponential function, the Taylor series is

$$\exp\theta = 1 + \theta + \frac{\theta^2}{2} + \frac{\theta^3}{3} + \frac{\theta^4}{4} + \frac{\theta^5}{5} + ... = \sum_{n=0}^{\infty} \frac{1}{i} x^i., \tag{14.7}$$

and when we plug in a purely imaginary argument $x = i\theta$, this becomes

$$\exp(i\theta) = 1 + i\theta - \frac{\theta^2}{2} - i\frac{\theta^3}{3} + \frac{\theta^4}{4} + i\frac{\theta^5}{5} + ... \tag{14.8}$$

Comparing this to Eqs. (12.13) and (12.17),

$$\sin\theta = \theta - \frac{\theta^3}{3} + \frac{\theta^5}{5} + ... = \sum_{n=0}^{\infty} \frac{(-1)^2}{(2n+1)} x^{2n+1} \tag{14.9}$$

$$\cos\theta = 1 - \frac{\theta^2}{2} + \frac{\theta^4}{4} + ... = \sum_{n=0}^{\infty} \frac{(-1)^2}{(2n)} x^{2n}, \tag{14.10}$$

we notice we can pair off the even terms, which are real, and the odd terms, which are imaginary, such that

Euler's formula.

$$e^{i\theta} = \cos\theta + i\sin\theta. \tag{14.11}$$

For $\theta = \pi$,

$$e^{i\pi} + 1 = 0, \tag{14.12}$$

which is one of the most compact equations that contains all the important constants in mathematics.

14.1.2 Complex conjugate

Notice that if we replace θ with $-\theta$, only the imaginary part is "flipped",

$$e^{-i\theta} = \cos\theta - i\sin\theta. \tag{14.13}$$

This is called *complex conjugation*. The *complex conjugate* of a complex number $z = x + iy = re^{i\theta}$ is

$$\overline{z} = x - iy \tag{14.14}$$

$$= r\cos\theta - ir\sin\theta = re^{-i\theta}. \tag{14.15}$$

An alternative notation for complex conjugation is z^*. This operation can be understood as a reflection in the complex plane with respect to the real axis, shown in red in Fig. 14.1. The conjugate of $\overline{z}$ is again z:

$$\overline{\overline{z}} = z. \tag{14.16}$$

With this definition, we can actually define the modulus of a complex number as

$$|z| = \sqrt{z\overline{z}} \tag{14.17}$$

It is easy to show this is consistent with Eq. (14.5). Notice that $\left|e^{i\theta}\right| = 1$.

The conjugate also allows us to cancel the real or imaginary parts; it's easy to show that

$$\mathrm{Re}[z] = \frac{z+\overline{z}}{2} \tag{14.18}$$

$$\mathrm{Im}[z] = \frac{z-\overline{z}}{2i}. \tag{14.19}$$

A complex number is "purely" real if $z = \overline{z}$ and thus the imaginary part is $\mathrm{Im}[z] = 0$.

14.1.3 Complex form of goniometric functions

These sine and cosine functions are one of the reasons why complex numbers are useful for studying oscillations and waves. Notice that we can use Euler's formula to write

$$\cos\theta = \frac{e^{i\theta}+e^{-i\theta}}{2} \tag{14.20}$$

$$\sin\theta = \frac{e^{i\theta}-e^{-i\theta}}{2i} \tag{14.21}$$

Remember that sine is an odd function, while cosine is an even one. This works out in our formulas above:

$$\sin(-\theta) = \frac{e^{-i\theta}-e^{+i\theta}}{2i} = -\sin\theta \tag{14.22}$$

$$\cos(-\theta) = \frac{e^{-i\theta}+e^{+i\theta}}{2} = \cos\theta \tag{14.23}$$

Furthermore, it's easy to show from the above, or directly from Fig. 14.1, that

$$e^{in\pi} = \begin{cases} 1 & \text{for even } n \\ -1 & \text{for odd } n \end{cases} \tag{14.24}$$

14.1.4 Extra: Angle-sum formula

One neat trick with the complex form of sine and cosine is to quickly derive some goniometric formulas that are hard to remember. All you need is to remember Euler's formula Eq. (14.11) to write

$$e^{i(\alpha+\beta)} = \cos(\alpha+\beta) + i\sin(\alpha+\beta) \tag{14.25}$$

$$= e^{i\alpha}e^{i\beta} = \big(\cos\alpha + i\sin\alpha\big)\big(\cos\beta + i\sin\beta\big). \tag{14.26}$$

Reshuffling a bit, we can identify the real and imaginary parts

$$\cos(\alpha+\beta) + i\sin(\alpha+\beta) = \big(\cos\alpha\cos\beta - \sin\alpha\sin\beta\big) + i\big(\sin\alpha\cos\beta + \sin\alpha\cos\beta\big), \tag{14.27}$$

and one can for example see that the imaginary parts yield:

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \sin\alpha\cos\beta. \tag{14.28}$$

14.2 Quadratic equations

Complex numbers are useful for solving quadratic equations. For example,

$$z^2 - 2z + 2 = 0. \tag{14.29}$$

has no *real* solution because the discriminant $d = 2^2 - 4 \cdot 1 \cdot 2 = -4$ is negative, so the square root is negative. Therefore, there are two *complex* solutions

$$z_\pm = \frac{2 \pm \sqrt{-4}}{2} = 1 \pm i. \tag{14.30}$$

A quadratic equation can either have one real root ($d = 0$), two real roots ($d > 0$), or two complex roots ($d < 0$). The roots of an equation of the form $ax^2 + c = 0$ are purely imaginary if $d = -4ac < 0$, so if $ac > 0$, which means a and c have the same sign (both positive, or both negative).

14.3 Solving second order differential equations

14.3.1 Characteristic equation

A homogeneous second-order differential equation is of the form

$$a_2 \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + a_1 \frac{\mathrm{d}y}{\mathrm{d}t} + a_0 y = 0, \tag{14.31}$$

with real constants $a_0, a_1, a_2 \in \mathbb{R}$. Suppose for a second that one solution has the form $y(t) = e^{rt}$, where r is a complex number. Plugging in our ansatz, we find

$$a_2 r^2 e^{rt} + a_1 r e^{rt} + a_0 e^{rt} = 0, \tag{14.32}$$

or, because $e^{rt} \neq 0$, and for $r \neq 0$,

Characteristic equation for second-order differential equation.

$$a_2 r^2 + a_1 r + a_0 = 0. \tag{14.33}$$

If the roots of this *characteristic equation* are two complex numbers r_1 and r_2, then there are two linearly independent solutions that form the most general solution

$$y(t) = Ae^{r_1 t} + Be^{r_2 t}, \tag{14.34}$$

where A and B are complex constants that are determined by the initial conditions. This general solution can be rewritten in several forms depending on the solutions of characteristic equation (14.33), which may have a complex form, $r = \alpha + i\beta$ with $\alpha, \beta \in \mathbb{R}$. Let's look at some case distinctions.

Two real solutions

If the discriminant $d = a_1^2 - 4a_0a_2 > 0$ is strictly positive, then there are two real solutions for y in Eq. (14.33);

$$\alpha_\pm = \frac{-a_1^2 \pm \sqrt{a_1^2 - 4a_0a_2}}{2a_0}. \tag{14.35}$$

The general solution (14.34) has the form of two exponentials

$$y(t) = A_+e^{\alpha_+ t} + A_-e^{\alpha_- t}, \tag{14.36}$$

with some complex constants $A_\pm$ determined by initial conditions.

One real solution

If the discriminant $d = a_1^2 - 4a_0a_2 = 0$, then there is only one root,

$$\alpha = -\frac{a_1^2}{2a_0}. \tag{14.37}$$

and so, as discussed in Section 12.2.2, one extra solution to the second-order differential equation that is linearly independent to $e^{\alpha t}$ has to exist. It turns out this is $te^{\alpha t}$. The most general solution is therefore

$$y(t) = Ae^{\alpha t} + Bte^{\alpha t}. \tag{14.38}$$

Two complex solutions

If the discriminant $d = a_1^2 - 4a_0a_2 < 0$ is strictly negative, and $a_1 \neq 0$, then there are two non-real solution of the form

$$\alpha \pm i\beta = \frac{-a_1^2 \pm i\sqrt{4a_0a_2 - a_1^2}}{2a_0}. \tag{14.39}$$

The solution to the differential equation is

$$y(t) = e^{\alpha t}\big(B_+e^{i\beta t} + B_-e^{-i\beta t}\big). \tag{14.40}$$

By substituting Euler's formula (14.11), we find another form that is like an underdamped harmonic oscillator Eq. (12.85):

$$y(t) = e^{\alpha t}\big(A\cos(\beta t) + B\sin(\beta t)\big), \tag{14.41}$$

where $A = B_+ + B_-$ and $B = iB_+ - iB_-$ are new constants.

Two imaginary solutions

If $a_1 = 0$ and the discriminant $d = -4a_0a_2 < 0$ is strictly negative, there will be two imaginary roots of the form

$$\pm i\beta = \pm i\sqrt{\frac{a_2}{a_0}}, \tag{14.42}$$

and the general solution to the differential equation is of the form

$$y(t) = B_+e^{i\beta t} + B_-e^{-i\beta t}, \tag{14.43}$$

or equivalently, a simple harmonic oscillator,

$$y(t) = A\cos(\beta t) + B\sin(\beta t). \tag{14.44}$$

14.3.2 Differential operator

A different formalism for the method discussed in the previous section relies on the so-called *linear differential operator*, defined as

$$\mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}t}, \tag{14.45}$$

which acts on t-dependent functions y. Notice that this is a linear operator acting on functions, meaning that for some functions y, y_1, y_2,

$$\mathrm{D}\big(y_1 + y_2\big) = \mathrm{D}y_1 + \mathrm{D}y_2 \tag{14.46}$$

$$\mathrm{D}\big(ay\big) = a\mathrm{D}y, \tag{14.47}$$

where a is a constant (see Section 3.8). Higher-order derivatives is like applying D several times, for example,

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = \mathrm{D}\big(\mathrm{D}y\big) = \mathrm{D}^2 y. \tag{14.48}$$

Notice that D^2 is still a linear operator.

We can rewrite the second-order equation Eq. (14.31) as

$$\big(a_2\mathrm{D}^2 + a_1\mathrm{D} + a_0\big)y = 0, \tag{14.49}$$

where the terms between the parentheses are a linear combination of linear operators. (The linear operator with a_0, the *identity operator* $\mathbb{1}y = y$, is hidden.) Equivalent to the characteristic equation Eq. (14.33), the algebraic form of this differential equation is

$$a_2\mathrm{D}^2 + a_1\mathrm{D} + a_0 = 0, \tag{14.50}$$

which can again be used to find the general solutions for different roots of a second order equation.

14.3.3 Simple harmonic oscillator

Let's take the example of the simple harmonic oscillator in Eq. (12.21),

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + \omega^2 y = 0, \tag{14.51}$$

and rewrite this with the differential operator

$$0 = \mathrm{D}^2 y + \omega^2 y \tag{14.52}$$

$$= (\mathrm{D}^2 + \omega^2)y \tag{14.53}$$

$$= (\mathrm{D} + i\omega)(\mathrm{D} - i\omega)y. \tag{14.54}$$

There are thus two imaginary roots $\pm i\omega$, which result in two independent differential equations

$$(\mathrm{D} + i\omega)y = 0, \tag{14.55}$$

$$(\mathrm{D} - i\omega)y = 0. \tag{14.56}$$

The second equation leads to

$$\frac{\mathrm{d}y}{\mathrm{d}t} = i\omega y, \tag{14.57}$$

and therefore

$$\int \frac{\mathrm{d}y}{y} = \int i\omega \mathrm{d}t. \tag{14.58}$$

The integrals are

$$\ln y = (i\omega + C), \tag{14.59}$$

with an integration constant C determined by initial conditions. Taking the logarithm to obtain y,

$$y(t) = Ae^{i\omega t}, \tag{14.60}$$

with a new integration constant $A = e^C$. This function does behave like a harmonic oscillator. Namely, due to Euler's formula,

$$y(t) = A\cos(\omega t) + iA\sin(\omega t), \tag{14.61}$$

which is like a vector of constant length A rotating uniformly in the complex plane in the counterclockwise direction, as shown in Fig. 14.2. It has a real cosine and imaginary sine component.

For the second differential equation (14.55), we similarly find

$$y(t) = Be^{-i\omega t}, \tag{14.62}$$

with some other integration constant B. This rotates clockwise in the complex plane. In general, the solution to the simple harmonic oscillator (14.51) is therefore

Complex solutions to the simple harmonic oscillator equation.

$$y(t) = Ae^{i\omega t} + Be^{-i\omega t}. \tag{14.63}$$

This is consistent with the general solution (14.43) we found earlier for purely imaginary roots of the characteristic equation. Again, we can rewrite this complex form to retrieve a sine and cosine

$$y(t) = A\cos(\omega t) + B\sin(\omega t). \tag{14.64}$$

as in Eqs. (12.36) and (14.44). As discussed in Section 12.2.2, this can be rewritten as a single cosine with a phase

$$y(t) = A\cos(\omega t - \phi). \tag{14.65}$$

14.3.4 Initial conditions & real solutions

The constants A, B, C and ϕ in the above equations, can be obtained by imposing initial conditions. Consider again form (14.63) and take for example

$$y(0) = 0 \tag{14.66}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = v(0) = v_0. \tag{14.67}$$

In order for $y(t)$ to be physical, it has to be real. So we need to get rid of the imaginary part. The imaginary part of y is vanishes if $y = \overline{y}$ for any value of t. Thus,

$$0 = (A - \overline{B})e^{i\omega t} + (B - \overline{A})e^{-i\omega t}. \tag{14.68}$$

Because $e^{i\omega t}$ and $e^{-i\omega t}$ are linearly independent functions, this equation only holds when

$$A - \overline{B} = 0 = B - \overline{A}, \tag{14.69}$$

so

$$A = \overline{B}. \tag{14.70}$$

Now, because $y(0) = 0$, Eq. (14.63) at $t = 0$ becomes

$$y(0) = 0 = A + \overline{A}, \tag{14.71}$$

and therefore A has to be purely imaginary, $\text{Re}[A] = 0$. The solutions is

$$y(t) = Ae^{i\omega t} - Ae^{-i\omega t}. \tag{14.72}$$

Consider now the second initial condition Eq. (14.67):

$$\frac{\mathrm{d}y}{\mathrm{d}t} = v_0 = 2i\omega A. \tag{14.73}$$

Therefore,

$$A = \frac{v_0}{2i\omega}. \tag{14.74}$$

After a bit of rearranging, we find the 100% guaranteed real, and therefore physical, solution

$$y(t) = \frac{v_0}{\omega} \sin \omega t. \tag{14.75}$$

14.3.5 Extra: Analytic representation and complex phase

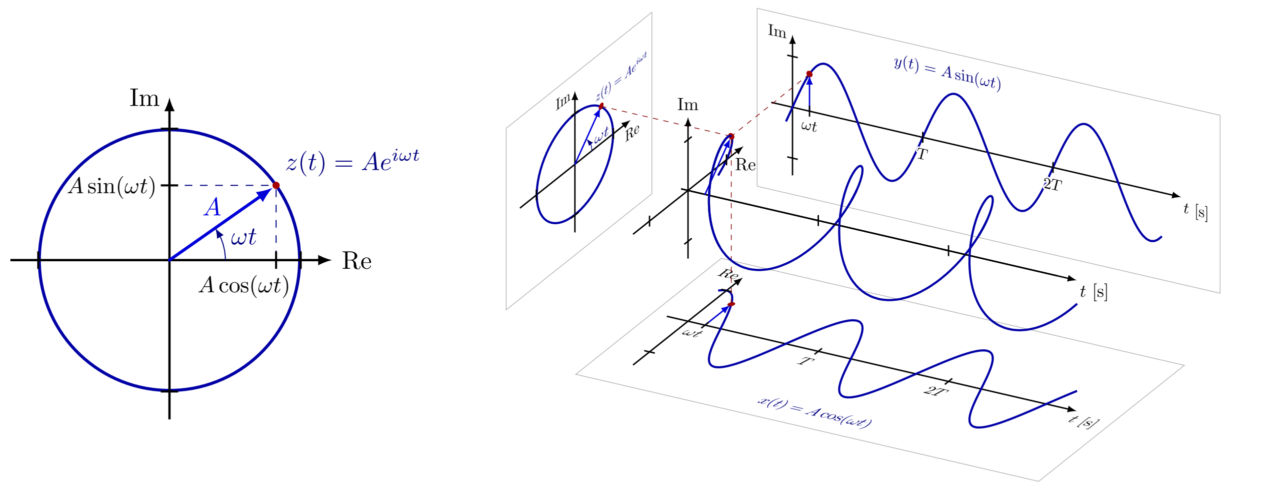
This section is only for the interested reader, and will not be covered on the exam.

Say you have a real harmonic oscillator of the form

$$y(t) = A\cos(\omega t - \phi). \tag{14.76}$$

The complex form of this oscillator is called the *analytic representation*:

$$z(t) = A\cos(\omega t - \phi) + Ai\sin(\omega t - \phi). \tag{14.77}$$



(a) The oscillation in the complex plane is a circle.

(b) The oscillation $z = Ae^{i\omega t}$ has a real and imaginary component. With the time axis, it forms a helix.

Figure 14.2: A simple harmonic oscillation with amplitude A and angular frequency ω can be represented by a complex number $z = Ae^{i\omega t}$.

such that the "physical" form is the real part $y = \text{Re}[z]$. One of the reasons this is interesting, is because the sine and cosine can readily be converted to an exponential thanks to Euler,

$$z(t) = Ae^{i(\omega t - \phi)}. \tag{14.78}$$

This allows us to more easily manipulate the argument, where the phase becomes a multiplicative term because it can be factorized as

$$z(t) = Ae^{i\omega t}e^{-i\phi}. \tag{14.79}$$

This new term $\left|e^{i\phi}\right| = 1$ has unit length. Factorizing the complex phase term is something that will return in fields like quantum mechanics, where the modulus of wave functions of this form are often squared such that the phase disappears.

14.3.6 Extra: Rotation in the complex plane

This section is only for the interested reader, and will not be covered on the exam. Some knowledge of linear algebra is assumed.

Notice that an extra phase $e^{i\phi}$ effectively rotates the complex number z in the complex plane by an angle ϕ as illustrated in Fig. 14.3b. Because it has modulus 1, $\left|e^{i\phi}\right| = 1$, it does not change the modulus $r = |z|$. It's immediately clear by writing z in Euler's form,

$$ze^{i\phi} = re^{i\theta}e^{i\phi} = re^{i(\theta+\phi)}. \tag{14.80}$$

But it we can make another connection if we use Cartesian coordinates:

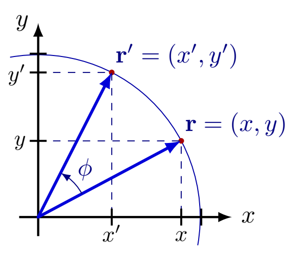
$$ze^{i\phi} = (x + iy)\,(\cos\phi + i\sin\phi) \tag{14.81}$$
$$= (x\cos\phi - y\sin\phi) + i(x\sin\phi + y\cos\phi). \tag{14.82}$$

If you have read the extra Section 10.3.2 on rotated coordinates, you will recognize the similarity with a counter-clockwise rotation in the real plane,

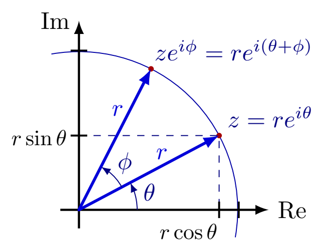
$$R(-\phi)\mathbf{r} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\phi - y\sin\phi \\ x\sin\phi + y\cos\phi \end{pmatrix}. \tag{14.83}$$

So not only do complex numbers behave as two-dimensional vectors in the complex plane, multiplying other complex numbers can act as transformations.

Any complex number $z' = e^{i\phi}$ with modulus 1 acts analogous to orthogonal matrices in the plane, which do not change the scalar product after their transformation. Later in your studies you might learn about symmetries in quantum physics, where this connection becomes important. You will see that in terms of linear algebra and group theory, the rotation matrices in a plane belong to the special orthogonal group of degree 2, SO(2), and the complex numbers with modulus 1 belong to the unitary group of degree 1, U(1).



(a) Rotation of a vector in the real plane (See Fig. 10.4).



(b) Rotation of a complex number in the complex plane.

Figure 14.3: Comparison of rotation in the real and complex plane.

14.4 Damped harmonic oscillator

Consider a harmonic oscillator with some damping that opposes velocity as discussed in Section 12.4. A new term with the damping factor b will enter our usual differential equation:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t} + \frac{b}{m}\frac{\mathrm{d}y}{\mathrm{d}t} + \omega_0^2 y = 0. \tag{14.84}$$

The characteristic equation is given by

$$\left(\mathrm{D}^2 + \frac{b}{m}\mathrm{D} + \omega_0^2\right) y = 0, \tag{14.85}$$

and has solutions given by

$$\mathrm{D} = \frac{-b \pm \sqrt{b^2 - 4m^2\omega_0^2}}{2m}. \tag{14.86}$$

Assuming $b \neq 0$, there are three different situations, depending on the sign of the discriminant $b^2/m^2 - 4\omega_0^2$. We have already seen these in Section 14.3.1 and identify them with the types of damping discussed in Section 12.4.4:

1. $b^2 > 4m^2\omega_0^2$: two real roots, *overdamped*;
2. $b^2 = 4m^2\omega_0^2$: one real roots, *critically damped*;
3. $b^2 < 4m^2\omega_0^2$: two complex roots, *underdamped.*

Here, $b_\mathrm{c} = 2m\omega_0$, is again the critical damping coefficient.

Let's look at the underdamped case (3) with two complex roots

$$\mathrm{D} = \frac{-b \pm i\sqrt{4m^2\omega_0^2 - b^2}}{2m} = -\frac{b}{2m} \pm i\omega, \tag{14.87}$$

with

$$\omega = \sqrt{m^2\omega_0^2 - \frac{b^2}{4m}}. \tag{14.88}$$

The general solution for y, then, has form

$$y(t) = Ae^{-\frac{b}{2m}t}\cos(\omega t - \phi), \tag{14.89}$$

with constant A and ϕ to be determined by initial conditions. This is consistent with our previous results, Eq. (12.85).

Chapter 15

Fourier Analysis

Fourier analysis studies the way in which functions can be represented by a linear combination of goniometric functions like sine and cosine. It was originally developed by Joseph Fourier (1768–1830) in his study of heat transfer and oscillations. The applications of Fourier analysis are very broad: solving differential equations, electronics, acoustics, spectroscopy, signal processing, data compression, and much more.

First, we will have a look at *Fourier series*, which can be used to approximate *periodic* functions with a *discrete* sum of sine and cosines. After that, we will discuss *Fourier transforms*, which are used to represent *aperiodic* functions with a *continuous* sum (i.e. integral) instead.

15.1 Interlude: Integration by parts

Before diving into Fourier analysis, recall the integration by parts method, which is a consequence of the product rule in differentation:

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(f(x)g(x)\big) = \frac{\mathrm{d}f}{\mathrm{d}x}g(x) + \frac{\mathrm{d}g}{\mathrm{d}x}f(x). \tag{15.1}$$

Integrating both sides over $\mathrm{d}x$,

$$f(x)g(x) = \int g(x)\frac{\mathrm{d}f}{\mathrm{d}x}\mathrm{d}x + \int f(x)\frac{\mathrm{d}g}{\mathrm{d}x}\mathrm{d}x. \tag{15.2}$$

A short-hand notation is often used.

Integration by parts.

$$\int u\,\mathrm{d}v = uv - \int v\,\mathrm{d}u\,. \tag{15.3}$$

Take for example,

$$I = \int x \sin x \mathrm{d}x,$$

and identify

$$\begin{cases} u = x \\ \mathrm{d}v = \sin x \mathrm{d}x \end{cases} \Rightarrow \begin{cases} \mathrm{d}v = \mathrm{d}x \\ v = -\cos x \mathrm{d}x \end{cases}$$

Therefore, we can easily solve the integral as

$$\begin{aligned} I &= -x\cos x + \int \cos x \mathrm{d}x, \\ &= -x\cos x - \sin x. \end{aligned}$$

15.2 Interlude: Averaging functions

For Fourier analysis, we will also need to know how to compute the average of a function over some interval $[a, b]$. It can be defined as

$$\langle f(x) \rangle_{[a,b]} = \frac{1}{b-a} \int_a^b f(x)\,\mathrm{d}x, \tag{15.4}$$

which is like the continuous version of the usual average $\langle x \rangle = \sum_i^N x_i/N$. A few trivial examples are

$$\langle k \rangle_{[a,b]} = k \tag{15.5}$$

$$\langle kx \rangle_{[a,b]} = \frac{k}{2}(b+a), \tag{15.6}$$

where $k \in \mathbb{R}$ is a constant. The average of sine over one half and one full period is

$$\langle \sin x \rangle_{[0,\pi]} = \frac{2}{\pi} \tag{15.7}$$

$$\langle \sin x \rangle_{[0,2\pi]} = 0. \tag{15.8}$$

Now consider the average of functions $\sin^2 x$ and $\cos^2 x$ over a full period. Even though sine and cosine are out of phase, over an interval of length $T = 2\pi$, we notice that they have the same integral,

$$\int_0^{2\pi} \sin^2 x\,\mathrm{d}x = \int_0^{2\pi} \cos^2 x\,\mathrm{d}x, \tag{15.9}$$

and the integral of their sum is simply

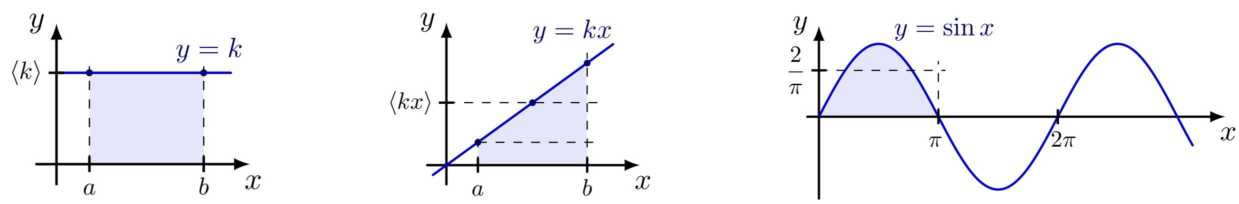
$$\int_0^{2\pi} \left(\sin^2 x + \cos^2 x\right)\mathrm{d}x = \int_0^{2\pi} \mathrm{d}x = 2\pi. \tag{15.10}$$

This result is equally shared, so

$$\int_0^{2\pi} \sin^2 x\,\mathrm{d}x = \pi = \int_0^{2\pi} \cos^2 x\,\mathrm{d}x. \tag{15.11}$$

The averages then, are these integrals divided by $b - a = 2\pi$,

$$\left\langle \sin^2 x \right\rangle_{[0,\pi]} = \frac{1}{2} = \left\langle \cos^2 x \right\rangle_{[0,\pi]}. \tag{15.12}$$



(a) Constant function $y = k$. **(b)** Linear function $y = kx$. **(c)** Sine function has an average of 0 over $[0, 2\pi]$.

Figure 15.1: Averaging basic functions over some range $[a, b]$.

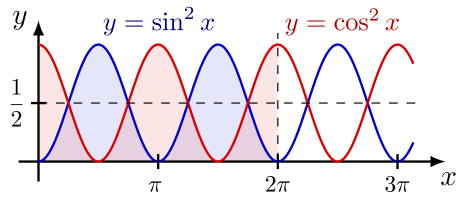


Figure 15.2: Average of $y = \sin^2 x$ and $y = \cos^2 x$.

Note that $\sin^2(nx) + \cos^2(nx) = 1$ holds for any number n, so

$$\left\langle \sin^2(nx) \right\rangle_{[a,b]} = \frac{1}{2} = \left\langle \cos^2(nx) \right\rangle_{[a,b]} . \tag{15.13}$$

In the next section, we need the following averages over the cycle of 2π:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \sin(mx) \cos(nx) \, \mathrm{d}x = 0 \tag{15.14}$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \sin(mx) \sin(nx) \, \mathrm{d}x = \begin{cases} 0 & m \neq n \\ \frac{1}{2} & m = n \neq 0 \\ 0 & m = n = 0 \end{cases} \tag{15.15}$$

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(mx) \cos(nx) \, \mathrm{d}x = \begin{cases} 0 & m \neq n \\ \frac{1}{2} & m = n \neq 0 \, , \\ 1 & m = n = 0 \end{cases} \tag{15.16}$$

with some general integers m and n. Let's show Eq. (15.14) is true as an illustration. Yet again complex algebra will prove us useful. We observe that for any integer k, $e^{ik\pi} = \cos(k\pi) = \pm 1$ because $\sin(k\pi) = 0$. Therefore, we know the following simple integral

$$\int_{-\pi}^{\pi} e^{ikx} \mathrm{d}x = \left[\frac{e^{ikx}}{ik} \right]_{-\pi}^{\pi} = 0. \tag{15.17}$$

We can now rewrite Eq. (15.14) as complex numbers using Eqs. (14.21) and (14.20):

$$\begin{aligned} \int_{-\pi}^{\pi} \sin(mx) \cos(nx) \mathrm{d}x &= \int_{-\pi}^{\pi} \left(\frac{e^{imx} - e^{-imx}}{2i} \right) \left(\frac{e^{inx} + e^{-inx}}{2} \right) \mathrm{d}x \\ &= \frac{1}{4i} \int_{-\pi}^{\pi} \left(e^{i(m+n)x} - e^{-i(n-m)x} - e^{i(m-n)x} + e^{-i(m+n)x} \right) \mathrm{d}x \end{aligned}$$

Each of these terms are like Eq. (15.17) because $m \pm n$ are still integers! Therefore, we have proven that Eq. (15.14) is indeed zero for any integer m and n. Equations (15.15) and (15.16) can be proven in the same way.

15.3 Fourier analysis

We have seen in Section 13.7.2 that when we pluck a string, we get a superposition of standing waves, which are a sum of its harmonics with different amplitudes. These are represented by terms like $a_n \sin(n\omega_0 t)$ and $b_n \cos(n\omega_0 t)$ for some amplitudes a_n and b_n and with the fundamental harmonic ω_0. In this section, we will derive these amplitudes for a given signal that is periodic. This is called *Fourier analysis*.

15.3.1 Period 2π

Take a function that repeats every 2π, meaning

$$f(x) = f(x + 2\pi) = f(x + 2n\pi) \tag{15.18}$$

for any integer n. Fourier analysis poses this 2π-periodic function can be written as

$$\begin{aligned} f(x) = \frac{1}{2}a_0 + a_1 \cos x + a_2 \cos 2x + a_3 \cos 3x + ... \\ + b_1 \sin x + b_2 \sin 2x + b_3 \sin 3x + ..., \end{aligned} \tag{15.19}$$

with real coefficients a_n and b_n. We will show that this is possible by deriving a_n and b_n. First, note several things:

1. The right-hand side of Eq. (15.19) also has a period 2π.
2. Each new sine or cosine term has a higher frequency $f_n = n/2\pi$ (or shorter period $T_n = 2\pi/n$) than, which is an integer multiple of the "fundamental" frequency $1/2\pi$.
3. Each coefficient a_n and b_n acts as an amplitude for cosine and sine with period $2n\pi$, respectively.
4. For $n = 0$, the cosine term becomes a simple constant $a_0/2$, while it vanishes for the sine, $b_0 = 0$. The factor $1/2$ in $a_0/2$ is arbitrary, but will be convenient later.

How do we get a_n and b_n for a some periodic function $f(x)$? Start with a_0 by taking the average over one period:

$$\begin{aligned} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\mathrm{d}x = \quad & \frac{a_0}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} 1\,\mathrm{d}x \\ & + a_1\frac{1}{2\pi}\int_{-\pi}^{\pi} \cos x\,\mathrm{d}x + b_1\frac{1}{2\pi}\int_{-\pi}^{\pi} \sin x\,\mathrm{d}x \\ & + a_2\frac{1}{2\pi}\int_{-\pi}^{\pi} \cos 2x\,\mathrm{d}x + b_2\frac{1}{2\pi}\int_{-\pi}^{\pi} \sin 2x\,\mathrm{d}x \\ & + ... \end{aligned} \tag{15.20}$$

The average of sine and cosine over one period is always 0, so only the first term survives. For any 2π-periodic function $f(x)$ therefore,

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\mathrm{d}x. \tag{15.21}$$

This term with a_0 serves as a constant, vertical offset corresponding to the average of f.

For a_1, we use a clever trick: Multiply both sides by $\cos x$, and then take the average

$$\begin{aligned} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\cos x\,\mathrm{d}x = \quad & \frac{a_0}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} \cos x\,\mathrm{d}x \\ & + a_1\frac{1}{2\pi}\int_{-\pi}^{\pi} \cos^2 x\,\mathrm{d}x + b_1\frac{1}{2\pi}\int_{-\pi}^{\pi} \sin x \cos x\,\mathrm{d}x \\ & + a_2\frac{1}{2\pi}\int_{-\pi}^{\pi} \cos(2x)\cos x\,\mathrm{d}x + b_2\frac{1}{2\pi}\int_{-\pi}^{\pi} \sin(2x)\cos x\,\mathrm{d}x \\ & + ... \end{aligned} \tag{15.22}$$

Using Eqs. (15.14) and (15.16), this time, all terms vanish except for the one with a_1. We are only left with

$$a_1 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos x\,\mathrm{d}x. \tag{15.23}$$

Similarly for a_2, we multiply both sides by $\cos 2x$, and take the average. Only the term with a_2 survives. We notice a pattern. In general, we get the formula for a_n by multiplying both sides by $\cos nx$, and then taking the average:

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\,\mathrm{d}x = \;& \frac{a_0}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos(nx)\,\mathrm{d}x \\ & +a_1\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos x\cos(nx)\,\mathrm{d}x + b_1\frac{1}{2\pi}\int_{-\pi}^{\pi}\sin x\cos(nx)\,\mathrm{d}x \\ & +a_2\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos(2x)\cos(nx)\,\mathrm{d}x + b_2\frac{1}{2\pi}\int_{-\pi}^{\pi}\sin(2x)\cos(nx)\,\mathrm{d}x \\ & +...\end{aligned}$$

The only non-zero terms are when

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos^2(nx)\,\mathrm{d}x = \frac{1}{2}. \tag{15.24}$$

Therefore, in general

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\,\mathrm{d}x. \tag{15.25}$$

By multiplying $f(x)$ with $\cos nx$ and averaging, we "project out" this component in $f(x)$, while the integral of all other components vanish. The amplitude a_n is a measure of "how much $\cos nx$ there is in $f(x)$". Analogously for the b_n terms, one multiplies by $\sin nx$ and averages to select this component:

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\,\mathrm{d}x. \tag{15.26}$$

In summary:

Fourier series expansion for 2π-periodic function. *Any 2π-periodic function f can be written as*

$$\begin{aligned} f(x) &= \frac{1}{2}a_0 + a_1\cos x + a_2\cos 2x + a_3\cos 3x + ... \\ &\quad + b_1\sin x + b_2\sin 2x + b_3\sin 3x + ... \\ &= \sum_{n\in\mathbb{N}} a_n\cos nx + \sum_{n\in\mathbb{N}} b_n\sin nx. \end{aligned} \tag{15.27}$$

where

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(nx)\,\mathrm{d}x \tag{15.28}$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin(nx)\,\mathrm{d}x. \tag{15.29}$$

Notice that the 1/2 factor in front of a_0 was good foresight, because it is consistent with the general formula for a_n.

Now we know how to decompose a given function with period $T = 2\pi$ into a discrete sum of sines and cosines by computing the coefficients, or amplitudes, a_n and b_n.

15.3.2 Any period T

Let us now make the previous a bit more general. Say you have a (time-dependent) function with period T,

$$f(t) = f(t + nT) \tag{15.30}$$

for any integer n. We need to transform our sines and cosines so they have the same period;

$$\cos nt \to \cos\frac{2\pi nt}{T} = \cos n\omega t \tag{15.31}$$

$$\sin nt \to \sin\frac{2\pi nt}{T} = \sin n\omega t, \tag{15.32}$$

with $\omega = 2\pi/T$ as usual. Some textbooks write $\omega_n = n\omega$. So everywhere in the previous sections, we need to perform the substitution $x \to 2\pi t/T = \omega t$, which is simple rescaling the horizontal axis. The differential in each integral becomes

$$\mathrm{d}x \to \frac{2\pi}{T}\mathrm{d}t = \omega\mathrm{d}t. \tag{15.33}$$

This leads to

Fourier series expansion. *A function f with period $T = 2\pi/\omega$ can be expanded as*

$$f(x) = \sum_{n\in\mathbb{N}} a_n \cos(n\omega t) + \sum_{n\in\mathbb{N}} b_n \sin(n\omega t). \tag{15.34}$$

with real coefficients a_n and b_n, or with complex coefficients c_n, where

$$a_n = \frac{2}{T}\int_0^T f(x)\cos(n\omega t)\,\mathrm{d}t \tag{15.35}$$

$$b_n = \frac{2}{T}\int_0^T f(x)\sin(n\omega t)\,\mathrm{d}t \tag{15.36}$$

When $T = 2\pi$, and thus $\omega = 1$, we retrieve the result in the previous section.

In the derivation of previous section, the integral ranged from $-\pi$ to π, while in this section, we went from 0 to T. The limits of the integral of a periodic function is arbitrary as long as the range is over the same period T. This is easily proven by substitution. Say you want to shift the limits by some constant $t_0 > 0$; split the integral as

$$\int_0^T f(t)\,\mathrm{d}t = \int_0^{t_0} f(t)\,\mathrm{d}t + \int_{t_0}^T f(t)\,\mathrm{d}t. \tag{15.37}$$

Use substitution $t \to t' = t + T$ in the first integral on the right-hand side, and the fact f has a period T, and you will find

$$\int_0^T f(t)\mathrm{d}t = \int_{t_0}^{t_0+T} f(t)\mathrm{d}t. \tag{15.38}$$

The proof for $t_0 < 0$ is similar. Therefore, the limits do not matter, as long as the range is T.

Fourier series are very useful and have many applications. It allows us to approximate any random periodic signal. In some sense it is like Taylor expansion, but instead of a polynomial

$$T_n(x) = \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i. \tag{15.39}$$

(linear sum of terms with increasing order x^i) approximating a function a some point $x = a$, we have a linear sum of sines and cosine of increasing frequency $n\omega t$,

$$s_n(x) = \sum_{i=0}^{n} a_i \cos(n\omega t) + \sum_{i=0}^{n} b_i \sin(n\omega t), \tag{15.40}$$

approximating a periodic function over its full range. Each new "higher order" term improves the approximation. Rebuilding a function with sines and cosines in this way is called *Fourier synthesis*. Each periodic function has its own unique set of coefficients a_n and b_n that is called its *spectrum*. Conversely, under the right mathematical conditions, a Fourier spectrum uniquely defines a periodic function. It contains all the information of the original function.

Let's make Fourier series a bit more concrete with some examples.

15.3.3 Example 1: Cosine

First a trivial example. The spectrum of a cosine

$$f(t) = A\cos\omega t \tag{15.41}$$

is simply one component with the same period,

$$b_n = 0, \quad a_n = \begin{cases} A & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases}. \tag{15.42}$$

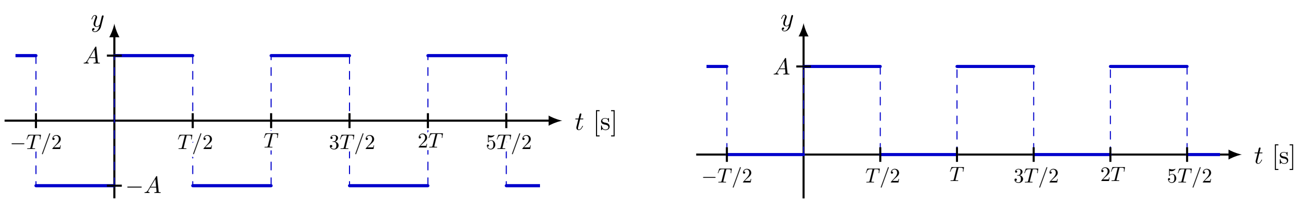
15.3.4 Example 2: Square wave

A classic example is the so-called square wave shown in Fig. 15.3. One form is

$$f(x) = \begin{cases} +1 & \text{for } 2\pi n < x \leqslant \pi(n+1),\ n = 0, \pm 2, \pm 4, \pm 6, ... \\ -1 & \text{for } \pi n < x \leqslant 2\pi(n+1),\ n = \pm 1, \pm 3, \pm 5, \pm 7, ... \end{cases} \tag{15.43}$$

Note that f has a period $T = 2\pi$, amplitude $A = 1$, and is an odd function, i.e. $f(-x) = -f(x)$, just like a sine. We therefore expect that all even cosine components will disappear. Indeed, breaking up the integral into its two pieces over $[0, 2\pi]$,

$$a_n = \frac{1}{\pi}\int_{-\pi}^{0} (-1)\cos(nx)\,\mathrm{d}x + \frac{1}{\pi}\int_{0}^{\pi} (+1)\cos(nx)\,\mathrm{d}x. = 0$$



(a) Odd square wave between $-A$ and A. **(b)** Square wave between 0 and A.

Figure 15.3: Square waves are simple periodic functions with period T and amplitude A.

Similarly,

$$b_n = \frac{1}{\pi}\int_0^{\pi}(-1)\sin(nx)\,\mathrm{d}x + \frac{1}{\pi}\int_{\pi}^{0}(+1)\sin(nx)\,\mathrm{d}x = \begin{cases} 0 & \text{for even } n = 0, 2, 4, 6, \dots \\ \dfrac{4}{n\pi} & \text{for odd } n = 1, 3, 5, 7, \dots \end{cases} \tag{15.44}$$

So we find

$$f(x) = \frac{4}{\pi}\left(\sin x + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \dots\right). \tag{15.45}$$

Since f is an odd function, it is not surprising that no cosine terms are left, but only sines.

Notice that each new term b_n with a higher frequency $n/2\pi$ in this series gets smaller as shown in Fig. 15.4b, which means that with each term we get a better approximation to f, as shown in Fig. 15.4a. One can mathematically show that this series (i.e. infinite sum) converges, and so the result is finite like f.

15.3.5 Example 3: Square wave with amplitude A and period T

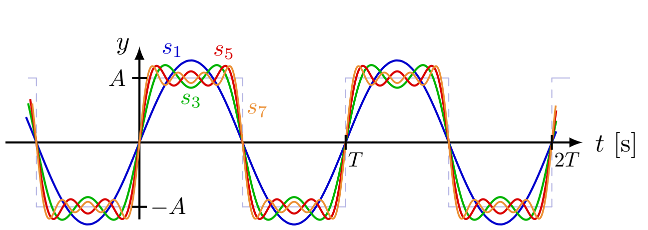
More generally, a square wave with general amplitude A and period T can have the form

$$f(t) = \begin{cases} A & \text{for } nT < t \leqslant (n+1)\dfrac{T}{2},\ n = 0, \pm 2, \pm 4, \pm 6, \dots \\ -A & \text{for } n\dfrac{T}{2} < t \leqslant (n+1)T,\ n = \pm 1, \pm 3, \pm 5, \pm 7, \dots \end{cases}$$

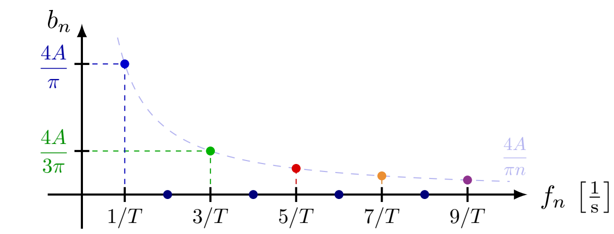
It is easy to show that this yields

$$f(t) = \frac{4A}{\pi}\left(\sin \omega t + \frac{\sin 3\omega t}{3} + \frac{\sin 5\omega t}{5} + \dots\right).$$

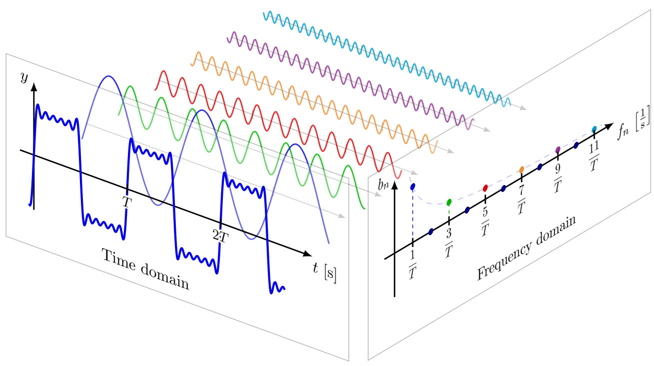
with $\omega = 2\pi/T$.



(a) Comparing Fourier series s_n of different order n. The higher the order, the better the approximation.



(b) Spectrum of the square wave showing the amplitudes as a function of frequencies $f_n = nf = n/T$.



(c) Illustration of Fourier synthesis: Square waves can be approximated by adding up sine waves of frequencies $f_n = nf = n/T$ (i.e. harmonics) with respective amplitudes A_n.

Figure 15.4: Fourier analysis of square wave with period T in Fig. 15.3a.

15.3.6 Example 4: Square wave with offset

A small variation on the previous square wave is one with an offset, shown in Fig. 15.3b:

$$f(x) = \begin{cases} 1 & \text{for } 2\pi n < x \leqslant \pi(n+1),\ n = 0, \pm 2, \pm 4, \pm 6, ... \\ 0 & \text{for } \pi n < x \leqslant 2\pi(n+1),\ n = \pm 1, \pm 3, \pm 5, \pm 7, ... \end{cases}$$

It is a repeated signal of 0's and 1's, and is used a lot in electronics. It can for example be used to keep time as it defines fixed intervals.

This time, a_0 is not zero, but all other a_n's are:

$$f(x) = \frac{1}{2} + \frac{2}{\pi}\left(\sin x + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + ...\right).$$

As mentioned before, the first constant term serves to vertically shift the sines up.

15.3.7 Example 5: Sawtooth wave

Another common example is the sawtooth wave shown in Fig. 15.5a. It is also a piece-wise, periodic function. One way to define it with an amplitude A and period T is as

$$f(t) = \frac{2A}{T}(t - nT) \quad \text{for } (2n-1)\frac{T}{2} < t \leqslant (2n+1)\frac{T}{2},$$

where n can be any integer. This as well, is an odd function, so $a_n = 0$ for each n. The odd components b_n are most easily computed by integrals from $-T$ to T:

$$\begin{aligned} b_n &= \frac{4A}{T}\int_{-T}^{T} t\sin(n\omega t)\,\mathrm{d}t \\ &= \frac{4A}{T^2}\left[\frac{\sin n\omega t}{(n\omega)^2} - \frac{t\cos n\omega t}{n\omega}\right]_{-T/2}^{T/2} = \begin{cases} +\dfrac{2A}{n\pi} & \text{for even } n = 0, 2, 4, 6, ... \\ -\dfrac{2A}{n\pi} & \text{for odd } n = 1, 3, 5, 7, ... \end{cases}. \end{aligned}$$

15.3.8 Example 6: Triangle wave

Finally, the triangle wave in Fig. 15.6a is another example of an odd function.

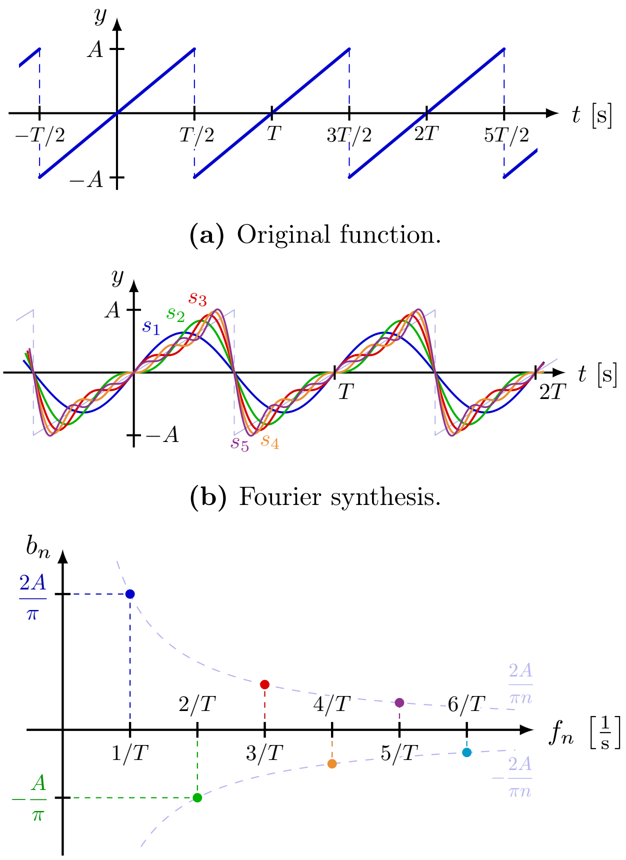
$$f(t) = \begin{cases} \dfrac{4A}{T}\left(t - n\dfrac{T}{2}\right) & \text{for } (2n-1)\dfrac{T}{4} < t \leqslant (2n+1)\dfrac{T}{4},\ n = 0, \pm 2, \pm 4, ... \\ \dfrac{4A}{T}\left(n\dfrac{T}{2} - 1\right) & \text{for } (2n-1)\dfrac{T}{4} < t \leqslant (2n+1)\dfrac{T}{4},\ n = \pm 1, \pm 3, \pm 5, ... \end{cases}.$$

The only components are given by

$$b_n = \frac{8A}{T}\int_{-T/4}^{T/4} t\sin(n\omega t)\,\mathrm{d}t + \frac{8A}{T}\int_{-T/4}^{3T/4}\left(\frac{T}{2} - t\right)\sin(n\omega t)\,\mathrm{d}t.$$

After a bit of algebra, left as an exercise,

$$b_n = \begin{cases} 0 & \text{for even } n = 0, 2, 4, 6, ... \\ (-1)^{\frac{n-1}{2}}\dfrac{8A}{n^2\pi^2} & \text{for odd } n = 1, 3, 5, 7, ... \end{cases}.$$

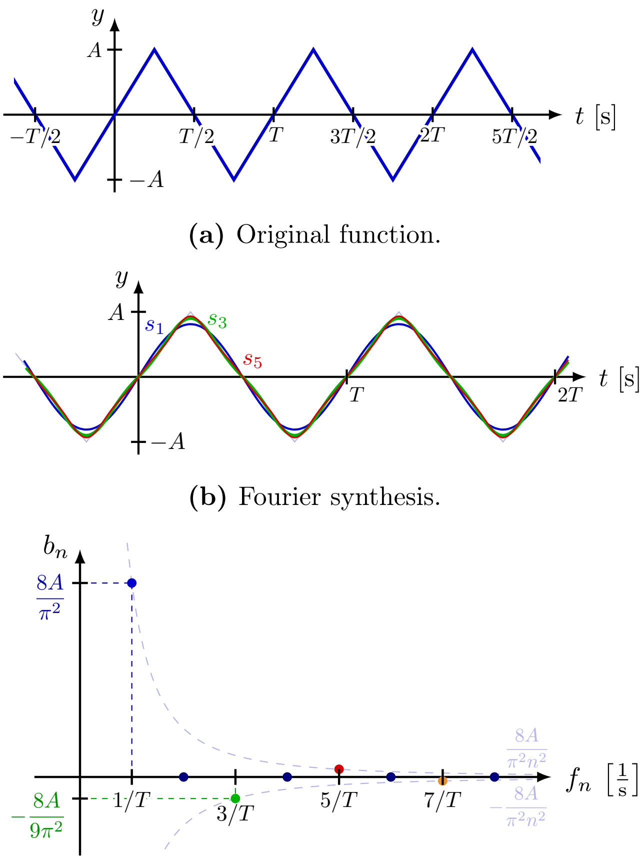


(a) Original function.

(b) Fourier synthesis.

(c) Spectrum.

Figure 15.5: Sawtooth wave.



(a) Original function.

(b) Fourier synthesis.

(c) Spectrum.

Figure 15.6: Triangle wave.

15.3.9 Extra: Amplitude-phase form

In Eq. (15.34), the expansion was in terms of sines and cosines, but this is not the only form one can use. As discussed in Sections 12.2.2 and 14.3.3, the superposition of a sine and cosine, which have the same frequency and phase, but different amplitude, can be rewritten as a single cosine (or sine), but with a new phase,

$$a_n \cos(n\omega t) + b_n \in (n\omega t) = A_n \cos\left(n\omega t - \phi_n\right).$$

where a_n and b_n are defined as before. So a different formulation of Fourier series becomes

Amplitude-phase form of Fourier series expansion. *Any periodic function f can be written as*

$$f(x) = \sum_{n\in\mathbb{N}} A_n \cos\left(n\omega t - \phi_n\right), \tag{15.46}$$

with coefficients

$$\begin{aligned} A_n &= \sqrt{a_n + b_n} \\ \tan\phi_n &= \frac{b_n}{a_n}. \end{aligned}$$

So now we have one frequency spectrum A_n and one phase spectrum ϕ_n.

15.3.10 Extra: Complex form

Yet another form is using Eqs. (14.21) and (14.20) again to write

$$\sin n\omega x = \frac{e^{in\omega x} - e^{-in\omega x}}{2i} \tag{15.47}$$

$$\cos n\omega x = \frac{e^{in\omega x} + e^{-in\omega x}}{2} \tag{15.48}$$

If one substitutes this into Eq. (15.34), it is easy to rearrange things into a more simplified form;

$$a_n \cos(n\omega t) + b_n \cos(n\omega t) = \frac{a_n - ib_n}{2} e^{in\omega t} + \frac{a_n + ib_n}{2} e^{-in\omega t}.$$

Define a new coefficient with $e^{in\omega t}$.

$$c_n = \frac{a_n - ib_n}{2} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\big(\cos(n\omega t) - i\sin(n\omega t)\big)\mathrm{d}t. \tag{15.49}$$

We see that the coefficient with $e^{-in\omega t}$ is actually the complex conjugate, such that it can conveniently be written as

$$c_{-n} = c_n^* = \frac{a_n + ib_n}{2}. \tag{15.50}$$

Therefore, f becomes

$$\begin{aligned} f(t) = c_0 &+ c_1 e^{i\omega t} + c_2 e^{i2\omega t} + c_3 e^{i3\omega t} + \ldots \\ &+ c_{-1} e^{-i\omega t} + c_{-2} e^{-i2\omega t} + c_{-3} e^{-i3\omega t} + \ldots \end{aligned} \tag{15.51}$$

In summary,

Complex form of Fourier series expansion. *Any periodic function f can be written as*

$$f(t) = \sum_{n\in\mathbb{Z}} c_n e^{int} \tag{15.52}$$

with complex coefficients

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int} \mathrm{d}t. \tag{15.53}$$

It turns out that this complex form is a very useful form of Fourier expansion. More generally, the coefficients c_n are complex numbers as well, but if f is a real function, as assumed up until now, one can simply impose $c_n = c_{-n}^*$, such that all the imaginary terms cancel, because for any sum of the form

$$\begin{aligned} c_n e^{i\theta} + c_{-n} e^{-i\theta} &= (c_n + c_n^*)\cos\theta + (c_n - c_n^*) i \sin\theta \\ &= 2\,\mathrm{Re}[c_n]\cos\theta - 2\,\mathrm{Im}[c_n]\sin\theta, \end{aligned}$$

The last line is manifestly real. We recognize the coefficients $a_n = 2\,\mathrm{Re}[c_n]$ belonging to the odd cosines, and $b_n = 2\,\mathrm{Im}[c_n]$ to the even sines in our real Fourier expansion, consistent with the derivation at the beginning of this section. Therefore, all the information is encoded in one single complex number c_n. Note that c_0 serves as a vertical offset.

15.3.11 Example 6: Square wave (revisited)

Let's see this in work with our previous square wave example with $A = 1$ and $T = 2\pi$. If $n \neq 0$,

$$\begin{aligned} c_n &= -\frac{1}{2\pi}\int_{-\pi}^{0} e^{-inx}\mathrm{d}x + \frac{1}{2\pi}\int_{0}^{\pi} e^{-inx}\mathrm{d}x \\ &= \begin{cases} \dfrac{2}{in\pi} & \text{for odd } n \\ 0 & \text{for even } n \end{cases} \end{aligned}$$

The coefficient c_n is purely imaginary, and thus $c_n = ib_n$, and our result is consistent with Eq. (15.44). The square wave can there for be expanded as

$$\begin{aligned} f(x) &= \frac{2}{2} + \frac{1}{\pi}\left[e^{ix} + \frac{e^{i3x}}{3} + \frac{e^{i5x}}{5} + ...\right] - \frac{2}{\pi}\left[e^{-ix} + \frac{e^{-i3x}}{3} + \frac{e^{-i5x}}{5} + ...\right] \\ &= \frac{1}{2} + \frac{2}{\pi}\left(\sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + ...\right). \end{aligned}$$

Again, exactly the same result as Eq. (15.45).

15.3.12 Application: Spectra of music instruments

As mentioned in Section 13.6.2, if one plays the same note on different music instruments, we still notice that it somehow sounds different. This is because each instrument, and the way it is played, has its own unique spectrum, which gives its own "feeling", called the *tone color*, or *timbre*. Using oscilloscopes, one can plot the spectrum of any sounding note and analyze which frequencies with which amplitudes are present. For a given note, you will typically not only find the fundamental frequency $f = f_1$, but also some set of the other excited harmonics $f_n = nf$ (the overtones), and often even a small contributions from other frequencies.

15.4 Fourier transforms

So far, we have expanded periodic functions into a sum of terms with different frequencies. But what if a shape does not repeat? We cannot use Fourier series anymore. This can be done with *Fourier transforms*. The derivation is out of the scope of this course, but the result is

Fourier transform. *A function f can be expanded as*

$$f(t) = \int_{-\infty}^{\infty} g(\omega)e^{i\omega t}\,\mathrm{d}\omega\,, \tag{15.54}$$

where the Fourier transform is a continuous function of a real variable ω,

$$g(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x)e^{-i\omega x}\mathrm{d}t. \tag{15.55}$$

Here, the $g(\omega)$ is the analogue of the coefficient c_n in the Fourier series expansion. The integral over the real variable ω is the analogue of the sum over discrete integers n. We usually say that g is the *Fourier transform* of f. But we see that the integrals are of exactly the same form, except for the minus sign in the exponent, so people often say f and g are Fourier transforms of each other.

In case of a signal $f(t)$ changing with time, the Fourier transform provides you with the (continuous) amplitude of the frequencies ω. This is called the *frequency* or *power spectrum*. When we analyze the function f versus time t, this is called the *time domain*, whereas we study g versus ω, it is in the *frequency domain*.

Some books use the notation $g = \hat{f}$. The position of the $1/2\pi$ factor in Eq. (15.55) is not important. Other books use different conventions where g has no $1/2\pi$ factor in front, but f in Eq. (15.54) does. Even other authors use a factor $1/\sqrt{2\pi}$ in front of both equations for f and g, which makes the quation more symmetric.

The Fourier transform works for every infinite interval if the following conditions on f are met

1. f is *absolutely integrable*; $\int_{-\infty}^{\infty} |f(t)|\mathrm{d}t$ is finite. This often means that at infinity, the function dampens, $\lim_{t\to\pm\infty} f(t) = 0$.

2. f has a finite number of maxima, minima and discontinuities.

3. f is a single-valued function.

15.4.1 Even/odd properties

Notice that the Fourier transform can be rewritten as

$$g(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t)(\cos\omega t - i\sin\omega t)\mathrm{d}t. \tag{15.56}$$

One can see the interesting property that if f is even, this integral reduces to

$$g(\omega) = \frac{1}{\pi}\int_{0}^{\infty} f(t)\cos(\omega t)\mathrm{d}t, \tag{15.57}$$

because cosine is also an even function, while sine is odd. From this last equation, it is obvious that if f is even, so is g. In the same way, if f is odd, so is g, because

$$g(\omega) = \frac{i}{\pi}\int_{0}^{\infty} f(t)\sin(\omega t)\mathrm{d}t, \tag{15.58}$$

15.4.2 Example: Rectangular pulse

As an example, take the rectangular pulse function shown in Fig. 15.7a

$$f(t) = \begin{cases} A & \text{for } -T < t \leqslant T \\ 0 & \text{otherwise} \end{cases} \tag{15.59}$$

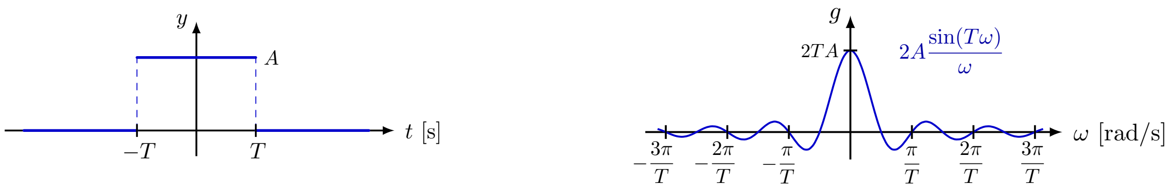
This function does not repeat, so we can use the Fourier transform, which is only non-zero between $-T$ and T:

$$g(\omega) = \int_{-T}^{T} Ae^{i\omega t} = 2A\frac{\sin\omega T}{\omega}. \tag{15.60}$$

This is the so-called *sin cardinalis (sinc) function*. Let's draw $g(\omega)$. Remember from calculus that

$$\lim_{x\to 0}\frac{\sin x}{x} = 1, \tag{15.61}$$

such that $g(\omega) = 2TA$. When is $g(\omega) = 0$? Simply when $\sin(\omega T) = 0$, so when $\omega = \pi/T, 2\pi/T, 3\pi/T, ...$ Putting this all together, we get Fig. 15.7b.



(a) Function f in time domain. **(b)** Fourier transform of f in frequency domain.

Figure 15.7: Fourier analysis of rectangular pulse with width T and amplitude A in Eq. (15.59).

15.4.3 Dirac delta function

Consider the step function, sometimes called the *Heaviside step function* shown in Fig. 15.8a,

$$f(x) = \begin{cases} 1 & \text{for } x > 0 \\ 1/2 & \text{for } x = 0 \\ 0 & \text{for } x < 0 \end{cases} \tag{15.62}$$

This equation is not continuous and it is not differentiable in $x = 0$. Although it is not well defined, we can intuitively say that the change that happens in $x = 0$ is infinite, while zero everywhere else. We therefore introduce the Dirac delta function:

Dirac delta function.

$$\delta(x) = \begin{cases} \infty & \textit{for } x = 0 \\ 0 & \textit{otherwise} \end{cases} \tag{15.63}$$

which has a unit integral (i.e. normalized*),*

$$\int_{-\infty}^{\infty} \delta(x)\mathrm{d}x = 1. \tag{15.64}$$

Because the Dirac delta function has an infinity, it is not well-defined in real calculus, but it is still a useful and convenient tool for physicists and has many applications. More rigorous definitions exist, using so-called *Schwartz distributions* or *measures* in mathematical analysis, but this falls outside the scope of this course.

The delta function is often depicted as a single spike, or with an arrow pointing to 1 as in Fig. 15.8b. Some books define it intuitively by taking the limit of some narrowing function that has a fixed integral of 1, like those illustrated in Fig. 15.9.

The delta function can be shifted by a real constant $a \in \mathbb{R}$,

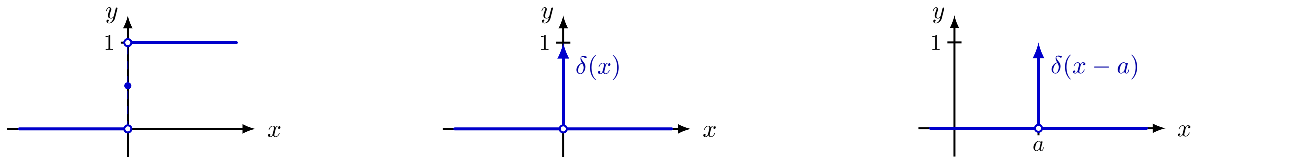
$$\delta(x-a) = \begin{cases} \infty & \text{for } x = a \\ 0 & \text{otherwise} \end{cases} \tag{15.65}$$

still with an integral of one;

$$\int_{-\infty}^{\infty} \delta(x-a)\mathrm{d}x = 1. \tag{15.66}$$

The most useful property of this function is that it can "project" out a single value of f when combined in an integral:

$$\int_{-\infty}^{\infty} f(x)\delta(x-a)\mathrm{d}x = f(a). \tag{15.67}$$



(a) Heaviside step function. **(b)** Dirac delta function $\delta(x)$. **(c)** Delta function $\delta(x-a)$ shifted by a.

Figure 15.8: The Heaviside step function and Dirac delta function.

(a) Rectangular function with width $\epsilon \to 0$. **(b)** Normal distribution with standard deviation $\sigma \to 0$.

Figure 15.9: Sometimes the Dirac delta function is defined as the limit of some narrow spike function with fixed integral 1.

What has this to do with Fourier transformations? Well, suppose there is a function f with the Dirac delta function as Fourier transform,

$$g(\omega) = \delta(\omega - a). \tag{15.68}$$

The corresponding time function then must be

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \delta(\omega - a) e^{-i\omega t}\,\mathrm{d}\omega \tag{15.69}$$

$$= \frac{1}{2\pi} e^{-iat}. \tag{15.70}$$

If we are talking about real, physical observable, we only need the real part (see Section 14.3.5),

$$\mathrm{Re}[f(t)] = \frac{1}{2\pi}\cos at. \tag{15.71}$$

Therefore, the spectrum of time-dependent cosine with frequency a is a single spike at a, similar to what the trivial example 15.3.3 of a Fourier series.

15.5 Summary

Table 15.1 gives a summary of the different methods of expanding and representing certain functions that we have seen in this course. A Taylor expansion is often used to approximate functions around some point with a polynomial. Fourier series are used to decompose a periodic function into a discrete sum sine and cosines, and can also be used as an approximation with limited terms. A Fourier transformation is the continuous version of a Fourier expansion, and is used to represent an aperiodic function in frequency space.

Table 15.1: Comparison of different ways of expanding function.

Method	Function conditions	Formula	Result
Taylor expansion	Differentiable in a	$f(x) = \sum_{n\in\mathbb{N}} \frac{f^{(n)}(x)}{n!}(x-a)^n$	Polynomial in a
Fourier expansion	Periodic (Piecewise) continuous	$f(x) = \sum_{n\in\mathbb{N}} a_n \cos(n\omega t) + \sum_{n\in\mathbb{N}} b_n \sin(n\omega t)$	Discrete sum of sines and cosines
Fourier transform	Aperiodic Absolutely integrable (Piecewise) continuous	$f(x) = \int_{-\infty}^{\infty} g(\omega) e^{-i\omega t}\,\mathrm{d}\omega$	Continuous integral Aperiodic

Part III

Fluid dynamics

Chapter 16

Hydrostatics & Pressure

This part of the course will study the basics of fluid dynamics. A fluid is a substance that has no fixed shape. They can be liquid or gas that occupy some volume and have some pressure. Ideal liquids are incompressible, while gasses are compressible; gasses tend to occupy the volume of their container, while liquids do not. Gases are a bit more complicated as their temperature can also vary with pressure and volume, as we will see next semester during the lectures on thermodynamics in PHY121. For now we will focus more on basic properties such as pressure, density and velocity. In this chapter we will first look at the basic concepts of hydrostatics, such as pressure.

What is pressure? Pressure P is when a total force F is applied to some surface with are A,

Pressure.

$$P = \frac{F}{A}. \tag{16.1}$$

It carries units of Pascal named after Blaise Pascal (1623–1662),

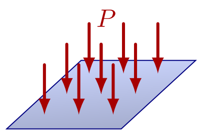
$$1\,\mathrm{Pa} = 1\frac{\mathrm{N}}{\mathrm{m}^2}. \tag{16.2}$$

16.1 Atmospheric pressure

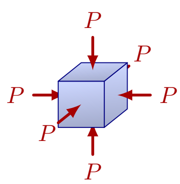
Otto von Guericke demonstrated the existence of atmospheric pressure in 1654 in Magdeburg. He sealed two large copper hemispheres and pumped out the air to create a vaccuum inside. Even two teams of horses could not pull apart the hemispheres, demonstrating the immense size of the atmospheric pressure.

The atmospheric pressure at sea level is about

$$P_{\mathrm{atm}} = 101.325\,\mathrm{kPa} = 1013.25\,\mathrm{hPa} = 1\,\mathrm{atm}, \tag{16.3}$$



(a) Pressure $P = FA$ on a surface with area A.



(b) Pressure is the same in all directions.

Figure 16.1: Pressure.

Table 16.1: Units of pressure. (* Per definition.)

	Pa	bar	atm	torr	psi
1 Pa	1	1×10^{-5}	9.8697×10^{-6}	7.5006×10^{-3}	1.4504×10^{-4}
1 hPa	$1 \times 10^{2} = 100$*	0.001	9.8697×10^{-4}	0.750 06	0.014 504
1 kPa	$1 \times 10^{3} = 1000$*	0.01	9.8697×10^{-3}	7.5006	0.145 03
1 mbar	$1 \times 10^{2} = 100$*	0.001	9.8697×10^{-4}	0.750 06	0.014 504
1 bar	$1 \times 10^{5} = 100\,000$*	1	0.986 97	750.06	14.5038
1 atm	$1.013\,25 \times 10^{4} = 101\,325$*	1.0132	1	760*	14.6952
1 torr	133.32	1.3332×10^{-3}	1.3160×10^{-3}	1	0.019 336
1 psi	6894.76	0.068 948	0.068 049	51.715	1

where atm is one of the alternative units for pressure, see Table 16.1. P_{atm} is the weight (i.e. a force) of all the air above us that is spread out over some area.

How much weight does the atmosphere "feel" like on an area of $A = 1\,\text{cm}^2$? The total force would be $F = P_{\text{atm}}A \approx 10.1\,\text{N}$, which is like having a 1 kg mass weighing on every square centimeter. However, we do not notice this, because our bodies are internally pressurized: The pressure inside our body is the same as outside. We only experience it when we find ourselves in a lower-pressure (like on a high mountain top), or high-pressure environment (like when diving underwater), for which our lungs and bodies were not built. Deep sea fish, like the so-called "blob fish", live at depths where the pressure is 60 to 120 times higher than at sea level. At this high-pressure, they look like normal fish, but once you take them to the surface, their bodies, which have evolved for high-pressure environments, decompress and they become very ugly, memeable blobs.[1]

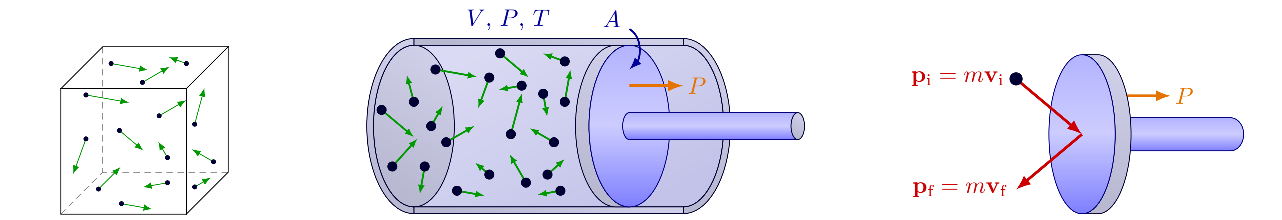
At any given height, a pressure pushes in all directions with the same value, as shown in Fig. 16.1b. Pressure is a scalar; it has a particular value in each point in space, but it does not have a particular direction. Pressure does exert a force orthogonal to a surface, but the important thing is the difference of pressure on either side of some surface. This is evident when you hold up the hand and feel no net force from atmospheric pressure pushing on the top, bottom, or sides. If you put your hand on the tube of a running vacuum cleaner, there will be a pressure difference, and the your hand will be "sucked" onto the tube's opening.

Another experiment you can try at home is holding a filled glass upside down with a coaster blocking off the opening. The coaster will stay put and block the water from running out. This happens because the atmosphere applies a pressure to the coaster, even if it is upside down. Say the beer glass has an opening of radius 4 cm, then the net force on the coaster from the atmosphere is about $F_{\text{atm}} \approx 510\,\text{N}$. Compare this to the weight of the water push down on the other side: $mg \approx 5\,\text{N}$ for half a liter of beer, which is about half a kilogram. Clearly, the atmospheric pressure is much larger. Later in this chapter we will see how pressure varies over height due to the added fluid weight.

16.2 Microscopic description

Why is there pressure? Where does it come from? Fluids are made out of small molecules that can freely move around. Each one constantly collides with others, but also with the

[1] In fact, something similar likely happens to an astronaut who is thrown into the vacuum of space without a suit: Beside suffocating, receiving horrible radiation, having their fluids boil and bubble form in their blood, they will swell up because all internal gas expands due to the lack of external pressure. Their body will not explode as in some movies, because the skin is elastic enough to withstand the increased internal pressure.



(a) A box of colliding gas particle has a pressure.

(b) A container with gas closed of with a piston can measure the pressure.

(c) Gas particles collides off piston causes an change in momentum.

Figure 16.2: Pressure is a macroscopic phenomena caused by collisions of gas particles between each other and on the wall at the microscopic level.

walls of its surrounding container. When it does, it has a change in momentum

$$\Delta \mathbf{p} = \mathbf{p}_\mathrm{f} - \mathbf{p}_\mathrm{f}, \tag{16.4}$$

which will be normal to the wall, as in Fig. 16.2c and this causes a small force

$$\mathbf{F}_i = \frac{\Delta \mathbf{p}_i}{\Delta t}. \tag{16.5}$$

There are actually many gas molecules constantly transferring small amounts of momentum to the wall, roughly of the order of Avogadro's number, N_A $O(10^23)$. The net force will be experienced as a constant pressure P. We will study this in more detail as part of the topic of statistical mechanics and thermodynamics in PHY121.

16.3 Bulk modulus

Even though we often assume solids and liquids are incompressible, in reality, they will be somewhat compressed by a large enough pressure P. As a measure of a material's compressibility, we define the bulk modulus

Bulk modulus.

$$B = \frac{-P}{\Delta V/V}, \tag{16.6}$$

where the total change in volume is ΔV. The minus sign makes B positive, as $\Delta V < 0$.

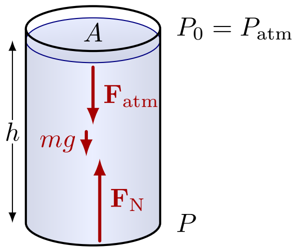
For gases the bulk modulus is very small and depends on temperature. For liquids and solids, which are harder to compress, the bulk modulus are larger. Some examples are listed in Table 16.2.

16.4 Pressure variation with depth

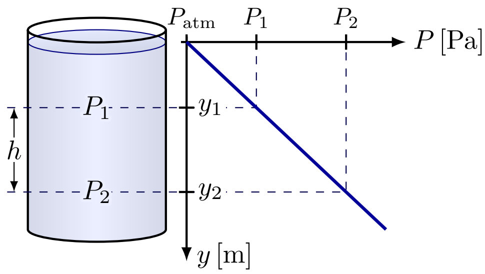
What is the pressure inside a liquid? Suppose you have a cylindrical container which is closed at the bottom, but open at the top as in Fig. 16.3a. First consider the total force

Table 16.2: Some examples of the bulk modulus B.

Material	B [Pa]
Iron	100
Lead	8
Water	2



(a) The force on the bottom of a container of fluid with height h, area A is $PA = P_0A + \rho Ahg$.



(b) Pressure increases linearly with depth y. The difference between y_1 and y_2 is $\Delta P = \rho g(y_2 - y_1)$.

Figure 16.3: The pressure in a fluid varies with depth due to weight.

on the bottom of the container. The fluid has a mass density ρ and a height h, and the cross-sectional area of the container is A, so the total weight is

$$F_g = (\rho V)g = \rho Ahg. \tag{16.7}$$

At the same time, the atmosphere pushes down on the top with a pressure $P_0 = P_{atm}$,

$$F_{atm} = P_0A. \tag{16.8}$$

At the bottom there is a normal force holding everything in place, so at equilibrium,

$$F_N = F_{atm} + F_g. \tag{16.9}$$

Call the total pressure on the bottom $P = F_N/A$, then

$$PA = P_0A + \rho Ahg, \tag{16.10}$$

and so the total pressure on the bottom is

Pressure variation with depth (at hydrostatic equilibrium).

$$P = P_0 + \rho gh. \tag{16.11}$$

Note that the cross-sectional area A cancels out, so the shape of the area really does not matter, it can be a circle, rectangle, heart, or whatever. It can even vary over the height, as the force is always "spread" out over the total area, $P = F/A$.

The argumentation and result above do not only hold at the bottom of the container, but also in every point in the fluid and on the walls at any depth. So the pressure increases linearly with depth, *but*, the pressure is the same in all points as the same depth and in all directions at equilibrium. So the pressure difference between any two depths, say y_1 and y_2 is

$$\Delta P = \rho g(y_2 - y_1) = \rho gh, \tag{16.12}$$

if $h = y_2 - y_1$ as in Fig. 16.3b.

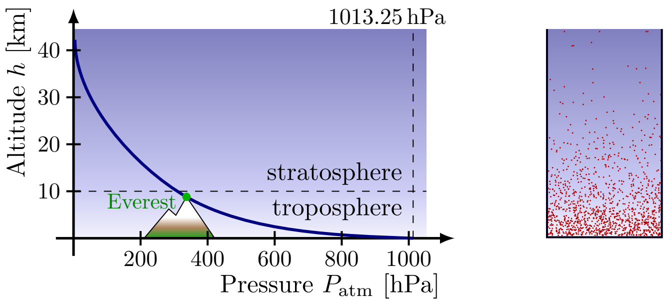


Figure 16.4: The atmospheric pressure varies strongly with altitude. At sea-level, it is typically about $P = 1013.25\,\mathrm{hPa}$. Around $h = 5.5\,\mathrm{km}$, the pressure is already halved.

16.4.1 Air pressure variation with altitude and weather

Air pressure in reality varies by altitude. The higher you go, the less air that weighs down on you is above you. It is more or less exponential as shown in Fig. 16.4 instead of linear, because the density and pressure of a gas also strongly depends on the temperature. The pressure at the top of Mount Everest at 8850 m is already one third of that at sea-level, making it extremely hard to breathe for most people. The difference in pressure is noticeable when you take a sealed bag of chips up a mountain, or by the feeling on your eardrums when you take off in an airplane (even though the cabin inside is partially pressurized).

16.4.2 Air pressure variation with weather

Local air pressure can vary wildly throughout the day from $P_{\mathrm{atm}} = 1013.35\,\mathrm{hPa}$ due to weather, with typically changes by less than 10 hPa. This is mostly caused by changes in temperature and creates so-called pressure systems that allow meteorologists to forecast the weather. Namely, air typically flows from high to low pressure regions. Humid air from low-pressure regions rise up, and can condense into rain clouds, while sinking air in low-pressure regions often brings dryer and nicer weather.

Beside variations due to weather, sea-level air pressure rises and falls up to 3 hP (closer to the equator) in 12 hour cycles due to atmospheric tides. The tides are mainly caused by thermal radiation by the sun, and to a lesser extend, the gravitational pull from the moon.

16.5 Measuring pressure

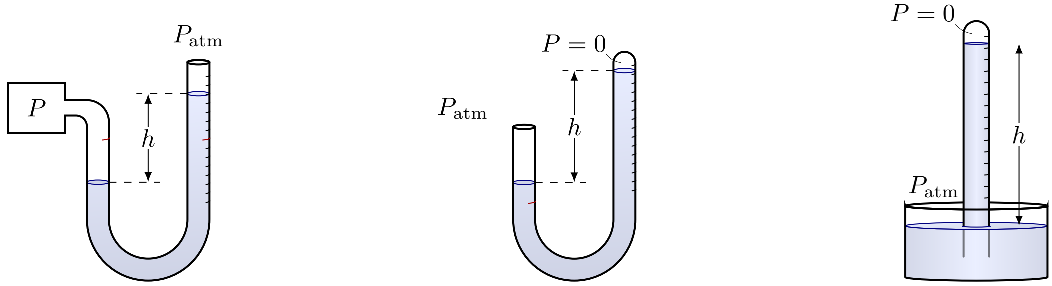
16.5.1 Manometer

A *manometer* is an instrument that can measure the pressure. Figure Fig. 16.5a shows how a U-shaped tube with one open end can measure the pressure P on the closed end by simply measuring the height difference of the fluid columns. If $P > P_{\mathrm{atm}}$, the column on the open end will be higher by a height h and vice versa if $P < P_{\mathrm{atm}}$. If $P < P_{\mathrm{atm}}$, then the pressure P on the close end must be

$$P = P_{\mathrm{atm}} + \rho g h, \tag{16.13}$$

where $\rho g h$ corresponds to the extra weight of the highest column. We measure the height, which can be expressed as

$$h = \frac{P - P_{\mathrm{atm}}}{\rho g}. \tag{16.14}$$



(a) A U-tube manometer can measure the pressure P in the closed end.

(b) A barometer measures the air pressure.

(c) Torricelli's experiment. The weight of the column creates a vacuum.

Figure 16.5: Pressure can be measured with the height of a fluid column.

16.5.2 Barometer

A *barometer* measures the air pressure. The open end is again exposed to atmospheric pressure, while the closed end is completely filled with a fluid as in Fig. 16.5b. The atmospheric pressure cannot compensate for the weight of the column if it is too high, causing the column to drop, leaving behind a vacuum where $P = 0$. The air pressure then can be measured by total height of the fluid's column:

$$P_{\text{atm}} = \rho g h. \tag{16.15}$$

16.5.3 Torricelli's experiment

This is in fact how Evangelista Torricelli (1608–1647) performed his famous experiment in 1643. He filled a tube closed on one end with mercury, which is more than 13 times denser than water, and put it upside down in a bowl as in Fig. 16.5c. This creates a vacuum on the closed end. He used a tube of one meter, such that the mercury column height was about

$$h = \frac{P_{\text{atm}}}{\rho g} \approx 0.76\,\text{m}. \tag{16.16}$$

This is why the unit of pressure named after him is defined as

$$1\,\text{torr} = \frac{1}{760}\text{atm}, \tag{16.17}$$

where 760 mm is the nicely rounded column height. A different, but closely related unit is the mmHg, which is almost exactly the same as one torr. This type of units are still used in medicine as well as in weather reporting and scuba diving.

Table 16.3: Some examples of volumetric mass densities ρ. Some metals are indicated with atom number Z.

Material	ρ [kg/m^3]
Gold ($Z = 79$)	19 320
Lead ($Z = 82$)	11 340
Mercury ($Z = 80$)	13 534
Iron ($Z = 26$)	7870
Cooking oil	910–930
Ice	916.7
Fresh water	1000
Salt water	1030
Wood	400–1000
Air (at sea-level)	1.2

In the 4th century B.C., Aristotle postulate that "nature abhors a vacuum", meaning that a vacuum cannot exist stably in Nature, as surrounding matter will always fill it. With his experiment, Torricelli proved otherwise. His work led to the first speculations of atmospheric pressure and invention of the barometer.

16.5.4 Drinking from a long straw

So what about water? The density of water is

$$\rho = 1000\,\mathrm{kg/m^3} = 1\,\mathrm{kg/L} \tag{16.18}$$

by the old definition of kilogram. How long does the tube have to be to create a vacuum? Well, the minimum length is given by the condition Eq. (16.15), so about $h = 10.3\,\mathrm{m}$.

When you drink cola through a straw, you create a lower pressure at the end in your mouth by sucking. The higher atmospheric pressure then "pushes" up the cola. But if you have straw longer than 10.3 meters, you cannot create a pressure lower than vacuum anymore, so you will not be able to drink!

Until Torricelli's work, people were very puzzled why their suction pumps could not pump water from wells higher than 10 meters.

16.6 Pascal's principle

The following is very useful.

Pascal's principle. *A change in pressure in any point of an enclosed, incompressible fluid is transmitted undiminished to every other point in the the fluid and the walls of the container.*

This is best understood with the example of the *hydraulic lift* shown in Fig. 16.6. A force F_1 pushes down on the left piston with are A_1. By doing so, it creates a change in pressure

$$\Delta P = \frac{F_1}{A_1}, \tag{16.19}$$

which must equal the change in pressure on the piston on the right, so

$$\Delta P = \frac{F_1}{A_1} = \frac{F_2}{A_2}, \tag{16.20}$$

where there is a force on the right piston,

$$F_2 = \frac{A_2}{A_1}F_1. \tag{16.21}$$

This has many applications. It allows us to lift heavy things such as cars or the arm of an excavator.

Say you want to lift a car of $M = 1500\,\mathrm{kg}$ from a circular hydraulic lift platform of radius $r_2 = 1\,\mathrm{m}$. If the smaller circular piston has a radius $r_1 = 0.1\,\mathrm{m}$, then what force do we need to hold the car in place? Using $F_2 = Mg$ and $A = \pi r^2$, we roughly find

$$F_1 = \frac{A_1}{A_2}F_2 = \frac{r_1^2}{r_2^2}F_2 = 150\,\mathrm{N}, \tag{16.22}$$

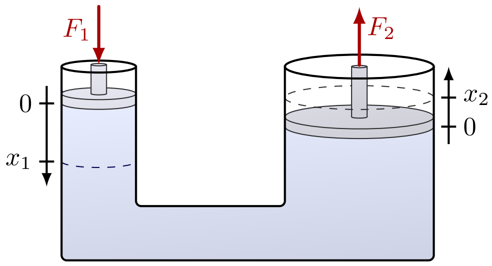


Figure 16.6: A hydraulic press demonstrating Pascal's principle. A force F_1 of the left piston with area A_1 causes an force $F_2 = (A_2/A_1)F_1$ on the right piston with area A_2.

which is the equivalent to a weight of only about 15 kg! So a 15 kg weight can compensate for the total weight of the car, thanks to the factor of r_1^2/r_2^2.

So what if we want to lift the car by a height $x_2 = 1\,\text{cm}$? The work to accomplish this is

$$W = F_2 x_2, \tag{16.23}$$

where F_2 is the work done by the average force F_2 on the large piston. The smaller piston performs exactly this amount of work with some average force F_1, and so

$$F_1 x_1 = F_2 x_2. \tag{16.24}$$

The smaller piston has to move a distance

$$x_1 = \frac{F_2}{F_1} x_2 = \frac{A_1}{A_2} x_2 = 1\,\text{m}, \tag{16.25}$$

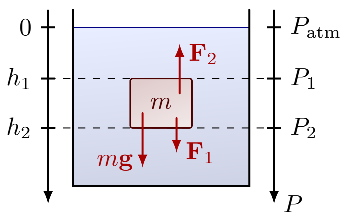
to move the car by a one meter with a force of about 150 N!

16.7 Buoyancy

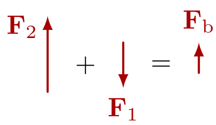
Suppose you fully submerge an object of density ρ_{obj} in a fluid with density ρ_ℓ. Everyone knows intuitively that the object will sink if $\rho_{\text{obj}} > \rho_\ell$, and float if $\rho_{\text{obj}} < \rho_\ell$. Even if it sinks, the object will appear lighter; it will have "less weight". The force that makes an object float or feel lighter is called the *buoyant force* F_b, and it is caused by the difference in pressure with depth.

Let's have a closer look at what exactly happens in Fig. 16.7. The fully submerged object feels a difference in pressure between the top and bottom. The pressure on the sides cancel. The net upward force due to this pressure difference is

$$F_\text{b} = F_2 - F_1 = P_2 A - P_1 A, \tag{16.26}$$



(a) Mass submerged at some depth. The total force due to pressure on the top is $\mathbf{F}_1$, and $\mathbf{F}_2$ on the bottom.



(b) The net upward force due to pressure is the buoyant force $\mathbf{F}_\text{b}$.

Figure 16.7: A mass m that is completely submerged in a fluid at some depth experiences a pressure differential. The pressure P_2 on the bottom is larger than the pressure P_1 on the top. Pressure on the sides balance.

where A is the top and bottom area of the object. From Eq. (16.12),

$$F_b = (\rho_\ell g h_{obj})A, \tag{16.27}$$

where h_{obj} is the object's total height. The object's total volume is then $V_{obj} = Ah$, so the buoyant force for a fully submerged object is

> **Buoyant force (fully submerged).**
>
> $$F_b = \rho_\ell V_{obj} g. \tag{16.28}$$

Clearly, if $F_b > mg$, then the object will float to the top, while if $F_b < mg$, it will sink. The mass of the object is given by $m = \rho_{obj} V_{obj}$, so the weight is given by

$$W_{obj} = mg = \rho_{obj} V_{obj} g. \tag{16.29}$$

Comparing to $F_b = \rho_\ell V_{obj} g$, we see that as expected the object floats when $\rho_{obj} < \rho_\ell$ and sinks when $\rho_{obj} > \rho_\ell$.

16.7.1 Archimedes' principle

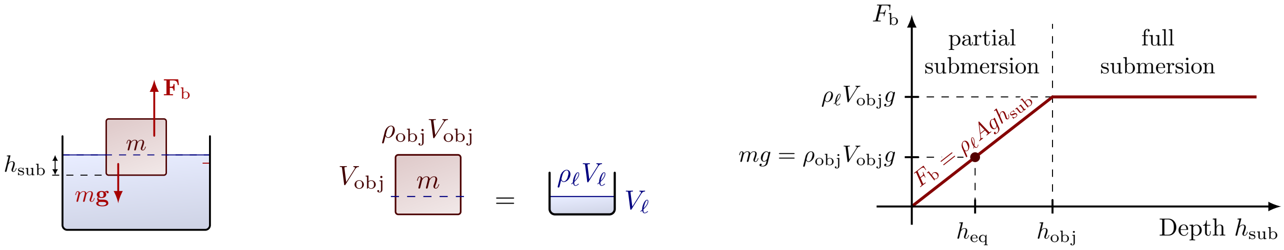
What if the object floats and is only partially submerged as in Fig. 16.8a? The pressure on the top of the object is P_{atm}, while on the bottom there is a pressure difference $\rho_\ell g h_{sub}$, where h_{sub} is the height of the object that is submerged. Therefore, we can write the buoyant force as

$$F_b = \rho_\ell V_\ell g, \tag{16.30}$$

where $V_\ell = h_{sub} A$ is the total volume of the object that is submerged. Notice that V_ℓ is actually the volume of the fluid that was displaced! This volume of the liquid has weight $W_\ell = \rho_\ell V_\ell g$. This is Archimedes principle:

> **Archimedes' principle.** *Body (partially or completely) submerged in fluid is forced up by a buoyant force equal to the weight of fluid that is displaced;* $F_b = \rho_\ell V_\ell g$.

If the body if fully submerged, $h = h_{sub}$ and $V_{obj} = V_\ell$, and we retrieve the previous Eq. (16.28). Before it is fully submerged, the buoyant force varies linearly with depth h_{sub}, as shown in Fig. 16.8c.



(a) The object is partially submerged and floats if $\rho < \rho_\ell$, such that $F_b \geqslant mg$.

(b) At equilibrium, the weight of displaced fluid in (a) equals the object's weight $\rho_\ell V_\ell g = mg$.

(c) Buoyant force varies with submersion h_{sub}. Partially submerged, $F_b = \rho_\ell A g h_{sub}$. Once fully submersed, the buoyant force is $F_b = \rho_\ell V_{obj} g$.

Figure 16.8: Archimedes' principle says that the buoyant force equals the weight of the displaced liquid, $F_b = \rho_\ell V_\ell g$. If $\rho_{obj} < \rho_\ell$, the object will float, with $\rho_\ell V_\ell g = mg = \rho_{obj} V_{obj} g$ at hydrostatic equilibrium.

At *hydrostatic equilibrium*, the buoyant force F_b must exactly balance the object's weight,

$$F_\mathrm{b} = \rho_\mathrm{obj} V_\mathrm{obj} g, \tag{16.31}$$

or,

Condition for floating at hydrostatic equilibrium.

$$\rho_\mathrm{obj} V_\mathrm{obj} = \rho_\ell V_\ell, \tag{16.32}$$

where V_ℓ is the volume of the displaced fluid, or equivalently of the part of the object that is submerged.

16.7.2 Example: Iceberg

What fraction of a iceberg floating freely in salt water is submerged? Ice has a density of $\rho_\mathrm{i} = 916.7\,\mathrm{kg/m^3}$, smaller than that of salt water with $\rho_\mathrm{w} = 1030\,\mathrm{kg/m^3}$. The volume V_sub of the iceberg that is submerged is given by Eq. (16.32):

$$V_\mathrm{sub} = \frac{\rho_\mathrm{i}}{\rho_\mathrm{w}} V_\mathrm{i}, \tag{16.33}$$

where V_i is the full volume of the iceberg. Therefore, the fraction that is submerged is

$$\frac{V_\mathrm{sub}}{V_\mathrm{i}} = \frac{\rho_\mathrm{i}}{\rho_\mathrm{w}} = 89\%.$$

So the tip of the iceberg you see above sea water is only about 11% of the whole iceberg.

16.7.3 Archimedes' trick: Measuring density

The famous legend tells the story that the king of Syracuse asked Archimedes (ca. 287–212 BC) to determine if his crown was made out of pure gold without damaging it. Archimedes accepted the challenge and started thinking. When he visited one of the public bath houses and lowered himself in the tub of water, he noticed the water level rising. This gave him a great idea to find out the density of the crown. He was so excited, he ran out into the streets, still naked, yelling "Eureka!" ("I have found it!"). It turned out that the goldsmith had indeed deceived the king and mixed in the cheaper and less dense silver.

Let's study this problem. Say you have a mass m that is denser than water, $\rho_\mathrm{obj} > \rho_\ell$. You submerge it fully in a container with water and hold it in the middle with a wire like in Fig. 16.9. Without the water, the wire would have a tension $T = mg$ equal to the weight, but now there is the additional buoyant force making the mass appear lighter:

$$T = mg - F_\mathrm{b}. \tag{16.34}$$

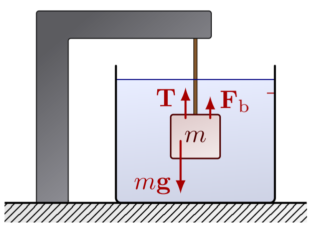


Figure 16.9: Object with higher mass density than the fluid, $\rho > \rho_\ell$ is held in place by a wire.

With a scale we can measure the weight $W_{\text{obj}} = mg$ and the new apparent weight T. Since the object is fully submerged, $F_{\text{b}} = \rho_\ell V_{\text{obj}} g$, which is the weight

$$W_\ell = m_\ell g = \rho_\ell V_\ell g \tag{16.35}$$

of the total volume of displaced fluid, $V_\ell = V_{\text{obj}}$, with mass $m_\ell = \rho_\ell V_\ell$. Therefore, we can write the object's density relative to water as

$$\frac{\rho_{\text{obj}}}{\rho_\ell} = \frac{m/V_{\text{obj}}}{m_\ell/V_\ell}. \tag{16.36}$$

The volumes cancel and we can substitute the weights

$$\frac{\rho_{\text{obj}}}{\rho_\ell} = \frac{m}{m_\ell} = \frac{mg}{W_\ell}. \tag{16.37}$$

By Archimedes' principle, $F_{\text{b}} = W_\ell$, such that

$$\frac{\rho_{\text{obj}}}{\rho_\ell} = \frac{mg}{mg - T}. \tag{16.38}$$

There are two ways to find the density ratio: Either you weigh the displaced water to find m_ℓ, *or* you measure the tension T. The displaced water can be weighed by pouring off the water above the normal level and placing it on a scale, or if you know the container's area and ρ_ℓ simply by measuring difference of the water level. The tension can be measured with a Newton meter, like a spring scale, or an old-fashioned balance scale.

Chapter 17

Fluids in Motion

17.1 Continuity equation

Suppose you have a pipe as in Fig. 17.1a that has two sections with different cross-sectional areas A_1 and A_2. Water flows through the pipe, and now suppose some section of water of volume V_1 enters the first section with some velocity v_1. The volume can be expressed as

$$V_1 = A_1 v_1 \Delta t, \tag{17.1}$$

where $\ell_1 = v_1 \Delta t$ is the distance traveled in time interval Δt. By the same argument, the volume in the second section is

$$V_2 = A_2 v_2 \Delta t. \tag{17.2}$$

We assume that the fluid is incompressible, such that for the same Δt, $V_1 = V_2$, and

$$A_1 v_1 \Delta t = A_2 v_2 \Delta t, \tag{17.3}$$

or,

$$A_1 v_1 = A_2 v_2 = \text{const.} \tag{17.4}$$

We define the *volume flow rate*, or *current flow*,

Current flow.

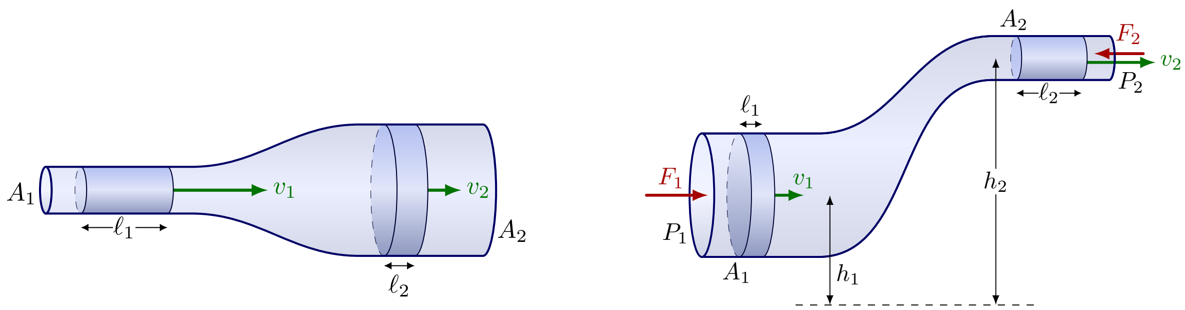
$$I_v = Av, \tag{17.5}$$

which has units m^3/s. We now arrive at a special equation that deserves its own box:

Continuity equation.

$$I_v = Av = const. \tag{17.6}$$

This equation simply tells us that the flow that goes in a pipe on what end, must come out on the other. So if the area A gets bigger, then the velocity gets bigger, then velocity decreases.



(a) Pipe with two segments of cross-sectional area A_1 and A_2.

(b) Pipe with two segments of volumes V_1 and V_2 and cross-sectional area A_1 and A_2 at different heights.

Figure 17.1: Fluids in pipes with different shapes and cross-sectional areas.

17.2 Bernouilli's equation

What if the pipe changes height in a gravitational field as in Fig. 17.1b? If the fluid is raised by some height h, it will lose some kinetic energy and gain potential energy, and vice versa if the fluid is lowered by some height. In some Δt, some small volume ΔV of the fluid with mass $\Delta m = \rho \Delta V$ gets lifted by a height $h = h_2 - h_1$. The change in potential energy for such a volume is

$$\Delta U = \Delta m g h = \rho \Delta V g h, \tag{17.7}$$

which can be negative or positive depending on the sign of $h = h_2 - h_1$. The corresponding change in kinetic energy is,

$$\Delta K = \frac{1}{2}\Delta m v_2^2 - \frac{1}{2}\Delta m v_1^2 \tag{17.8}$$

The work-energy theorem states that the total work done by the fluid is

$$W = \Delta U + \Delta K, \tag{17.9}$$

or,

$$W = \rho \Delta V g h + \frac{1}{2}\rho \Delta V (v_2^2 - v_1^2). \tag{17.10}$$

But we know that work comes from force times distance, $W = F\Delta x$, so where does the force come from? In the case of these fluids, there has to some pressure that pushes the fluid through the pipe. The force at the bottom of the pipe, F_1, comes from some pressure P_1, and the opposing force F_2 at the top of the pipe comes from pressure P_2. At the bottom, we have

$$W_1 = F_1 \Delta x_1 = P_1 A_1 \Delta x_1 = P_1 \Delta V, \tag{17.11}$$

while at the top for $\Delta V = A_2 \Delta x_2$,

$$W_2 = -P_2 \Delta V, \tag{17.12}$$

where the minus sign indicates that F_2 is opposite to Δx_2 and F_1. Then, the total work done is

$$W = W_1 + W_2 = (P_1 - P_2)\Delta V. \tag{17.13}$$

We compare this equation to Eq. (17.13), and find

$$P_1 - P_2 = \rho g h + \frac{1}{2}\rho (v_2^2 - v_1^2) \tag{17.14}$$

Reshuffling, we arrive at *Bernoulli's equation*:

Bernoulli's equation. *Between any two points in a pipe,*

$$P_1 + \rho g h_1 + \frac{1}{2}\rho v_1^2 = P_2 + \rho g h_2 + \frac{1}{2}\rho v_2^2. \tag{17.15}$$

In other words, the sum of quantities on either side always stay constant throughout the pipe. Note that this is basically the equivalent of energy conservation for fluids. The $\rho g h$ terms corresponds to gravitational potential energy and $\frac{1}{2}\rho v^2$ corresponds to kinetic energy. The P_1 and P_2 terms correspond to the work done by forces pushing the fluid. As with potential energy, we will see that differences $\rho g(h_2 - h_1)$ and $P_2 - P_1$ are important than their individual values.

At rest, the kinetic terms drop, and we find our previous result Eq. (16.11):

$$P_2 = P_1 + \rho g(h_1 - h_2). \tag{17.16}$$

17.2.1 Torricelli's law

As an example, consider a tank of fluid with holes in it, as in Fig. 17.2. A water stream will be pouring out the holes. The pressure is higher the deeper you go in the fluid, so the lower the hole, the higher the initial velocity of the stream. Apply Bernoulli's equation (17.15) to the top of the fluid and on hole. We can simplify this equation by setting $h_1 = h$ on the top and $h_2 = 0$ at the hole:

$$P_1 + \rho g h + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2. \tag{17.17}$$

The top and the hole are exposed to the atmosphere, and therefore we also know that $P_1 = P_2 = P_{\text{atm}}$. Furthermore, if the tank is very large, we can assume that the tank level stays constant, neglecting the velocity of fluid at the top as the water level lowers. This means that $v_1 \approx 0$. We are left with

$$\rho g h = \frac{1}{2}\rho v_2^2. \tag{17.18}$$

Rewriting this, give us

Torricelli's law.

$$v = \sqrt{2gh}. \tag{17.19}$$

Notice that this is the same result as for an object falling freely over a height h, see Section 4.3.1.

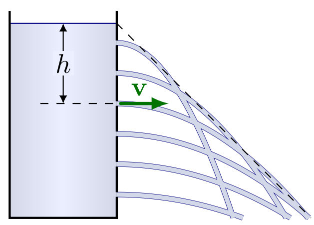


Figure 17.2: The initial velocity of streams in leaks depends on the height.

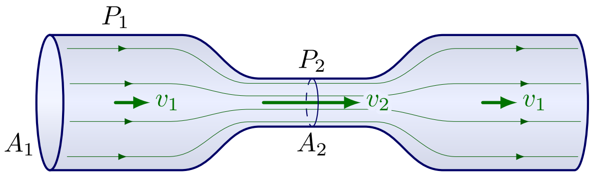


Figure 17.3: Pipe with two segments of cross-sectional area A_1 and A_2.

17.2.2 Venturi effect

Now consider a fluid moving trough a pipe with a cross-sectional area A_1 that has a narrow middle section with area $A_2 < A_1$ as in Fig. 17.3. We assume that the pressure P_1 is the same on either end, and that the height does not change. We are therefore left with the continuity equation,

$$v_2 = \frac{A_1}{A_2} v_2. \tag{17.20}$$

If A_2 is more narrow, $A_2 < A_1$, then the velocity has to be larger there, $v_2 > v_1$. What is the pressure there? Bernouilli's equation reduces to

$$P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2. \tag{17.21}$$

Since $v_2 > v_1$, the pressure will drop in the narrow section, $P_2 < P_1$. This is called the Venturi effect.

Venturi effect. *The pressure of a fluid decreases when its speed increases.*

When large vehicles, like trucks or trains, move by very fast, it can feel as if they are sucking you in due to the pressure drop. Similarly, when a train enters a narrow tunnel at high speed, the air between the train and tunnel walls suddenly speed up, and you can sometimes feel the difference in pressure on your ear drums.

An experiment that you can try at home is the following. Hold a piece of paper by the corners of a short end close to your mouth, and blow above it. Because the fast air passing above, the paper will lifted upward. This is similar to how some airplane wings can create a lift: Due to its shape, the air above the wing will travel faster than the air below.

17.3 Current resistance & viscous flow

Bernouilli's equation states pressure is the same anywhere in a pipe at constant height and constant area. However, in practice, we see that there is a pressure drop ΔP across the pipe. This can be explained by a drag force on the fluid coming from the pipe's surface and from each the fluid itself. To see why the fluid exerts a drag force on itself, imagine different concentric layers of the fluid as in Fig. 17.4a. The outer layer is dragged by the surface, and gets slowed down. This layer in turn, slows down the next layer due to internal friction, etc. Typically, the fluid will move faster in the center than close to the pipe's surface. There is a pressure difference that is proportional to the current flow,

Pressure drop due to drag.

$$\Delta P = P_1 - P_2 = R I_v, \tag{17.22}$$

with $I_v = vA$ again, and where R is a *constant of resistance*. The larger R, the larger the drag, and the larger the pressure drop. The resistance for steady flow in a pipe is

Resistance for steady flow in a cylindrical pipe.

$$R = \frac{8\eta L}{\pi r^4}, \tag{17.23}$$

where L and r are the length and radius of the pipe, respectively, and η is the *coefficient of viscosity* that depends on the material. A high viscosity means a large drag (like honey or crude oil), while low viscosity means the fluid can easily flow with too much drag (like water).

To define the viscosity η, take two plates that each have an area A and are separated by a height z. Suppose there is a fluid in between, that we imagine has different layers as in Fig. 17.4b. We push the top plate with some force F, such that it has a constant horizontal velocity v. The fluid below will also start to move due to drag, but the drag is larger closer to either plates. We see that the velocity continuously varies with z. Experimentally, it is found that

$$F = \frac{\eta v A}{z}. \tag{17.24}$$

By measuring the speed v for a given force F, we obtain η. Notice that the units of η are $\mathrm{N\,s/m^2}$ or $\mathrm{Pa\,s}$. It even has its own units of Poise (P),

$$1\,\mathrm{P} = 0.1\,\mathrm{Pa}. \tag{17.25}$$

From Eqs. 17.22 and 17.23 we see that the flow for a cylindrical pipe is

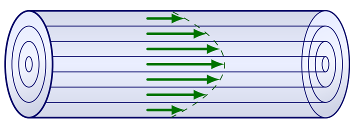
$$I_v = \Delta P \left(\frac{\pi r^4}{8\eta L} \right). \tag{17.26}$$

There are several interesting conclusions one can draw from this equation. For the same pressure difference ΔP, the flow rate depends on r^4. For the same pressure difference ΔP and radius r, the flow rate is inverse proportional to the pipe's length L.

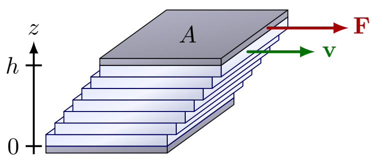
17.4 Laminar & turbulent flow

A flow that can be thought of as moving in separate, adjacent layers that do not mix as in Fig. 17.4a, is called *laminar*.

Consider you put an obstacle in a fluid that is in laminar flow. For small velocities, the fluid will smoothly flow around the object without mixing its layers, such that after passing



(a) Pipe with laminar flow: Layer close to the wall go slower due to drag forces. The dashed line is the *laminar boundary layer*.



(b) Defining the viscosity η by moving the the top plate.

Figure 17.4: In laminar flow, you can cleanly divide up the moving fluids into layers moving with their own speed.

(a) Laminar flow with an obstacle. The layers do not mix.

(b) Under some conditions, an object causes turbulence after.

Figure 17.5: An obstacle in a flow of fluid can cause turbulence.

the obstacles, it will still continue as laminar flow. This is shown in Fig. 17.5a. (Note that the velocity of flow is relative: The fluid can move, while the object is stationary, but the object can also be moving in a stationary fluid.)

At large velocities, however, there will be a transition from laminar to *turbulent flow*, depicted in Fig. 17.5b. Here the layers of the fluids start mixing and the moving in a chaotic, unpredictable pattern. This happens at some critical velocity, which can be understood with *Reynold's number*,

Reynold's number in a cylindrical pipe.

$$N_{\mathrm{R}} = \frac{2r\rho v}{\eta}. \tag{17.27}$$

The Reynold's number does not have any units. It gives the condition of type of flow. If $N_{\mathrm{R}} < 2000$, then the flow will be laminar. If $N_{\mathrm{R}} > 3000$, then the flow will be turbulent. Between 2000 and 3000 there will be some transition regime. So if a laminar fluid starts moving very fast, it will become more turbulent.

If the pressure difference ΔP across a pipe is higher, the velocity v and Reynold's number N_{R} will become larger and the flow will become turbulent. The turbulence will reduce the overall flow as shown in Fig. 17.6.

17.5 Magnus effect

A object that is rotating and moving relative to a fluid will experience a net force. This is called the *Magnus effect* and is depicted in Fig. 17.7.

Remember the Venturi effect: If the fluid's velocity increases, then the pressure decreases. Because the object is rotating, the relative velocity between the object's surface and the fluid will be larger on one side than the other. This generates a pressure difference, and therefore a net force that tends to push the object from the high pressure toward the low pressure.

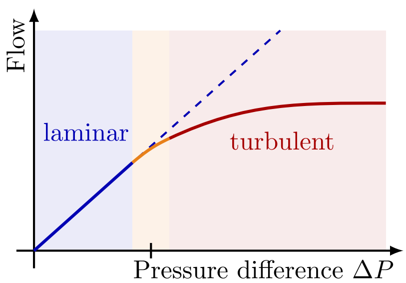


Figure 17.6: Flow becomes turbulent at large pressure differences ΔP. At higher ΔP, the velocity v and Reynold's number N_{R} will become larger.

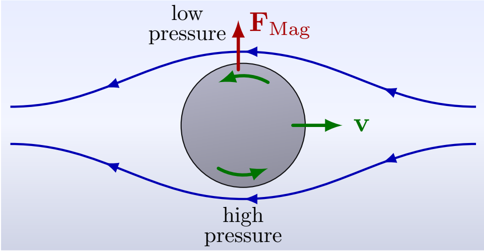


Figure 17.7: A rotating object moving relative to a fluid will feel a difference in pressure, causing a net force $\mathbf{F}_{\text{Mag}}$.

The Magnus effect allows players in baseball and football to curve a ball by giving it the right spin when throwing or kicking it. Similarly, a golf ball with backspin will spin higher into the air than with the same velocity but no spin.

Chapter 18

Surface Tension

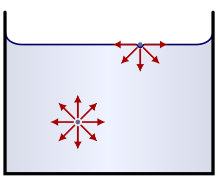
As we have seen in the last few chapters, a random point inside a fluid feels pressure from all sides due to molecules constantly colliding into each other and their container. However, the molecules of the fluid can in addition be attracted to each other due to *cohesive forces*. A random molecule *inside* a liquid will feel an equal attractive force from all directions, causing them to cancel. A molecule that is on the surface of the liquid, however, will only feel a cohesive force from this liquid below and from the sides, shown in Fig. 18.1a. The forces from the liquid molecules underneath are therefore unbalanced, and causes a pull inward. This generates some internal pressure that will deform the surface to minimize the area. The so-called *surface tension* is the total unbalanced force from the molecules of liquid below. The liquid's surface, as a consequence, acts as a stretchable membrane that can carry the weight of a light object that would otherwise sink ($\rho_{\text{obj}} > \rho_\ell$), like a needle or insect.

If you look more closely at the cross section of a floating needle on a water surface, you would see that the surface, otherwise flat, is indented as in Fig. 18.1b. Over the whole area where the needle is touching the water, there is a force F_{st} which is balancing the force of gravity $F_{\text{g}} = mg$. In general, the surface tension F_{st} on a thin object like a needle, is proportional to the total length L of the object,

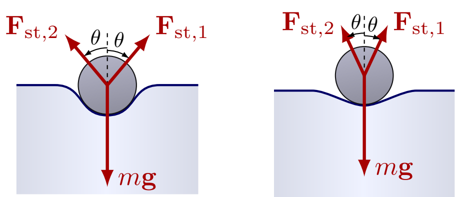
Surface tension on a thin, long object.

$$F_{\text{st}} = \gamma L, \tag{18.1}$$

where γ is a constant that depends on the liquid and its temperature. It is defined as the



(a) The attractive force on a molecule inside the liquid cancel, but not on the surface. The surface will act as a stretchable membrane.



(b) The total surface tension F_{st} from the points of contact balances the weight of the needle, $F_{\text{g}} = mg$. The weaker the surface tension, the deeper the indent.

Figure 18.1: The surface tension of a liquid can balance a metal needle with $\rho_{\text{obj}} > \rho_\ell$.

(a) Convex meniscus with $\theta_C > 90°$ has stronger cohesion than adhesion.

(b) Concave meniscus with $\theta_C < 90°$ has stronger adhesion than cohesion.

Figure 18.2: Capillary action

amount of force per unit length,

$$\gamma = \frac{F_{st}}{2L}. \tag{18.2}$$

The needle indents the surface, and there is a surface tension form either side, as shown in Fig. 18.1b. Therefore, the vertical component of the total surface tension is

$$F_{st,y} = F_{st,1}\cos\theta + F_{st,2}\cos\theta = (\gamma L)\cos\theta. \tag{18.3}$$

The needle floats as long as $F_{st,y} > mg$, and the surface will change shape to find equilibrium.

The angle θ decreases for larger needle's mass m. The maximum mass m_{max} of a needle with length L that the surface tension can carry is when $\theta = 0$, such that

$$m_{max} = \frac{2\gamma L}{g}. \tag{18.4}$$

18.1 Surfactants & soap bubbles

Soap, whose molecule has one hydrophilic end that attracts water, can lower the surface tension of water, causing an insect of needle to sink through. Soap is therefore called a *surfactant* (from "surface-active agent").

Surfactants in combination with water can create foam and bubbles. Without a surfactant, the water surface tension is too large to form stable bubbles. Soapy water will create a *film*, which is a thin layer of water that is between two layer of soap. Soap molecules have one greasy, hydrophobic end that stick out to the outside of the film and protects the water from evaporation, prolonging the bubble's lifetime. When a soap film encloses some air, it will form a sphere to minimize the surface tension, and therefore the area. The pressure of the trapped air is slightly increased to balance the internal pressure due to the surface tension.

18.2 Capillary action

One consequence of surface tension is *capillary action*. This comes from the attractive *adhesive forces* between the liquid and walls of the container. If the adhesive forces with the wall are larger than the cohesive forces between the molecules, the surface will make a "U" shape (i.e. *concave*). If the cohesive forces are larger than the adhesive forces, the surface will curve upward in the middle (i.e. *convex*). A curved surface of a liquid is called a *meniscus*. This can be described by the *contact angle* between the surface and the wall, θ_C, shown in Fig. 18.2.

The contact angle θ_C depends on the fluid and container materials. For water in a glass container, $\theta_C = 0$, creating a very concave-shaped meniscus. For mercury with glass, it is $\theta_C = 140°$, creating a concave meniscus.