INTRODUCTORY

PHYSICS

Building Models to Describe Our World

Ryan Martin • Emma Neary • Olivia Woodman

License

Preface

About this textbook

This textbook is written to fill several needs that we believe were not already met by the many existing introductory physics textbooks. First, we wanted to ensure that the textbook is free to use for students and professors. Second, we wanted to design a textbook that is mindful of the new pedagogies being used in introductory physics, by writing it in a way that is adapted to a flipped-classroom approach where students complete readings, think about the readings, and then discuss the material in class. Third, we wanted to create a textbook that also addresses the experimental aspect of physics, by proposing experiments to be conducted at home or in the lab, as well as providing guidelines for designing experiments and reporting on experimental results. Finally, we wanted to create a textbook that is a sort of "living document", that professors can edit and re-mix for their own needs, and to which students can contribute material as well. The textbook is hosted on GitHub, which allows anyone to make suggestions, point out issues and mistakes, and contribute material.

This textbook is meant to be paired with the accompanying "Question Library", which contains many practice problems, many of which were contributed by students.

This textbook would not have been possible without the support of Queen's University and the Department of Physics, Engineering Physics & Astronomy at Queen's University, as well as the many helpful discussions with the students, technicians and professors at Queen's University.

Hello from the authors

Ryan Martin I am a professor of physics at Queen's University. My main research is in the field of particle astrophysics, particularly in studying the properties of neutrinos. I grew up in Switzerland, obtained my Bachelor's, Master's and Ph.D. at Queen's University. I was then a postdoctoral fellow at Lawrence Berkeley National Laboratory, a faculty at the University of South Dakota, before returning to Queen's. I am particularly passionate about education, and I am always seeking opportunities to involve students in helping to make education more accessible. I also like to cook and to play volleyball.

Emma Neary I am currently a second year physics major and QuARMS (Queen's University Accelerated Route to Medical School) student, as well as a native of St. John's, Newfoundland. Uniting the perspectives of students and professors in an accessible way is important to me. I strongly believe in the importance of building physical models; whether it be in physics, medicine, sciences or the arts. It has been my goal to infuse the textbook with the theme of modelling in a creative and engaging way. Aside from doing physics, I enjoy hiking, dancing, reading and doing research in gastroenterology and neuropsychiatry.

Olivia Woodman I am a currently a third year undergraduate student at Queen's Univeristy, majoring in physics. The flipped classroom approach has been beneficial to my own learning, and I think that we have created a textbook that really complements this learning style. Throughout this book, I have shared my thoughts on various topics in physics, as well as some useful tips and tricks. I hope that students enjoy using this book and continue to contribute to it in the future. Working on this textbook has also allowed me to combine my love of physics with my love of doodling, so I hope you enjoy the drawings!

How to use this textbook

This textbook is designed to be used in a flipped-classroom approach, where students complete readings at home, and the material is then discussed in class. The material is thus presented fairly succinctly, and contains **Checkpoint Questions** throughout that are meant to be answered as the students complete the reading. We suggest including these Checkpoint Questions as part of a quiz in a reading assignment (marked based on completion, not correctness), and then using these questions as a starting point for discussions in class.

For topics that are particularly difficult, we have included **Thought Boxes** written by students that try to present the material in a different light. We are always happy if students (or professors) wish to contribute additional thought boxes.

Chapters start with a set of **Learning outcomes** and an **Opening question** to help students have a sense of the chapter contents. The chapters have **Examples** throughout, as well as additional practice problems at the end. The **Question Library** should be consulted for additional practice problems. At the end of the chapter, a **Summary** presents the key points from the chapter. We suggest that students carefully read the summaries to make sure that they understand the contents of the chapter (and potentially identify, before reading the chapter, if the content is review to them). At the end of the chapters, we also present a section to **Think about the material**. This includes questions that can be assigned in reading assignments to research applications of the material or historical context. The thinking about the material section also includes experiments that can be done at home (as part of the reading assignment) or in the lab.

Appendices cover the main background in mathematics (Calculus and Vectors), as well as present an introduction to programming in python, which we feel is a useful skill to have in science. There is also an Appendix that is intended to guide work in the lab, by providing examples of how to write experimental proposals and reports, as well as guidelines for reviewing proposals and reports. We believe that introductory laboratories should not be be "recipe-based", but rather that students should take an approach similar to that of a researcher in designing (proposing) an experiment, conducting it, and reviewing the proposals and results of their peers.

Credits

This textbook, and especially the many questions in the Question Library would not have been possible without the many contributions from students, teaching assistants and other professors. Below is a list of the people that have contributed material that have made this textbook and Question Library possible.

Adam McCaw
Ali Pirhadi
Alexis Brossard
Amy Van Nest
Ceaira Heimstra
Damara Gagnier
Daniel Barake
Daniel Tazbaz
David Cutler
Emily Darling
Emily Mendelson
Emily Wener
Emma Lanciault
Genevieve Fawcett
Gregory Love
Haoyuan Wang
Jack Fitzgerald
James Godfrey
Jenna Vanker
Jesse Fu
Jesse Simmons
Jessica Grennan
Joanna Fu
Jonathan Abott
Josh Rinaldo
Kate Fenwick
Madison Facchini
Marie Vidal
Matt Routliffe
Maya Gibb
Nicholas Everton
Nick Brown
Nicole Gaul
Noah Rowe
Olivia Bouaban
Patrick Singal
Qiqi Zhang
Quentin Sanders
Robin Joshi
Ryan Underwood
Sam Connolly
Sara Stephens
Shona Birkett
Stephanie Ciccone
Talia Castillo
Tamy Puniani
Thomas Faour
Troy Allen
Wei Zhuolin
Yumian Chen
Zifeng Chen
Zoe Macmillan

Contents

1 The Scientific Method and Physics

Learning Objectives

- Understand the Scientific Method.
- Define the scope of Physics.
- Understand the difference between theory and model.
- Have a sense of how a physicist thinks.

Think About It

A scientific theory...

A) must explain the physical world, and it may or may not be experimentally verifiable.
B) proves our models to be correct, and it must be experimentally verifiable.
C) describes the physical world, and must be experimentally verifiable.
D) must disprove other theories, and may or may not be experimentally verifiable.

1.1 Science and the Scientific Method

Science is the process of *describing* the world around us. It is important to note that describing the world around us is not the same as *explaining* the world around us. Science aims to answer the question "How?" and not the question "Why?". As we develop our description of the physical world, you should remember this important distinction and resist the urge to ask "Why?".

The Scientific Method is a prescription for coming up with a description of the physical world that anyone can challenge and improve through performing experiments. If we come up with a description that can describe many observations, or the outcome of many different experiments, then we usually call that description a "Scientific Theory". We can get some insight into the Scientific Method through a simple example.

Imagine that we wish to describe how long it takes for a tennis ball to reach the ground after being released from a certain height. One way to proceed is to describe how long it takes for a tennis ball to drop 1 m, and then to describe how long it takes for a tennis ball to drop 2 m, etc. We could generate a giant table showing how long it takes a tennis ball to drop from any given height. Someone would then be able to perform an experiment to measure how long a tennis ball takes to drop from 1 m or 2 m and see if their measurement disagrees with the tabulated values. If we collected the descriptions for all possible heights, then we would effectively have a valid and testable scientific theory that describes how long it takes tennis balls to drop from any height.

Suppose that a budding scientist, let's call her Chloë, then came along and noticed that there is a pattern in the theory that can be described much more succinctly and generally than by using a giant table. In particular, suppose that she notices that, mathematically, the time, t, that it takes for a tennis ball to drop a height, h, is proportional to the square root of the height:

$$t \propto \sqrt{h}$$

Example 1-1

Use Chloë's Theory ($t \propto \sqrt{h}$) to determine how much longer it will take for an object to drop by 2 m than it would to drop by 1 m.

Solution

When we have a proportionality law (with a $\propto$) sign, we can always change this to an equal sign by introducing a constant, which we will call k:

$$t \propto \sqrt{h}$$
$$\rightarrow t = k\sqrt{h}$$

Let t_1 be the time to fall a distance $h_1 = 1\,\text{m}$, and t_2 be the time to fall a distance $h_2 = 2\,\text{m}$. In terms of our unknown constant, k, we have:

$$t_1 = k\sqrt{h_1} = k\sqrt{(1\,\text{m})}$$
$$t_2 = k\sqrt{h_2} = k\sqrt{(2\,\text{m})}$$

By taking the ratio, $\frac{t_1}{t_2}$, our unknown constant k will cancel:

$$\begin{aligned}\frac{t_1}{t_2} &= \frac{\sqrt{(1\,\mathrm{m})}}{\sqrt{(2\,\mathrm{m})}} = \frac{1}{\sqrt{2}} \\ \therefore t_2 &= \sqrt{2}t_1\end{aligned}$$

and we find that it will take $\sqrt{2} \sim 1.41$ times longer to drop by 2 m than it will by 1 m.

Chloë's "Theory of Tennis Ball Drop Times" is appealing because it is succinct, and it also allows us to make **verifiable predictions**. That is, using this theory, we can predict that it will take a tennis ball $\sqrt{2}$ times longer to drop from 2 m than it will from 1 m, and then perform an experiment to verify that prediction. If the experiment agrees with the prediction, then we conclude that Chloë's theory adequately describes the result of that particular experiment. If the experiment does not agree with the prediction, then we conclude that the theory is not an adequate description of that experiment, and we try to find a new theory.

Chloë's theory is also appealing because it can describe not only tennis balls, but the time it takes for other objects to fall as well. Scientists can then set out to continue testing her theory with a wide range of objects and drop heights to see if it describes those experiments as well. Inevitably, they will discover situations where Chloë's theory fails to adequately describe the time that it takes for objects to fall (can you think of an example?).

We would then develop a new "Theory of Falling Objects" that would include Chloë's theory that describes most objects falling, and additionally, a set of descriptions for the fall times for cases that are not described by Chloë's theory. Ideally, we would seek a new theory that would also describe the new phenomena not described by Chloë's theory in a succinct manner. There is of course no guarantee, ever, that such a theory would exist; it is just an optimistic hope of physicists to find the most general and succinct description of the physical world. This is a general difference between physics and many of the other sciences. In physics, one always tries to arrive at a succinct theory (e.g. an equation) that can describe many phenomena, whereas the other sciences are often very descriptive. For example, there is no succinct formula for how butterflies look; rather, there is a giant collection of observations of different butterflies.

This example highlights that applying the Scientific Method is an iterative process. Loosely, the prescription for applying the Scientific Method is:

1. Identify and describe a process that is not currently described by a theory.
2. Look at similar processes to see if they can be described in a similar way.
3. Improve the description to arrive at a "Theory" that can be generalized to make predictions.
4. Test predictions of the theory on new processes until a prediction fails.

5. Improve the theory.

Checkpoint 1-1

Fill in the blanks:

Physics is a branch of science that __________ the behaviour of the universe. When doing physics, we attempt to answer the question of __________ things work the way they do.

A) explains
B) describes
C) how
D) why

1.2 Theories and models

For the purpose of this textbook, we wish to introduce a distinction in what we mean by "theory" and by "model". We will consider a "theory" to be a set of statements that gives us a broad description, applicable to several phenomena and that allow us to make verifiable predictions. We will consider a "model" to be a situation-specific description of a phenomenon *based on a theory*, that allows to make a specific prediction. Using the example from the previous section, our theory would be that the fall time of an object is proportional to the square root of the drop height, and a model would be applying that theory to describe a tennis ball falling by $4.2\,\text{m}$.

This textbook will introduce the theories from Classical Physics, which were mostly established and tested between the seventeenth and nineteenth centuries. We will take it as given that readers of this textbook are not likely to perform experiments that challenge those well-established theories. The main challenge will be, given a theory, to define a model that describes a particular situation, and then to test that model. This introductory physics course is thus focused on thinking of "doing physics" as the task of correctly modelling a situation.

Emma's Thoughts

What's the difference between a model and a theory?
"Model" and "Theory" are sometimes used interchangeably among scientists. In physics, it is particularly important to distinguish between these two terms. A model provides an immediate understanding of something based on a theory.

For example, if you would like to model the launch of your toy rocket into space, you might run a computer simulation of the launch based on various theories of propulsion that you have learned. In this case, the model is the computer simulation, which describes what will happen to the rocket. This model depends on various theories that have been extensively tested such as Newton's Laws of motion, Fluid dynamics, etc.

- "Model": Your homemade rocket computer simulation
- "Theory": Newton's Laws of motion, Fluid dynamics

With this analogy, we can quickly see that the "model" and "theory" are not interchangeable. If they were, we would be saying that all of Newton's Laws of Motion depend on the success of your piddly toy rocket computer simulation!

Checkpoint 1-2

Models cannot be scientifically tested, only theories can be tested.

A) True
B) False

1.3 Fighting intuition

It is important to remember to fight one's intuition when applying the scientific method. Certain theories, such as Quantum Mechanics, are very counter-intuitive. For example, in Quantum Mechanics, an object can be described as being in two locations at the same time. In the Theory of Special Relativity, it is possible for two people to disagree on whether two events occurred at the same time. These particular prediction from these theories have not been invalidated by any experiment.

There is no requirement in science that a theory be "pretty" or intuitive. The only requirement is that a theory describe experimental data. One should then take care in not forcing one's preconceived notions into interpreting a theory. For example, Quantum Mechanics does not actually predict that objects can be in two locations at once, only that objects behave *as if* they were in two locations at once. A famous example is Schrödinger's cat, which can be modelled as being both alive and dead at the same time. However, just because we model it that way does not mean that it really is alive and dead at the same time.

1.4 The scope of Physics

Physics describes a wide range of phenomena within the physical sciences, ranging from the behaviour of microscopic particles that make up matter to the evolution of the entire Universe. We often distinguish between "classical" and "modern" physics depending on when the theories were developed, and we can further subdivide these areas of physics depending on the scale or the type of the phenomena that they describe.

The word physics comes from Ancient Greek and translates to "nature" or "knowledge of nature". The goal of physics is to develop theories from which mathematical models can be derived to describe our observations. One of the ambitious goals of physicists is to develop a single theory that describes all of nature, instead of having multiple theories to describe different categories of phenomena. This is in stark contrast to other fields of science, as Rutherford famously quipped: "All science is either physics or stamp collecting". That is, physicists hope that there exists one single mathematical theory (like Chloë's theory of falling objects) that describes the entire physical world. In Biology, for example, this would not be a reasonable goal, as one needs to describe every single living being, and there is no overarching "theory of what all living things look like". Currently, physicists have been able to narrow down the number of theories required to describe all of the physical world to only three, which is impressive (the theory of gravity, the theory of the strong nuclear force, and physicists have now further unified the weak nuclear force with electromagnetism to make the "electroweak force").

1.4.1 Classical Physics

This textbook is focused on classical physics, which corresponds to the theories that were developed before 1905.

Mechanics

Mechanics describes most of our everyday experiences, such as how objects move, including how planets move under the influence of gravity. Isaac Newton was the first to formally develop a theory of mechanics, using his "Three Laws" to describe the behaviour of objects in our everyday experience. His famous work published in 1687, "Philosophiae Naturalis Principia Mathematica" ("The Principia") also included a theory of gravity that describes the motion of celestial objects.

Following the 1781 discovery of the planet Uranus by William Herschel, astronomers noticed that the orbit of the planet was not well described by Newton's theory. This led Urbain Le Verrier (in Paris) and John Couch Adams (in Cambridge) to predict the location of a new planet that was disturbing the orbit of Uranus rather than to claim that Newton's theory was incorrect. The planet Neptune was subsequently discovered by Le Verrier in 1846, one year after the prediction, and seen as a resounding confirmation of Newton's theory.

In 1859, Urbain Le Verrier also noted that Mercury's orbit around the Sun is different than that predicted by Newton's theory. Again, a new planet was proposed, "Vulcan", but that planet was never discovered and the deviation of Mercury's orbit from Newton's prediction remained unexplained until 1915, when Albert Einstein introduced a new, more complete, theory of gravity, called "General Relativity". This is a good example of the scientific method; although the discovery of Neptune was consistent with Newton's theory, it did not prove that the theory is correct, only that it correctly described the motion of Uranus. The discrepancy that arose when looking at Mercury ultimately showed that Newtons' theory of gravity fails to provide a proper description of planetary orbits in the proximity of very massive objects (Mercury is the closest planet to the Sun).

Checkpoint 1-3

What did the inability to find the planet Vulcan show:

A) It showed that Newton's model of Mercury was correct.
B) It showed that Newton's theory did not correctly describe the orbits of all planets.
C) It showed that the technology at the time was inadequate.
D) It showed that Einstein's theory of General Relativity was correct.

Electromagnetism

Electromagnetism describes electric charges and magnetism. At first, it was not realized that electricity and magnetism were connected. Charles Augustin de Coulomb published in 1784 the first description of how electric charges attract and repel each other. Magnetism was discovered in the ancient world, when people noticed that lodestone (rocks made from magnetized magnetite mineral) could attract iron tools. In 1819, Oersted discovered that moving electric charges could influence a compass needle, and several subsequent experiments were carried out to discover how magnets and moving electric charges interact.

In 1865, James Clerk Maxwell published "A Dynamical Theory of the Electromagnetic Field", wherein he first proposed a theory that unified electricity and magnetism as two facets of the same phenomenon. One important concept from Maxwell's theory is that light is an electromagnetic wave with a well-defined speed. This uncovered some potential issues with the theory as it required an absolute frame of reference in which to describe the propagation of light. Experiments in the late 1800s failed to detect the existence of this frame of reference.

1.4.2 Modern Physics

In 1905, Albert Einstein published three major papers that set the foundation for what we now call "Modern Physics". These papers covered the following areas that were not well-described by classical physics:

- A description of Brownian motion that implied that all matter is made of atoms.
- A description of the photoelectric effect that implied that light is made of particles.
- A description of the motion of very fast objects that implied that mass is equivalent to energy, and that time and distance are relative concepts.

In order to accommodate Einstein's descriptions, physicists had to dramatically re-formulate new theories.

Quantum mechanics and particle physics

Quantum mechanics is a theory that was developed in the 1920s to incorporate Einstein's conclusion that light is made of particles (or rather, quantized lumps of energy called quanta) and describe nature at the smallest scales. This could only be done at the expense of determinism, the idea that we can predict how particular situations evolve in time. This led to a theory that could only provide the *probabilities* that certain outcomes will be realized. Quantum mechanics was further refined during the twentieth century into Quantum Field Theory, which led to the Standard Model of particle physics that describes our current understanding of matter through the theories of the electroweak and strong forces.

The Special and General Theories of Relativity

In 1905, Einstein published his "Special Theory of Relativity", which describes how light propagates at a constant speed without the need for an absolute frame of reference, thus solving the problem introduced by Maxwell. This required physicists to consider space and time on an equal footing ("space-time"), rather than two independent aspects of the natural world, and led to a flurry of odd, but verified, experimental predictions. One such prediction is that time flows slower for objects that are moving fast, which has been experimentally verified by flying precise atomic clocks on airplanes and satellites. In 1915, Einstein further refined his theory into General Relativity, which is our best current description of gravity and includes a description of Mercury's orbit which was not described by Newton's theory.

Checkpoint 1-4

Special relativity can be applied to which of these science fiction plots?

A) An eccentric duo travel back in time to alter the past.
B) An astronaut travelling near light speed for many years comes home to find that he has aged less than his family on Earth.
C) A superhero harnesses lightning to use as a weapon.

Cosmology and astrophysics

Cosmology describes processes at the largest scales and is mostly based on applying General Relativity to the scale of the Universe. For example, cosmology describes how our Universe started from the Big Bang and how large scale structures, such as galaxies and clusters of galaxies, have formed and evolved into our present day Universe.

Figure 1.1: A galaxy in the Coma cluster of galaxies (credit:NASA).

Astrophysics is focused on describing the formation and the evolution of stars, galaxies, and other "astrophysical objects" such as neutron stars and black holes.

Particle astrophysics
Particle astrophysics is a relatively new field that makes use of subatomic particles produced by astrophysical objects to learn both about the objects *and* about the particles. For example, the 2015 Nobel Prize in Physics was awarded to Art McDonald (a Canadian physicist from Queen's University) for using neutrinos[1] produced by the Sun to both learn about the nature of neutrinos and about how the Sun works.

1.5 Thinking like a physicist

In a sense, physics can be thought of as the most fundamental of the sciences, as it describes the interactions of the smallest constituents of matter. In principle, if one can precisely describe how protons, neutrons, and electrons interact, then one can completely describe how a human brain thinks. In practice, the theories of particle physics lead to equations that are too difficult to solve for systems that include as many particles as a human brain. In fact, they are too difficult to solve exactly for even rather small systems of particles such as atoms bigger than helium (containing several protons, neutrons and electrons).

We have a number of other fields of science to cover complex systems of particles interacting. Chemistry can be used to describe what happens to systems consisting of many atoms and molecules. In a living being, it is too difficult to keep track of systems of atoms and molecules, so we use Biology to describe living systems.

One of the key qualities required to be an effective physicist is an ability to understand how to apply a theory and develop a model to describe a phenomenon. Just like any other skill, it takes practice to become good at developing models. Students that graduate with a physics degree are thus often sought for jobs that require critical thinking and the ability to develop quantitative models, which covers many fields from outside of physics such as finance or Big Data. This textbook thus tries to emphasize practice with developing models, while also providing a strong background in the theories of classical physics.

[1]Neutrinos are the lightest subatomic particles that we know of

1.6 Summary

Key Takeaways

Science attempts to *describe* the physical world (it answers the question "How?", not "Why?").

The Scientific Method provides a prescription for arriving at theories that describe the physical world and can be experimentally verified. The Scientific Method is necessarily an iterative process where theories are continuously updated as new experimental data are acquired. An experiment can only disprove a theory, not confirm it in any general sense.

Physics covers a wide scale of phenomena ranging from the Universe down to subatomic particles. Classical physics encompasses the theories developed before 1905, when Einstein introduced the need for Quantum Mechanics and the Theorie(s) of Relativity. One of the main goals of physics is to arrive at a single theory that describes all of our natural world. Currently, physicists require three theories to describe the natural world.

1.7 Thinking about the Material

Reflect and research

1. What particle helps to give mass to all of the massive elementary particles?
2. Name that physicist! Who was the first to propose that the universe is expanding?
3. Before discovering the CMBR (Cosmic Microwave Background Radiation), scientists Arno Penzias and Robert Wilson were trying to detect radio waves with very sensitive antennae. The very first time they heard a consistent, low noise on their detectors they discovered that it was (mostly) not the CMBR. What was causing most of this noise?
4. Physicist Lene Hau first slowed a beam of light to $17\,\mathrm{m/s}$ using a very cold, dilute gas of bosons. In 2001, how fast was she able to slow down the beam of light?
5. Think of two theories that you use in your every day life. (For example, when we wash our hands, we do so because of the germ theory of disease!)

1.8 Sample problems and solutions

1.8.1 Problems

Problem 1-1: Your friend Martin loves to explore "conspiracy theories". His favourite theory involves "Chem Trails". He tells you that the government is secretly using airliners to spread chemicals in the atmosphere for some unknown reason. (Solution)

a) Think of 2 ways in which you could objectively test Martin's theory.
b) After proposing your experiment to Martin, he claims that his theory cannot be invalidated by any experiment, no matter how scientifically rigorous the experiment is. Is Martin correct?

1.8.2 Solutions

Solution to problem 1-1:

a) You could do an investigation to see if the government is spreading chemicals, and try to find out why. You could make measurements of the contents in the atmosphere before and after an airline passes to see if any unexpected chemicals show up.

b) No he is not, as you just proposed two experiments that could invalidate his theory.

2 Comparing Model and Experiment

In this chapter, we will learn about the process of doing science and lay the foundations for developing skills that will be of use throughout your scientific careers. In particular, we will start to learn how to test a model with an experiment, as well as learn to estimate whether a given result or model makes sense.

Learning Objectives

- Be able to estimate orders of magnitude.
- Understand units.
- Understand the process of building a model and performing an experiment.
- Understand uncertainties in experiments.

Think About It

Newton's Universal Theory of Gravity predicts that objects near the surface of the Earth will fall with an acceleration of $9.8\,\text{m/s}^2$. Your friend reports that they have measured the acceleration of a falling ball and found that it was $(9.0 \pm 0.5)\,\text{m/s}^2$. Does their result invalidate the prediction from Newton's Theory?

A) Yes, since the range $(9.0 \pm 0.5)\,\text{m/s}^2$ does not include $9.8\,\text{m/s}^2$.
B) Not necessarily, as it depends on whether your friend correctly determined the uncertainty in their measurement.
C) Definitely not, since Newton's Universal Theory of Gravity has been confirmed by many experiments.

2.1 Orders of magnitude

Although you should try to fight intuition when building a model to describe a particular phenomenon, you should not abandon critical thinking and should always ask if a prediction from your model makes sense. One of the most straightforward ways to estimate if a model makes sense is to ask whether it predicts the correct order of magnitude for a quantity. Usually, the order of magnitude for a quantity can be determined by making a very simple model, ideally one that you can work through in your head. When we say that a prediction gives the right "order of magnitude", we usually mean that the prediction is within a factor of "a few" (up to a factor of 10) of the correct answer. For example, if a measurement gives a value of 2000, then we would consider that a model prediction of 8000 gave the right order of magnitude (it differs from the correct answer by a factor of 4), whereas a prediction of 24000 would not (it differs by a factor of 12).

Example 2-1

How many ping pong balls can you fit into a school bus? Is it of order 10,000, or 100,000, or more?

Solution

Our strategy is to estimate the volumes of a school bus and of a ping pong ball, and then calculate how many times the volume of the ping pong ball fits into the volume of the school bus.

We can model a school bus as a box, say $20\,\text{m}\times 2\,\text{m}\times 2\,\text{m}$, with a volume of $80\,\text{m}^3 \sim 100\,\text{m}^3$. We can model a ping pong ball as a sphere with a diameter of $0.03\,\text{m}$ ($3\,\text{cm}$). When stacking the ping pong balls, we can model them as little cubes with a side given by their diameter, so the volume of a ping pong ball, for stacking, is $\sim 0.000\,03\,\text{m}^3 = 3\times 10^{-5}\,\text{m}^3$. If we divide $100\,\text{m}^3$ by $3\times 10^{-5}\,\text{m}^3$, using scientific notation:

$$\frac{100\,\text{m}^3}{3\times 10^{-5}\,\text{m}^3} = \frac{1\times 10^2}{3\times 10^{-5}} = \frac{1}{3}\times 10^7 \sim 3\times 10^6$$

Thus, we expect to be able to fit about three million ping pong balls in a school bus.

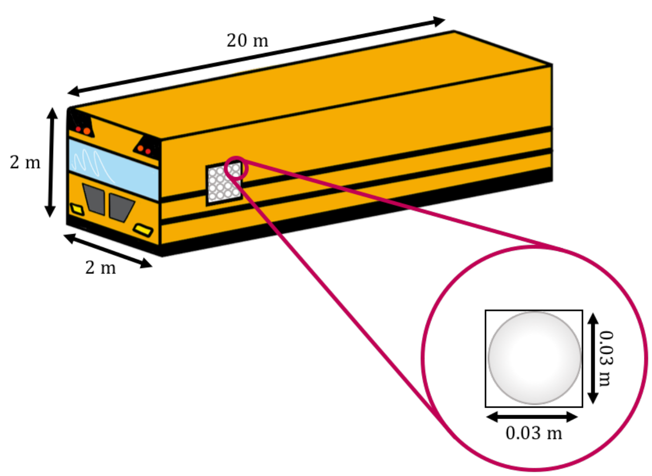


Figure 2.1: A school bus and ping pong balls modelled as boxes.

Checkpoint 2-1

Fill in the following table, giving the order of magnitude (in meters) of the sizes of different physical objects. Feel free to look these up on the internet!

Object	**Order of magnitude**
Proton	
Nucleus of atom	
Hydrogen atom	
Virus	
Human skin cell	
Width of human hair	
Human	1 m
Height of Mt. Everest	
Radius of the Earth	
Radius of the Sun	
Radius of the Milky Way	

2.2 Units and dimensions

In 1999, the NASA Mars Climate Orbiter disintegrated in the Martian atmosphere because of a mixup in the units used to calculate the thrust needed to slow the probe and place it in orbit about Mars. A computer program provided by a private manufacturer used units of pounds seconds to calculate the change in momentum of the probe instead of the Newton seconds expected by NASA. As a result, the probe was slowed down too much and disintegrated in the Martian atmosphere. This example illustrates the need for us to **use and specify units** when we describe the properties of a physical quantity, and it also demonstrates the difference between a dimension and a unit.

"Dimensions" can be thought of as types of measurements. For example, length and time are both dimensions. A unit is the standard that we choose to quantify a dimension. For example, meters and feet are both units for the dimension of length, whereas seconds and jiffys[1] are units for the dimension of time.

When we compare two numbers, for example a prediction from a model and a measurement, it is important that both quantities have the same dimension *and* be expressed in the same units.

[1]A jiffy is a unit used in electronics and generally corresponds to either 1/50 or 1/60 seconds.

Checkpoint 2-2

The speed limit on a highway...

A) has the dimension of length over time and can be expressed in units of kilometers per hour.
B) has the dimension of length can and be expressed in units of kilometers per hour.
C) has the dimension of time over length and can be expressed in units of meters per second.
D) has the dimension of time and can be expressed in units of meters.

2.2.1 Base dimensions and their SI units

In order to facilitate communication of scientific information, the International System of units (SI for the french, Système International d'unités) was developed. This allows us to use a well-defined convention for which units to use when describing quantities. For example, the SI unit for the dimension of length is the meter and the SI unit for the dimension of time is the second.

In order to simplify the SI unit system, a fundamental (base) set of dimensions was chosen and the SI units were defined for those dimensions. Any other dimension can always be re-expressed in terms of the base dimensions shown in Table 2.1 and its units in terms of the corresponding combination of the base SI units.

Dimension	**SI unit**
Length [L]	meter [m]
Time [T]	seconds[s]
Mass [M]	kilogram [kg]
Temperature [Θ]	kelvin [K]
Electric current [I]	ampère [A]
Amount of substance [N]	mole [mol]
Luminous intensity [J]	candela [cd]
Dimensionless [1]	unitless []

Table 2.1: Base dimensions and their SI units with abbreviations.

From the base dimensions, one can obtain "derived" dimensions such as "speed" which is a measure of how fast an object is moving. The dimension of speed is L/T (length over time) and the corresponding SI unit is m/s (meters per second)[2] Many of the derived dimension have corresponding derived SI units which can be expressed in terms of the base SI units.

[2]Note that we can also write meters per second as $\text{m}\cdot\text{s}^{-1}$, but we often use a divide by sign if the power of the unit in the denominator is 1.

Table 2.2 shows a few derived dimensions and their corresponding SI units and how those SI units are obtained from the base SI units.

Dimension	**SI unit**	**SI base units**
Speed [L/T]	meter per second [m/s]	[m/s]
Frequency [1/T]	hertz [Hz]	[1/s]
Force [M·L·T^{-2}]	newton [N]	[kg·m·s^{-2}]
Energy [M·L^2·T^{-2}]	joule [J]	[N·m=kg·m^2·s^{-2}]
Power [M·L^2·T^{-3}]	watt [W]	[J/s=kg·m^2·s^{-3}]
Electric Charge [I· T]	coulomb [C]	[A· s]
Voltage [M·L^2·T^{-3}·I^{-1}]	volt [V]	[J/C=kg·m^2·s^{-3}·A^{-1}]

Table 2.2: Example of derived dimensions and their SI units with abbreviations.

By convention, we can indicate the dimension of a quantity, X, by writing it in square brackets, $[X]$. For example, $[X] = I$, would mean that the quantity X has the dimension I, so it has the dimension of electric current. Similarly, we can indicate the SI units of X with $SI[X]$. Referring to Table 2.1, since X has the dimension of current, $SI[X] = A$.

2.2.2 Dimensional analysis

We call "dimensional analysis" the process of working out the dimensions of a quantity in terms of the base dimensions and a model prediction for that quantity. A few simple rules allow us to easily work out the dimensions of a derived quantity. Suppose that we have two quantities, X and Y, both with dimensions. We then have the following rules to find the dimension of a quantity that depends on X and Y:

1. Addition/Subtraction: You can only add or subtract two quantities if they have the same dimension: $[X + Y] = [X] = [Y]$
2. Multiplication: The dimension of the product, $[XY]$, is the product of the dimensions: $[XY] = [X] \cdot [Y]$
3. Division: The dimension of the ratio, $[X/Y]$, is the ratio of the dimensions: $[X/Y] = [X]/[Y]$

The next two examples show how to apply dimensional analysis to obtain the unit or dimension of a derived quantity.

Example 2-2

Acceleration has SI units of ms^{-2} and force has the dimension of mass multiplied by acceleration. What are the dimensions and SI units of force, expressed in terms of the base dimensions and units?

Solution

We can start by expressing the dimension of acceleration, since we know from its SI units that it must have the dimension of length over time squared.

$$[acceleration] = \frac{L}{T^2}$$

Since force has the dimension of mass times acceleration, we have:

$$[force] = [mass] \cdot [acceleration] = M\frac{L}{T^2}$$

and the SI units of force are thus:

$$SI[force] = \mathrm{kg} \cdot \mathrm{m/s}^2$$

Force is such a common dimension that it, like many other derived dimensions, has its own derived SI unit, the Newton [N].

Example 2-3

Use Table 2.2 to show that voltage has the same dimension as force multiplied by speed and divided by electric current.

Solution

According to Table 2.2, voltage has the dimension:

$$[voltage] = M \cdot L^2 \cdot T^{-3} \cdot I^{-1}$$

while force, speed and current have dimensions:

$$\begin{aligned} [force] &= M \cdot L \cdot T^{-2} \\ [speed] &= L \cdot T^{-1} \\ [current] &= I \end{aligned}$$

The dimension of force multiplied by speed divided by electric charge

$$\left[\frac{force \cdot speed}{current}\right] = \frac{[force] \cdot [speed]}{[current]} = \frac{M \cdot L \cdot T^{-2} \cdot L \cdot T^{-1}}{I}$$
$$= M \cdot L^2 \cdot T^{-3} \cdot I^{-1}$$

where, in the last line, we combined the powers of the same dimensions. By inspection, this is the same dimension as voltage.

When you build a model to predict the value of a physical quantity, you should always use dimensional analysis to ensure that the dimension of the quantity your model predicts is correct.

Example 2-4

Your model predicts that the speed, v, of an object of mass m, after having fallen a distance h on the surface of a planet with mass M and radius R is given by:

$$v = \frac{mMh}{R}$$

Is this a reasonable prediction?

Solution

First, we can see that the speed will be larger if h is bigger, which makes sense, since we expect the speed to be greater if the object fell a greater distance. Similarly, we expect that the speed would be higher if the mass of the planet, M, is larger, as it would exert a larger gravitational force, as given by this model. We also expect that the object will have a greater speed if it has a larger mass, m, if the drag from the atmosphere on the planet is significant. Finally, if the radius of the planet R is larger, we would expect the speed to be smaller, as the planet would be less dense and exert less gravitational force at its surface. However, if we verify the dimensions for the prediction of v, we find the model does not predict dimensions of speed:

$$[v] = \frac{[m][M][h]}{[R]}$$
$$= \frac{MML}{L} = M^2$$

and our model predicts a speed with dimensions of mass squared. By performing simple dimensional analysis, we can easily confirm that our model is definitely wrong. You should always check the dimensions of any model prediction, to make sure it is correct.

Olivia's Thoughts

In this section, we were given three rules for combining dimensions. You'll notice that these rules are the same as the rules for algebra, except you're using dimensions instead of x's and y's. So, you can really just approach dimensional analysis problems as you would algebra problems.

There are some basic steps you can follow when you are trying to find the SI units for a value/variable in your equation. I'll go through Example 2-2 in a bit of a different way. Let's say that you have the equation $F = ma$ and this time, you know the dimensions of F and m, and you want to find the dimensions of a:

1. Rewrite the values/variables in your equation in terms of their dimensions, leaving all other operations (multiplication, exponents, etc.) as is: $F = m \cdot a \rightarrow [F] = [m] \cdot [a]$
2. Rearrange for your unknown dimension: $[a] = \frac{[F]}{[m]}$
3. Substitute in your known dimensions: $[a] = \frac{[F]}{[m]} \rightarrow [a] = \frac{MLT^{-2}}{M} = \frac{ML}{MT^2}$
4. Solve using the rules of algebra: $[a] = \frac{L}{T^2}$ (where we just cancelled out the M's)
5. Replace the dimensions with their corresponding SI units: $[a] = \frac{L}{T^2} \rightarrow SI[a] = \frac{m}{s^2}$

Checkpoint 2-3

In Chloë's theory of falling objects from Chapter 1, the time, t, for an object to fall a distance, x, was given by $t = k\sqrt{x}$. What must the SI units of Chloë's constant, k, be?

A) $\mathrm{T\,L^{\frac{1}{2}}}$
B) $\mathrm{T\,L^{-\frac{1}{2}}}$
C) $\mathrm{s\,m^{\frac{1}{2}}}$
D) $\mathrm{s\,m^{-\frac{1}{2}}}$

Dimensional analysis can also be used to determine formulas (usually to within an order of magnitude). One famous example of this is when a British physicist named G.I. Taylor was able to determine a formula that showed how the blast radius of an atomic bomb scaled with time. Using pictures of the first atomic bomb explosion, he was able to determine the amount of energy released in the explosion, which was classified information at the time.

Example 2-5

Find a formula that shows how the blast radius, r, scales with the time since the explosion, t, where the radius also depends on the energy released in the explosion, E, and the density of the medium into which the bomb explodes, ρ.

Solution

We want to find out how the blast radius scales with time, so we want an expression that relates r to some combination of E, ρ, and t:

$$r \sim E^x \rho^y t^z$$

where x, y, and z are our unknown exponents, since we don't know yet how we will combine E, ρ, and t. However, we do know that when we combine these quantities, we have to get the correct dimension (length) for the radius:

$$[r] = [E]^x [\rho]^y [t]^z$$

We know the dimensions for radius and time, and the dimension for E can be found in Table 2.2. Density is mass divided by volume, so its dimension is M/L^3. Our equation then becomes:

$$\begin{aligned} L &= (ML^2T^{-2})^x (ML^{-3})^y (T)^z \\ L &= (M^x L^{2x} T^{-2x})(M^y L^{-3y})(T^z) \end{aligned}$$

We have three unknowns, so we need three equations. We can recognize that the left hand side (with dimension of length, L) is equivalent to $L^1 \cdot M^0 \cdot T^0$. We can then separate the above expression into three equations, one for each of M, L, and T:

$$\begin{aligned} M^0 &= M^x M^y \to 0 = x + y \\ L^1 &= L^{2x} L^{-3y} \to 1 = 2x - 3y \\ T^0 &= T^{-2x} T^z \to 0 = z - 2x \end{aligned}$$

Solving the sytem of equations, we find that $x = 1/5$, $y = -1/5$, and $z = 2/5$. So, the combination of E, ρ, and t that gives us the dimension of length is:

$$\begin{aligned} r \sim E^{1/5} \rho^{-1/5} t^{2/5} \\ \therefore r \propto t^{2/5} \end{aligned}$$

You can also write this equation as:

$$r \sim \sqrt[5]{\frac{Et^2}{\rho}}$$

Thus, by measuring the blast radius at some time, and knowing the density of the air, you can estimate the energy that was released during the explosion.

2.3 Making measurements

Having introduced some tools for the modelling aspect of physics, we now address the other side of physics, namely performing experiments. Since the goal of developing theories and models is to describe the real world, we need to understand how to make meaningful measurements that test our theories and models.

Suppose that we wish to test Chloë's theory of falling objects from Chapter 1:

$$t = k\sqrt{x}$$

which states that the time, t, for any object to fall a distance, x, near the surface of the Earth is given by the above relation. The theory assumes that Chloë's constant, k, is the same for any object falling any distance on the surface of the Earth.

One possible way to test Chloë's theory of falling objects is to measure k for different drop heights to see if we always obtain the same value. Results of such an experiment are presented in Table 2.3, where the time, t, was measured for a bowling ball to fall distances of x between 1 m and 5 m. The table also shows the values computed for $\sqrt{x}$ and the corresponding value of $k = t/\sqrt{x}$:

x [m]	**t** [s]	$\sqrt{x}$ [$\mathrm{m}^{\frac{1}{2}}$]	**k** [$\mathrm{s\,m}^{-\frac{1}{2}}$]
1.00	0.33	1.00	0.33
2.00	0.74	1.41	0.52
3.00	0.67	1.73	0.39
4.00	1.07	2.00	0.54
5.00	1.10	2.24	0.49

Table 2.3: Measurements of the drop times, t, for a bowling ball to fall different distances, x. We have also computed $\sqrt{x}$ and the corresponding value of k.

When looking at Table 2.3, it is clear that each drop height gave a different value of k, so at face value, we would claim that Chloë's theory is incorrect, as there does not seem to be a value of k that applies to all situations. However, we would be incorrect in doing so unless we understood *the precision of the measurements* that we made. Suppose that we **repeated** the measurement multiple times at a **fixed** drop height of $x = 3$ m, and obtained the values in Table 2.4.

x [m]	**t** [s]	$\sqrt{x}$ [$\mathrm{m}^{\frac{1}{2}}$]	**k** [$\mathrm{s\,m}^{-\frac{1}{2}}$]
3.00	1.01	1.73	0.58
3.00	0.76	1.73	0.44
3.00	0.64	1.73	0.37
3.00	0.73	1.73	0.42
3.00	0.66	1.73	0.38

Table 2.4: Repeated measurements of the drop time, t, for a bowling ball to fall a distance $x = 3\,\mathrm{m}$. We have also computed $\sqrt{x}$ and the corresponding value of k.

This simple example highlights the critical aspect of making any measurement: it is impossible to make a measurement with infinite precision. The values in Table 2.4 show that if we repeat the exact same experiment, we are likely to measure different values for a single quantity. In this case, for a fixed drop height, $x = 3\,\mathrm{m}$, we obtained a spread in values of the drop time, t, between roughly $0.6\,\mathrm{s}$ and $1.0\,\mathrm{s}$. Does this mean that it is hopeless to do science, since we can never repeat measurements? Thankfully, no! It does however require that we deal with the inherent imprecision of measurements in a formal manner.

2.3.1 Measurement uncertainties

The values in Table 2.4 show that for a fixed experimental setup (a drop height of $3\,\mathrm{m}$), we are likely to measure a spread in the values of a quantity (the time to drop). We can quantify this "uncertainty" in the value of the measured time by quoting the measured value of t by providing a "central value" and an "uncertainty":

$$t = (0.76 \pm 0.15)\,\mathrm{s}$$

where $0.76\,\mathrm{s}$ is called the "central value" and $0.15\,\mathrm{s}$ the "uncertainty" or the "error'. Note that we use the word error as a synonym for uncertainty, not "mistake". When we present a number with an uncertainty, we mean that we are "pretty certain" that the true value is in the range that we quote. In this case, the range that we quote is that t is between $0.61\,\mathrm{s}$ and $0.91\,\mathrm{s}$ (given by $0.76\,\mathrm{s}$ - $0.15\,\mathrm{s}$ and $0.76\,\mathrm{s} + 0.15\,\mathrm{s}$). When we say that we are "pretty sure" that the value is within the quoted range, we usually mean that there is a 68% chance of this being true and allow for the possibility that the true value is actually outside the range that we quoted. The value of 68% comes from statistics and the normal distribution.

Emma's Thoughts

"Precision", "Accuracy" and "Uncertainty" - what's the difference?

Have you ever started writing a lab report and wondered whether or not you should describe your measurement in terms of "accuracy" or "precision"? What about describing the error in your experiment as a measure of "accuracy" or "uncertainty"?

You're not alone! Precision, accuracy and uncertainty all relate to error, but have different meanings. To clarify these terms, I think it is useful to study them side-by-side.

Precision refers to how close your measurements are to each other when you repeat a measurement multiple times. If the values obtained are close to one another, your measurements are precise. For example, say you were measuring the rebound height of a basketball, dropped from a fixed height. After performing the measurement multiple times, you find that the measured rebound heights are very close in value to each other. You could then report that "After repeating our measurement multiple times, the values that we obtained were very close together. Our measurements were precise!". Of course, you have to specify what you mean by "close" (perhaps in terms of the divisions on the ruler that you used to measure rebound height).

Accuracy measures the agreement between a measured value and its true value. If the measured value is close to the true value, your measured value is accurate. For example, say that you developed a model for the distance covered by a rock thrown with a slingshot. If you find that the measured value is close to the predicted value, you would say that your model is accurate, "Our model value was very close to the value that we measured - our model was accurate." Again, you have to specify what you mean by "close", usually in terms of the uncertainty on your measured value.

Uncertainty is an estimate of the amount that a measurement will differ from a true value. In science, we aim to lower the uncertainty in our measurements, so that we can test models and theories with more precision. Let's say that you are measuring the number of rotations of a spinning top during a certain period of time. Your measurements are close together, but have a fixed range of values. This would be an example where you could calculate the uncertainty in your measurements. It would be sensible to say "After multiple measurements, we've found that our values are similar and our uncertainty captures the range of values that we measured."

Determining the central value and uncertainty

The tricky part when performing a measurement is to decide how to assign a central value and an uncertainty. For example, how did we come up with $t = (0.76 \pm 0.15)\,\text{s}$ from the values in Table 2.4?

Determining the uncertainty and central value on a measurement is greatly simplified when one can repeat the same measurement multiple times, as we did in Table 2.4. With repeatable measurements, a reasonable choice for the central value and uncertainty is to use the **mean** and **standard deviation** of the measurements, respectively.

If we have N measurements of some quantity t, $\{t_1, t_2, t_3, \dots t_N\}$, then the mean, $\bar{t}$, and standard deviation, σ_t, are defined as:

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{i=N} t_i = \frac{t_1 + t_2 + t_3 + \cdots + t_N}{N} \tag{2.1}$$

$$\sigma_t^2 = \frac{1}{N-1}\sum_{i=1}^{i=N}(t_i - \bar{t})^2 = \frac{(t_1-\bar{t})^2 + (t_2-\bar{t})^2 + (t_3-\bar{t})^2 + \cdots + (t_N-\bar{t})^2}{N-1} \tag{2.2}$$

$$\sigma_t = \sqrt{\sigma_t^2} \tag{2.3}$$

The mean is just the arithmetic average of the values, and the standard deviation, σ_t, requires one to first calculate the mean, then the variance (σ_t^2, the square of the standard deviation). You should also note that for the variance, we divide by $N-1$ instead of N. The standard deviation and variance are quantities that come from statistics and are a good measure of how spread out the values of t are about their mean, and are thus a good measure of the uncertainty.

Example 2-6

Calculate the mean and standard deviation of the values for k from Table 2.4.

Solution

In order to calculate the standard deviation, we first need to calculate the mean of the $N = 5$ values of k: $\{0.58, 0.44, 0.37, 0.42, 0.38\}$. The mean is given by:

$$\bar{k} = \frac{0.58 + 0.44 + 0.37 + 0.42 + 0.38}{5} = 0.44\,\mathrm{s\,m^{-\frac{1}{2}}}$$

We can now calculate the variance using the mean:

$$\begin{aligned}\sigma_k^2 = &\frac{1}{4}[(0.58-0.44)^2 + (0.44-0.44)^2 \\ &+ (0.37-0.44)^2 + (0.42-0.44)^2 + (0.38-0.44)^2] = 7.3\times10^{-3}\,\mathrm{s^2\,m}\end{aligned}$$

and the standard deviation is then given by the square root of the variance:

$$\sigma_k = \sqrt{0.0073} = 0.09\,\mathrm{s\,m^{-\frac{1}{2}}}$$

Using the mean and standard deviation, we would quote our value of k as :

$$k = (0.44 \pm 0.09)\,\mathrm{s\,m^{-\frac{1}{2}}}$$

Any value that we measure will always have an uncertainty. In the case where we can easily repeat the measurement, we should do so to evaluate how reproducible it is, and the standard deviation of those values is usually a good first estimate of the uncertainty in a value[3]. Sometimes, the measurements cannot easily be reproduced; in that case, it is still important to determine a reasonable uncertainty, but in this case, it usually has to be estimated. Table 2.6 shows a few common types of measurements and how to determine the uncertainties in those measurements.

Type of measurement	**How to determine central value and uncertainty**
Repeated measurements	Mean and standard deviation
Single measurement with a graduated scale (e.g. ruler, digital scale, analogue meter)	Closest value and half of the smallest division
Counted quantity	Counted value and square root of the value

Table 2.5: Different types of measurements and how to assign central values uncertainties.

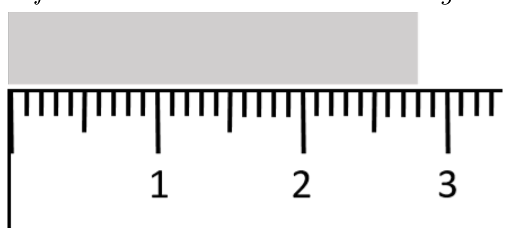


Figure 2.2: The length of the grey rectangle would be quoted as $L = (2.80 \pm 0.05)\,\mathrm{cm}$ *using the rule of "half the smallest division".*

For example, we would quote the length of the grey object in Figure 2.2 to be $L = (2.80 \pm 0.05)\,\mathrm{cm}$ based on the rules in Table 2.6, since $2.8\,\mathrm{cm}$ is the closest value on the ruler that matches the length of the object and $0.5\,\mathrm{mm}$ is half of the smallest division on the ruler. Using half of the smallest division of the ruler means that our uncertainty range covers one full division. Note that it is usually better to reproduce a measurement to evaluate the uncertainty instead of using half of the smallest division, although half of the smallest division should be the lower limit on the uncertainty. That is, by repeating the measurements and obtaining the standard deviation, you should see if the uncertainty is *larger* than half of the of the smallest division, not smaller.

The **relative uncertainty** in a measured value is given by dividing the uncertainty by the central value, and expressing the result as a percent. For example, the relative uncertainty in $t = (0.76 \pm 0.15)\,\mathrm{s}$ is given by $0.15/0.76 = 20\%$. The relative uncertainty gives an idea of how precisely a value was determined. Typically, a value above 10% means that it was not a very precise measurement, and we would generally consider a value smaller than 1% to correspond to quite a precise measurement.

Random and systematic sources of error/uncertainty

It is important to note that there are two possible sources of uncertainty in a measurement. The first is called "statistical" or "random" and occurs because it is impossible to exactly

[3]In practice, the standard deviation is an overly conservative estimate of the error and we would use the error on the mean, which is the standard deviation divided by the square root of the number of measurements.

reproduce a measurement. For example, every time you lay down a ruler to measure something, you might shift it slightly one way or the other which will affect your measurement. The important property of random sources of uncertainty is that if you reproduce the measurement many times, these will tend to cancel out and the mean can usually be determined to high precision with enough measurements.

The other source of uncertainty is called "systematic". Systematic uncertainties are much more difficult to detect and to estimate. One example would be trying to measure something with a scale that was not properly tarred (where the 0 weight was not set). You may end up with very small random errors when measuring the weights of object (very repeatable measurements), but you would have a hard time noticing that all of your weights were offset by a certain amount unless you had access to a second scale. Some common examples of systematic uncertainties are: incorrectly calibrated equipment, parallax error when measuring distance, reaction times when measuring time, effects of temperature on materials, etc.

As a reminder, we want to emphasized the difference between "error" and "mistake" in the context of making measurements. "Uncertainty" or "error" in a measurement comes from the fact that it is impossible to measure anything to infinite accuracy. A "mistake" also affects a measurement, but is preventable. If a "mistake" occurs in physics, the experiment is generally re-done and the previous data are discarded. The term "human error" should never be used in a lab report as it implies that a mistake was made. Instead, if you think that you measured time imprecisely, for example, refer to human reaction time, not "human error".

Table 2.6 shows examples of sources of error that students often call "human error" but that should be instead described more precisely.

Situation	**Source of Error**
While taking measurements, your line of sight was not completely parallel to the measuring device.	This is parallax error - a type of systematic error.
You incorrectly performed calculations.	Mistake! Redo the calculations.
A draft of wind in the lab slightly altered the direction of your ball rolling down an incline.	This is an environmental effect/error - it could be random or systematic, depending on whether it always had the same effect.
Your hand slipped while holding the ruler - the object was measured to be twice its original size!	Mistake! Redo this experiment and discard the data.
When timing an experiment, you don't hit the "STOP" button exactly when the experiment stops.	Reaction time error - usually a systematic error (time is usally measured longer than it is).

Table 2.6: Don't use the term "human error", instead, use these.

Propagating uncertainties

Going back to the data in Table 2.4, we found that for a known drop height of $x = 3\,\text{m}$, we measured different values of the drop time, which we found to be $t = (0.76 \pm 0.15)\,\text{s}$ (using the mean and standard deviation). We also calculated a value of k corresponding to each value of t, and found $k = (0.44 \pm 0.09)\,\text{s}\,\text{m}^{-\frac{1}{2}}$ (Example 2-6).

Suppose that we did not have access to the individual values of t, but only to the value of $t = (0.76 \pm 0.15)\,\text{s}$ with uncertainty. How do we calculate a value for k with uncertainty? In order to answer this question, we need to know how to "propagate" the uncertainties in a measured value to the uncertainty in a value derived the measured value. We briefly present different methods for propagating uncertainties, before advocating for the use of computers to do the calculations for you.

1. Estimate using relative uncertainties

The relative uncertainty in a measurement gives us an idea of how precisely a value was determined. Any quantity that depends on that measurement should have a precision that is similar; that is, we expect the relative uncertainty in k to be similar to that in t. For t, we saw that the relative uncertainty was approximately 20%. If we take the central value of k to be the central value of t divided by $\sqrt{x}$, we find:

$$k = \frac{(0.76\,\text{s})}{\sqrt{(3\,\text{m})}} = 0.44\,\text{s}\,\text{m}^{-\frac{1}{2}}$$

Since we expect the relative uncertainty in k to be approximately 20%, then the absolute uncertainty is given by:

$$\sigma_k = (0.2)k = 0.09\,\text{s}\,\text{m}^{-\frac{1}{2}}$$

which is close to the value obtained by averaging the five values of k in Table 2.4.

2. The Min-Max method

A pedagogical way to determine k and its uncertainty is to use the "Min-Max method". Since $k = t/\sqrt{x}$, k will be the biggest when t is the biggest, and the smallest when t is the smallest. We can thus determine "minimum" and "maximum" values of k corresponding to the minimum value of t, $t^{min} = 0.61\,\mathrm{s}$ and the maximum value of t, $t^{max} = 0.91\,\mathrm{s}$:

$$\begin{aligned} k^{min} &= \frac{t^{min}}{\sqrt{x}} = \frac{0.61\,s}{\sqrt{(3\,m)}} = 0.35\,\mathrm{s\,m^{-\frac{1}{2}}} \\ k^{max} &= \frac{t^{max}}{\sqrt{x}} = \frac{0.91\,s}{\sqrt{(3\,m)}} = 0.53\,\mathrm{s\,m^{-\frac{1}{2}}} \end{aligned}$$

This gives us the range of values of k that correspond to the range of values of t. We can choose the middle of the range as the central value of k and half of the range as the uncertainty:

$$\begin{aligned} \bar{k} &= \frac{1}{2}(k^{min} + k^{max}) = 0.44\,\mathrm{s\,m^{-\frac{1}{2}}} \\ \sigma_k &= \frac{1}{2}(k^{max} - k^{min}) = 0.09\,\mathrm{s\,m^{-\frac{1}{2}}} \\ \therefore k &= (0.44 \pm 0.09)\,\mathrm{s\,m^{-\frac{1}{2}}} \end{aligned}$$

which, in this case, gives the same value as that obtained by averaging the individual values of k. While the Min-Max method is useful for illustrating the concept of propagating uncertainties, we usually do not use it in practice as it tends to overestimate the uncertainty.

3. The derivative method

In the example above, we assumed that the value of x was known precisely (and we chose $3\,\mathrm{m}$), which of course is not realistic. Let us suppose that we have measured x to within $1\,\mathrm{cm}$ so that $x = (3.00 \pm 0.01)\,\mathrm{m}$. The task is now to calculate $k = \frac{t}{\sqrt{x}}$ when both x and t have uncertainties.

The derivative method lets us propagate the uncertainty in a general way, so long as the relative uncertainties on all quantities are "small" (less than 10-20%). If we have a function, $F(x, y)$ that depends on multiple variables with uncertainties (e.g. $x \pm \sigma_x$, $y \pm \sigma_y$), then the central value and uncertainty in $F(x, y)$ are given by:

$$\begin{aligned} \bar{F} &= F(\bar{x}, \bar{y}) \\ \sigma_F &= \sqrt{\left(\frac{\partial F}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial F}{\partial y}\sigma_y\right)^2} \end{aligned} \tag{2.4}$$

That is, the central value of the function F is found by evaluating the function at the central values of x and y. The uncertainty in F, σ_F, is found by taking the quadrature sum of the partial derivatives of F evaluated at the central values of x and y multiplied by the uncertainties in the corresponding variables that F depends on. The uncertainty will contain one term in the sum per variable that F depends on.

In appendix D, we will show you how to calculate this easily with a computer, so do not worry about getting comfortable with partial derivatives (yet!). Note that the partial derivative, $\frac{\partial F}{\partial x}$, is simply the derivative of $F(x, y)$ relative to x evaluated as if y were a constant. Also, when we say "add in quadrature", we mean square the quantities, add them, and then take the square root (same as you would do to calculate the hypotenuse of a right-angle triangle).

Example 2-7

Use the derivative method to evaluate $k = \frac{t}{\sqrt{x}}$ for $x = (3.00 \pm 0.01)\,\mathrm{m}$ and $t = (0.76 \pm 0.15)\,\mathrm{s}$.

Solution

Here, $k = k(x, t)$ is a function of both x and t. The central value is easily found using the central values for x and t:

$$\bar{k} = \frac{t}{\sqrt{x}} = \frac{(0.76\,\mathrm{s})}{\sqrt{(3\,\mathrm{m})}} = 0.44\,\mathrm{s\,m}^{-\frac{1}{2}}$$

Next, we need to determine and evaluate the partial derivative of k with respect to t and x:

$$\begin{aligned}
\frac{\partial k}{\partial t} &= \frac{1}{\sqrt{x}}\frac{d}{dt}t = \frac{1}{\sqrt{x}} = \frac{1}{\sqrt{(3\,\mathrm{m})}} = 0.58\,\mathrm{m}^{-\frac{1}{2}} \\
\frac{\partial k}{\partial x} &= t\frac{d}{dx}x^{-\frac{1}{2}} = -\frac{1}{2}tx^{-\frac{3}{2}} = -\frac{1}{2}(0.76\,\mathrm{s})(3.00\,\mathrm{m})^{-\frac{3}{2}} = -0.073\,\mathrm{s\,m}^{-\frac{3}{2}}
\end{aligned}$$

And finally, we plug this into the quadrature sum to get the uncertainty in k:

$$\begin{aligned}
\sigma_k &= \sqrt{\left(\frac{\partial k}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial k}{\partial t}\sigma_t\right)^2} \\
&= \sqrt{\left((0.073\,\mathrm{s\,m}^{-\frac{3}{2}})(0.01\,\mathrm{m})\right)^2 + \left((0.58\,\mathrm{m}^{-\frac{1}{2}})(0.15\,\mathrm{s})\right)^2} \\
&= 0.09\,\mathrm{s\,m}^{-\frac{1}{2}}
\end{aligned}$$

So we find that:

$$k = (0.44 \pm 0.09)\,\mathrm{s\,m^{-\frac{1}{2}}}$$

which is consistent with what we found with the other two methods.

Discussion: We should ask ourselves if the value we found is reasonable, since we also included an uncertainty in x and would expect a bigger uncertainty than in the previous calculations where we only had an uncertainty in t. The reason that the uncertainty in k has remained the same is that the relative uncertainty in x is very small, $\frac{0.01}{3.00} \sim 0.3\%$, so it contributes very little compared to the 20% uncertainty from t.

The derivative method leads to a few simple short cuts when propagating the uncertainties for simple operations, as shown in Table 2.9. A few rules to note:

1. Uncertainties should be combined in quadrature
2. For addition and subtraction, add the absolute uncertainties in quadrature
3. For multiplication and division, add the relative uncertainties in quadrature

Operation to get z	**Uncertainty in z**
$z = x + y$ (addition)	$\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$
$z = x - y$ (subtraction)	$\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$
$z = xy$ (multiplication)	$\sigma_z = xy\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$
$z = \frac{x}{y}$ (division)	$\sigma_z = \frac{x}{y}\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$
$z = f(x)$ (a function of 1 variable)	$\sigma_z = \left\lvert\frac{df}{dx}\sigma_x\right\rvert$

Table 2.7: How to propagate uncertainties from measured values $x \pm \sigma_x$ and $y \pm \sigma_y$ to a quantity $z(x, y)$ for common operations.

Checkpoint 2-4

We have measured that a llama can cover a distance of $(20.0 \pm 0.5)\,\mathrm{m}$ in $(4.0 \pm 0.5)\,\mathrm{s}$. What is the speed (with uncertainty) of the llama?

2.3.2 Using graphs to visualize and analyse data

Table 2.8 below reproduces our measurements of how long it took (t) for an object to drop a certain distance, x. Chloë's Theory of gravity predicted that the data should be described by the following model:

$$t = k\sqrt{x}$$

where k was an undetermined constant of proportionality.

x [m]	**t** [s]	$\sqrt{x}$ [$\mathrm{m}^{\frac{1}{2}}$]	**k** [$\mathrm{s\,m}^{-\frac{1}{2}}$]
1.00	0.33	1.00	0.33
2.00	0.74	1.41	0.52
3.00	0.67	1.73	0.39
4.00	1.07	2.00	0.54
5.00	1.10	2.24	0.49

Table 2.8: Measurements of the drop times, t, for a bowling ball to fall different distances, x. We have also computed $\sqrt{x}$ and the corresponding value of k.

The easiest way to visualize and analyse these data is to plot them on a graph. In particular, if we plot (graph) t versus $\sqrt{x}$, we expect that the points will fall on a straight line that goes through zero, with a slope of k (if the data are described by Chloë's Theory). In Appendix D, we show you how you can easily plot these data using the Python programming language as well as find the slope and offset of the line that best fits the data, as show in Figure 2.3.

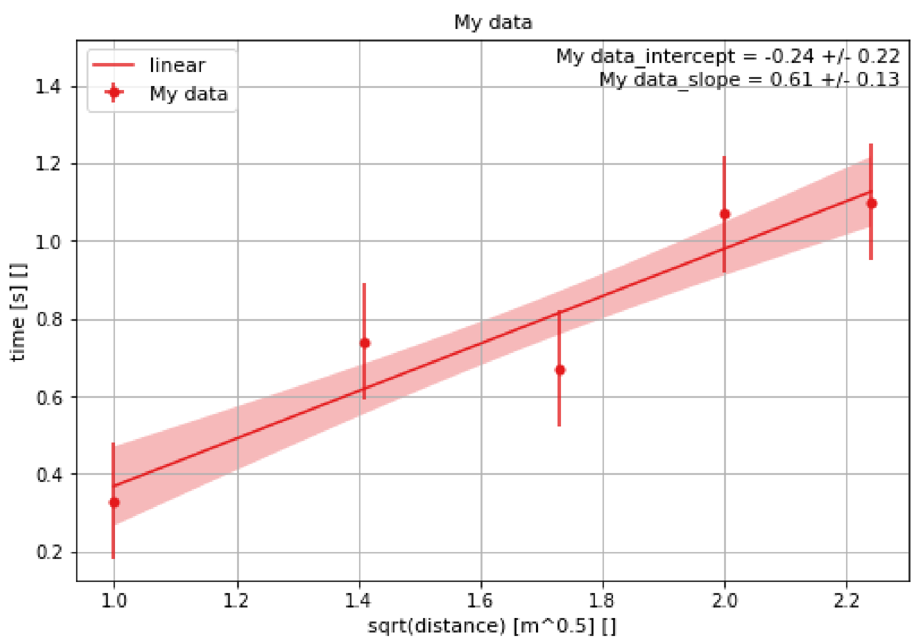


Figure 2.3: Graph of t versus $\sqrt{x}$ and line of best fit.

When plotting data and fitting them to a line (or other function), it is important to make sure that the values have at least an uncertainty in the quantity that is being plotted on the y axis. In this case, we have assumed that all of the measurements of time have an uncertainty of $0.15\,\mathrm{s}$ and that the measurements of the distance have no (or negligible) uncertainties.

Since we expect the slope of the data to be k, finding the line of best fit provides us a with method to determine k by using all of the data points. In this case, we find that $k = (0.61 \pm 0.13)\,\mathrm{s\,m^{-\frac{1}{2}}}$. **Performing a linear fit of the data is the best way to determine a constant of proportionality between the measurements**. Note that we expect the intercept to be equal to zero according to our model, but the best fit line has an intercept of $(-0.24 \pm 0.22)\,\mathrm{s}$, which is slightly below, but consistent, with zero. From these data, we would conclude that our measurements are consistent with Chloë's Theory. Again, remember that we can never confirm a theory, we can only exclude it; in this case, we cannot exclude Chloë's Theory.

2.3.3 Reporting measured values

Now that you know how to attribute an uncertainty to a measured quantity and then propagate that uncertainty to a derived quantity, you are ready to present your measurement to the world. In order to conduct "good science", your measurements should be reproducible, clearly presented, and precisely described. Here are general rules to follow when reporting a measured number:

1. Indicate the units, preferably SI units (use derived SI units, such as newtons, when appropriate).
2. Include a description of how the uncertainty was determined (if it is a direct measurement, how did you choose the uncertainty? If it is a derived quantity, how did you propagate the uncertainty?).
3. Show no more than 2 "significant digits"[4] in the uncertainty and format the central value to the same decimal as the uncertainty.
4. Use scientific notation when appropriate (usually numbers bigger than 1000 or smaller than 0.01).
5. Factor out the power 10 from the central value and uncertainty (e.g. $(10\,123 \pm 310)\,\mathrm{m}$ would be better presented as $(10.12 \pm 0.31) \times 10^3\,\mathrm{m}$ or $(101.2 \pm 3.1) \times 10^2\,\mathrm{m}$).

Checkpoint 2-5

Someone has measured the average height of tables in the laboratory to be $1.0535\,\mathrm{m}$ with a standard deviation of $0.0525\,\mathrm{m}$. What is the best way to present this measurement?

A) $(1.0535 \pm 0.0525)\,\mathrm{m}$
B) $(1.054 \pm 0.053)\,\mathrm{m}$
C) $(105.4 \pm 5.3) \times 10^{-2}\,\mathrm{m}$
D) $(105.35 \pm 5.25)\,\mathrm{cm}$

2.3.4 Comparing model and measurement - discussing a result

In order to advance science, we make measurements and compare them to a theory or model prediction. We thus need a precise and consistent way to compare measurements with each other and with predictions. Suppose that we have measured a value for Chloë's constant $k = (0.44 \pm 0.09)\,\mathrm{s\,m^{-\frac{1}{2}}}$. Of course, Chloë's theory does not predict a value for k, only that fall time is proportional to the square root of the distance fallen. Isaac Newton's Universal

[4] Significant digits are those excluding leading and trailing zeroes.

Theory of Gravity does predict a value for k of $0.45\,\mathrm{s\,m^{-\frac{1}{2}}}$ with negligible uncertainty. In this case, since the model (theoretical) value easily falls within the range given by our uncertainty, we would say that our measurement is consistent (or compatible) with the theoretical prediction.

Suppose that, instead, we had measured $k = (0.55 \pm 0.08)\,\mathrm{s\,m^{-\frac{1}{2}}}$ so that the lowest value compatible with our measurement, $k = 0.55\,\mathrm{s\,m^{-\frac{1}{2}}} - 0.08\,\mathrm{s\,m^{-\frac{1}{2}}} = 0.47\,\mathrm{s\,m^{-\frac{1}{2}}}$, is not compatible with Newton's prediction. Would we conclude that our measurement invalidates Newton's theory? The answer is: it depends... What "it depends on" should always be discussed any time that you present a measurement (even if it happened that your measurement is compatible with a prediction - maybe that was a fluke). Below, we list a few common points that should be addressed when presenting a measurement that will guide you into deciding whether your measurement is consistent with a prediction:

- How was the uncertainty determined and/or propagated? Was this reasonable?
- Are there systematic effects that were not taken into account when determining the uncertainty? (e.g. reaction time, parallax, something difficult to reproduce).
- Are the relative uncertainties reasonable based on the precision that you would reasonable expect?
- What assumptions were made in calculating your measured value?
- What assumptions were made in determining the model prediction?

In the above, our value of $k = (0.55 \pm 0.08)\,\mathrm{s\,m^{-\frac{1}{2}}}$ is the result of propagating the uncertainty in t which was found by using the standard deviation of the values of t. It is thus conceivable that the true value of t, and therefore of k, is outside the range that we quote. Since our value of k is still quite close to the theoretical value, we would not claim to have invalidated Newton's theory with this measurement. Our uncertainty in k is $\sigma_k = 0.08\,\mathrm{s\,m^{-\frac{1}{2}}}$, and the difference between our measured and the theoretical value is only $1.25\sigma_k$, so very close to the value of the uncertainty.

In a similar way, we would discuss whether two different measurements, each with an uncertainty, are compatible. If the ranges given by uncertainties in two values overlap, then they are clearly consistent and compatible. If, on the other hand, the ranges do not overlap, they could be inconsistent or the discrepancy might instead be the result of how the uncertainties were determined and the measurements could still be considered consistent.

2.4 Summary

Key Takeaways

Measurable quantities have dimensions and units. A physical quantity should always be reported with units, preferably SI units.

When you build a model to predict a physical quantity, you should always ask if the prediction makes sense (Does it have a reasonable order of magnitude? Does it have the right dimensions?).

Any quantity that you measure will have an uncertainty. Almost any quantity that you determine from a model or theory will also have an uncertainty.

The best way to determine an uncertainty is to repeat the measurement and use the mean and standard deviation of the measurements as the central value and uncertainty. If we have N measurements of some quantity t, $\{t_1, t_2, t_3, \dots t_N\}$, then the mean, $\bar{t}$, and standard deviation, σ_t, are defined as:

$$\begin{aligned}
\bar{t} &= \frac{1}{N}\sum_{i=1}^{i=N} t_i = \frac{t_1 + t_2 + t_3 + \dots + t_N}{N} \\
\sigma_t^2 &= \frac{1}{N-1}\sum_{i=1}^{i=N} (t_i - \bar{t})^2 = \frac{(t_1 - \bar{t})^2 + (t_2 - \bar{t})^2 + (t_3 - \bar{t})^2 + \dots + (t_N - \bar{t})^2}{N-1} \\
\sigma_t &= \sqrt{\sigma_t^2}
\end{aligned}$$

You have to pay special attention to systematic uncertainties, which are difficult to determine. You should always think of ways that your measured values could be wrong, even after repeated measurements. Relative uncertainties tell you whether your measurement is precise.

There are multiple ways to propagate uncertainties. You can estimate the uncertainty using relative uncertainties or use the Min-Max method, which tends to overestimate the uncertainties. The preferred way to propagate uncertainties is with the derivative method, which you can use so long as the relative uncertainties on the measurements are small. If we have a function, $F(x, y)$ that depends on multiple variables with uncertainties (e.g. $x \pm \sigma_x$, $y \pm \sigma_y$), then the central value and uncertainty in $F(x, y)$ are given by:

$$\begin{aligned}
\bar{F} &= F(\bar{x}, \bar{y}) \\
\sigma_F &= \sqrt{\left(\frac{\partial F}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial F}{\partial y}\sigma_y\right)^2}
\end{aligned}$$

This can be easily calculated using a computer.

If you expect two measured quantities to be linearly related (one is proportional to the other), plot them to find out! Use a computer to do so!

Important Equations

Central value and uncertainty:

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{i=N} t_i = \frac{t_1 + t_2 + t_3 + \cdots + t_N}{N}$$
$$\sigma_t^2 = \frac{1}{N-1}\sum_{i=1}^{i=N}(t_i - \bar{t})^2 = \frac{(t_1-\bar{t})^2 + (t_2-\bar{t})^2 + (t_3-\bar{t})^2 + \cdots + (t_N-\bar{t})^2}{N-1}$$
$$\sigma_t = \sqrt{\sigma_t^2}$$

Derivative method:

$$\bar{F} = F(\bar{x}, \bar{y})$$
$$\sigma_F = \sqrt{\left(\frac{\partial F}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial F}{\partial y}\sigma_y\right)^2}$$

Operation to get z	Uncertainty in z
$z = x + y$ (addition)	$\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$
$z = x - y$ (subtraction)	$\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}$
$z = xy$ (multiplication)	$\sigma_z = xy\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$
$z = \frac{x}{y}$ (division)	$\sigma_z = \frac{x}{y}\sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$
$z = f(x)$ (a function of 1 variable)	$\sigma_z = \left\lvert\frac{df}{dx}\sigma_x\right\rvert$

Table 2.9: How to propagate uncertainties from measured values $x \pm \sigma_x$ and $y \pm \sigma_y$ to a quantity $z(x, y)$ for common operations.

2.5 Thinking about the material

Reflect and research

1. Often, physicists will report a measured number with a "standard" uncertainty and indicate that there is a 68% that the true value lies within the range covered by the uncertainty. Where does the number 68% come from?
2. Why can the derivative method only be used when the relative uncertainties are small?
3. How would you estimate the height of a tall building?

Experiments to try at home

1. Estimate the volume of your room, and how many people could be piled into the room. State your assumptions and how you determined the values.

Experiments to try in the lab

1. Newton's Universal Theory of gravity predicts that the distance, x, covered by an object that has fallen for a length of time, t, is given by:

$$x = \frac{1}{2}gt^2$$

Determine the value of g (with uncertainty) by performing an experiment that will allow you to determine g by determining the slope of a line of best fit.

2.6 Sample problems and solutions

2.6.1 Problems

Problem 2-1: During a physics lecture, you look under your seat and find a sheet containing data from an experiment on throwing balls vertically (perhaps a juggling experiment). The following equation is shown at the bottom of the sheet:

$$= \frac{v_2^2 - v_1^2}{2a}$$

along with the following description:

- v_1 = initial measured velocity of the ball m/s - various measurements.
- v_2 = final measured velocity of the ball m/s - seems to be zero every time.
- a = acceleration of the ball ($-9.8\,\text{m/s}^2$).

Unfortunately, the students spilled ketchup on the left hand side of their equation, making it illegible. Luckily, you are proficient in dimensional analysis. What were the students trying to calculate, based on this model? (Solution)

Problem 2-2: Chelsea is preparing meticulously for her upcoming trip to Europe. Being a self-proclaimed "shop-a-holic" and physics lover, she wants to figure out how many pairs of shoes she can buy on vacation that will physically fit in her closet. Her closet is a walk-in closet with two entrance doors. Estimate the number of pairs of shoes that can fit in Chelsea's closet. (Solution)

2.6.2 Solutions

Solution to problem 2-1: We can use their equation to determine the dimension of the quantity on the left hand side:

$$[?] = \frac{[v_2^2] - [v_1^2]}{[a]} = \frac{\frac{L}{T}^2 - \frac{L}{T}^2}{\frac{L}{T^2}} = L$$

Thus, the dimension of the unknown quantity is length. Given the context, they were likely attempting to model the height at which a vertically thrown ball would travel before stopping.

Solution to problem 2-2: We start by estimating the volume of Chelsea's closet as well as that of a pair of shoes. Chelsea's closet is a "walk-in closet' with two double doors. If we know the dimensions of the door, we can estimate the width and height of the closet. Estimating the average size of a large door to be $1\,\text{m} \times 2\,\text{m}$, one face of the close will have an area of $4\,\text{m}^2$. If we estimate the depth of Chelsea's closet to be about $3\,\text{m}$, the volume of her closet is $12\,\text{m}^3$

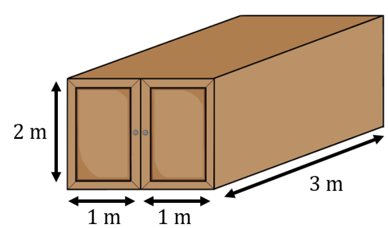


Figure 2.4: Chelsea's closet.

Next, we can estimate the size of an average pair of shoes, by modelling a shoe as a rectangular box. A single shoe has a height and width of about $5\,\text{cm}$ and a length of about $25\,\text{cm}$. A pair of shoes will thus be equivalent to box with dimensions $5\,\text{cm} \times 10\,\text{cm} \times 25\,\text{cm} = 1250\,\text{cm}^3$. This is equivalent to $0.001\,25\,\text{m}^3$. We can now determine how many pairs of shoes, N, would fit in the closet:

$$N = \frac{(12\,\text{m}^3)}{(0.001\,25\,\text{m}^3)} = 9600 \approx 10,000$$

We find that Chelsea can buy about 10,000 new pairs of shoes on her trip, and still fit them all into her closet. Time to get shopping, Chelsea!

3 Describing motion in one dimension

In this chapter, we will introduce the tools required to describe motion in one dimension. In later chapters, we will use the theories of physics to model the motion of objects, but first, we need to make sure that we have the tools to describe the motion. We generally use the word "kinematics" to label the tools for describing motion (e.g. speed, acceleration, position, etc), whereas we refer to "dynamics" when we use the laws of physics to predict motion (e.g. what motion will occur if a force is applied to an object).

Learning Objectives

- Describe motion in 1D using functions and defining an axis.
- Define position, velocity, speed, and acceleration.
- Use calculus to describe motion.
- Be able to describe motion in different frames of reference.

Think About It

You throw a ball upwards with an initial speed v. Assume there is no air resistance. When you catch the ball, its speed will be...

A) greater than v.
B) equal to v.
C) less than v.
D) in the opposite direction.

The most simple type of motion to describe is that of a particle that is constrained to move along a straight line (one-dimensional motion); much like a train along a straight piece of track. When we say that we want to describe the motion of the particle (or train), what we mean is that we want to be able to say where it is at what time. Formally, we want to know the particle's **position as a function of time**, which we will label as $x(t)$. The function will only be meaningful if:

- we specify an x-axis and the direction that corresponds to increasing values of x
- we specify an origin where $x = 0$
- we specify the units for the quantity, x.

That is, unless all of these are specified, you would have a hard time describing the motion of an object to one of your friends over the phone.

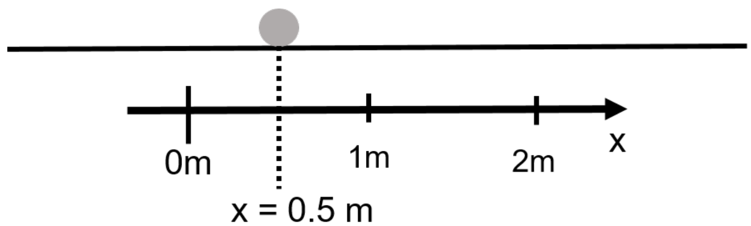


Figure 3.1: In order to describe the motion of the grey ball along a straight line, we introduce the x-axis, represented by an arrow to indicate the direction of increasing x, and the location of the origin, where $x = 0\,\text{m}$. Given our choice of origin, the ball is currently at a position of $x = 0.5\,\text{m}$.

Consider Figure 3.1 where we would like to describe the motion of the grey ball as it moves along a straight line. In order to quantify where the ball is, we introduce the "x-axis", illustrated by the black arrow. The direction of the arrow corresponds to the direction where x increases (i.e. becomes more positive). We have also chosen a point where $x = 0$, and by convention, we choose to express x in units of meters (the S.I. unit for the dimension of length).

Note that we are completely free to choose both the direction of the x-axis and the location of the origin. The x-axis is a mathematical construct that we introduce in order to describe the physical world; we could just as easily have chosen for it to point in the opposite direction with a different origin. Since we are completely free to choose where we define the x-axis, we should choose the option that is most convenient to us.

3.1 Motion with constant speed

Now suppose that the ball in Figure 3.1 is rolling, and that we recorded its x position every second in a table and obtained the values in Table 3.1 (we will ignore measurement uncertainties and pretend that the values are exact).

Time [s]	**X position [m]**
0.0 s	0.5 m
1.0 s	1.0 m
2.0 s	1.5 m
3.0 s	2.0 m
4.0 s	2.5 m
5.0 s	3.0 m
6.0 s	3.5 m
7.0 s	4.0 m
8.0 s	4.5 m
9.0 s	5.0 m

Table 3.1: Position of a ball along the x-axis recorded every second.

The easiest way to visualize the values in the table is to plot them on a graph, as in Figure 3.2. Plotting position as a function of time is one of the most common graphs to make in physics, since it is often a complete description of the motion of an object.

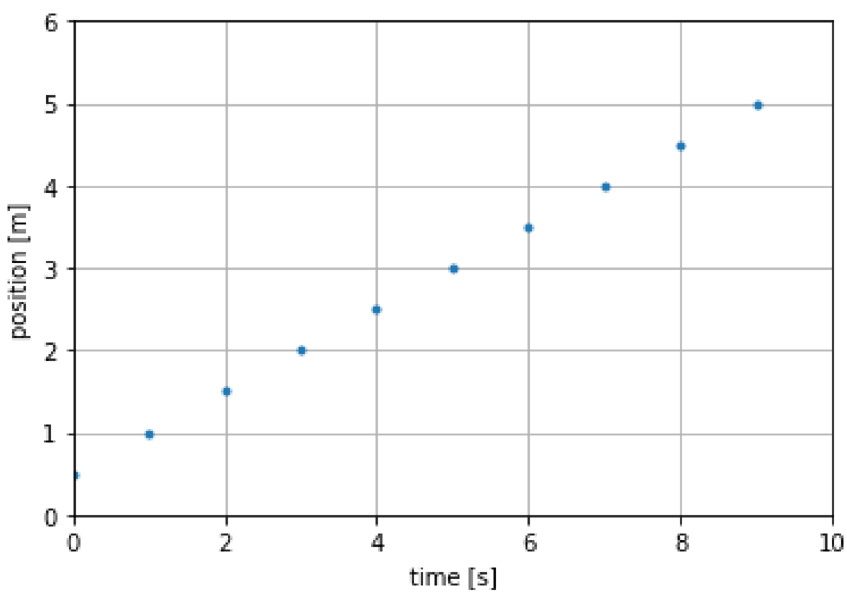


Figure 3.2: Plot of position as a function of time using the values from Table 3.1.

The data plotted in Figure 3.2 show that the x position of the ball increases linearly with time (i.e. it is a straight line and the position increases at a constant rate). This means that in equal time increments, the ball will cover equal distances. Note that we also had the liberty to choose when we define $t = 0$; in this case, we chose that time is zero when the ball is at $x = 0.5\,\text{m}$.

Checkpoint 3-1

Using the data from Table 3.1, at what position along the x-axis will the ball be when time is $t = 9.5\,\text{s}$, if it continues its motion undisturbed?

A) $5.0\,\text{m}$
B) $5.25\,\text{m}$
C) $5.75\,\text{m}$
D) $6.0\,\text{m}$

Since the position as a function of time for the ball plotted in Figure 3.2 is linear, we can summarize our description of the motion using a function, $x(t)$, instead of having to tabulate the values as we did in Table 3.1. The function will have the functional form:

$$\boxed{x(t) = x_0 + v_x t} \tag{3.1}$$

The constant x_0 is the "offset" of the function; the value that the function has at $t = 0\,\text{s}$. We call x_0 the "initial position" of the object (its position at $t = 0$). The constant v_x is the "slope" of the function and gives the rate of change of the position as a function of time. We call v_x the "velocity" of the object.

The initial position is simply the value of the position at $t = 0$, and is given from the table as:

$$x_0 = 0.5\,\text{m}$$

The velocity, v_x, is simply the difference in position, Δx, between any two points divided by the amount of time, Δt, that it took the object to move between those to points ("rise over run" for the graph of $x(t)$):

$$v = \frac{\Delta x}{\Delta t}$$

By looking at any two rows from Table 3.1, we can see that the object travels a distance $\Delta x = 0.5\,\text{m}$ in a time $\Delta t = 1\,\text{s}$. Its velocity is thus:

$$v = \frac{\Delta x}{\Delta t} = \frac{(0.5\,\text{m})}{(1\,\text{s})} = 0.5\,\text{m/s}$$

The position of the object as a function of time is thus

$$x(t) = (0.5\,\text{m}) + (0.5\,\text{m/s})t$$

If v_x is large, then the object covers more distance in a given time, i.e. it moves faster. If v_x is a negative number, then the object moves in the negative x direction. The **speed** of the object is the absolute value of its velocity. Thus objects moving in different directions will have different velocities, but can have the same speed if they cover the same amount of distance in the same amount of time.

Checkpoint 3-2

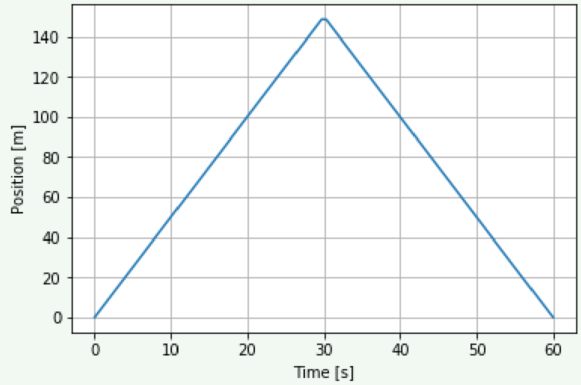


Figure 3.3: Position as a function of time for an object.

Referring to Figure 3.3, what can you say about the motion of the object?

A) The object moved faster and faster between $t = 0\,\text{s}$ and $t = 30\,\text{s}$, then slowed down to a stop at $t = 60\,\text{s}$.
B) The object moved in the positive x-direction between $t = 0\,\text{s}$ and $t = 30\,\text{s}$, and then turned around and moved in the negative x-direction between $t = 30\,\text{s}$ and $t = 60\,\text{s}$.
C) The object moved faster between $t = 0\,\text{s}$ and $t = 30\,\text{s}$ than it did between $t = 30\,\text{s}$ and $t = 60\,\text{s}$.

Checkpoint 3-3

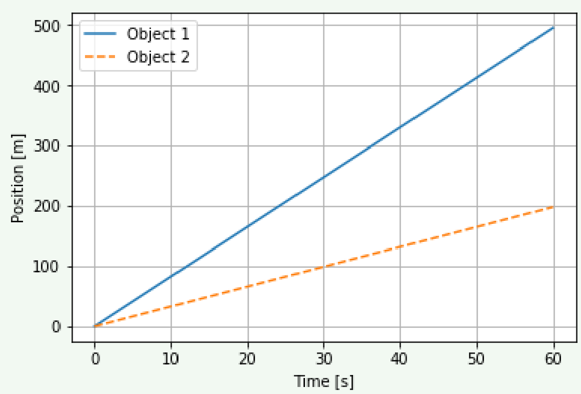


Figure 3.4: Positions as a function of time for two objects.

Referring to Figure 3.4, what can you say about the motion of the two objects?

A) Object 1 is slower than Object 2
B) Object 1 is more than twice as fast as Object 2
C) Object 1 is less than twice as fast as Object 2

3.2 Motion with constant acceleration

Until now, we have considered motion where the velocity is a constant (i.e. where velocity does not change with time and the position of an object is a linear function of time). Suppose that we wish to describe the position of a falling object that we released from rest at time $t = 0\,\text{s}$. The object will start with a velocity of $0\,\text{m/s}$ and it will **accelerate** as it falls. We say that an object is "accelerating" if its velocity is not constant. As we will see in later chapters, objects that fall near the surface of the Earth experience a constant acceleration (their velocity changes at a constant rate).

Formally, we define acceleration as the rate of change of velocity. Recall that velocity is the rate of change of position, so acceleration is to velocity what velocity is to position. In particular, we saw that if the velocity, v_x, is constant, then position as a function of time is given by:

$$x(t) = x_0 + v_x t \tag{3.1}$$

In analogy, if the acceleration is constant, then the velocity as a function of time is given by:

$$\boxed{v_x(t) = v_{0x} + a_x t} \tag{3.2}$$

where a_x is the "acceleration" and v_{0x} is the velocity of the object at $t = 0$. We can work out the dimensions of acceleration for this equation to make sense. Since we are adding v_{0x} and $a_x t$, we need the dimensions of $a_x t$ to be velocity:

$$\begin{aligned}
[a_x t] &= \frac{L}{T} \\
[a_x] &= \frac{L}{T^2}
\end{aligned}$$

Acceleration thus has dimensions of length over time squared, with corresponding S.I. units of m/s^2 (meters per second squared or meters per second per second). In order to describe the position of an object that is accelerating, we cannot use equation 3.1, since it is only correct if the velocity is constant.

In Section 3.3.2, we will show that the position as a function of time, $x(t)$, of an object with **constant acceleration**, a_x, is given by:

$$\boxed{x(t) = x_0 + v_{0x} t + \frac{1}{2} a_x t^2} \tag{3.3}$$

where, at $t = 0$, the object was at position $x = x_0$ and had a velocity v_{0x}.

Example 3-1

A ball is thrown upwards with a velocity of $10\,\mathrm{m/s}$. After what distance will the ball stop before falling back down? Assume that gravity causes a constant downwards acceleration of $9.8\,\mathrm{m/s^2}$.

Solution

We will solve this problem in the following steps:

1. Setup a coordinate system (define the x-axis).
2. Identify the condition that corresponds to the ball stopping its upwards motion and falling back down.
3. Determine the distance at which the ball stopped.

Since we throw the ball upwards with an initial velocity upwards, it makes sense to choose an x-axis that points up and has the origin at the point where we release the ball. With this choice, referring to the variables in equation 3.3, we have:

$$\begin{aligned} x_0 &= 0 \\ v_{0x} &= +10\,\mathrm{m/s} \\ a_x &= -9.8\,\mathrm{m/s^2} \end{aligned}$$

where the initial velocity is in the positive x-direction, and the acceleration, a_x, is in the negative direction (the velocity will be getting smaller and smaller, so its rate of change is negative).

The condition for the ball to stop at the top of the trajectory is that its velocity will be zero (that is what it means to stop). We can use equation 3.2 to find what time that corresponds to:

$$\begin{aligned} v(t) &= v_{0x} + a_x t \\ 0 &= (10\,\mathrm{m/s}) + (-9.8\,\mathrm{m/s^2})t \\ \therefore t &= \frac{(10\,\mathrm{m/s})}{(9.8\,\mathrm{m/s^2})} = 1.02\,\mathrm{s} \end{aligned}$$

Now that we know that it took $1.02\,\mathrm{s}$ to reach the top of the trajectory, we can find how much distance was covered:

$$\begin{aligned} x(t) &= x_0 + v_{0x}t + \frac{1}{2}a_x t^2 \\ x &= (0\,\mathrm{m}) + (10\,\mathrm{m/s})(1.02\,\mathrm{s}) + \frac{1}{2}(-9.8\,\mathrm{m/s^2})(1.02\,\mathrm{s})^2 = 5.10\,\mathrm{m} \end{aligned}$$

and we find that the ball will rise by $5.10\,\mathrm{m}$ before falling back down.

3.2.1 Visualizing motion with constant acceleration

When an object has a constant acceleration, its velocity and position as a function of time are described by the two following equations:

$$v(t) = v_{0x} + a_x t$$
$$x(t) = x_0 + v_{0x} t + \frac{1}{2} a_x t^2$$

where the velocity changes linearly with time, and the position changes quadratically with time (it goes as t^2). Figure 3.5 shows the position and the speed as a function of time for the ball from Example 3-1 for the first three seconds of the motion.

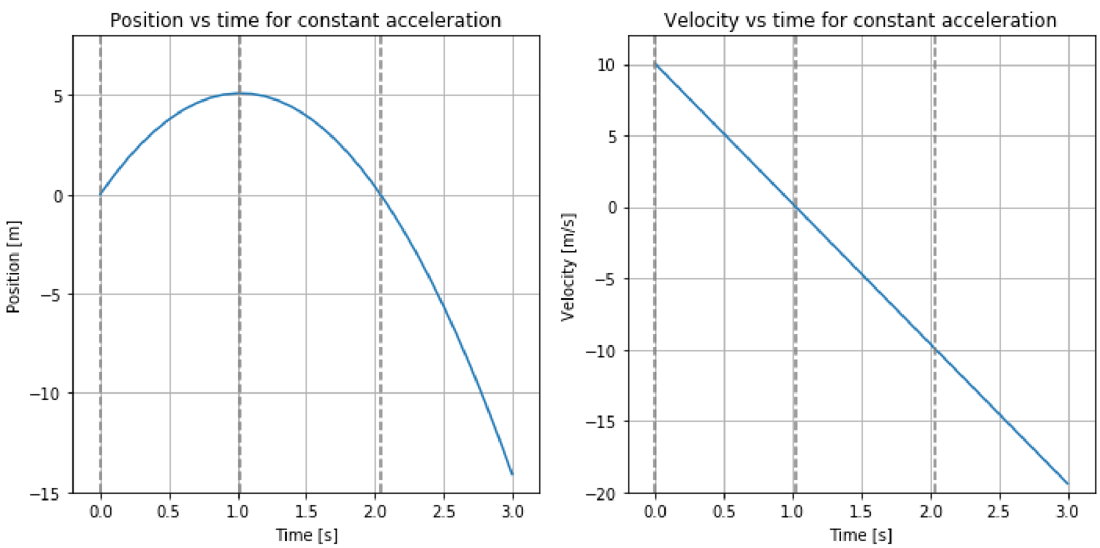


Figure 3.5: Position and velocity as a function of time for the ball in Example 3-1.

We can divide the motion into three parts (shown by the vertical dashed lines in Figure 3.5):

1) Between $t = 0\,\text{s}$ and $t = 1.02\,\text{s}$

At time $t = 0\,\text{s}$, the ball starts at a position of $x = 0\,\text{m}$ (left panel) and has a velocity of $v_{0x} = 10\,\text{m/s}$ (right panel). During the first second of motion, the position, (t), increases (the ball is moving up), until the position stops increasing at $t = 1.02\,\text{s}$, as found in example 3-1. During that time, the velocity decreases linearly from $10\,\text{m/s}$ to $0\,\text{m/s}$ due to the constant negative acceleration from gravity. At $t = 1.02\,\text{s}$, the velocity is instantaneously $0\,\text{m/s}$ and the ball is momentarily at rest (as it reaches the top of the trajectory before falling back down).

2) Between $t = 1.02\,\text{s}$ and $t = 2.04\,\text{s}$

At $t = 1.02\,\text{s}$, the velocity continues to decrease linearly (it becomes more and more negative) as the ball start to fall back down faster and faster. The position also starts decreasing just after $t = 1.02\,\text{s}$, as the ball returns back down to the point of release. At $t = 2.04\,\text{s}$, the ball returns to the point from which it was thrown, and the ball is going with the same speed ($10\,\text{m/s}$) as when it was released, but the velocity is negative (downwards motion).

3) After $t = 2.04\,\text{s}$

If nothing is there to stop the ball, it continues to move downwards with ever decreasing velocity. The position continues to become more negative and the speed continues to increase.

Checkpoint 3-4

Make a sketch of the acceleration as a function of time corresponding to the position and velocity shown in Figure 3.5.

3.3 Using calculus to describe motion

Objects do not necessarily have a constant velocity or acceleration. We thus need to extend our description of the position and velocity of an object to a more general case. This can be done in much the same way as we introduced accelerated motion; namely by pretending that during a very small interval in time, Δt, the velocity and acceleration are constant, and then considering the motion as the sum over many small intervals in time. In the limit that Δt tends to zero, this will be an accurate description.

3.3.1 Instantaneous and average velocity

Suppose that an object is moving with a non constant velocity, and covers a distance Δx in an amount of time Δt. We can define an **average velocity**, v^{avg}:

$$v^{avg} = \frac{\Delta x}{\Delta t}$$

That is, regardless of our choice of time interval, Δt, we can always calculate the average velocity, v^{avg}, of an object over a particular distance. If we shrink the length of the time interval used to measure the velocity, and take the limit $\Delta t \to 0$, we can define the **instantaneous velocity**:

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

The instantaneous velocity is the velocity only in that small instant in time where we choose Δx and Δt. Another way to read this equation is that the velocity, v, is the slope of the graph of $x(t)$. Recall that the slope is the "rise over run", in other words, the change in x divided by the corresponding change in t. Indeed, when we had no acceleration, the position as a function of time, equation 3.1, explicitly had the velocity as the slope of a linear function:

$$x(t) = v_{0x} + v_x t$$

If we go back to Figure 3.5, where velocity was no longer constant, we can indeed see that the graph of the velocity versus time, $v(t)$, corresponds to the instantaneous slope of the graph of position versus time, $x(t)$. For $t < 1.02\,\text{s}$, the slope of the $x(t)$ graph is positive but decreasing (as is $v(t)$). At $t = 1.02\,\text{s}$, the slope of $x(t)$ is instantaneously $0\,\text{m/s}$ (as is the velocity). Finally, for $t > 1.02\,\text{s}$, the slope of $x(t)$ is negative and increasing in magnitude, as is $v(t)$.

Leibniz and Newton were the first to develop mathematical tools to deal with calculations that involve quantities that tend to zero, as we have here for our time interval Δt. Nowadays, we call that field of mathematics "calculus", and we will make use of it here. Using the vocabulary of calculus, rather than saying that "instantaneous velocity is the slope of the graph of position versus time at some point in time", we say that "instantaneous velocity is the time derivative of position as a function of time". We also use a slightly different notation so that we do not have to write the limit $\lim_{\Delta t\to 0}$:

$$\boxed{v(t) = \lim_{\Delta t\to 0}\frac{\Delta x}{\Delta t} = \frac{dx}{dt} = \frac{d}{dt}x(t)} \tag{3.4}$$

where we can really think of dt as $\lim_{\Delta t\to 0}\Delta t$, and dx as the corresponding change in position over an *infinitesimally* small time interval dt.

Similarly, we introduce the **instantaneous acceleration**, as the time derivative of $v(t)$:

$$\boxed{a_x(t) = \frac{dv}{dt} = \frac{d}{dt}v(t)} \tag{3.5}$$

Olivia's Thoughts

When looking at a graph of position versus time, it is sometimes hard to tell at first glance whether the speed of the object is increasing or decreasing. This section gives us an easy way to figure it out. The velocity is the instantaneous slope of the graph $x(t)$, so the speed is the "steepness" of that graph. Simply draw a few lines that are tangent to (meaning just touching) the curve, and see what happens as time increases. If the lines get steeper, the object is speeding up. If they are getting flatter, the object is slowing down.

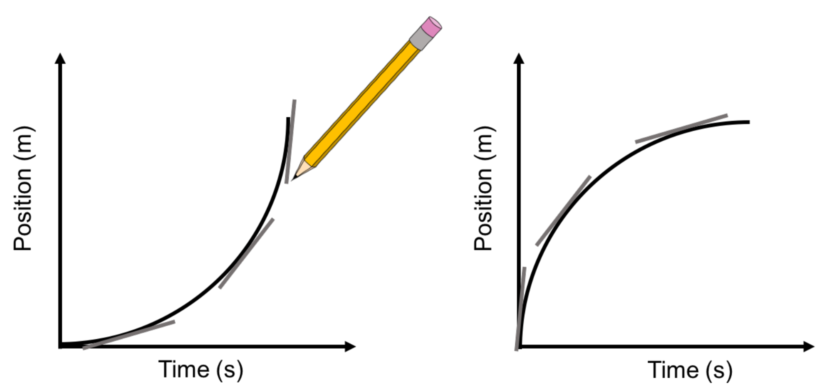


Figure 3.6: Two graphs of $x(t)$ showing tangent lines. Left: the object is speeding up (positive velocity, positive acceleration). Right: the object is slowing down (positive velocity, negative acceleration).

From here, you can also figure out what the direction of the acceleration is. If an object is speeding up, the acceleration and velocity must be in the same direction (i.e. both positive or both negative). If the object is slowing down, they must be in opposite directions. Imagine the graphs in Figure 3.6 are describing the motion of a person running in heavy wind. In the graph on the left, the person is running with the wind and accelerating ($v(t)$ and $a(t)$ positive), and in the second graph the person is running against the wind and decelerating ($v(t)$ positive and $a(t)$ negative).

3.3.2 Using calculus to obtain acceleration from position

Suppose that we know the function for position as a function of time, and that it is given by our previous result (for the case when the acceleration a_x is constant):

$$x(t) = x_0 + v_{0x}t + \frac{1}{2}a_x t^2$$

The velocity is given by taking the derivative of $x(t)$ with respect to time:

$$\begin{aligned} v(t) &= \frac{dx}{dt} = \frac{d}{dt}\left(x_0 + v_{0x}t + \frac{1}{2}a_x t^2\right) \\ &= v_{0x} + a_x t \end{aligned}$$

as we found before, in equation 3.2. The acceleration is then given by the time-derivative of the velocity:

$$\begin{aligned} a_x &= \frac{dv}{dt} = \frac{d}{dt}(v_{0x}t + a_x t) \\ &= a_x \end{aligned}$$

as expected.

Checkpoint 3-5

Chloë has been working on a detailed study of how vicuñas[a] run, and found that their position as a function of time when they start running is well modelled by the function $x(t) = (40\,\mathrm{m/s^2})t^2 + (20\,\mathrm{m/s^3})t^3$. What is the acceleration of the vicuñas?

A) $a_x(t) = 40\,\mathrm{m/s^2}$
B) $a_x(t) = 80\,\mathrm{m/s^2}$
C) $a_x(t) = 40\,\mathrm{m/s^2} + (20\,\mathrm{m/s^3})t$
D) $a_x(t) = 80\,\mathrm{m/s^2} + (120\,\mathrm{m/s^3})t$

[a] Never heard of vicuñas? Internet!

3.3.3 Using calculus to obtain position from acceleration

Now that we saw that we can use derivatives to determine acceleration from position, we will see how to do the reverse and use acceleration to determine position. Let us suppose that we have a constant acceleration, $a_x(t) = a_x$, and that we know that at time $t = 0\,\text{s}$, the object had a speed of v_{0x} and was located at a position x_0.

Since we only know the acceleration as a function of time, we first need to find the velocity as a function of time. We start with:

$$a_x(t) = \frac{d}{dt}v(t)$$

which tells us that we know the slope (derivative) of the function $v(t)$, but not the actual function. In this case, we must do the opposite of taking the derivative, which in calculus is called taking the "anti-derivative" with respect to t and has the symbol $\int dt$. In other words, if:

$$\frac{d}{dt}v(t) = a_x(t)$$

then:

$$v(t) = \int a_x(t)dt$$

Since in this case, $a_x(t)$ is a constant, a_x, the anti-derivative is easily found:

$$\int a_x dt = a_x t + C$$

The velocity is thus given by:

$$v(t) = \int a_x dt = a_x t + C$$

The constant C is determined by what we call our "initial conditions". In this case, we stated that at time $t = 0$, the velocity should be v_{0x}. The constant C is thus v_{0x}:

$$v(t) = C + a_x t = v_{0x} + a_x t$$

and we recover the formula for velocity when the acceleration is constant. Now that we know the velocity as a function of time, we can take one more anti-derivative with respect to time to obtain the position:

$$\begin{aligned} v(t) &= \frac{dx}{dt} \\ \therefore x(t) &= \int v(t)dt \end{aligned}$$

In the case where acceleration is constant, this gives:

$$\begin{aligned} x(t) &= \int v(t)dt \\ &= \int (v_{0x} + a_x t)dt \\ &= v_{0x}t + \frac{1}{2}a_x t^2 + C' \end{aligned}$$

where C' is a different constant than the one we had when determining velocity. The constant is given by our initial conditions. If the object was located at position $x = x_0$ at time $t = 0$, then $C' = x_0$ and we recover the equation for position as a function of time for constant acceleration:

$$x(t) = x_0 + v_{0x}t + \frac{1}{2}a_x t^2$$

Checkpoint 3-6

Choose the graph of $x(t)$ for the case when acceleration is given by $a(t) = A\omega^2 \cos(\omega t)$, where ω and A are positive constants. The velocity and position are zero at $t = 0$.

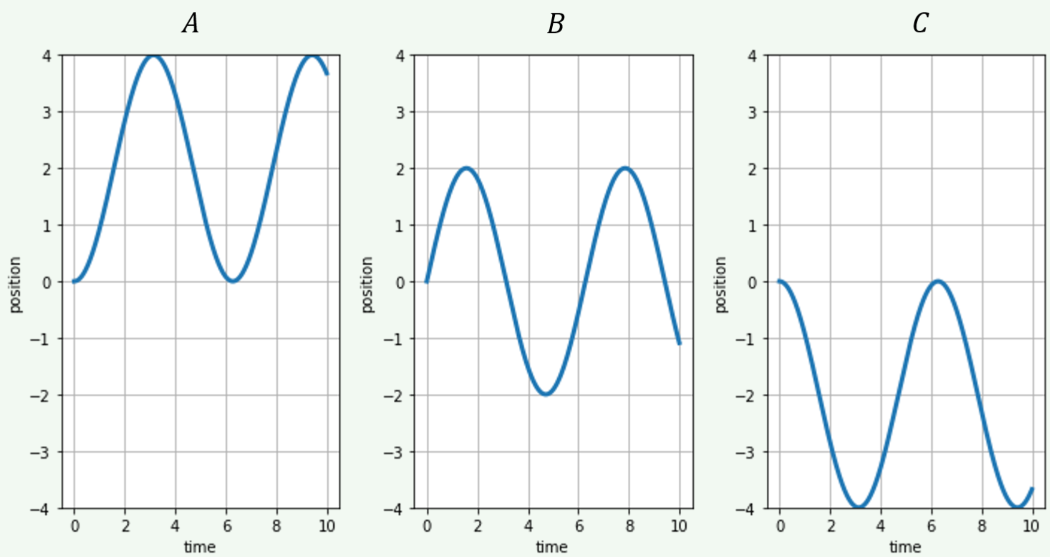


Figure 3.7: Choose the correct position versus time graph.

A) Figure A
B) Figure B
C) Figure C

Checkpoint 3-7

The acceleration of a cricket jumping sideways is observed to increase linearly with time, that is, $a_x(t) = a_0 + jt$, where a_0 and j are constants. What can you say about the velocity of the cricket as a function of time?

A) it is constant
B) it increases linearly with time ($v(t) \propto t$)
C) it increases quadratically with time ($v(t) \propto t^2$)
D) it increases with the cube of time ($v(t) \propto t^3$)

3.4 Relative motion

In order to describe the motion of an object confined to a straight line, we introduced an axis (x) with a specified direction (in which x increases) and an origin (where $x = 0$). Sometimes, it can be more convenient to use an axis that is *moving*. For example, consider a person, Alice, moving inside of a train headed for the French town of Nice. The train is moving with a constant speed, v'^B as measured from the ground. Suppose that another person riding the train, Brice, describes Alice's position using the function $x^A(t)$ using an x-axis defined inside of the train car ($x = 0$ where Brice is sitting, and positive x is in the direction of the train's motion), as depicted in Figure 3.8 below. As long as any person is in the train with Brice, they will easily be able to describe Alice's motion using the x-axis that is moving with the train. Suppose that the train goes through the French town of Hossegor, where a third person, Igor, watches the train go by. If Igor wishes to describe Alice's motion, it is easier for him to use a different axis, say x', that is fixed to the ground and not moving with the train.

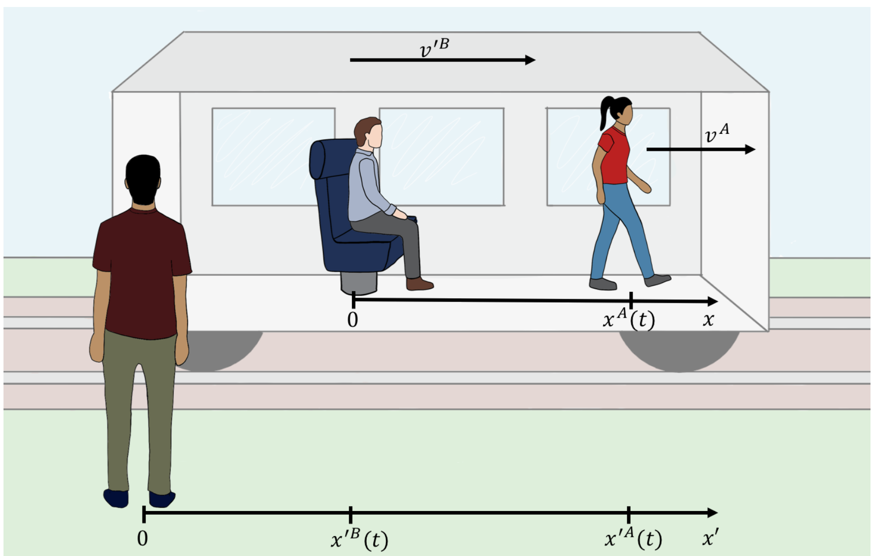


Figure 3.8: Alice is walking in the train and her position is described by both Brice, who is sitting in the train (using the x axis), and Igor, who is at rest on the ground (using the x' axis).

Since Brice already went through the work of determining the function $x^A(t)$ in the **reference frame** of the train, we wish to determine how to *transform* $x^A(t)$ into the reference frame of the train station, $x'^A(t)$, so that Igor can also describe Alice's motion. In other words, we wish to describe Alice's motion in two different *reference frames*.

A reference frame is simply a choice of coordinates, in this case, a choice of x-axis. Ideally, in physics, we prefer to use *inertial* reference frames, which are reference frames that are either "at rest" or that are moving at a constant speed relative to a frame that we consider at rest.

In principle, if you blocked out all of the windows in the train, it would not be possible for

Alice and Brice to determine if the train is moving at constant speed or if it is stopped. Thus, the concept of a "rest frame" is itself arbitrary. It is not possible to define a frame of reference that is truly at rest. Even Igor's frame of reference, the train station, is on the planet Earth, which is moving around the Sun with a speed of 108 000 km/h.

Referring to Figure 3.8, we wish to use Brice's description of Alice's motion, $x^A(t)$, and convert it into a description, $x'^A(t)$, that Igor can use in the train station. Since Brice is at rest in the train, the speed of Brice *relative* to Igor is $v'^B(t)$ (the speed of the train, or the speed of the x frame of reference relative to the x' frame of reference). The first step is for Igor to describe Brice's position, $x'^B(t)$, (that is, the position of Brice's origin).

Assume that we choose $t = 0$ to be the point in time where the two origins are aligned. Since the train is moving at a constant speed, v'^B (as measured by Igor), then the position of Brice's origin, $x'^B(t)$, as measured from Igor's origin is given by:

$$x'^B(t) = v'^B t$$

Now that Igor can describe the position of the origin of Brice's coordinate system, he can use Brice's description of Alice's motion. Recall that $x^A(t)$ is Brice's measure of Alice's distance from his origin. Similarly, $x'^B(t)$, is Igor's measure of the distance from his origin to Brice's origin. Thus, to obtain Alice's distance from Igor's origin, we simply add the distance, $x'^B(t)$, from Igor's origin to Brice's origin, and then add, $x^A(t)$, the distance from Brice's origin to Alice. Thus:

$$\boxed{x'^A(t) = x'^B(t) + x^A(t) = v'^B t + x^A(t)} \tag{3.6}$$

which tells us how to obtain the position of object A in the x' reference frame, when $x^A(t)$ is the description the object's position in the x reference frame which is moving with a velocity v'^B relative to the x' reference frame.

Since we know the position of Alice as measured in Igor's frame of reference, we can now easily find her velocity and her acceleration, as measured by Igor. Her velocity as measured by Igor, v'^A, is given by the time-derivative of her position measured in Igor's frame of reference:

$$\begin{aligned}
v'^A(t) &= \frac{d}{dt}x'^A(t) && (3.7)\\
&= \frac{d}{dt}(v'^B t + x^A(t)) && (3.8)\\
&= v'^B + \frac{d}{dt}x^A(t) && (3.9)\\
&= v'^B + v^A(t) && (3.10)
\end{aligned}$$

where $v^A(t) = \frac{d}{dt}x^A(t)$ is Alice's speed as measured by Brice, in the train. That is, the velocity of Alice as measured by Igor is the sum of the velocity of the train relative to the ground and the velocity of Alice relative to the train, which makes sense. If we now determine Alice's acceleration, $a'^A(t)$, as measured by Igor, we find:

$$\begin{aligned} a'^A(t) &= \frac{d}{dt}v'^A(t) && (3.11)\\ &= \frac{d}{dt}(v'^B + v^A) && (3.12)\\ &= 0 + \frac{d}{dt}v^A(t) && (3.13)\\ &= a^A && (3.14) \end{aligned}$$

where we have explicitly used the fact that the train is moving at constant velocity ($\frac{d}{dt}v'^B = 0$). Here we find that both Brice and Igor will measure the same number when referring to Alice's acceleration (if the train is moving at a constant velocity). This is a particularity of "inertial" frame of references: accelerations do not depend on the reference frame, as long as the reference frames are moving with a constant velocity relative to each other. As we will see later, forces exerted on an object are directly related to the acceleration experienced by that object. Thus, the forces on an object do not depend on the choice of inertial reference frame.

Example 3-2

A large boat is sailing North at a speed of $v'^B = 15\,\text{m/s}$ and a restless passenger is walking about on the deck. Chloë, another passenger on the boat, finds that the passenger is walking at a constant speed of $v^A = 3\,\text{m/s}$ towards the South (opposite the direction of the boat's motion). Marcel is watching the boat pass by from the shore. What velocity (magnitude and direction) does Marcel measure for the restless passenger?

Solution

First, we must choose coordinate systems in the boat and on the shore. On the boat, let us define an x axis that is positive in the North direction and has an origin such that the position of the restless passenger was $x^A(t = 0) = 0$ at time $t = 0$. In Chloë's reference frame, the passenger is thus described by:

$$x^A(t) = v^A t = (-3\,\text{m/s})t$$

where we note that v^A is negative since the passenger is moving in the negative x direction (the passenger is walking towards the South, but we chose positive x to be in the North direction). On shore, we choose an x' axis that also is positive in the North direction. We can choose the origin such that the position of the origin of the boat's coordinate system was at $x' = 0$ at time $t = 0$. The position of the origin of the boat's coordinate system, $x'^B(t)$, as measured by Marcel (on shore) is thus:

$$x'^B(t) = v'^B t = (15\,\text{m/s})t$$

The position of the passenger, $x'^A(t)$, as measured by Marcel, is then given by adding the position of the boat's origin and the position of the passenger as measured from the boat's origin:

$$\begin{aligned}
x'^A(t) &= x'^B(t) + x^A(t)\\
&= v'^B t + v^A t\\
&= (v'^B + v^A)t\\
&= ((15\,\text{m/s}) + (-3\,\text{m/s}))t\\
&= (12\,\text{m/s})t
\end{aligned}$$

To find the velocity of the passenger as measured by Marcel, we take the time derivative:

$$\begin{aligned}
v'^A &= \frac{d}{dt}x'^A(t)\\
&= \frac{d}{dt}\Big((v'^B + v^A)t\Big)\\
&= (v'^B + v^A)\\
&= ((15\,\text{m/s}) + (-3\,\text{m/s}))\\
&= 12\,\text{m/s}
\end{aligned}$$

Since this is a positive number, Marcel still sees the passenger moving in the North direction (the direction of his positive x' axis), but with a speed of $12\,\text{m/s}$, which is less than that of the boat. On the boat, the passenger appears to be walking towards the South, but the net motion of the passenger relative to the ground is still in the North direction, as their speed is less than that of the boat.

3.5 Summary

Key Takeaways

To describe motion in one dimension, we must define an axis with:

1. An origin (where $x = 0$).
2. A direction (the direction in which x increases).
3. A unit for the length.

We describe the position of an object with a function $x(t)$ that *depends* on time. The rate of change of position is called "velocity", $v_x(t)$, and the rate of change of velocity is called "acceleration", $a_x(t)$:

$$v_x(t) = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$$
$$a_x(t) = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv_x}{dt}$$

Given the acceleration, one can find the velocity and position:

$$v_x(t) = \int a_x(t)dt$$
$$x(t) = \int v_x(t)dt$$

With a constant acceleration, $a_x(t) = a_x$, if the object had velocity v_{0x} and position x_0 at $t = 0$:[a]

$$v_x(t) = v_{0x}t + a_x t$$
$$x(t) = x_0 + v_{0x}t + \frac{1}{2}a_x t^2$$
$$v^2 - v_0^2 = 2a(x - x_0)$$

An inertial frame of reference is one that is moving with a constant velocity. It is impossible to define a frame of reference that is truly "at rest", so we consider inertial frames of reference only relative to other frames of reference that we also consider to be inertial. If an object has position x^A as measured in a frame of reference x that is moving at constant speed v'^B as measured in a second frame of reference x', then in the x' reference frame, the kinematic quantities for the object are obtained by the Galilean transformation:

$$x'^A(t) = v'^B t + x^A(t)$$
$$v'^A(t) = v'^B + v^A(t)$$
$$a'^A(t) = a(t)$$

[a] We did not derive the third of these kinematic equations in this chapter, but it is derived in problem 3-1.

Important Equations

Position, Velocity, and Acceleration:

$$v_x(t) = \lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$$

$$a_x(t) = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv_x}{dt}$$

$$v_x(t) = \int a_x(t) dt$$

$$x(t) = \int v_x(t) dt$$

Kinematic Equations:

$$v_x(t) = v_{0x}t + a_x t$$

$$x(t) = x_0 + v_{0x}t + \frac{1}{2}a_x t^2$$

$$v^2 - v_0^2 = 2a(x - x_0)$$

Relative Motion:

$$x'^A(t) = v'^B t + x^A(t)$$

$$v'^A(t) = v'^B + v^A(t)$$

$$a'^A(t) = a(t)$$

3.6 Thinking about the material

Reflect and research

1. Look up the depth of a competition diving pool. What is the relationship between the height of the diving platform and the minimum pool depth? Why? If the designers of the pool assumed that every diver drops straight down off the diving board, would the pool still be safe for divers that jump up first?
2. When did Galileo Galilei first describe his principles of Galilean Relativity?
3. In Galileo's "Dialogue Concerning the Two Chief World Systems", what example did he use to describe relative motion?
4. Imagine that you are a judge, trying to charge an irresponsible driver for speeding on the highway. In the courtroom, he argues that in his own frame of reference, he was sitting still with respect to his car. In fact, he says that it was the officer, parked on the side of the highway that was speeding. You realize that in his reference frame, he is indeed correct - but that's not what matters! How do you explain the relative motion of driving laws to this sneaky offender, in order to serve him justice?

To try at home

1. Find a way to measure the value of g (the acceleration from Earth's gravity) and describe what you did.

To try in the lab

1. Measure the value of g (the acceleration from Earth's gravity) by measuring the time it takes for an object to drop from different heights. Analyse your data in way that you perform a linear fit to your data and determine g from the slope of that fit.

3.7 Sample Problems and Solutions

3.7.1 Problems

Problem 3-1: Show that one can use equations 3.2 and 3.3 to derive the following equation:

$$v^2 - v_0^2 = 2a(x - x_0)$$

which is independent of time. (Solution)

Problem 3-2: Rob is riding his bike at a speed of $8\,\mathrm{m/s}$. He passes by a velociraptor, as one often does, who is eating by the side of the road. The velociraptor begins chasing him. The velociraptor accelerates from rest at a rate of $4\,\mathrm{m/s^2}$. (Solution)

a) Assuming it takes 3 seconds for the velociraptor to react, how long does it take from the moment Rob passes by for the velicoraptor to catch up to him?
b) If there is a safe place 70 metres from where Rob passes the velociraptor, will Rob make it there in time to escape being eaten?

Problem 3-3: Figure 3.9 shows a graph of the acceleration, $a(t)$, of a particle moving in one dimension. Draw the corresponding velocity and position graphs. Assume that $v(0) = 0$ and $x(0) = 0$, and be as quantitative as possible. (Solution)

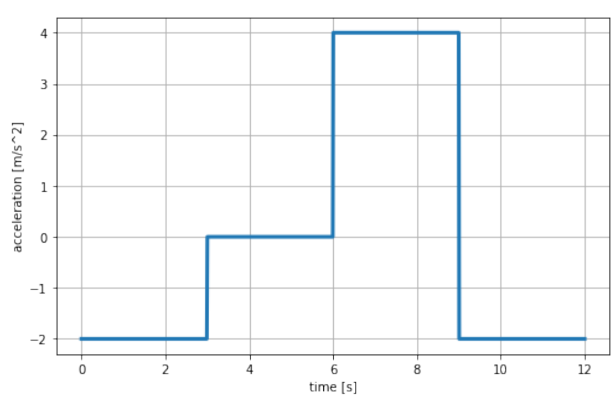


Figure 3.9: A graph of acceleration as a function of time.

3.7.2 Solutions

Solution to problem 3-1: We start with the equations for position and velocity that we derived in this chapter:

$$x = x_0 + v_0 t + \frac{1}{2}at^2$$
$$v = v_0 + at$$

The first equation can be written as:

$$(x - x_0) = v_0 t + \frac{1}{2}at^2$$

Our goal is to find an equation that is independent of time t. We start by isolating t in our equation for velocity:

$$v = v_0 + at$$
$$t = \frac{v - v_0}{a}$$

We then substitute this value of t into our equation for $(x - x_0)$:

$$(x - x_0) = v_0 t + \frac{1}{2}at^2$$
$$(x - x_0) = v_0 \left(\frac{v - v_0}{a}\right) + \frac{1}{2}a\left(\frac{v - v_0}{a}\right)^2$$

We want the left hand side to be $2a(x - x_0)$, so we multiply each term by $2a$:

$$2a(x - x_0)x = (2a)v_0 \left(\frac{v - v_0}{a}\right) + (2a)\frac{1}{2}a\left(\frac{v - v_0}{a}\right)^2$$
$$2a(x - x_0) = (2v_0)a\left(\frac{v - v_0}{a}\right) + a^2\left(\frac{v - v_0}{a}\right)^2$$
$$2a(x - x_0) = 2v_0(v - v_0) + (v - v_0)^2$$

We distribute $2v_0$ into the brackets. Then we expand the third term and get:

$$2a(x - x_0) = (2v_0 v - 2v_0^2) + (v_0 - v^2)(v_0 - v^2)$$
$$2a(x - x_0) = (2v_0 v - 2v_0^2) + (v_0^2 - 2v_0 v + v^2)$$

All that's left to do is collect like terms, and we get the formula we are looking for:

$$2a(x - x_0) = 2v_0 v - 2v_0^2 + v_0^2 - 2v_0 v + v^2$$
$$2a(x - x_0) = (v^2) + (2v_0 v - 2v_0 v) + (v_0^2 - 2v_0^2)$$
$$2a(x - x_0) = v^2 - v_0^2$$
$$\therefore v^2 - v_0^2 = 2a(x - x_0)$$

If you choose a coordinate system such that x_0, this equation becomes $v^2 - v_0^2 = 2ax$.

Solution to problem 3-2: We start by choosing our coordinate system. The solution is simplest if the x axis is positive in the direction of motion and has an origin at the point where Rob passes the velociraptor. We also choose $t = 0$ to be the moment the velociraptor starts running.

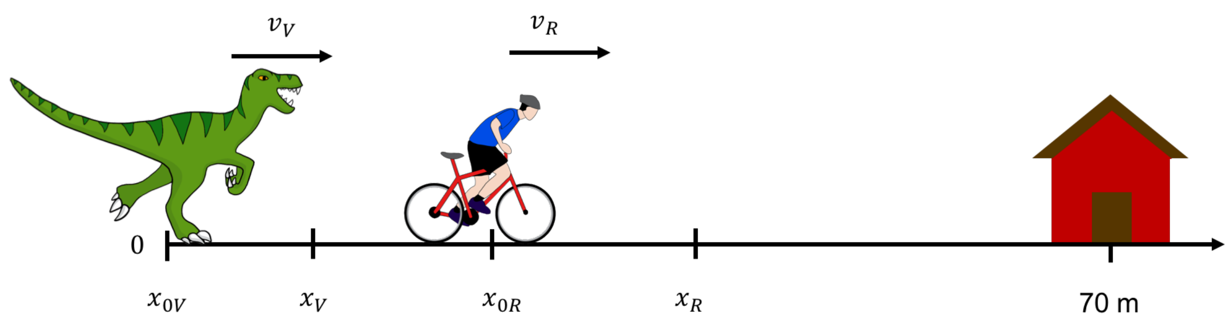


Figure 3.10: Rob is being chased by a velociraptor. At $t = 0$, Rob is a distance x_{0R} from the velociraptor. Safety is 70 m away from the origin.

(a) What do we mean by "catch up"? It means that Rob and the velociraptor will have the same position at the same time. So, we are interested in the value of t when $x_R = x_V$, where x_R is the position of Rob, and x_V is the position of the velociraptor. We need two equations, one describing Rob's position and one describing the position of the velociraptor. Rob is moving at a constant velocity, so his position is described by:

$$x_R = x_{0R} + v_R t$$

The velociraptor has a constant acceleration, so its position is described by:

$$x_V = x_{0V} + v_{0V} t + \frac{1}{2} a_V t^2$$

We can use a table to list the numerical values that we know:

Rob	**Velociraptor**
$x_{0R} = ?$	$x_{0V} = 0\,\mathrm{m}$
$v_R = 8\,\mathrm{m/s}$	$v_{0V} = 0\,\mathrm{m/s}$
	$a_V = 4\,\mathrm{m/s^2}$

x_{0R} is Rob's position at the instant the velociraptor starts running. The value of x_{0R} is unknown but can be easily solved for. It takes 3 seconds for the velociraptor to react, so at $t = 0$, Rob has moved $(8\,\mathrm{m/s}) \times (3\,\mathrm{s}) = 24\,\mathrm{m} = x_{0R}$ (where we used the formula $x = vt$).
Since $v_{0V} = 0$ (the velociraptor starts running from rest) and $x_{0V} = 0$ (the velociraptor

starts at the origin), we can write our equations for the position as:

$$\begin{aligned} x_R &= x_{0R} + v_R t \\ x_V &= \frac{1}{2} a_V t^2 \end{aligned}$$

Remember that we want to find t when x_R=x_V. Setting the above equations equal to one another gives:

$$\begin{aligned} x_R &= x_V \\ x_{0R} + v_R t &= \frac{1}{2} a_V t^2 \\ \therefore \frac{1}{2} a_V t^2 - v_R t - x_{0R} &= 0 \end{aligned}$$

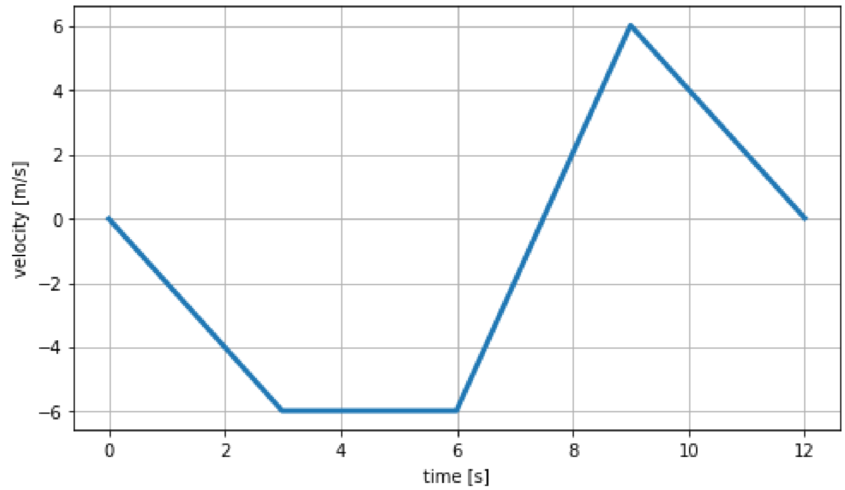
which is a quadratic equation for t. Substituting in numerical values, and solving for t:

$$\begin{aligned} \frac{1}{2}(4\,\mathrm{m/s^2})t^2 - (8\,\mathrm{m/s})t - (24\,\mathrm{m}) &= 0 \\ 2t^2 - 8t - 24 &= 0 \\ \therefore t &= \frac{8 \pm \sqrt{256}}{4} = 6.0\,\mathrm{s} \end{aligned}$$

Where we chose the positive root of the quadratic, since the time must be a positive quantity. This doesn't quite give us the answer we want, since we want to know how long it takes the velociraptor to catch up *from the moment Rob passes by.* We thus have to add the $3\,\mathrm{s}$ reaction time, giving a total time of $9\,\mathrm{s}$.

(b) We can use this solution to figure out whether Rob makes it to safety. The velociraptor catches up after 9 seconds. In 9 seconds, Rob has travelled a distance of $(8\,\mathrm{m/s}) \times (9\,\mathrm{s}) = 72\,\mathrm{m}$. The shelter is only $70\,\mathrm{m}$ away, so Rob gets to safety in time!

Solution to problem 3-3:



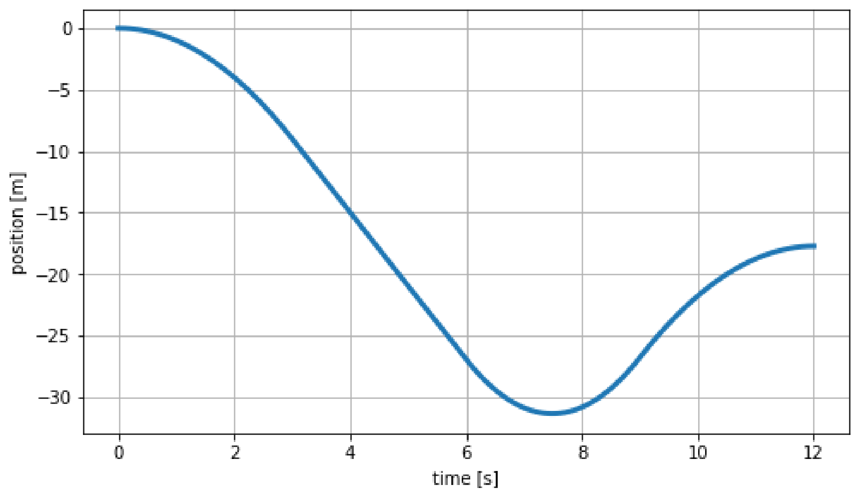


Figure 3.11: Graphs of $v(t)$ and $x(t)$ corresponding to the accleration versus time graph given in the question.

We start by drawing the graph of $v(t)$ from the graph of $a(t)$. Solutions may vary, but a few key features must be present:

- Between $t = 0\,\text{s}$ and $t = 3\,\text{s}$, the velocity decreases linearly, since the acceleration is constant and negative.
- Between $t = 3\,\text{s}$ and $t = 6\,\text{s}$, the velocity remains constant, since the acceleration is zero.
- Between $t = 6\,\text{s}$ and $t = 9\,\text{s}$, the velocity increases linearly, since the acceleration is positive. Since the acceleration is twice as large as in the first interval, the velocity increases at twice the rate that it decreased in the first interval. The object changes direction during this interval, since the velocity changes sign.
- Between $t = 9\,\text{s}$ and $t = 12\,\text{s}$, the velocity decreases linearly with the same rate as in the first interval, and is zero at the end of this interval.

We can get the graph of $x(t)$ from the graph of $v(t)$. The graph of $x(t)$ should have these features:

- Between $t = 0\,\mathrm{s}$ and $t = 3\,\mathrm{s}$, position decreases quadratically, as the velocity is negative and the decreasing.
- Between $t = 3\,\mathrm{s}$ and $t = 6\,\mathrm{s}$, position decreases linearly, since the velocity is negative and constant.
- Between $t = 6\,\mathrm{s}$ and $t = 9\,\mathrm{s}$, the position continues to decrease, but at a lesser rate ans the velocity approaches zero. When the velocity is zero, the position stop changing, and starts to increase quadratically as the velocity becomes positive and increasing.
- Between $t = 9\,\mathrm{s}$ and $t = 12\,\mathrm{s}$, the position continues to increase, but at a lesser rate as the velocity decreases back to zero.

4 Describing motion in multiple dimensions

In this chapter, we will learn how to extend our description of an object's motion to two and three dimensions by using vectors. We will also consider the specific case of an object moving along the circumference of a circle.

Learning Objectives

- Describe motion in a 2D plane.
- Describe motion in 3D space.
- Describe motion along the circumference of a circle.

Think About It

Jake and Madi are riding a carousel that spins at a constant rate. Madi is closer to the centre of the carousel than Jake is. What can you say about their accelerations?

A) Both of their accelerations are zero.
B) Madi's acceleration is greater than Jake's.
C) Jake's acceleration is greater than Madi's.
D) Madi and Jake have the same non-zero acceleration.

4.1 Motion in two dimensions

4.1.1 Using vectors to describe motion in two dimensions

We can specify the location of an object with its coordinates, and we can describe any displacement by a vector. First, consider the case of an object moving with a constant velocity in a particular direction. We can specify the position of the object at any time, t, using its **position vector**, $\vec{r}(t)$, which is a function of time. The position vector is a vector that goes from the origin of the coordinate system to the position of the object. We can describe the x and y components of the position vector with independent functions, $x(t)$, and $y(t)$, that correspond to the x and y coordinates of the object at time t, respectively:

$$\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = x(t)\hat{x} + y(t)\hat{y}$$

Suppose that in a period of time Δt, the object goes from a position described by the position vector $\vec{r}_1$ to a position described by the position vector $\vec{r}_2$, as illustrated in Figure 4.1.

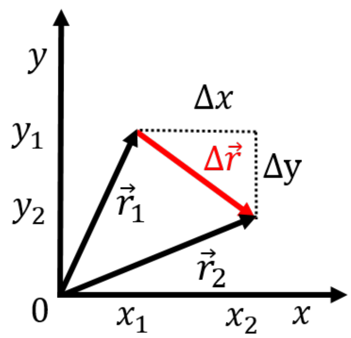


Figure 4.1: Illustration of a displacement vector, $\Delta\vec{r} = \vec{r}_2 - \vec{r}_1$, for an object that was located at position $\vec{r}_1$ at time t_1 and at position $\vec{r}_2$ at time $t_2 = t_1 + \Delta t$.

We can define a **displacement vector**, $\Delta\vec{r} = \vec{r}_2 - \vec{r}_1$, and by analogy to the one dimensional case, we can define an **average** velocity vector, $\vec{v}$ as:

$$\vec{v} = \frac{\Delta\vec{r}}{\Delta t} \tag{4.1}$$

The average velocity vector will have the same direction as $\Delta\vec{r}$, since it is the displacement vector divided by a scalar (Δt). The magnitude of the velocity vector, which we call "speed", will be proportional to the length of the displacement vector. If the object moves a large distance in a small amount of time, it will thus have a large velocity vector. This definition of the velocity vector thus has the correct intuitive properties (points in the direction of motion, is larger for faster objects).

For example, if the object went from position (x_1, y_1) to position (x_2, y_2) in an amount of time Δt, the average velocity vector is given by:

$$\begin{aligned}
\vec{v} &= \frac{\Delta\vec{r}}{\Delta t} \\
&= \frac{1}{\Delta t}\begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix} \\
&= \frac{1}{\Delta t}\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}
\end{aligned}$$

$$= \begin{pmatrix} \frac{\Delta x}{\Delta t} \\ \frac{\Delta y}{\Delta t} \end{pmatrix}$$
$$= \begin{pmatrix} v_x \\ v_y \end{pmatrix}$$
$$\therefore \vec{v} = v_x\hat{x} + v_y\hat{y}$$

That is, the x and y components of the average velocity vector can be found by separately determining the average velocity in each direction. For example, $v_x = \frac{\Delta x}{\Delta t}$ corresponds to the average velocity in the x direction, and can be considered independent from the velocity in the y direction, v_y. The magnitude of the average velocity vector (i.e. the average speed), is given by:

$$||\vec{v}|| = \sqrt{v_x^2 + v_y^2} = \frac{1}{\Delta t}\sqrt{\Delta x^2 + \Delta y^2} = \frac{\Delta r}{\Delta t}$$

where Δr is the magnitude of the displacement vector. Thus, the average speed is given by the distance covered divided by the time taken to cover that distance, in analogy to the one dimensional case.

Checkpoint 4-1

A llama runs in a field from a position $(x_1, y_1) = (2\,\text{m}, 5\,\text{m})$ to a position $(x_2, y_2) = (6\,\text{m}, 8\,\text{m})$ in a time $\Delta t = 0.5\,\text{s}$, as measured by Marcel, a llama farmer standing at the origin of the Cartesian coordinate system. What is the average speed of the llama?

A) $1\,\text{m/s}$
B) $5\,\text{m/s}$
C) $10\,\text{m/s}$
D) $15\,\text{m/s}$

If the velocity of the object is not constant, then we define the **instantaneous velocity vector** by taking the limit $\Delta t \to 0$:

$$\vec{v}(t) = \lim_{\Delta t \to 0} \frac{\Delta \vec{r}}{\Delta t} = \frac{d\vec{r}}{dt} \tag{4.2}$$

which gives us the time derivative of the position vector (in one dimension, it was the time derivative of position). Writing the components of the position vector as functions $x(t)$ and

$y(t)$, the instantaneous velocity becomes:

$$\begin{aligned}
\boxed{\vec{v}(t) = \frac{d}{dt}\vec{r}(t)} \qquad (4.3)\\
&= \frac{d}{dt}\begin{pmatrix} x(t) \\ y(t) \end{pmatrix}\\
&= \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix}\\
&= \begin{pmatrix} v_x(t) \\ v_y(t) \end{pmatrix}\\
\therefore \vec{v}(t) &= v_x(t)\hat{x} + v_y(t)\hat{y}
\end{aligned}$$

where, again, we find that the components of the velocity vector are simply the velocities in the x and y direction. This means that we can treat motion in two dimensions as two times one-dimensional motion: a motion along x and a separate motion along y. This highlights the usefulness of the vector notation for allowing us to use one vector equation ($\vec{v} = \frac{d}{dt}\Delta\vec{r}$) to represent two equations (one for x and one for y).

Similarly the acceleration vector is given by:

$$\begin{aligned}
\boxed{\vec{a}(t) = \frac{d}{dt}\vec{v}(t)} \qquad (4.4)\\
&= \begin{pmatrix} \frac{dv_x}{dt} \\ \frac{dv_y}{dt} \end{pmatrix}\\
&= \begin{pmatrix} a_x(t) \\ a_y(t) \end{pmatrix}\\
\therefore \vec{a}(t) &= a_x(t)\hat{x} + a_y(t)\hat{y}
\end{aligned}$$

If an object is at position $\vec{r}_0 = (x_0, y_0)$ with a velocity vector $\vec{v}_0 = v_{0x}\hat{x} + v_{0y}\hat{y}$ at time $t = 0$, and has a **constant acceleration vector**[1], $\vec{a} = a_x\hat{x} + a_y\hat{y}$, then the velocity vector at some later time t, $\vec{v}(t)$, is given by:

$$\vec{v}(t) = \vec{v}_0 + \vec{a}t$$

Or, if we write out the components explicitly:

$$\begin{pmatrix} v_x(t) \\ v_y(t) \end{pmatrix} = \begin{pmatrix} v_{0x} \\ v_{0y} \end{pmatrix} + \begin{pmatrix} a_x t \\ a_y t \end{pmatrix}$$

[1]Where a constant vector means that both the magnitude and direction are constant in time.

these be considered as two independent equations for the components of the velocity vector:

$$v_x(t) = v_{0x} + a_x t$$
$$v_y(t) = v_{0y} + a_y t$$

which is the same equation that we had for one dimensional kinematics, but once for each coordinate. The position vector is given by:

$$\vec{r}(t) = \vec{r}_0 + \vec{v}_0 t + \frac{1}{2}\vec{a}t^2$$

with components:

$$x(t) = x_0 + v_{0x}t + \frac{1}{2}a_x t^2$$
$$y(t) = y_0 + v_{0y}t + \frac{1}{2}a_y t^2$$

which again shows that two dimensional motion can be considered as separate and independent motions in each direction.

Example 4-1

An object starts at the origin of a coordinate system at time $t = 0\,\text{s}$, with an initial velocity vector $\vec{v}_0 = (10\,\text{m/s})\hat{x} + (15\,\text{m/s})\hat{y}$. The acceleration in the x direction is $0\,\text{m/s}^2$ and the acceleration in the y direction is $-10\,\text{m/s}^2$.

(a) Write an equation for the position vector as a function of time.
(b) Determine the position of the object at $t = 10\,\text{s}$.
(c) Plot the trajectory of the object for the first $5\,\text{s}$ of motion.

Solution

a)We can consider the motion in the x and y direction separately. In the x direction, the acceleration is 0, and the position is thus given by:

$$\begin{aligned} x(t) &= x_0 + v_{0x}t \\ &= (0\,\text{m}) + (10\,\text{m/s})t \\ &= (10\,\text{m/s})t \end{aligned}$$

In the y direction, we have a constant acceleration, so the position is given by:

$$\begin{aligned}
y(t) &= y_0 + v_{0y}t + \frac{1}{2}a_y t^2 \\
&= (0\,\text{m}) + (15\,\text{m/s})t + \frac{1}{2}(-10\,\text{m/s}^2)t^2 \\
&= (15\,\text{m/s})t - \frac{1}{2}(10\,\text{m/s}^2)t^2
\end{aligned}$$

The position vector as a function of time can thus be written as:

$$\begin{aligned}
\vec{r}(t) &= \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} \\
&= \begin{pmatrix} (10\,\text{m/s})t \\ (15\,\text{m/s})t - \frac{1}{2}(10\,\text{m/s}^2)t^2 \end{pmatrix}
\end{aligned}$$

b) Using $t = 10\,\text{s}$ in the above equation gives:

$$\begin{aligned}
\vec{r}(t = 10\,\text{s}) &= \begin{pmatrix} (10\,\text{m/s})(10\,\text{s}) \\ (15\,\text{m/s})(10\,\text{s}) - \frac{1}{2}(10\,\text{m/s}^2)(10\,\text{s})^2 \end{pmatrix} \\
&= \begin{pmatrix} (100\,\text{m}) \\ (-350\,\text{m}) \end{pmatrix}
\end{aligned}$$

c) We can plot the trajectory using python:

Python Code 4.1: Trajectory in xy plane

```
#import modules that we need
import numpy as np #for arrays of numbers
import pylab as pl #for plotting

#define functions for the x and y positions:
def x(t):
    return 10*t

def y(t):
    return 15*t-0.5*10*t**2

#define 10 values of t from 0 to 5 s:
tvals = np.linspace(0,5,10)

#calculate x and y at those 10 values of t using the functions
#we defined above:
xvals = x(tvals)
```

```
yvals = y(tvals)

#plot the result:
pl.plot(xvals,yvals, marker='o')
pl.xlabel("x [m]",fontsize=14)
pl.ylabel("y [m]",fontsize=14)
pl.title("Trajectory in the xy plane",fontsize=14)
pl.grid()
pl.show()
```

Output 4.1:

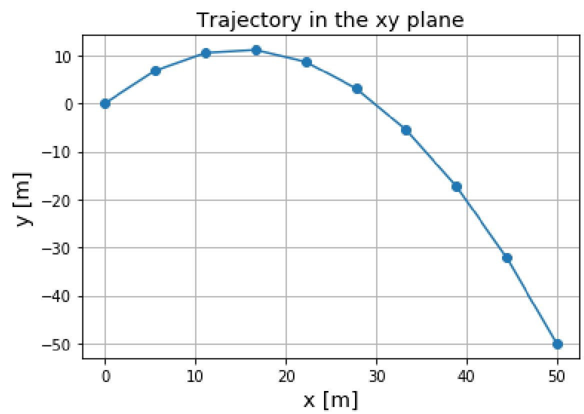


Figure 4.2: Parabolic trajectory of an object with no acceleration in the x direction and a negative acceleration in the y direction.

As you can see, the trajectory is a parabola, and corresponds to what you would get when throwing an object with an initial velocity with upwards (positive y) and horizontal (positive x) components. If you look at only the y axis, you will see that the object first goes up, then turns around and goes back down. This is exactly what happens when you throw a ball upwards, independently of whether the object is moving in the x direction. In the x direction, the object just moves with a constant velocity. The points on the graph are drawn for constant time intervals (the time between each point, Δt is constant). If you look at the distance between points projected onto the x axis, you will see that they are all equidistant and that along x, the motion corresponds to that of an object with constant velocity.

Checkpoint 4-2

In example 4-1, what is the velocity vector exactly at the top of the parabola in Figure 4.2?

A) $\vec{v} = (10\,\text{m/s})\hat{x} + (15\,\text{m/s})\hat{y}$
B) $\vec{v} = (15\,\text{m/s})\hat{y}$
C) $\vec{v} = (10\,\text{m/s})\hat{x}$
D) None of the above.

Example 4-2

A monkey is hanging from a tree branch and you want to feed the monkey by throwing it a banana (Figure 4.3). You know that the monkey is easily frightened and will let go of the tree branch the instant you throw the banana. The monkey is a horizontal distance d away and a height h above the point from which you release the banana when you throw it. At what angle with respect to the horizontal should you throw the banana so that the banana reaches the monkey?

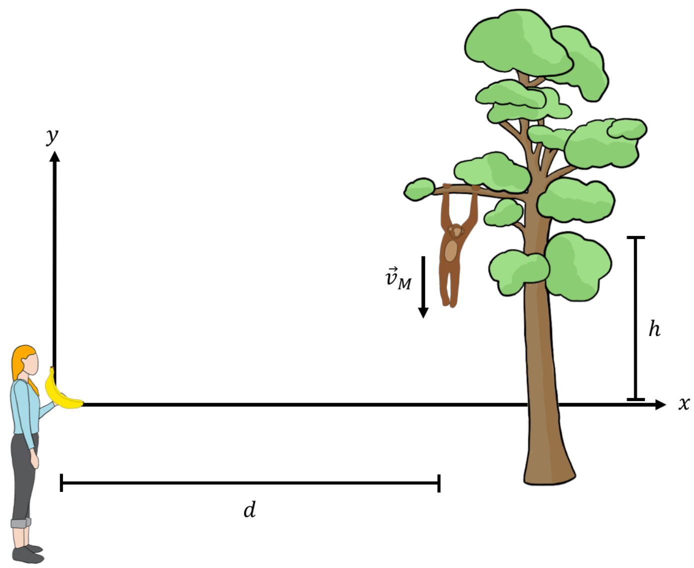


Figure 4.3: Feeding a monkey in a tree.

Solution

This question is asking us to find the angle, θ, between the banana's initial velocity vector, $\vec{v}_{0B}$, and the horizontal for the banana to hit the monkey. This angle is given by the horizontal (v_{B0x}) and vertical (v_{B0y}) components of the initial velocity vector of the banana:

$$\tan\theta = \frac{v_{B0y}}{v_{B0x}}$$

In order for the banana to hit the monkey, and the banana and the monkey must be **in the same place at the same time** at some time, t. Our approach will be as follows: we will start by finding equations that describe the x and y position of the monkey and of the banana. Then, we will use our conditions for a successful "hit" to find the ratio ($\tan\theta = v_{B0y}/v_{B0x}$) that we want for our initial throw, and use that to find θ.

First, we define a coordinate system. We choose the origin to be where the banana is released. We let y be in the vertical direction (positive upwards) and let x be in the horizontal direction (positive towards the monkey), as shown in Figure 4.3.

We treat the x and y components of the banana and monkey's velocity and position vectors as independent. The monkey's motion has only a vertical component. The y component of the monkey's acceleration is the acceleration due to gravity, $a_y = -9.8\,\text{m/s}^2 = -g$, which is negative, since gravity produces an acceleration in the negative y direction. The y component of the monkey's initial position is $y_{M0} = h$ and the y component of its initial velocity is $v_{M0y} = 0$. The y component of the monkey's position as a function of time, $y_M(t)$, is given by:

$$\begin{aligned} y_M(t) &= y_{M0} + v_{My0}t + \frac{1}{2}a_y t^2 \\ &= h + (0) - \frac{1}{2}gt^2 \end{aligned}$$

The horizontal position of the monkey is constant, and is equal to $x_M(t) = d$.

The banana's motion has both x and y components. There is no acceleration in the x direction, so the x component of the banana's velocity is v_{B0x} and constant. We defined the banana's initial x coordinate to be $x_{B0} = 0$, so the x position of the banana as a function of time, $x_B(t)$ is given by:

$$\begin{aligned} x_B(t) &= x_{B0} + v_{B0x} \\ &= (0) + v_{B0x}t \end{aligned}$$

We defined the initial y position of the banana to be $y_{B0} = 0$. The y position of the banana as a function of time, $y_B(t)$, can thus be described by:

$$\begin{aligned} y_B(t) &= y_{B0} + v_{B0y}t + \frac{1}{2}a_y t^2 \\ &= (0) + v_{B0y}t - \frac{1}{2}gt^2 \end{aligned}$$

where v_{B0y} is the y component of the banana's initial velocity and $a_y = -g$ is the y component of the banana's acceleration (due to gravity). Now that we have equations that describe the position of both the banana and the monkey, we can use our conditions for the banana and monkey to be at the same position at the same time. For the monkey and the banana to be in the same position, we need $y_M(t) = y_B(t)$ and $x_B(t) = x_M(t) = d$ at some time t.

Setting our equations for $y_M(t)$ and $y_B(t)$ equal to one another gives:

$$h - \frac{1}{2}gt^2 = v_{0yB}t - \frac{1}{2}gt^2$$
$$\therefore h = v_{0yB}t$$

And setting $x_M(t) = d$ equal to $x_B(t)$ gives:

$$\therefore d = v_{xB}$$

We can just divide one equation by the other to find:

$$\frac{h}{d} = \frac{v_{0yB}t}{v_{xB}t}$$
$$\frac{h}{d} = \frac{v_{0yB}}{v_{xB}}$$

This gives us the ratio we are looking for, so we now know that

$$\tan\theta = \frac{h}{d}$$
$$\therefore \theta = \tan^{-1}\left(\frac{h}{d}\right)$$

This is a somewhat surprising result, as it means that you only need to thrown the banana in the direction of the monkey (that is, aim at the monkey, and throw!). Thus, it will not matter how fast you throw the banana, and you will always hit the monkey if you aimed correctly. When you throw the banana faster, you will hit the monkey higher in its trajectory. If there is no ground for the monkey to hit, you can throw the banana as slowly as you like, and it will eventually catch up with the monkey when the banana reaches $x = d$.

4.1.2 Relative motion

In the previous chapter, we examined how to convert the description of motion from one reference frame to another. Recall the one dimensional situation where we described the position of an object, A, using an axis x as $x^A(t)$. Suppose that the reference frame, x, is moving with a constant speed, v'^B, relative to a second reference frame, x'. We found that the position of the object is described in the x' reference frame as:

$$x'^A(t) = v'^B t + x^A(t)$$

if the origins of the two systems coincided at $t = 0$. The equation above simply states that the distance of the object to the x' origin is the sum of the distance from the x' origin to the x origin **and** the distance from the x origin to the object.

In two dimensions, we proceed in exactly the same way, but use vectors instead:

$$\vec{r}'^A(t) = \vec{v}'^B t + \vec{r}^A(t)$$

where $r^A(t)$ is the position of the object as described in the xy reference frame, $\vec{v}'^B$, is the velocity vector describing the motion of the origin of the xy coordinate system relative to an $x'y'$ coordinate system and $\vec{r}'^A(t)$ is the position of the object in the $x'y'$ coordinate system. We have assumed that the origins of the two coordinate systems coincided at $t = 0$ and that the axes of the coordinate systems are parallel (x parallel to x' and y parallel to y').

Note that the velocity of the object in the $x'y'$ system is found by adding the velocity of xy relative to $x'y'$ and the velocity of the object in the xy frame ($\vec{v}^A(t)$):

$$\begin{aligned}\frac{d}{dt}\vec{r}'^A(t) &= \frac{d}{dt}(\vec{v}'^B t + \vec{r}^A(t)) \\ &= \vec{v}'^B + \vec{v}^A(t)\end{aligned}$$

As an example, consider the situation depicted in Figure 4.4. Brice is on a boat off the shore of Nice, with a coordinate system xy, and is describing the position of a boat carrying Alice. He describes Alice's position as $\vec{r}^A(t)$ in the xy coordinate system. Igor is on the shore and also wishes to describe Alice's position using the work done by Brice. Igor sees Brice's boat move with a velocity $\vec{v}'^B$ as measured in his $x'y'$ coordinate system. In order to find the vector pointing to Alice's position $\vec{r}'^A(t)$, he adds the vector from his origin to Brice's origin ($\vec{v}'^B t$) and the vector from Brice's origin to Alice $\vec{r}^A(t)$.

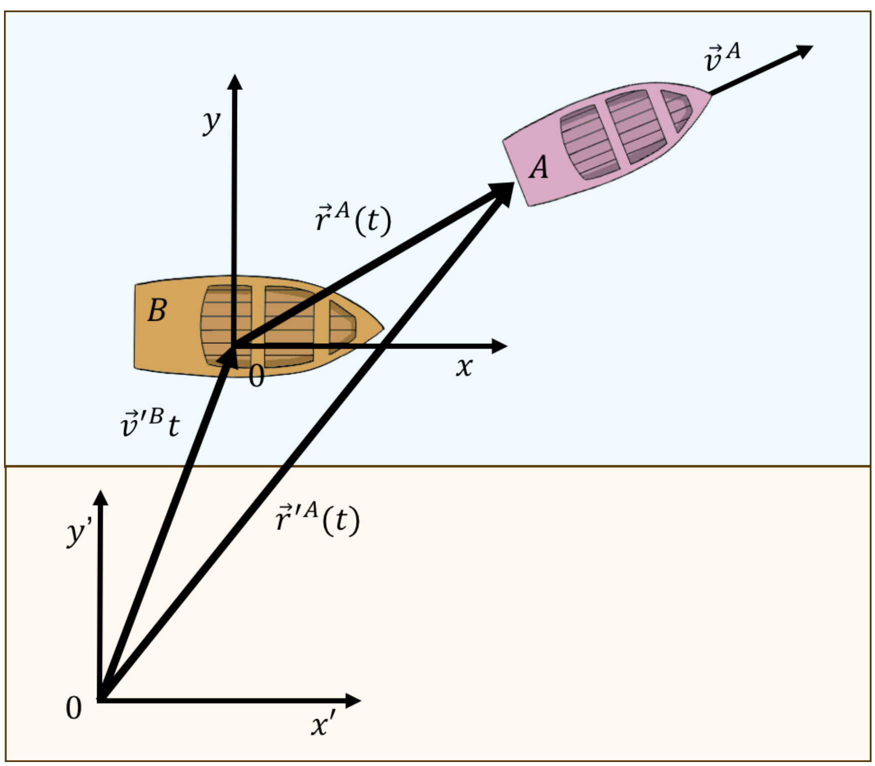


Figure 4.4: Example of converting from one reference frame to another in two dimensions using vector addition.

Writing this out by coordinate, we have:

$$x'^A(t) = v'^B_x t + x^A(t)$$
$$y'^A(t) = v'^B_y t + y^A(t)$$

and for the velocities:

$$v'^A_x(t) = v'^B_x + v^A_x(t)$$
$$v'^A_y(t) = v'^B_y + v^A_y(t)$$

Checkpoint 4-3

You are on a boat and crossing a North-flowing river, from the East bank to the West bank. You point your boat in the West direction and cross the river. Chloë is watching your boat cross the river from the shore, in which direction does she measure your velocity vector to be?

A) In the North direction.
B) In the West direction.
C) A combination of North and West directions.

4.2 Motion in three dimensions

The big challenge was to expand our description of motion from one dimension to two. Adding a third dimension ends up being trivial now that we know how to use vectors. In three dimensions, we describe the position of a point using three coordinates, so all of the vectors simply have three independent components, but are treated in exactly the same way as in the two dimensional case. The position of an object is now described by three independent functions, $x(t)$, $y(t)$, $z(t)$, that make up the three components of a position vector $\vec{r}(t)$:

$$\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}$$
$$\therefore \vec{r}(t) = x(t)\hat{x} + y(t)\hat{y} + z(t)\hat{z}$$

The velocity vector now has three components and is defined analogously to the 2D case:

$$\vec{v}(t) = \frac{d\vec{r}}{dt} = \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \\ \frac{dz}{dt} \end{pmatrix} = \begin{pmatrix} v_x(t) \\ v_y(t) \\ v_z(t) \end{pmatrix}$$
$$\therefore \vec{v}(t) = v_x(t)\hat{x} + v_y(t)\hat{y} + v_z(t)\hat{z}$$

and the acceleration is defined in a similar way:

$$\vec{a}(t) = \frac{d\vec{v}}{dt} = \begin{pmatrix} \frac{dv_x}{dt} \\ \frac{dv_y}{dt} \\ \frac{dv_z}{dt} \end{pmatrix} = \begin{pmatrix} a_x(t) \\ a_y(t) \\ a_z(t) \end{pmatrix}$$
$$\therefore \vec{a}(t) = a_x(t)\hat{x} + a_y(t)\hat{y} + a_z(t)\hat{z}$$

In particular, if an object has a constant acceleration, $\vec{a} = a_x\hat{x} + a_y\hat{y} + a_z\hat{z}$, and started at $t = 0$ with a position $\vec{r}_0$ and velocity $\vec{v}_0$, then its velocity vector is given by:

$$\vec{v}(t) = \vec{v}_0 + \vec{a}t = \begin{pmatrix} v_{0x} + a_x t \\ v_{0y} + a_y t \\ v_{0z} + a_z t \end{pmatrix}$$

and the position vector is given by:

$$\vec{r}(t) = \vec{r}_0 + \vec{v}_0 t + \frac{1}{2}\vec{a}t^2 = \begin{pmatrix} x_0 + v_{0x}t + \frac{1}{2}a_x t^2 \\ y_0 + v_{0y}t + \frac{1}{2}a_y t^2 \\ z_0 + v_{0z}t + \frac{1}{2}a_z t^2 \end{pmatrix}$$

where again, we see how writing a single vector equation (e.g. $\vec{v}(t) = \vec{v}_0 + \vec{a}t$) is really just a way to write the three independent equations that are true for each component.

4.3 Accelerated motion when the velocity vector changes direction

One key difference with one dimensional motion is that, in two dimensions, it is possible to have an acceleration even when the speed is constant. Recall, the acceleration **vector** is defined as the time derivative of the velocity **vector** (equation 4.4). This means that if the velocity vector changes with time, then the acceleration vector is non-zero. If the length of the velocity vector (speed) is constant, it is still possible that the **direction** of the velocity vector changes with time, and thus, that the acceleration vector is non-zero. This is, for example, what happens when an object goes around in a circle with a constant speed (the direction of the velocity vector changes).

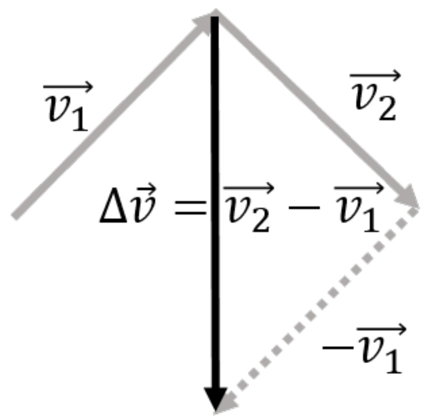


Figure 4.5: Illustration of how the direction of the velocity vector can change when speed is constant.

Figure 4.5 shows an illustration of a velocity vector, $\vec{v}(t)$, at two different times, $\vec{v}_1$ and $\vec{v}_2$, as well as the vector difference, $\Delta\vec{v} = \vec{v}_2 - \vec{v}_1$, between the two. In this case, the length of the velocity vector did not change with time ($||\vec{v}_1|| = ||\vec{v}_2||$). The acceleration vector is given by:

$$\vec{a} = \lim_{\Delta t \to 0} \frac{\Delta\vec{v}}{\Delta t}$$

and will have a direction parallel to $\Delta\vec{v}$, and a magnitude that is proportional to Δv. Thus, even if the velocity vector does not change amplitude (speed is constant), the acceleration vector can be non-zero if the velocity vector changes *direction*.

Let us write the velocity vector, $\vec{v}$, in terms of its magnitude, v, and a unit vector, $\hat{v}$, in the direction of $\vec{v}$:

$$\begin{aligned}
\vec{v} &= v_x\hat{x} + v_y\hat{y} = v\hat{v}\\
v &= ||\vec{v}|| = \sqrt{v_x^2 + v_y^2}\\
\hat{v} &= \frac{v_x}{v}\hat{x} + \frac{v_y}{v}\hat{y}
\end{aligned}$$

In the most general case, both the magnitude of the velocity and its direction can change with time. That is, both the direction and the magnitude of the velocity vector are functions of time:

$$\vec{v}(t) = v(t)\hat{v}(t)$$

When we take the time derivative of $\vec{v}(t)$ to obtain the acceleration vector, we need to take the derivative of a product of two functions of time, $v(t)$ and $\hat{v}(t)$. Using the rules for taking the derivative of a product, the acceleration vector is given by:

$$\begin{aligned}\vec{a} &= \frac{d}{dt}\vec{v}(t) = \frac{d}{dt}v(t)\hat{v}(t)\\ \Aboxed{\vec{a} &= \frac{dv}{dt}\hat{v}(t) + v(t)\frac{d\hat{v}}{dt}}\end{aligned} \tag{4.5}$$

and has two terms. The first term, $\frac{dv}{dt}\hat{v}(t)$, is zero if the speed is constant ($\frac{dv}{dt} = 0$). The second term, $v(t)\frac{d\hat{v}}{dt}$, is zero if the direction of the velocity vector is constant ($\frac{d\hat{v}}{dt} = 0$). In general though, the acceleration vector has two terms corresponding to the change in speed, and to the change in the direction of the velocity, respectively.

The specific functional form of the acceleration vector will depend on the path being taken by the object. If we consider the case where speed is constant, then we have:

$$\begin{aligned}v(t) &= v\\ \frac{dv}{dt} &= 0\\ v_x^2(t) + v_y^2(t) &= v^2\\ \therefore v_y(t) &= \sqrt{v^2 - v_x(t)^2}\end{aligned}$$

In other words, if the magnitude of the velocity is constant, then the x and y components are no longer independent (if the x component gets larger, then the y component must get smaller so that the total magnitude remains unchanged). If the speed is constant, then the acceleration vector is given by:

$$\begin{aligned}\vec{a} &= \frac{dv}{dt}\hat{v}(t) + v\frac{d\hat{v}}{dt}\\ &= 0 + v\frac{d}{dt}\hat{v}(t)\\ &= v\frac{d}{dt}\left(\frac{v_x(t)}{v}\hat{x} + \frac{v_y(t)}{v}\hat{y})\right)\\ &= \frac{dv_x}{dt}\hat{x} + \frac{d}{dt}\sqrt{v^2 - v_x(t)^2}\hat{y}\\ &= \frac{dv_x}{dt}\hat{x} + \frac{1}{2\sqrt{v^2 - v_x(t)^2}}(-2v_x(t))\frac{dv_x}{dt}\hat{y}\\ &= \frac{dv_x}{dt}\hat{x} - \frac{v_x(t)}{\sqrt{v^2 - v_x(t)^2}}\frac{dv_x}{dt}\hat{y}\\ &= \frac{dv_x}{dt}\hat{x} - \frac{v_x(t)}{v_y(t)}\frac{dv_x}{dt}\hat{y}\\ \therefore \quad \Aboxed{\vec{a} &= \frac{dv_x}{dt}\left(\hat{x} - \frac{v_x(t)}{v_y(t)}\hat{y}\right)}\end{aligned} \tag{4.6}$$

where most of the algebra that we did was to separate the x and y components of the acceleration vector, and we used the Chain Rule to take the derivative of the square root. The resulting acceleration vector is illustrated in Figure 4.6 along with the velocity vector[2].

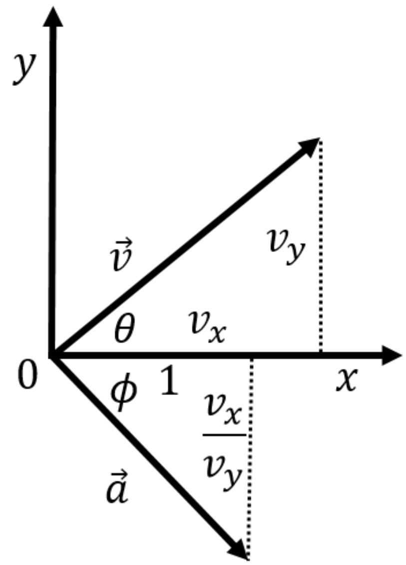


Figure 4.6: Illustration that the acceleration vector is perpendicular to the velocity vector if speed is constant.

The velocity vector has components v_x and v_y, which allows us to calculate the angle, θ that it makes with the x axis:

$$\tan(\theta) = \frac{v_y}{v_x}$$

Similarly, the vector that is parallel to the acceleration has components of 1 and $-\frac{v_x}{v_y}$, allowing us to determine the angle, ϕ, that it makes with the x axis:

$$\tan(\phi) = \frac{v_x}{v_y}$$

Note that $\tan(\theta)$ is the inverse of $\tan(\phi)$, or in other words, $\tan(\theta) = \cot(\phi)$, meaning that θ and ϕ are complementary and thus must sum to $\frac{\pi}{2}$ (90°). This means that **the acceleration vector is perpendicular to the velocity vector if the speed is constant and the direction of the velocity changes**.

In other words, when we write the acceleration vector, we can identify two components,

[2]Rather, it is a vector parallel to the acceleration vector that is illustrated, as the factor of $\frac{dv_x}{dt}$ was omitted (as you recall, multiplying by a scalar only changes the length, not the direction)

$\vec{a}_{\parallel}(t)$ and $\vec{a}_{\perp}(t)$:

$$\begin{aligned}\vec{a} &= \frac{dv}{dt}\hat{v}(t) + v(t)\frac{d\hat{v}}{dt} \\ &= \vec{a}_{\parallel}(t) + \vec{a}_{\perp}(t) \\ \therefore \vec{a}_{\parallel}(t) &= \frac{dv}{dt}\hat{v}(t) \\ \therefore \vec{a}_{\perp}(t) &= v\frac{d\hat{v}}{dt} = \frac{dv_x}{dt}\left(\hat{x} - \frac{v_x(t)}{v_y(t)}\hat{y}\right)\end{aligned}$$

where $\vec{a}_{\parallel}(t)$ is the component of the acceleration that is parallel to the velocity vector, and is responsible for changing its magnitude, and $\vec{a}_{\perp}(t)$, is the component that is perpendicular to the velocity vector and is responsible for changing the direction of the motion.

Checkpoint 4-4

A satellite moves in a circular orbit around the Earth with a constant speed. What can you say about its acceleration vector?

A) It has a magnitude of zero.
B) It is perpendicular to the velocity vector.
C) It is parallel to the velocity vector.
D) It is in a direction other than parallel or perpendicular to the velocity vector.

4.4 Circular motion

We often consider the motion of an object around a circle of fixed radius, R. In principle, this is motion in two dimensions, as a circle is necessarily in a two dimensional plane. However, since the object is constrained to move along the circumference of the circle, it can be thought of (and treated as) motion along a one dimensional axis that is curved.

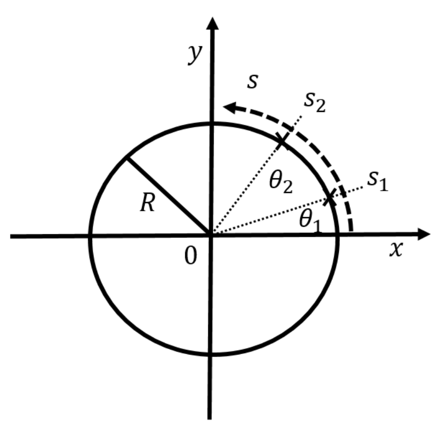


Figure 4.7: Describing the motion of an object around a circle of radius R.

Figure 4.7 shows how we can describe motion along a circle of radius, R. We could use $x(t)$ and $y(t)$ to describe the position on the circle, however, $x(t)$ and $y(t)$ are no longer independent since they have to correspond to the coordinates of points on a circle:

$$x^2(t) + y^2(t) = R^2$$

Instead of using x and y, we could think of an axis that is bent around the circle (as shown by the curved arrow in Figure 4.7, the s axis). The s axis is such that $s = 0$ where the circle intersects the x axis, and the value of s increases as we move counter-clockwise along the circle. Distance along the s axis thus corresponds to the distance along the circumference of the circle.

Another variable that could be used for position instead of s is the angle, θ, between the position vector of the object and the x axis, as illustrated in Figure 4.7. If we express the angle θ in radians, then it is easy to convert between s and θ. Recall, an angle in radians is defined as the length of an arc subtended by that angle divided by the radius of the circle. We thus have:

$$\boxed{\theta(t) = \frac{s(t)}{R}} \tag{4.7}$$

In particular, if the object has gone around the whole circle, then $s = 2\pi R$ (the circumference of a circle), and the corresponding angle is, $\theta = \frac{2\pi R}{R} = 2\pi$, namely $360°$.

By using the angle, θ, instead of x and y, we are effectively using polar coordinates, with a fixed radius. As we already saw, the x and y positions are related to θ by:

$$\begin{aligned} x(t) &= R\cos(\theta(t)) \\ y(t) &= R\sin(\theta(t)) \end{aligned}$$

where R is a constant. For an object moving along the circle, we can write its position vector, $\vec{r}(t)$, as:

$$\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = R \begin{pmatrix} \cos(\theta(t)) \\ \sin(\theta(t)) \end{pmatrix}$$

and the velocity vector is thus given by:

$$\begin{aligned} \vec{v}(t) = \frac{d}{dt}\vec{r}(t) &= \frac{d}{dt} R \begin{pmatrix} \cos(\theta(t)) \\ \sin(\theta(t)) \end{pmatrix} \\ &= R \begin{pmatrix} \frac{d}{dt}\cos(\theta(t)) \\ \frac{d}{dt}\sin(\theta(t)) \end{pmatrix} \\ &= R \begin{pmatrix} -\sin(\theta(t))\frac{d\theta}{dt} \\ \cos(\theta(t))\frac{d\theta}{dt} \end{pmatrix} \end{aligned}$$

where we used the Chain Rule to calculate the time derivatives of the trigonometric functions (since $\theta(t)$ is function of time). We can write this in component form:

$$\begin{aligned} v_x &= -R\sin(\theta(t))\frac{d\theta}{dt} \\ v_y &= R\cos(\theta(t))\frac{d\theta}{dt} \end{aligned} \tag{4.8}$$

The magnitude of the velocity vector is given by:

$$\begin{aligned} ||\vec{v}|| &= \sqrt{v_x^2 + v_y^2} \\ &= \sqrt{\left(-R\sin(\theta(t))\frac{d\theta}{dt}\right)^2 + \left(R\cos(\theta(t))\frac{d\theta}{dt}\right)^2} \\ &= \sqrt{R^2\left(\frac{d\theta}{dt}\right)^2 [\sin^2(\theta(t)) + \cos^2(\theta(t)]} \\ &= R\left|\frac{d\theta}{dt}\right| \end{aligned}$$

The position and velocity vectors are illustrated in Figure 4.8 for an angle θ in the first quadrant ($0 < \theta < \frac{\pi}{2}$).

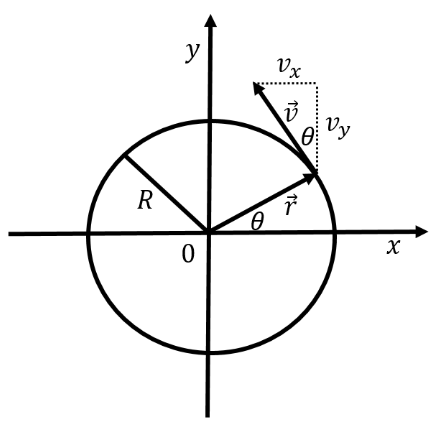


Figure 4.8: The position vector, $\vec{r}(t)$ is always perpendicular to the velocity vector, $\vec{v}(t)$, for motion on a circle.

In this case, you can note that the x component of the velocity is negative (from the diagram and from Equation 4.8). From Equation 4.8, you can also see that $\frac{|v_x|}{|v_y|} = \tan(\theta)$, which is illustrated in Figure 4.8, showing that **the velocity vector is tangent to the circle** and perpendicular to the position vector. This is always the case for motion along a circle.

We can simplify our description of motion along the circle by using either $s(t)$ or $\theta(t)$ instead of the vectors for position and velocity. If we use $s(t)$ to represent position along the circumference ($s = 0$ where the circle intersects the x axis), then the velocity along the

s axis is:

$$\begin{aligned}
v_s(t) &= \frac{d}{dt}s(t)\\
&= \frac{d}{dt}R\theta(t)\\
&= R\frac{d\theta}{dt}
\end{aligned}$$

where we used the fact that $\theta = s/R$ to convert from s to θ. The velocity along the s axis is thus precisely equal to the magnitude of the two-dimensional velocity vector (derived above), which makes sense since the velocity vector is tangent to the circle (and thus in the s "direction").

If the object has a **constant speed**, v_s, along the circle and started at a position along the circumference $s = s_0$, then its position along the s axis can be described using 1D kinematics:

$$s(t) = s_0 + v_s t$$

or, in terms of θ:

$$\begin{aligned}
\theta(t) &= \frac{s(t)}{R} = \frac{s_0}{R} + \frac{v_s}{R}t\\
&= \theta_0 + \frac{d\theta}{dt}t\\
&= \theta_0 + \omega t\\
&\boxed{\therefore \omega = \frac{d\theta}{dt}}
\end{aligned}$$

where we introduced θ_0 as the angle corresponding to the position s_0, and we introduced $\omega = \frac{d\theta}{dt}$, which is analogous to velocity, but for an angle. ω is called the **angular velocity** and is a measure of the rate of change of the angle θ (as it is the time derivative of the angle). The relation between the "linear" velocity v_s (the magnitude of the velocity vector, which corresponds to the velocity in the direction tangent to the circle) and ω is:

$$\boxed{v_s = R\frac{d\theta}{dt} = R\omega}$$

Similarly, if the object is accelerating, we can define an **angular acceleration**, $\alpha(t)$, as the rate of change of the angular velocity:

$$\alpha(t) = \frac{d\omega}{dt}$$

which can directly be related to the acceleration in the s direction, $a_s(t)$:

$$\begin{aligned} a_s(t) &= \frac{d}{dt}v_s \\ &= \frac{d}{dt}\omega R = R\frac{d\omega}{dt} \\ \Aboxed{a_s(t) &= R\alpha} \end{aligned}$$

Thus, the linear quantities (those along the s axis) can be related to the angular quantities by multiplying the angular quantities by R:

$$s = R\theta \tag{4.9}$$
$$v_s = R\omega \tag{4.10}$$
$$a_s = R\alpha \tag{4.11}$$

If the object started at $t = 0$ with a position $s = s_0$ ($\theta = \theta_0$), and an initial linear velocity v_{0s} (angular velocity ω_0), and has a **constant linear acceleration** around the circle, a_s (angular acceleration, α), then the position of the object can be described using either the linear or the angular quantities:

$$\begin{aligned} s(t) &= s_0 + v_{s0}t + \frac{1}{2}a_s t^2 \\ \theta(t) &= \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2 \end{aligned}$$

As you recall from section 4.3, we can compute the acceleration **vector** and identify components that are parallel and perpendicular to the velocity vector:

$$\begin{aligned} \vec{a} &= \vec{a}_{\parallel}(t) + \vec{a}_{\perp}(t) \\ &= \frac{dv}{dt}\hat{v}(t) + v\frac{d\hat{v}}{dt} \end{aligned}$$

The first term, $\vec{a}_{\parallel}(t) = \frac{dv}{dt}\hat{v}(t)$, is parallel to the velocity vector $\hat{v}$, and has a magnitude given by:

$$||\vec{a}_{\parallel}(t)|| = \frac{dv}{dt} = \frac{d}{dt}v(t) = \frac{d}{dt}R\omega = R\alpha$$

That is, the component of the acceleration vector that is parallel to the velocity is precisely the acceleration in the s direction (the linear acceleration). This component of the acceleration is responsible for increasing (or decreasing) the speed of the object and is zero if the object goes around the circle with a constant speed (linear or angular).

As we saw earlier, the perpendicular component of the acceleration, $\vec{a}_{\perp}(t)$, is responsible for changing the direction of the velocity vector (as the object continuously changes direction when going in a circle). When the motion is around a circle, this component of the acceleration vector is called "centripetal" acceleration (i.e. acceleration pointing towards the centre of the circle, as we will see). We can calculate the centripetal acceleration in terms of our angular variables, noting that the unit vector in the direction of the velocity is $\hat{v} = -\sin(\theta)\hat{x} + \cos(\theta)\hat{y}$:

$$\begin{aligned}
\vec{a}_\perp(t) &= v\frac{d\hat{v}}{dt} \\
&= (\omega R)\frac{d}{dt}\left[-\sin(\theta)\hat{x} + \cos(\theta)\hat{y}\right] \\
&= \omega R\left[-\frac{d}{dt}\sin(\theta)\hat{x} + \frac{d}{dt}\cos(\theta)\hat{y}\right] \\
&= \omega R\left[-\cos(\theta)\frac{d\theta}{dt}\hat{x} - \sin(\theta)\frac{d\theta}{dt}\hat{y}\right] \\
&= \omega R[-\cos(\theta)\omega\hat{x} - \sin(\theta)\omega\hat{y}] \\
&\boxed{\vec{a}_\perp(t) = \omega^2 R[-\cos(\theta)\hat{x} - \sin(\theta)\hat{y}]}
\end{aligned} \tag{4.12}$$

where you can easily verify that the vector $[-\cos(\theta)\hat{x} - \sin(\theta)\hat{y}]$ has unit length and points towards the centre of the circle (when the tail is placed on a point on the circle at angle θ). The centripetal acceleration thus points towards the centre of the circle and has magnitude:

$$a_c(t) = ||\vec{a}_\perp(t)|| = \omega^2(t)R = \frac{v^2(t)}{R} \tag{4.13}$$

where in the last equal sign, we wrote the centripetal acceleration in terms of the speed around the circle ($v = ||\vec{v}|| = v_s$).

If an object goes around a circle, it will always have a centripetal acceleration (since its velocity vector must change direction). In addition, if the object's speed is changing, it will also have a linear acceleration, which points in the same direction as the velocity vector (it changes the velocity vector's length but not its direction).

Checkpoint 4-5

A vicuña is going clockwise around a circle that is centred at the origin of an xy coordinate system that is in the plane of the circle. The vicuña runs faster and faster around the circle. In which direction does its acceleration vector point just as the vicuña is at the point where the circle intersects the positive y axis?

A) In the negative y direction.
B) In the positive y direction.
C) A combination of the positive y and positive x directions.
D) A combination of the negative y and positive x directions.
E) A combination of the negative y and negative x directions.

4.4.1 Period and frequency

When an object is moving around in a circle, it will typically complete more than one revolution. If the object is going around the circle with a constant speed, we call the motion "uniform circular motion", and we can define the **period and frequency** of the motion.

The period, T, is defined to be the time that it takes to complete one revolution around the circle. If the object has constant angular speed ω, we can find the time, T, that it takes to complete one full revolution, from $\theta = 0$ to $\theta = 2\pi$:

$$\omega = \frac{\Delta\theta}{T} = \frac{2\pi}{T}$$
$$\boxed{\therefore T = \frac{2\pi}{\omega}} \tag{4.14}$$

We would obtain the same result using the linear quantities; in one revolution, the object covers a distance of $2\pi R$ at a speed of v:

$$\begin{aligned} v &= \frac{2\pi R}{T} \\ T &= \frac{2\pi R}{v} = \frac{2\pi R}{\omega R} = \frac{2\pi}{\omega} \end{aligned}$$

The frequency, f, is defined to be the inverse of the period:

$$f = \frac{1}{T} = \frac{\omega}{2\pi}$$

and has SI units of $\text{Hz} = \text{s}^{-1}$. Think of frequency as the number of revolutions completed per second. Thus, if the frequency is $f = 1\,\text{Hz}$, the object goes around the circle once per second. Given the frequency, we can of course obtain the angular velocity:

$$\omega = 2\pi f$$

which is sometimes called the "angular frequency" instead of the angular velocity. The angular velocity can really be thought of as a frequency, as it represents the "amount of angle" per second that an object covers when going around a circle. The angular velocity does not tell us anything about the actual speed of the object, which depends on the radius $v = \omega R$. This is illustrated in Figure 4.9, where two objects can be travelling around two circles of radius R_1 and R_2 with the same angular velocity ω. If they have the same angular velocity, then it will take them the same amount of time to complete a revolution. However, the outer object has to cover a much larger distance (the circumference is larger), and thus has to move with a larger linear speed.

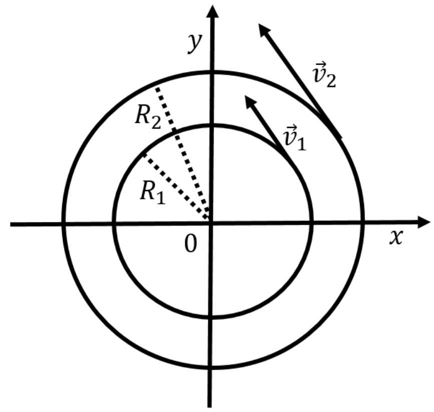


Figure 4.9: For a given angular velocity, the linear velocity will be larger on a larger circle ($v = \omega R$).

Checkpoint 4-6

A motor is rotating at 3000 rpm, what is the corresponding frequency in Hz?

A) 5 Hz
B) 50 Hz
C) 500 Hz

Olivia's Thoughts

There's a trick I like to use to remember how linear and angular velocities work. Figure 4.10 shows your hand in two positions, which we call (1) and (2).

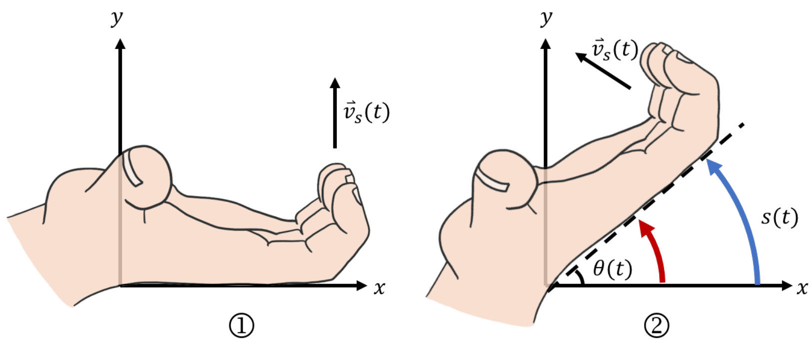


Figure 4.10: How to use your hand to better understand circular motion

Let's say you want to describe the location of your fingers in (2). Start by putting your hand in position (1). This is the position where $\theta = 0$ and $s = 0$. Imagine that your wrist (or your thumb, whichever you prefer) is fixed at the origin. If you keep your fingers perpendicular to your hand, they will always point in the positive s direction.

Imagine that you have a blue glob of paint on the back of your pinky. Rotate your hand until it is in position (2). The length of the curve that the paint makes is the value of s. The angle between the back of your hand and the positive x-axis is θ. Now, imagine that there is a red glob of paint at your palm. It takes the same amount of time for your palm to get from position (1) to position (2) as it does for your fingers. Since they both go through the same angle θ in the same amount of time, the **angular velocity**, ω must be the same for both. However, the blue line left by your fingers will be much longer than the red line left by your palm. Your fingers travelled a greater distance than your palm in the same amount of time, so they must have a greater **linear velocity**, v_s. The further you are from your thumb, the greater the linear velocity will be, which we know from the formula $v_s = R\omega$.

If you kept rotating your hand around the circle, you would see that your fingers always point in the same direction as your linear velocity. This means that if you are using cartesian coordinates, the direction of your linear velocity is always changing.

There are a couple of limitations to this trick. Remember that this only works for circular motion (the radius R must be constant) and that if you are moving in the negative s direction, your fingers will point antiparallel to the linear velocity.

4.5 Summary

Key Takeaways

When the motion of an object is in more than one dimension, we describe the position of the object using a vector, $\vec r$.

$$\vec r(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix} = x(t)\hat x + y(t)\hat y + z(t)\hat z$$

where $x(t)$, $y(t)$, and $z(t)$, are the position coordinates of the object. We treat the motion in each dimension as independent.

The instantaneous velocity vector and the acceleration vector are given by:

$$\begin{aligned}\vec v(t) &= \frac{d}{dt}\vec r(t)\\ \vec a(t) &= \frac{d}{dt}\vec v(t)\end{aligned}$$

If the acceleration vector is constant (in magnitude and direction), then the position and velocity of the object are described by:

$$\begin{aligned}\vec r(t) &= \vec r_0 + \vec v_0 t + \frac{1}{2}\vec a t^2\\ \vec v(t) &= \vec v_0 + \vec a t\end{aligned}$$

where each of these vector equations represents 3 independent equations, one for each of the x, y, and z component of the vectors.

If an object has position $\vec r^A$ as measured in a frame of reference xy that is moving at constant speed $\vec v'^B$ as measured in a second frame of reference $x'y'$, then in the $x'y'$ reference frame:

$$\begin{aligned}\vec r'^A(t) &= \vec v'^B t + \vec r^A(t)\\ \vec v'^A(t) &= \vec v'^B + \vec v^A(t)\\ \vec a'^A(t) &= \vec a^A(t)\end{aligned}$$

An acceleration can change the magnitude and/or the direction of the velocity vector.

1. The component of the acceleration vector that is parallel to the velocity vector changes the magnitude of the velocity.

2. The component of the acceleration vector that is perpendicular to the velocity vector changes the direction of the velocity.

The acceleration vector for motion in two dimensions can be written as the sum of vectors that are parallel ($\vec{a}_{\parallel}$) and perpendicular ($\vec{a}_{\perp}$) to the velocity vector:

$$\vec{a} = \frac{dv}{dt}\hat{v}(t) + v(t)\frac{d\hat{v}}{dt} = \vec{a}_{\parallel} + \vec{a}_{\perp}$$

If the position of an object moving in a circle of radius R is described by its position along the curved axis s, then its position along the circle can be described using an angle, θ, in radians:

$$\theta(t) = \frac{s(t)}{R}$$

For an object moving along a circle, we can write its position vector, $\vec{r}(t)$, as:

$$\vec{r}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = R\begin{pmatrix} \cos(\theta(t)) \\ \sin(\theta(t)) \end{pmatrix}$$

The angular velocity, ω, is the rate of change of the angle. The angular acceleration, α, is the rate of change of the angular velocity:

$$\omega = \frac{d\theta}{dt}$$
$$\alpha = \frac{d\omega}{dt}$$

The linear kinematic quantities can be found from the angular quantities:

$$s = R\theta$$
$$v_s = R\omega$$
$$a_s = R\alpha$$

For circular motion, the velocity vector is tangent to the circle and the perpendicular component of the acceleration is called the centripetal acceleration. The centripetal acceleration points towards the centre of the circle and has a magnitude of:

$$a_c(t) = \omega^2(t)R = \frac{v^2(t)}{R}$$

The centripetal acceleration vector can be written as:

$$\vec{a}_{\perp}(t) = \omega^2 R[-\cos(\theta)\hat{x} - \sin(\theta)\hat{y}]$$

Uniform circular is the motion of an object around a circle with a constant speed. The period, T, is the time that it takes for the object to complete one revolution. The frequency, f, is the inverse of the period, and can be thought of as the number of revolutions completed per second:

$$T = \frac{2\pi}{\omega}$$
$$f = \frac{1}{T} = \frac{\omega}{2\pi}$$

Important Equations

Motion in 2D:

$$\vec r(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = x(t)\hat x + y(t)\hat y$$
$$\vec v(t) = \frac{d}{dt}\vec r(t)$$
$$\vec a(t) = \frac{d}{dt}\vec v(t)$$

Relative Motion 2D:

$$\vec r'^A(t) = \vec v'^B t + \vec r^A(t)$$
$$\vec v'^A(t) = \vec v'^B + \vec v^A(t)$$
$$\vec a'^A(t) = \vec a^A(t)$$

Acceleration Vector 2D:

$$\vec a = \frac{dv}{dt}\hat v(t) + v(t)\frac{d\hat v}{dt}$$
$$\text{(constant speed:)} \quad \vec a = \frac{dv_x}{dt}\left(\hat x - \frac{v_x(t)}{v_y(t)}\hat y\right)$$

Circular Motion:

$$\vec r(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = R\begin{pmatrix} \cos(\theta(t)) \\ \sin(\theta(t)) \end{pmatrix}$$
$$\omega = \frac{d\theta}{dt}$$
$$\alpha = \frac{d\omega}{dt}$$
$$s = R\theta$$
$$v_s = R\omega$$
$$a_s = R\alpha$$
$$a_c(t) = \omega^2(t)R = \frac{v^2(t)}{R}$$
$$\vec a_\perp(t) = \omega^2 R[-\cos(\theta)\hat x - \sin(\theta)\hat y]$$
$$T = \frac{2\pi}{\omega}$$
$$f = \frac{1}{T} = \frac{\omega}{2\pi}$$

4.6 Thinking about the material

Reflect and research

1. It was once believed that there was an absolute reference frame called the "luminiferous aether". What was the name of the experiment that disproved the existence of this frame of reference?
2. Find the centripetal acceleration of the Earth around the Sun.

To try at home

1. Describe and carry out a small experiment to confirm that the amount of time that it takes for a projectile to fall a certain distance does not depend on the horizontal component of its velocity.

To try in the lab

1. Develop a proposal for measuring how fast you can throw a ball, and carry out the experiment.
2. Develop a proposal for measuring how far you can jump with a running start (e.g. a long jump).

4.7 Sample problems and solutions

4.7.1 Problems

Problem 4-1: Ethan is jumping hurdles. He gets a running start, moving with a speed of $3\,\text{m/s}$. The hurdle is $0.5\,\text{m}$ high and the maximum speed that he can have when he leaves the ground is $5\,\text{m/s}$. (You can assume Ethan is a point particle, and ignore air resistance). (Solution)

a) What is the closest distance from the hurdle at which Ethan can jump and still clear the hurdle?
b) What maximum height does he reach?

Problem 4-2: A cowboy swings a lasso above his head. The lasso moves at a constant speed in a circle of radius $1.5\,\text{m}$ in the horizontal plane. A hawk flies toward the lasso at $50\,\text{km/h}$. The hawk sees the end of the lasso moving at $60\,\text{km/h}$ when the lasso is directly in front of it (see Figure 4.11). In the reference frame of the cowboy ... (Solution)

a) How long does it take for the lasso to complete one revolution? (Hint: From the point of view of the hawk, the lasso is moving towards him in addition to moving in a circle. You will have to use your knowledge of relative motion to solve this problem!)
b) What is the centripetal acceleration of the end of the lasso?
c) What is the angular acceleration of the lasso?

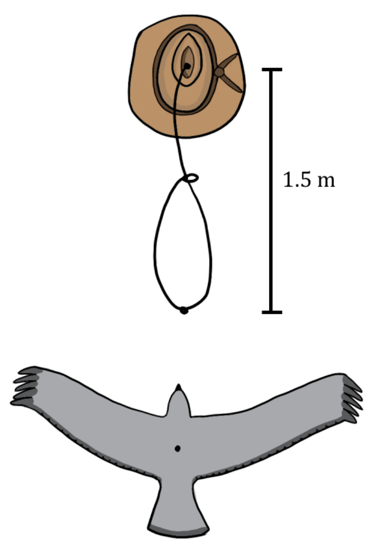


Figure 4.11: The problem as viewed from above. This diagram depicts the moment that the end of the lasso passes in front of the hawk.

4.7.2 Solutions

Solution to problem 4-1: Our approach will be to consider the x and y components of the motion separately. We start by drawing a diagram and choosing our coordinate system. We will choose y to be vertical and positive upwards and x to be in the direction that Ethan is running. We choose the origin to be the location where Ethan leaves the ground for the jump, as illustrated in Figure 4.12.

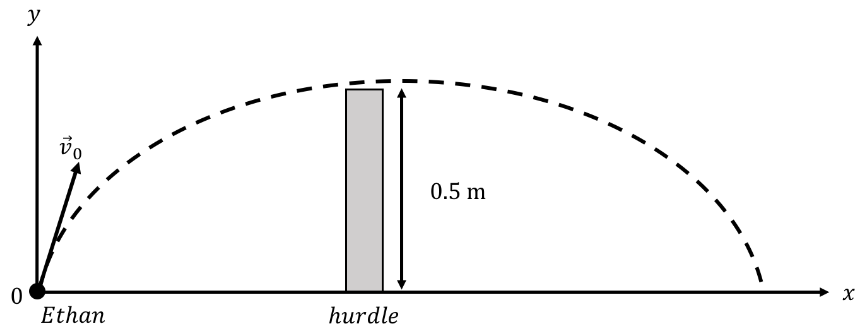


Figure 4.12: Ethan wants to clear a $0.5\,\text{m}$ *hurdle and has an initial velocity* $\vec{v}_0$ *with* x *and* y *components.*

a) Ethan's speed at the beginning of the jump is $v_0 = 5\,\text{m/s}$ and the horizontal (x) component of his velocity is $v_x = 3\,\text{m/s}$. The y component of his initial velocity, v_{0y}, is given by:

$$\begin{aligned}
v_x^2 + v_{0y}^2 &= v_0^2 \\
v_{0y} &= \sqrt{v_0^2 - v_x^2} \\
v_{0y} &= \sqrt{(5\,\text{m/s})^2 - (3\,\text{m/s})^2} = 4\,\text{m/s}
\end{aligned}$$

We chose the origin at the beginning of the jump, so that Ethan's x and y coordinates at time $t = 0$ are $x_0 = 0$ and $y_0 = 0$, respectively. Once Ethan is in the air, there will be no acceleration in the x direction, and the only acceleration is in the y direction and will be that due to gravity. Ethan's position at any time t can be described by the following equations:

$$\begin{aligned}
x(t) &= v_x t \\
y(t) &= v_{0y} t - \frac{1}{2} g t^2
\end{aligned}$$

where g is the acceleration due to gravity, $g = 9.8\,\text{m/s}^2$.
We want to determine the value of $x(t)$ when the vertical displacement, $y(t)$, is equal to the height of the hurdle, h. We thus find the value of t when $y = 0.5\,\text{m}$ and then find the value of x at that time.
We can re-arrange the equation for $y(t)$ and solve the resulting quadratic for t (we get

two solutions):

$$\begin{aligned}
0 &= -\frac{1}{2}gt^2 + v_{0y}t - h \\
0 &= \frac{1}{2}(-9.8\,\text{m/s}^2)t^2 + (4\,\text{m/s})t - 0.5\,\text{m} \\
t &= 0.15\,\text{s}, \quad 0.66\,\text{s}
\end{aligned}$$

The jump will be a parabola, and Ethan will cross a height of 0.5 m twice, once on the way up, and once on the way down. We want to know when Ethan reaches 0.5 m for the first time (on the way up), so we choose $t = 0.15\,\text{s}$. The horizontal displacement at this time is:

$$\begin{aligned}
x &= v_x t \\
&= (3\,\text{m/s})(0.15\,\text{s}) \\
&= 0.45\,\text{m}
\end{aligned}$$

Therefore, he can get as close as 0.45 m from the hurdle before he has to jump, if his initial horizontal velocity is 3 m/s.

b) Ethan's motion follows a parabolic shape. At the maximum height, Ethan's vertical velocity is equal to zero. We can model only the vertical part of the motion to solve for the value of y when $v_y = 0$. We know the following quantities:

$$\begin{aligned}
v_{0y} &= 4\,\text{m/s} \\
v_y &= 0\,\text{m/s} \\
g &= 9.8\,\text{m/s}^2
\end{aligned}$$

The easiest way to determine y is to use the formula,

$$\begin{aligned}
v_y^2 &= v_{0y}^2 - 2g(y - y_0) \\
\therefore y &= \frac{v_y^2 - v_{0y}^2}{(-2g)}
\end{aligned}$$

Substituting our values for v_y, v_{0y}, and g, we get:

$$\begin{aligned}
y_{max} &= \frac{(-4\,\text{m/s})^2}{(2)(-9.8\,\text{m/s}^2)} \\
y_{max} &= 0.82\,\text{m}
\end{aligned}$$

Ethan reaches a maximum height of 0.82 m.

Solution to problem 4-2:

a) We need to determine the speed of the end of the lasso in the cowboy's frame of reference, knowing its speed in the hawk's frame of reference and knowing the velocity of the hawk. Once we know the speed of the lasso in the cowboy's frame of reference we can easily determine how long it takes to complete one revolution (its period).

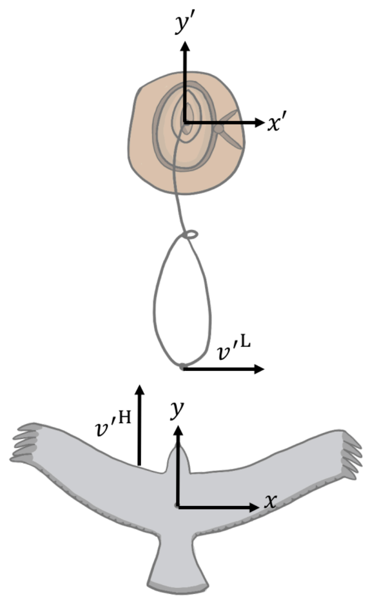


Figure 4.13: The two coordinate systems are aligned so that positive y' and positive y are in the same direction. The velocity vectors of the hawk and the lasso in the reference frame of the cowboy are shown.

We start by introducing coordinate systems for the hawk (xy) and the cowboy ($x'y'$), and choose for the x (y) and x' (y') axes to be parallel. We choose the axes such that x is to the right (when seen from above, as in Figure 4.13) and y is in the direction of motion of the hawk as seen in the cowboy's reference frame. The velocity vector of the hawk in the cowboy's frame of reference is:

$$\vec{v}'_H = v'_H\hat{y} = (50\,\text{km/h})\hat{y}$$

In the hawk's frame of reference, the lasso will have a y component of velocity in the negative y direction with the same magnitude as the speed of the hawk, and an unknown component, v_{Lx}, in the x direction. The velocity of the lasso in the hawk's frame of reference is:

$$\vec{v}_L = v_{Lx}\hat{x} - v'_H\hat{y}$$

However, we know the speed of the lasso in the hawk's frame of reference ($v_L = 60\,\text{km/h}$), so we can easily find v_{Lx}:

$$v_{Lx} = \sqrt{v_L^2 - v_H'^2} = \sqrt{(60\,\text{km/h})^2 - (50\,\text{km/h})^2} = 33.17\,\text{km/h}$$

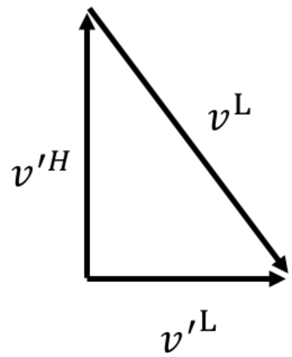


Figure 4.14: Vector addition to determine the velocity of the lasso in the cowboy's reference frame.

In the cowboy's frame of reference, the lasso will have a velocity vector (Figure 4.14), $\vec{v}'_L$, given by:

$$\begin{aligned}\vec{v}'_L &= \vec{v}'_H + \vec{v}_L\\ &= v'_H\hat{y} + v_{Lx}\hat{x} - v'_H\hat{y}\\ &= v_{Lx}\hat{x} = (33.17\,\text{km/h})\hat{x}\end{aligned}$$

That is, in the cowboy's frame of reference, the lasso has a velocity that is in the x direction. This corresponds to the speed, v_s, of the end of the lasso in uniform circular motion about a circle of radius $R = 1.5\,\text{m}$. We can thus find the time required for one revolution to be:

$$\begin{aligned}v_s &= \frac{2\pi R}{T}\\ \therefore T &= \frac{2\pi R}{v_s} = \frac{2\pi(1.5\,\text{m})}{(33.17\,\text{km/h})} = \frac{2\pi(1.5\,\text{m})}{(9.2\,\text{m/s})} = 1.02\,\text{s}\end{aligned}$$

where we converted the speed into m/s before determining the time.

b) The motion is uniform circular motion, so it has a centripetal acceleration given by

$$a_c(t) = \frac{v_s^2(t)}{R}$$

To find the centripetal acceleration of the end of the lasso, we just user our values for v_s and R.

$$a_c(t) = \frac{(9.2\,\text{m/s})^2}{1.5\,\text{m}} = 56\,\text{m/s}^2$$

c) The angular acceleration of the lasso is zero. The angular acceleration refers to the rate of change of the angular velocity (the rate at which the lasso rotates), which is constant for uniform circular motion.

5 Newton's Laws

In this chapter, we introduce Newton's Laws, which is a succinct theory of physics that describes an incredibly large number of phenomena in the natural world. Newton's Laws are one possible formulation of what we call "Classical Physics" (as opposed to "Modern Physics" which include Quantum Mechanics and Special Relativity). Newton's Laws make the connection between dynamics (the causes of motion) and the kinematics of motion (the description of that motion).

Learning Objectives

- Understand Newton's Three Laws.
- Understand the concept of force and how to identify a force.
- Understand the concepts of mass and inertia.
- Understand how to draw free-body diagrams.

Think About It

You are at the supermarket, pushing a cart full of groceries. To keep the cart moving, you notice that you have to keep applying a force to the cart. You conclude that a continuous force is needed for continuous motion. This statement is,

A) True, since the natural state of all objects is to be at rest. Eventually, all objects will be at rest, so to keep an object moving, a force needs to be applied.
B) False. The force you apply to keep an object moving is only to counteract a frictional force.

5.1 Newton's Three Laws

Newton's classical theory of physics is based on the three following laws:

- **Law 1**: An object will remain in its state of motion, be it at rest or moving with constant velocity, unless a net external force is exerted on the object.
- **Law 2**: An object's acceleration is proportional to the net force exerted **on the**

object, inversely proportional to the mass of the object, and in the same direction as the net force exerted on the object.
- **Law 3**: If one object exerts a force on another object, the second object exerts a force on the first object that is equal in magnitude and opposite in direction.

The three statements above are sufficient to describe almost all of the natural phenomena that we experience in our lives. Concepts such as energy, centre of mass, torque, etc, which you may have already encountered, are derived naturally from these three laws. In order to build models to describe specific experiments or observations using Newton's Laws, one needs to understand the two main mathematical concepts that are introduced by the theory: force and mass. A few comments on each of the three laws are first provided before the concepts of force and mass are developed further.

5.1.1 Newton's First Law

Newton's First Law is often referred to as the law of inertia which was originally stated by Galileo. The first law is counter-intuitive, as our experience is that if you push a block on a table and let it go, it will eventually stop. Indeed, Aristotle proposed that the natural state of objects is to be at rest. As a result of Newton's theory, we now understand that if you model a block sliding on a table, one must include a force of friction between the table and the block that acts to slow it down; a sliding block is thus not in a situation where no net external force is exerted on the object.

Newton's First Law is useful in defining what we call an "inertial frame of reference", which is a frame of reference in which Newton's First Law holds true. A frame of reference can be thought of as a coordinate system which can be moving. For example, if a train is moving with constant velocity, we can consider the train as an inertial frame of reference since objects in the train would follow Newton's First Law for observers that are in the train. If a train passenger placed an object on a table, they would observe that the object does not spontaneously start moving; if they slide an object on a frictionless table, they would observe that it keeps on sliding at constant velocity.

However, if the train is accelerating forwards, then an object placed on a frictionless table would appear, for observers in the train's frame of reference, to be accelerating in the direction opposite to that of the train, and violate Newton's First Law. An accelerating train is thus not an inertial frame of reference. To an observer on the ground, looking into the accelerating train through a window, the object placed on the table would appear to move with the same constant velocity as when it was placed on the table (the velocity of the train at the instant the object is placed on the table). In a similar way, when you are in a car, Newton's First Law holds if the car is going at constant velocity, but if the car goes around a curve (and thus accelerates even its speed is constant), you will find that all objects in the car suddenly appear to be pushed towards the outside of the curve, in conflict with Newton's First Law; this is because the accelerating car is not an inertial frame of reference and Newton's First Law is thus not expected to hold.

Newton's First Law thus allows us to define an inertial frame of reference; Newton's Three

Laws only hold in inertial frames of reference.

Checkpoint 5-1

You are in an elevator accelerating upwards.

A) The elevator is an inertial frame of reference.
B) The elevator is not an inertial frame of reference.

5.1.2 Newton's Second Law

Newton's Second Law is often written as a vector equation:

$$\sum \vec{F} = m\vec{a}$$

where $\sum \vec{F}$ is the vector sum of the forces exerted on an object, $\vec{a}$ is the acceleration vector of the object, and m is the "inertial mass" of the object. As we will see, a force is represented by a vector, and the sum of the force vectors on an object is often called the "net force". Recall that using vectors to write an equation is just a shorthand for writing the equation out for each component. In three dimensions, this would thus correspond to three independent scalar equations (one for each component of the force and acceleration vectors):

$$\begin{aligned}\sum F_x &= ma_x\\ \sum F_y &= ma_y\\ \sum F_z &= ma_z\end{aligned}$$

Newton's Second Law is the foundation for Classical Physics, in which we seek to quantitatively describe the motion of any object. The motion of an object is fully specified by its acceleration as long as we know the position and velocity at a specific point in time. That is, by knowing the position and velocity of the object at a point in time and its acceleration, we can describe its motion both in the future and in the in past; we call Classical Physics a deterministic theory (as opposed to, say, Quantum Mechanics, which would only tell us the probability that a particle would be at some particular position in the future). The right-hand side of Newton's Second Law thus contains the kinematic description of the object; if we know the acceleration, we know everything about the motion of the object.

The left-hand side of the equation contains all of the "dynamics" to describe the object; force is the tool that Newton introduced in order to be able to determine the acceleration of an object. Newton's Second Law thus tells how to determine the kinematics of an object by using the concept of forces; it relates the dynamics to the kinematics. Having already covered kinematics, we will now focus on understanding dynamics and how to develop models that allow us to calculate the net force on an object. The inertial mass, m, is a specific property of an object that tells us how large an acceleration it will experience based on a given net force. Thus, objects with different masses will experience different accelerations if subject to the same net force.

Checkpoint 5-2

Object 1 has twice the inertial mass of object 2. If both objects have the same acceleration vector.

A) The net force on both objects is the same.
B) The net force on object 1 is twice that on object 2.
C) The net force on object 1 is half of that on object 2.

5.1.3 Newton's Third Law

Newton's Third Law relates the forces that two objects exert on each other. It is important to understand that the forces that are mentioned in the Newton's Third Law are exerted on *different* objects. If object A exerts a force on object B, then object B will also exert a force on object A. The two forces have the same magnitude but opposite directions. Sometimes, the forces are called "action" and "reaction" forces, although this is misleading, because it makes it sound like the reaction force is "in response to" some voluntary action force. However, inanimate objects can exert forces, and so this can lead to needless confusion as to which force is the reaction force.

It does not matter which force you choose to call the action (reaction) force. If a block is pushing down on a table (action force), then the table is pushing up on the block (reaction force). However, one could just as well say that the table is pushing up on the block (action force) so the block is pushing down on the table (reaction force). It does not matter which force you call the action force. This can be confusing, because if you choose to push on a wall (exerting an action force), then the wall exerts a force on you (the reaction force). If you choose not to push on the wall (exerting no force), then the wall does not exert the reaction force. This leads to people thinking that the reaction force is in response to an action force exerted by a sentient being, which is not the case. You can call the force that you choose to exert on the wall the reaction force and Newton's Laws will still work just as well!

Newton's Third law often leads to confusion when Newton's Second Law is applied. Recall that Newton's Second Law involves the sum of the forces on a particular object (the "net force" on that object). The **two forces that are mentioned in Newton's Third Law are not exerted on the same object**, so they would never appear together in the sum of the forces from Newton's Second Law, and they never cancel each other.

Checkpoint 5-3

You push a heavy block in the North direction. The block is twice as heavy as you are. Which statement is true?

A) The block exerts half of the force on you, in the North direction.
B) The block exerts the same force on you, but in the South direction.
C) The block exerts double of the force on you, in the South direction.
D) The block is inanimate and thus does not exert a force on you.

5.2 Force

A force is a mathematical tool that is introduced in Newton's theory of physics. A force is not a real "thing"; there are no forces in the real world, you cannot give someone a force, or buy a force at the supermarket. A force is a purely mathematical tool, so it is important to fight your intuition about what a force is and to stick to well-defined rules for identifying forces to build models.

Mathematically, a **force is represented by a vector**, and thus has a magnitude and a direction. The SI unit for the magnitude of a force is the "Newton", abbreviated, N. A force is used to describe how the motion of an object is affected by external agents. It is important to note that a force can be exerted by an inanimate being; that is, there is no intent - no conscious decision to push or pull - associated with a force.

When you push a block along a horizontal surface, we would model the motion of the block as being related to a force that you exert on the block in the direction that you are pushing and with a magnitude that is proportional to how hard you are pushing. Newton's Third Law states that the block will exert a force on you that is of equal magnitude but in the opposite direction; if we want to model *your motion*, we will need to include that force exerted by the block *on you.*

If you are pulling on a cart, we would model the motion of the cart by including a force that is exerted on the cart by you. The force would be represented by a vector in the direction that you are pulling with a magnitude based on how hard you are pulling. Similarly, to model your motion, we would include a force vector that is equal in magnitude and opposite in direction to represent the force exerted by the cart on you. When modelling the motion of an object, it is important to consider only the forces exerted on that object.

One way to quantify a force is to use a spring scale. Springs have a natural "rest length" if not acted upon by external forces. If you try to stretch a spring, it will "want" to come back to its normal rest length; it exerts a force on your hand in the opposite direction from the one you are pulling on the spring. You may have noticed that the more you stretch a spring, the harder you have to pull on it. We can quantify the magnitude of a force by the distance that the forces causes a spring to stretch, since that distance increases with what we conceptualize as a force. For example, one could designate a "standard spring" to be one that extends (or compresses) by 1 cm when a force of 1 N is exerted on the spring in the direction co-linear with the axis of the spring. We could then use that "standard spring" to measure the magnitude of any force.

5.2.1 Types of forces

When modelling the dynamics of an object, we need to identify all of the forces exerted on that object. Some of the forces can be classified as "contact forces" as they arise from something making contact with the object (such as you pushing on the object). Other forces can be exerted "at a distance"; for example, the force of gravity from the Earth can be exerted on a bird in flight, even if the bird is not in contact with the Earth. In reality, contact forces arise because the electrons from two objects repel each other. When you push against a wall, the reason that you feel a resistance is because the electrons on your hand are repelled by the electrons on the wall; you never actually "touch" the wall[1]!

In this section, we list and describe the most common types of forces that arise when modelling the motion of an object. When determining the forces that are acting on an object, it is usually a good idea to run down this list to see if any of these forces should be included. Again, try to fight your intuition about what a force "feels" like and instead be objective in determining whether any of the forces below should be included based on their characteristics.

Weight

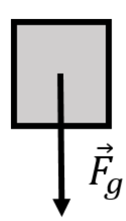
Weight is the force exerted by gravity. While all objects with mass exert an attractive force of gravity on all other objects with mass, that force is usually negligible unless the mass of one of the objects is very large. For an object near the surface of the Earth, we can, to a very good degree of approximation, assume that the only force of gravity on the object is from the Earth. We usually label the force of gravity on an object as $\vec F_g$. All objects near the surface of the Earth will experience a weight, as long as they have a mass. If an object has a mass, m, and is located near the surface of the Earth, it will experience a force (its weight) that is given by:

$$\vec F_g = m\vec g$$

where $\vec g$ is the Earth's "gravitational field" vector and **points towards the centre of the Earth**. Near the surface of the Earth, the magnitude of the gravitational field is approximately $g = 9.8\,\text{N/kg}$. The gravitational field is a measure of the strength of the force of gravity from the Earth (it is the gravitational force per unit mass). The magnitude of the gravitational field is weaker as you move further from the centre of the Earth (e.g. at the top of a mountain, or in Earth's orbit). The gravitational field is also different on different planets; for example, at the surface of the moon, it is approximately $g_m = 1.62\,\text{N/kg}$ (six times less) - thus the weight of an object is six times less at the surface of the moon (but its mass is still the same). As we will see, the magnitude of the gravitational field from any spherical body of mass M (e.g a planet) is given by:

$$g(r) = G\frac{M}{r^2}$$

where $G = 6.67\times 10^{-11}$ is Newton's constant of gravity, and r is the distance from the centre of the object.



[1]As a matter of fact, it is impossible to ever touch anything, you can just get really close!

Figure 5.1: The weight force on an object near the surface of the Earth points towards the centre of the Earth (downwards).

Although we have not yet introduced the concept of mass, it is worth emphasizing that mass and weight are different (they have different dimensions). Mass is an intrinsic property of an object, whereas weight is a force of gravity that is exerted on that object because it has mass and is located next to another object with mass (e.g. the Earth). On Earth, when we measure our weight, we usually do so by standing on a spring scale, which is designed to measure a force by compressing a spring. We are thus measuring mg, which can easily be related to our mass since, on Earth, weight and mass are related by a factor of $g = 9.8\,\text{N/kg}$; this is usually what leads to the confusion between mass and weight.

Checkpoint 5-4

A person standing on a scale finds that they weigh 80 kg.

A) They exert an upwards force on the Earth with a magnitude of 80 N.
B) They exert an upwards force on the Earth with a magnitude of 784 N.
C) They exert an downwards force on the Earth with a magnitude of 80 N.
D) They exert an downwards force on the Earth with a magnitude of 784 N.
E) They exert no force on the Earth.

Normal forces

Normal forces are exerted when two surfaces are in contact and "pushing" against each other. For example, if a block is resting on a horizontal table, the table will exert a normal force on the block that is upwards. The force is called "normal" because it is normal (i.e. perpendicular) to the interface between the two objects. The normal force exerted by a surface onto an object points in the direction **from the surface to the object** in such a way that it is perpendicular to the interface between the surface and the object. Because of Newton's Third Law, whenever an object experiences a normal force from a surface, the object also exerts a force of the same magnitude (in the opposite direction) on the surface. The magnitude of the normal force exerted by a surface onto an object, in general, depends on the other forces that are exerted on the object. For example, if a block is on a table, it will experience a stronger normal force if you exert a downwards force on the block.

Figure 5.2 shows two examples of the normal force on a block that is exerted by a surface (it is explicitly assumed that the block also experiences a downwards force from gravity that is not shown). In both cases, the normal force, $\vec{N}$, is perpendicular to the interface and in the direction that goes from the interface towards the object.

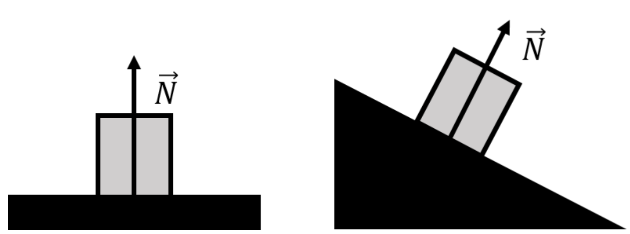


Figure 5.2: The normal force, $\vec{N}$, exerted by a horizontal surface on a block (left side) and by an inclined surface (right side). In both cases, the normal force on the object is perpendicular to the interface between the object and the surface and points in the direction from the interface towards the object.

Frictional forces

A frictional force can exist at the interface between two surfaces and is always perpendicular to the normal force that corresponds to that interface. A frictional force is used to model the resistance that is felt when one tries to slide an object along a surface. The frictional force is used to model the details of how two surfaces interact at a microscopic level; since surfaces are never perfectly flat, two surfaces will never slide without resistance as the various bumps and valleys of the two surfaces will interact (Figure 5.3). Furthermore, even if the two surfaces were perfectly smooth, the electrons on the two surfaces would still interact and lead to an effective force when one surface moves with respect to the other.

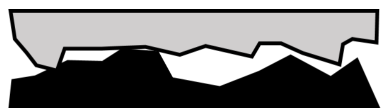


Figure 5.3: Illustration that the frictional force between surfaces can be thought of as arising from microscopic imperfection in the surfaces, although even two perfectly smooth surfaces would still interact.

One distinguishes between two types of frictional forces: kinetic and static, depending on whether the surfaces are sliding with respect to each other (kinetic) or not (static). Because of Newton's Third Law, the objects associated with each surface will both experience a frictional force (same magnitude, opposite direction).

The frictional force exerted on an object is always parallel to the surface of the object. For the kinetic force of friction, the force is exerted in the direction that is opposite to the motion of the object relative to the surface. For the static force of friction, the force is exerted in the direction that is opposite to the *impeding motion*. If a block is sliding towards the right on a table (Figure 5.4, left), it will experience a kinetic force of friction that is to the left. The table will then experience a force of friction that is to the right (Newton's Third Law). If there is a heavy crate on the ground which you try to push but does not move (Figure 5.4, right), there is a force of static friction exerted by the ground on the object that is in the opposite direction that you are pushing.

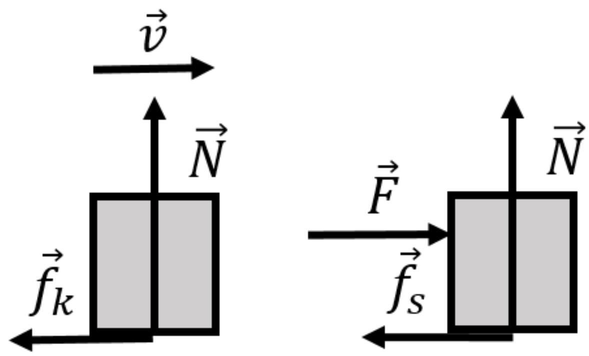


Figure 5.4: (Left:) A block sliding to the right on a horizontal surface (not shown). The force of kinetic friction, $\vec{f}_k$, is always perpendicular to the normal force and opposite of the direction of motion. (Right:) A block that is being acted upon by an external force $\vec{F}$ to the right. A force of static friction, $\vec{f}_s$, is perpendicular to the normal force and opposite the direction of "impeding motion" - without the force of static friction, the block would start to accelerate towards the right, so the force of static friction is to the left.

One key difference between the forces of static and kinetic friction is that the magnitude of the force of static friction can vary in magnitude; the force of static friction on the crate increases as you push harder, until you push hard enough to overcome the maximal force of static friction that can exist between the ground and the crate. Often, the force of kinetic friction is smaller than the static force of friction; you may have noticed that you have to push very hard to get an object sliding, but once it is sliding, you do not need to push as hard to keep it moving.

The magnitude of the kinetic force of friction between two surfaces, f_k, is modelled as being proportional to the normal force between the two surfaces:

$$f_k = \mu_k N$$

where μ_k is called the "coefficient of kinetic friction" and depends on the two surfaces. If you push down on an object, it is more difficult to slide it along a surface, because the normal force, and thus the kinetic friction force increases.

Similarly, the maximum magnitude of the force of static friction between to surfaces, f_s, is modelled as:

$$f_s \leq \mu_s N$$

where μ_s is called the "coefficient of static friction" and the inequality sign is used to indicate that the force of static friction has a maximum value, but that its magnitude depends on the other forces being exerted on the object. For example, if you do not push against a crate on a horizontal surface, there is no force of static friction on the crate (as long as no other forces are exerted that are parallel to the surface).

Tension forces

Tension forces are "pulling" forces that are applied by a rope or other non rigid media (e.g. a chain) which cannot usually be used to push[2]. If you attach a rope to a crate and use the rope to pull the crate, we call the force exerted by the rope onto the crate a force of tension.

When you pull on a rope that is attached to a wall at the other end, we say that the rope is under tension, or that the tension force is present throughout the rope. If you pull really hard on the rope, it is harder to displace the centre of the rope (or any other point) than if you did not pull on the rope at all. It thus makes sense to view the tension as being present throughout the rope. The force of tension that a rope can apply onto an object depends on what is pulling on the rope at the other end. A rope can be used to change the direction of a force, as illustrated in Figure 5.5, which shows a pulley and rope being used to lift a block vertically by applying a horizontal force, $\vec{F}$, to the rope.

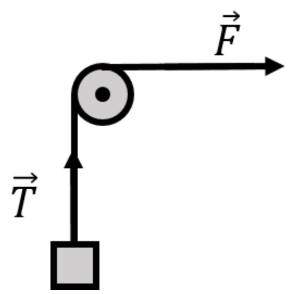


Figure 5.5: A force $\vec{F}$ is applied to a rope, which goes around a pulley and is attached to a crate. The rope exerts a force of tension $\vec{T}$ on the crate. If the pulley and rope are massless, then the magnitude of the applied force is equal to that of the tension force, and the rope and pulley effectively allow one to change the direction of the applied force vector.

The same tension is present throughout sections of the rope that can move freely. Imagine a rope lying on the ground and someone pressing down with their foot on the rope at its midpoint. If you pull on one end of the rope with your hand, there will be a tension in the section of the rope between your hand and the foot that is pressing on the rope, but the other side of the rope will be slack; the tension is thus different in different sections of the rope. As we will see in later chapters, if a rope goes around a pulley that is accelerating and has mass, then the tension in the rope on either side of the pulley is different; this is similar to the tension being different on either side of the foot pressing down on the rope.

Drag forces

Drag forces are exerted on an object that is moving through a fluid (a gas or a liquid). As an object moves through a fluid, the fluid must be displaced which results in a net force opposing the motion of the object. Drag forces are thus always in the opposite direction of the motion of the object relative to the fluid, similar to friction. Often, one hears the term "air friction" which refers to the drag force experienced by an object that is moving through

[2]If you attached a rigid rod to an object and pulled on the rigid rod, you could call the force exerted by the rod on the object a force of tension, even if the rod is rigid.

the air.

There is no good general model for calculating the magnitude of the drag force on any object moving through any fluid. This usually has to be measured; while good software exist for simulating drag, you will still ultimately need to test your new airplane design in a wind tunnel to measure the drag force.

The magnitude of the drag force generally depends on the cross-section of the object (the area of the object as seen when looking at the object in the direction of motion), the speed of the object, and the visocity of the fluid (how difficult it is to displace the fluid). For small objects moving relatively slowly through a fluid (e.g. pollen falling through the air), the drag force is usually proportional to the object's speed, whereas for larger objects moving faster through a fluid (e.g a car or airplane moving through the air) the drag force is usually proportional to the speed of the object squared.

Spring forces

Spring forces are those forces that are exerted by those materials and objects that can be compressed or extended. A common example is a simple coil spring, which has a natural rest length. If the spring is extended, the spring will exert "restoring forces" on both ends of the spring that are directed towards the centre of the spring. If the spring is compressed, the spring will exert restoring forces that point away from the centre of the spring. In either case, the spring will exert forces that would allow it to come back to its rest length.

Most springs, if they are not stretched or compressed too much, will exert a restoring force that is given by Hooke's Law:

$$\vec{F} = -kx\hat{x}$$

where $\vec{F}$ is the force exerted by the spring, k is called the "spring constant" of the spring, and x is the amount that the spring has been stretched or compressed. The negative sign indicates that the restoring force from the spring will be in the opposite direction that the spring length was changed, and the x axis is defined to be co-linear with the axis of the spring and the origin is located where the spring is at rest. This is illustrated in Figure 5.6.

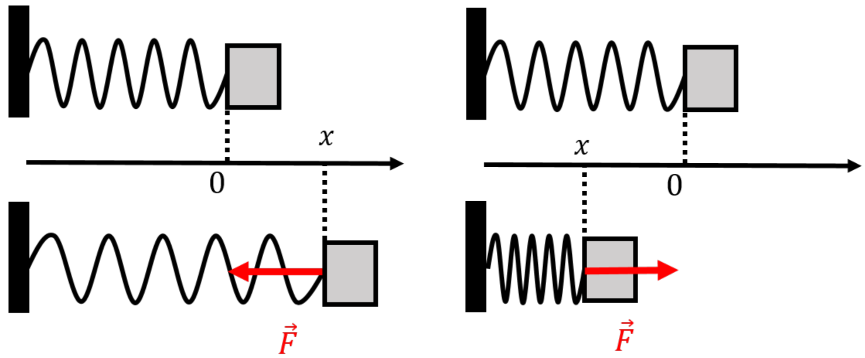


Figure 5.6: A spring is attached to a fixed wall on its left and to a movable block on its right. The x axis is chosen to describe the position of the end of the spring where the block is attached and the origin corresponds to the point where the spring is not extended or compressed (the top row). The x axis is chosen so that positive values of x correspond to the spring being extended. On the bottom left, the spring is extended by a distance x (the position of the block has positive x), and the force from the spring on the block is in the negative x direction. On the bottom right, the spring is compressed (the position of the block has negative x), and the force from the spring is in the positive x direction.

Checkpoint 5-5

In Figure 5.6, we chose the positive x axis to correspond to positions where the spring is extended and verified that Hooke's Law ($\vec{F} = -kx\hat{x}$) holds. If we had chosen the positive direction to correspond to compression (positive x to the left), would Hooke's Law still correctly describe the direction of the force exerted by the spring on the block?

A) Yes.
B) No.

Inertial forces

Inertial forces are exerted on an object when the forces on the object are modelled in a non-inertial frame of reference. For example, in the frame of reference of an accelerating elevator, or that of a car going around a curve, one can use Newton's Three Laws to model motion, if an additional inertial force is included. In a frame of reference that has an acceleration given by $\vec{a}$, an inertial force $-m\vec{a}$ is exerted on an object. This is the nature of the outwards force that is felt when your car goes around a curve, or the perception of being weightless in an elevator that has a large downwards acceleration. We will discuss inertial forces in more detail in section 5.6.

"Applied" forces

"Applied" forces is just a general "catch-all" term for specifying forces that are not described above. For example, the force applied by a person onto an object is often referred to as an applied force.

5.3 Mass and inertia

Mass is a property of an object that quantifies how much matter the object contains. In SI units, mass is measured in kilograms. One kilogram is defined to be the mass of a cylinder that is made of a platinum-iridium alloy that is kept at the international Bureau of Weights and Measures, in France. All other masses are obtained by comparison to this standard.

Newton's Second Law introduces the concept of mass as that property of the object that determines how large of an acceleration it will experience given a net force exerted on that object. In principle, one can compare the accelerations of different bodies to that of the international standard to determine their mass in kilograms. For example, under a given net force, if an object's acceleration is half of that of the standard kilogram, the object has a mass of 2 kg.

In the context of Newton's Second Law, mass is a measure of the inertia of an object; that is, it is a measure of how that particular object resists a change in motion due to a force (we can think of a large acceleration as a large change in motion, as the velocity vector of the object will change more). For this reason, the mass that appears in Newton's Second Law is referred to as "inertial mass".

As you recall, the weight of an object is given by the mass of the object multiplied by the strength of the gravitational field, $\vec{g}$. There is no reason that the mass that is used to calculate weight, $F_g = mg$, has to be the same quantity as the mass that is used to calculate inertia $F = ma$. Thus, people will sometimes make the distinction between "gravitational mass" (the mass that you use to calculate weight and the force of gravity) and "inertial mass" as described above. Very precise experiments have been carried out to determine if the gravitational and inertial masses are equal. So far, experiments have been unable to detect any difference between the two quantities. As we will see, both Newton's Universal Theory of Gravity and Einstein Theory of General Relativity assume that the two are indeed equal. In fact, it is a key requirement for Einstein's Theory that the two be equal (the assumption that they are equal is called the "Equivalence Principle"). You should however keep in mind that there is no physical reason that the two are the same, and that as far as we know, it is a coincidence!

Unless stated otherwise, we will not make any distinction between gravitational and inertial mass and assume that they are equal. We will simply use the term "mass" and only clarify the type of mass when relevant (e.g. when we cover gravity).

5.4 Applying Newton's Laws

Now that we have introduced all of the concepts from Newton's Theory of Classical Physics, we present some general strategies for building models that use the theory. Recall that if we can describe the motion of all objects of interest to us, we have described everything that we can. Newton's Second Law allows us to determine the acceleration of an object based on the net force acting on the object. Once we have determined the accelerations of all objects of interest we have built a complete model.

The most important step in applying Newton's Theory is to identify the forces that are exerted on an object. The most important step in applying Newton's Theory is to identify the forces that are exerted on an object. The most important step in applying Newton's Theory is to identify the forces that are exerted on an object. Now that you have read it three times, you realize this step is important, right?!

The strategy for building a model for the motion of an object using Newton's Theory is straightforward:

1. Identify an inertial frame of reference in which to build the model.
2. Identify the forces acting on the object (did we mention that this step is important?).
3. Draw a free-body diagram.
4. Apply Newton's Second Law.

5.4.1 Identifying the forces

The first step in applying Newton's theory is to identify all of the forces that are acting on an object. This can be done by asking yourself: "what could possibly be pushing or pulling on the object?", as well as running through the list of forces that we enumerated in section 5.2.1 to identify if any of them are relevant here. For easy reference, we reproduce the types of forces here and include some questions that you might ask yourself to decide whether or not to include the corresponding force:

- Weight (is the object near the surface of a planet?).
- Normal forces (is the object in contact with any surface? There could be more than one!).
- Frictional forces (are there static or kinetic friction forces associated with the normal forces?).
- Tension forces (is something like a rope pulling on the object?).
- Drag forces (is the object moving through a fluid?).
- Spring forces (is there a spring pushing or pulling on the object?).
- Applied forces (is anything else pushing or pulling on the object?).

Example 5-1

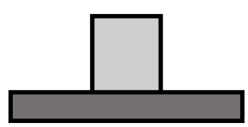


Figure 5.7: A block on a horizontal table.

A block of mass m is at rest on a horizontal table, as shown in Figure 5.7. What forces are exerted on the block?

Solution

The forces on the block are illustrated in Figure 5.8 and are:

1. $\vec{F}_g$, its weight.
2. $\vec{N}$, a normal force exerted by the plane. The normal force is perpendicular to the interface between the table and the block. It points upwards in "reaction" to the downwards force that the block exerts onto the table. The downwards force from the block onto the table is not shown, since that force is not exerted on the block but on the table.

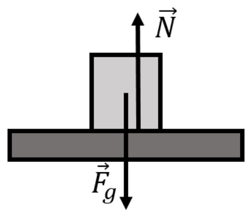


Figure 5.8: Forces on a block on a horizontal table.

Example 5-2

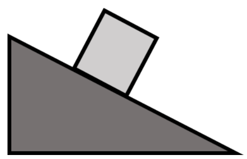


Figure 5.9: A block on an inclined surface.

A block of mass m is at rest on a inclined surface, as shown in Figure 5.9. What forces are exerted on the block?

Solution

The forces on the block are illustrated in Figure 5.10 and are:

1. $\vec{F}_g$, its weight.
2. $\vec{N}$, a normal force exerted by the inclined plane.
3. $\vec{f}_s$, a force of static friction exerted by the inclined plane. Without this force, the block would slide down. The force is in the direction opposite of impeding motion and is parallel to the interface (and perpendicular to the normal force).

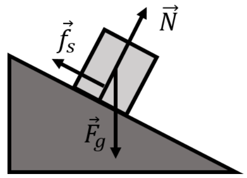


Figure 5.10: Forces on block on an inclined surface.

Example 5-3

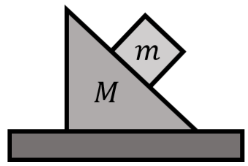


Figure 5.11: A block resting on a wedge-shaped block.

A block of mass m is at rest on a wedge-shaped block of mass M itself at rest on a horizontal table, as shown in Figure 5.11. What forces are exerted on each of the two blocks?

Solution

Since it will be too messy to draw all of the forces on the same diagram, we have drawn each block separately in Figure 5.12. Usually, when multiple blocks are stacked on each other, it is easiest to start with the forces on the top block. In this case, the top block is in the same condition as the block from Example 5-2. The forces exerted on the top block are:

1. $\vec{F}_g^m$, its weight.
2. $\vec{N}^m$, a normal force from the wedge-shaped block.
3. $\vec{f}_s^m$, a force of static friction exerted by the wedge-shaped block.

The forces exerted on the wedge-shaped block are:

1. $\vec{F}_g^M$, its weight.
2. $\vec{N}^M$, a normal force exerted by the small block. Note that this force is equal in magnitude and opposite in direction to $\vec{N}^m$ (the two forces, $\vec{N}^m$ and $\vec{N}^M$, which are on different objects, are an action/reaction pair of forces).
3. $\vec{f}_s^M$, a force of friction exerted by the small block (again, this forms an action/reaction pair of forces with $\vec{f}_s^m$).
4. N_2^M, a normal force exerted by the table.

The forces for both blocks are shown in Figure 5.12.

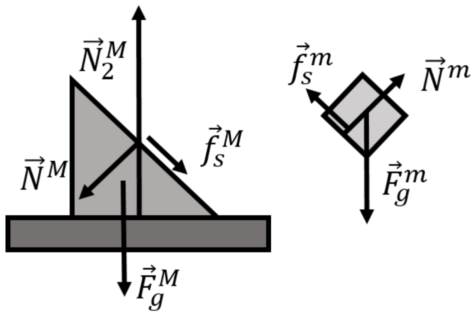


Figure 5.12: Forces on the block and the wedge-shaped block.

5.4.2 Free body diagrams

In order to analyse the forces on an object more clearly, it is a very good idea to draw a "Free-Body Diagram" (FBD). A free-body diagram is simply a diagram where we draw the forces on a single object and represent the object as a point. Because the object is a point, we do not worry where on the object the forces are exerted. In later chapters, we will see that for extended bodies, it does matter where the forces are applied. However, Newton's Laws as presented so far are only valid for objects that can be represented as a small point.

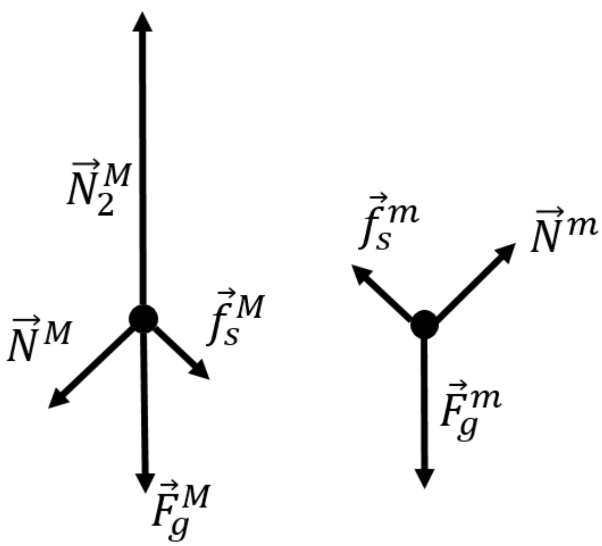


Figure 5.13: Free-body diagram for the block and the wedge-shaped block from Example 5-3.

In Example 5-3 above, we would draw one free-body diagram for each object (each mass), as shown in Figure 5.13.

Example 5-4

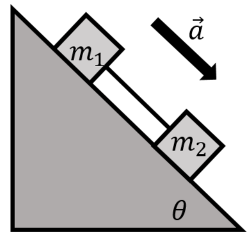


Figure 5.14: Two connected blocks sliding down an inclined plane.

Two blocks, of masses m_1 and m_2, are placed on an inclined plane that makes an angle θ with the horizontal. The blocks are connected by a massless string, as shown in Figure 5.14. The two blocks are sliding and accelerating downwards with an acceleration, $\vec{a}$, as shown. The coefficient of kinetic friction between the plane and either block is μ_k. Draw a free-body diagram for each block.

Solution

First, we identify the forces on each mass (each block), which we then use to make the free-body diagram shown in Figure 5.15. On mass m_1, the forces are:

1. $\vec{F}_{g1}$, its weight.
2. $\vec{N}_1$, a normal force exerted by the inclined plane.
3. $\vec{f}_{k1}$, a force of kinetic friction exerted by the inclined plane. The force is in the opposite direction of the motion, and has a magnitude given by $f_{k1} = \mu_k N_1$.
4. $\vec{T}$, a force of tension from the string.

On mass m_2, the forces are:

1. $\vec{F}_{g2}$, its weight.
2. $\vec{N}_2$, a normal force from the inclined plane.
3. $\vec{f}_{k2}$, a force of kinetic friction exerted by the inclined plane. The force is in the opposite direction of the motion, and has a magnitude given by $f_{k2} = \mu_k N_2$.
4. $-\vec{T}$, a force of tension from the string. This is the same force as on m_1, but in the opposite direction. We chose to label the force as $-\vec{T}$, instead of using a different variable, since it is just the negative of the vector that represents the tension force on m_1.

In Figure 5.15, we have shown the forces on each block using a free-body diagram. We also reproduced the vector for the acceleration (we drew the vector for the acceleration using a thicker arrow to indicate that it has a different dimension). We also reproduced the angle θ in the free-body diagram, as this is helpful once the free-body diagram is used with Newton's Second Law.

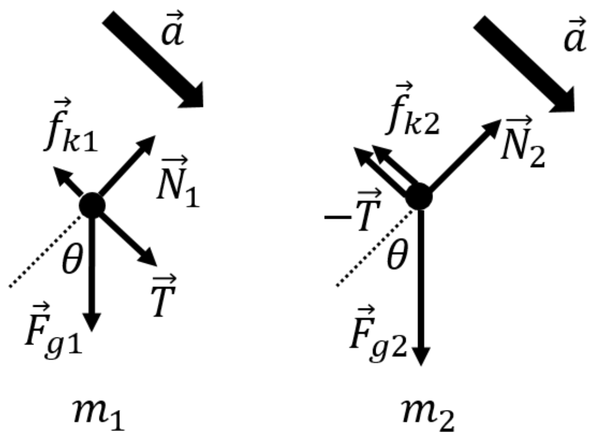


Figure 5.15: Free-body diagram for the blocks m_1 and m_2 from Figure 5.14.

5.4.3 Using Newton's Second Law

Applying Newton's Second Law is straightforward once all of the forces exerted on an object have been identified. You should thus make sure that you spend most of your time drawing a good and complete free-body diagram before proceeding.

Newton's Second Law is a vector equation that relates the vector sum of all forces exerted on an object and the acceleration vector of the object. This corresponds to one scalar equation per component of the vector.

$$\begin{aligned}\sum \vec{F} &= m\vec{a}\\ \sum F_x &= ma_x\\ \sum F_y &= ma_y\\ \sum F_z &= ma_z\end{aligned}$$

In order to use Newton's Second Law, we thus need to introduce a coordinate system so that we can work with the components of the vectors (forces and acceleration) in that coordinate system. Usually, a good choice of coordinate system is one where the x (or y) axis is parallel to the acceleration vector. Figure 5.16 shows the free-body diagram from the m_1 block from the previous example (Example 5-4) along with a good choice of coordinate system.

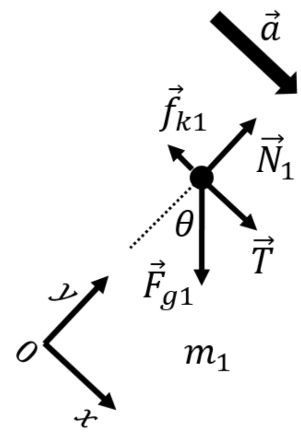


Figure 5.16: Free-body diagram and choice of coordinate system for the m_1 blocks from Figure 5.15, Example 5-4.

To apply Newton's Second Law using the free-body diagram and coordinate system from Figure 5.16, we first write out all of the vector and then identify their x and y components. The force vectors are:

$$\begin{aligned}
\vec{T} &= T\hat{x} + 0\hat{y} \\
\vec{f}_{k1} &= -f_{k1}\hat{x} + 0\hat{y} \\
\vec{F}_{g1} &= m_1 g(\sin\theta\hat{x} - \cos\theta\hat{y}) \\
\vec{N}_1 &= 0\hat{x} + N_1\hat{y}
\end{aligned}$$

We can now write out the x component of Newton's Second Law:

$$\begin{aligned}
\sum F_x &= T - f_{k1} - F_{g1}\sin\theta = m_1 a \\
\therefore T - f_{k1} - F_{g1}\sin\theta &= m_1 a
\end{aligned}$$

where we note that the normal force has no component in the x direction. The y component of Newton's Second Law for mass m_1 is given by:

$$\begin{aligned}
\sum F_y &= N_1 - F_{g1} = 0 \\
\therefore N_1 - F_{g1} &= 0
\end{aligned}$$

where we note that the forces of tension and friction have no y component. The two equations that we obtained above for x and y fully specify the motion of the m_1 block if all quantities are known[3].

A few notes on applying Newton's Second Law:

- When applying Newton's Second Law, analyze each mass in the problem separately. It does not matter that block m_1 is connected by a rope to block m_2. Once you have

[3] Since we have two equations, we technically only need to specify all but two quantities to be able to fully model the motion of the block.

determined all of the forces exerted on m_1, you can write Newton's Second Law for m_1.

- Newton's Second Law is a vector equation; this means that it is true for each (scalar) component of the vectors involved.
- You can choose the coordinate system, so choose one that makes it easy to write out the vector components. A good choice is to choose x to be parallel to the acceleration vector, so that you do not have to break the acceleration vector up into components. The choice of coordinate system is only made in order to allow you to write out the components of Newton's Second Law based on the free-body diagram.
- Treat each mass separately (since Newton's Second Law is only true for an individual mass). This means that each mass will have its own free-body diagram and that you can choose the coordinate system that is most convenient for a given free-body diagram. In particular, this means that you do not need to choose the same coordinate system for different masses in a problem.

The following example shows how to write Newton's Second Law for a system of two blocks.

Example 5-5

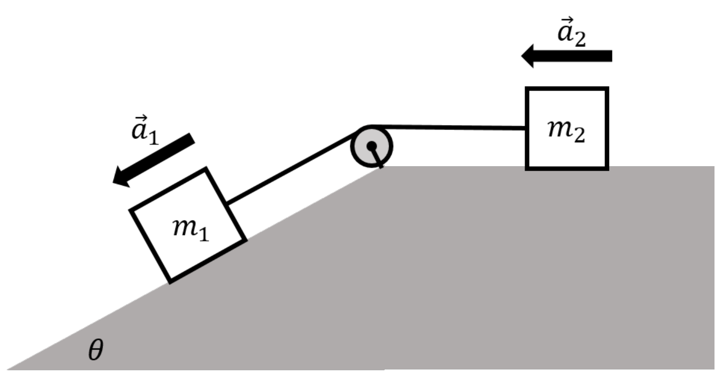


Figure 5.17: Two blocks connected by a massless string and massless pulley. Both blocks are accelerating.

A block of mass m_1 is placed on an incline that makes an angle of θ with the horizontal. The block of mass m_1 is connected by a massless string through a massless pulley to a second block of mass m_2, which rests on a horizontal surface. The blocks are accelerating in such a way that the block of mass m_1 is accelerating down the incline, as shown in Figure 5-5. The coefficient of kinetic friction between either block and the surface it is resting on is μ_k. Write Newton's Second Law for both blocks.

Solution

First, we identify the forces on each mass (each block). On mass m_1, the forces are:

1. $\vec{F}_{g1}$, its weight.
2. $\vec{N}_1$, a normal force exerted by the inclined plane.
3. $\vec{f}_{k1}$, a force of kinetic friction exerted by the inclined plane. The force is in the opposite direction of the motion, and has a magnitude given by $f_{k1} = \mu_k N_1$.
4. $\vec{T}_1$, a force of tension from the string.

On mass m_2, the forces are:

1. $\vec{F}_{g2}$, its weight.
2. $\vec{N}_2$, a normal force from the horizontal surface.
3. $\vec{f}_{k2}$, a force of kinetic friction exerted by the horizontal surface. The force is in the opposite direction of the motion, and has a magnitude given by $f_{k2} = \mu_k N_2$.
4. $\vec{T}_2$, a force of tension from the string. This force has the same magnitude as the tension force $\vec{T}_1$ exerted on mass m_1, because the pulley is massless.

We can then proceed to draw the free-body diagram for each mass, and use that to write out Newton's Second Law. For mass m_1, the free-body diagram is shown in Figure 5.18. We have chosen a coordinate system that has the x axis parallel to the acceleration of the block, and the y axis upwards and perpendicular to the x axis, as shown.

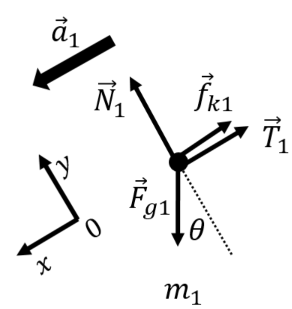


Figure 5.18: Free-body diagram for m_1.

For m_1, we can write Newton's Second Law, starting with the x components:

$$\begin{aligned}\sum F_x = F_{g1}\sin\theta - f_{k1} - T_1 = m_1 a_1 \\ \therefore m_1 g\sin\theta - \mu_k N_1 - T_1 = m_1 a_1\end{aligned}$$

where, in the second line, we used the magnitude of the weight ($F_{g1} = m_1 g$) and of the force of kinetic friction ($f_{k1} = \mu_k N_1$). For the y component of Newton's Second Law, in which the acceleration has no component, we have:

$$\begin{aligned}\sum F_y = N_1 - F_{g1}\cos\theta = 0 \\ \therefore N_1 = m_1 g\cos\theta\end{aligned}$$

which shows us that the magnitude of the normal force can easily be expressed in terms of the weight ($F_{g1} = m_1 g$) and the angle of the incline.

For m_2, we can proceed in much the same way, choosing a different coordinate system, since the acceleration vector for m_2 points in a different direction (we don't have to choose a different coordinate system, but we can if we find it makes things easier). The free-body diagram for m_2 is shown in Figure 5.19 along with our choice of coordinate system.

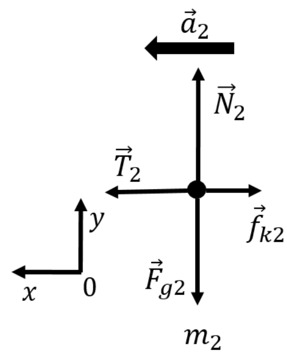


Figure 5.19: Free-body diagram for m_2.

We start by writing out the x component of Newton's Second Law for m_2:

$$\sum F_x = T_2 - f_{k2} = m_2 a_2$$
$$\therefore T_2 - \mu_k N_2 = m_2 a_2$$

where again, we expressed the kinetic force of friction using the normal force and the coefficient of kinetic friction. The y component of Newton's Second Law gives:

$$\sum F_y = F_{g2} - N_2 = 0$$
$$\therefore N_2 = m_2 g$$

where again, we expressed the weight in terms of the mass and g, and we find that the normal force has the same magnitude as the weight.

Now that we have written Newton's Second Law **for each mass**, we can write all four equations that we obtained to describe **the system of two masses**. We should also note that the magnitude of the tension forces are the same for the two masses ($T_1 = T_2 = T$), and that since the masses are connected by a string, the magnitude of their acceleration vectors are the same ($a_1 = a_2 = a$). Using this, we can describe the full system with the following 4 equations:

$$\begin{aligned}
m_1 g \sin\theta - \mu_k N_1 - T &= m_1 a \\
N_1 &= m_1 g \cos\theta \\
T - \mu_k N_2 &= m_2 a \\
N_2 &= m_2 g
\end{aligned}$$

Of the variables above (m_1, m_2, μ_k, T, N_1, N_2, a), one would only need to specify all but four of them to fully describe the motion of the system. For example, if one specifies the two masses and the coefficient of kinetic friction, all of the other variables can be determined.

5.5 The acceleration due to gravity

If you have studied some physics before reading this textbook, you may have been surprised by our choice of dimension for g to be force per unit mass rather than acceleration. This is indeed an unconventional choice as g is usually presented as "the acceleration due to Earth's gravity" instead of the "strength of Earth's gravitational field". Our choice comes from the potential difference between inertial mass, m_I, and gravitational mass, m_G, which we distinguish in this section.

Consider the simple model of a mass falling freely near the surface of the Earth in the absence of air-resistance. The only force exerted on the mass is its weight, $m_G\vec{g}$, which is given in terms of gravitational mass (the mass that determines how an object experiences gravity). Both the weight and the acceleration of the object point downwards. The free-body diagram for the mass is shown in Figure 5.20, where the y axis was chosen to be vertically upwards (co-linear with the acceleration).

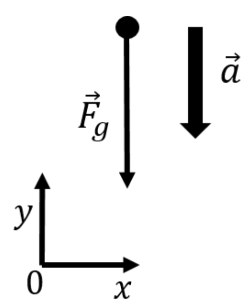


Figure 5.20: Free-body diagram for a mass that is free-falling in the absence of air resistance (drag).

Writing out the y component of Newton's Second Law, being careful to distinguish between inertial and gravitational mass, and noting that both the weight and the acceleration are in the negative y direction:

$$\begin{aligned}
\sum F_y = -F_g &= -m_I a \\
\therefore m_G g &= m_I a
\end{aligned}$$

This makes it clear that g is not necessarily the acceleration due to gravity. It is only the acceleration due to gravity in the limit that the inertial and gravitational masses are the same. If $m_G = m_I$, then we have:

$$a = g$$

and indeed, the acceleration of objects near the surface of the Earth has a magnitude of g. It is also clear that the dimensions of g can also be written as an acceleration, and in most cases, one writes that, near the surface of the Earth, $g = 9.8\,\text{m/s}^2$. You should however remember that this is only true when inertial and gravitational masses are the same, and that g really should be thought of as the strength of the gravitational field, not as an acceleration.

5.6 Non-inertial frames of reference and inertial forces

In the previous sections, we described how to use Newton's First Law to identify an inertial frame of reference (one where Newton's First Law holds true) in order to identify the forces exerted on an object so that Newton's Second Law could be applied. It is possible to apply Newton's Laws in a non-inertial frame of reference, **provided that one includes an additional "inertial force".**

Let us assume that we hang a mass, m, from the ceiling of our car using a string. If the car accelerates forwards with a constant acceleration, $\vec{a}$, the mass will swing towards the back of the car and the string will not be vertical as long as the car maintains its constant acceleration, as shown in Figure 5.21. As the car maintains its acceleration, the hanging mass will not move relative to the car.

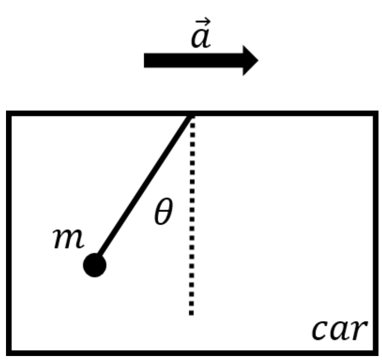


Figure 5.21: A mass hanging from the ceiling of a car accelerating to the right.

We can analyse this motion from the inertial frame of reference of the ground. In this frame of reference, there are two forces exerted on the mass:

1. $\vec{F}_g$, its weight, with magnitude mg.
2. $\vec{T}$, a force of tension exerted by the string, in the direction of the string.

The two forces are shown in the free-body diagram of Figure 5.22, along with a coordinate system chosen such that x points in the direction of the acceleration the mass (which is the same as the acceleration of the car, since the mass does not move relative to the car).

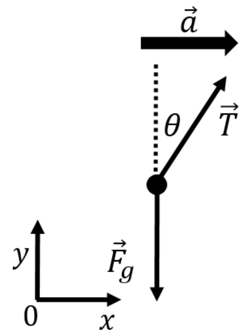


Figure 5.22: Free-body diagram for the forces acting on a mass suspended from the ceiling of accelerating car.

Writing out the x and y components of Newton's Second Law for the mass, we have:

$$\begin{aligned}\sum \vec{F} &= \vec{T} + \vec{F}_g = m\vec{a}\\ \therefore \sum F_x &= T\sin\theta = ma\\ \therefore \sum F_y &= T\cos\theta - F_g = 0\end{aligned}$$

We can, instead, model the motion of the mass in the frame of reference of the car, by pretending that we are sitting in the car. In the frame of reference of the car, the mass is immobile, and thus has no acceleration. In the non-inertial frame of reference of the car, we still have the weight and tension forces exerted on the mass; these have the same magnitude and direction as in the inertial frame of reference of the ground. One could replace the string with a spring scale, and observers in the car and on the ground would agree that the spring scale reads the same number. Those observers would also agree that the weight of the mass is the same. However, the two observers disagree on whether the mass is accelerating, since the observer in the car measures that the mass has no acceleration.

In the frame of reference of the car, the acceleration of the mass is zero. If we want Newton's Second Law to hold, this implies that, in the reference frame of the car, the sum of the forces on the mass must be zero:

$$\sum \vec{F} = 0 \qquad \text{(car reference frame)}$$

We know from analysing the motion from the frame of reference of the ground that the vector sum of the forces $\vec{T}$ and $\vec{F}_g$ is equal to $m\vec{a}$. The only way for the force in the frame of reference of the car to add up to zero is if there is an additional force, $\vec{F}_I$, that is exerted in that frame of reference:

$$\sum \vec{F} = \vec{T} + \vec{F}_g + \vec{F}_I = 0 \qquad \text{(car reference frame)}$$

Since we know that $\vec{T} + \vec{F}_g = m\vec{a}$, we can substitute this in the equation above:

$$\begin{aligned}\sum \vec{F} &= \vec{T} + \vec{F}_g + \vec{F}_I = 0 \qquad \text{(car reference frame)}\\ &= m\vec{a} + \vec{F}_I = 0\\ \therefore F_I &= -m\vec{a}\end{aligned}$$

and we find that this "inertial force", $\vec{F}_I$, must be exerted in the opposite direction from the acceleration of the frame of reference, with a magnitude given by ma. The free-body diagram for the mass, as viewed in the reference frame of the car, is illustrated in Figure 5.23.

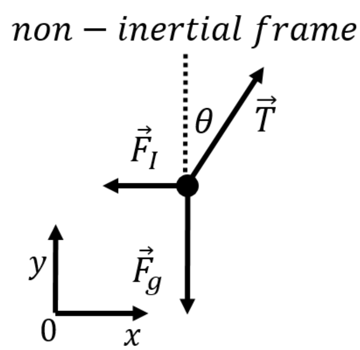


Figure 5.23: Free-body diagram for the forces acting on a mass suspended from the ceiling of accelerating car, in the frame of reference of the car. An additional inertial force, $\vec{F}_I = -m\vec{a}$, has to be included.

Example 5-6

You are in an elevator that is accelerating downwards with a constant acceleration $\vec{a}$. You are standing on a spring scale. What is the value of your weight as displayed on the spring scale? Assume that your mass is m. (The spring scale will display your weight as having the same magnitude as the normal force that the scale exerts on you).

Solution

We can model your motion in the non-inertial frame of reference of the elevator, where your acceleration is zero. The forces that are exerted on you are:

1. $\vec{F}_g$, your weight, with magnitude mg.
2. $\vec{N}$, the normal force exerted upwards by the spring scale, which is the weight as measured by the scale.
3. $\vec{F}_I$, an inertial force with magnitude ma that is exerted upwards (in the direction opposite of the acceleration of the frame of reference).

The forces in the frame of reference of the elevator are illustrated in Figure 5.24, along with a coordinate system that was chosen so that the forces are co-linear with one of the axes (since the acceleration is zero).

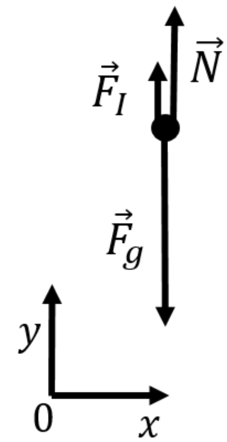


Figure 5.24: Free-body diagram for the forces exerted on a person as modelled in a frame of reference that is accelerating downwards.

All of the forces are in the vertical direction, so we only need to write out the y component of Newton's Second Law, which we can easily solve for the normal force:

$$\begin{aligned}\sum F_y &= N + F_I - F_g = 0\\ N + ma - mg &= 0\\ \therefore N &= m(g-a)\end{aligned}$$

Remember that you need to be careful about the signs. We have included the fact that F_I is exerted upwards with the plus sign in the first equation (the y component of $\vec{F}_I = 0\hat{x} + F_I\hat{y}$ is $+F_I$). We then used the fact that the magnitude of the inertial force is given by $F_I = ma$ in the second line.

You can easily verify that you would obtain the same result in the inertial frame of reference of the ground, where there is no inertial force, but the acceleration is non-zero (and in the negative y direction if we use the same coordinate system):

$$\sum F_y = N - mg = -ma \qquad \text{(ground frame of reference)}$$

The normal force, which corresponds to weight as read by the scale, is thus $N = m(g-a)$. We should ask ourselves if the result makes sense:

- Since the dimension of a and g are the same, $m(g-a)$ has the correct dimension of force.
- If the acceleration, a, is zero, then the magnitude is $N = mg$, as it should be if the elevator is at rest with respect to the ground.

- If the acceleration a is equal to g, the normal force exerted by the scale is exactly zero, and your measured weight is zero. This is what we call being "weightless", which is not a good description, since the force of weight is still applied, and it is the normal force which is zero.
- If the acceleration, a, is bigger than g, then the normal force would be negative. This corresponds to the elevator accelerating downwards faster than gravity, and the model breaks down, since in this case, you would first hit the ceiling of the elevator, which would then exert a downwards normal force with magnitude $m(a+g)$.

5.7 Summary

Key Takeaways

Newton's Three Laws are a theory of classical physics that allow the motion of an object to be fully described by introducing the concepts of force and mass.

Newton's First Law states that objects will not accelerate if no net force is exerted on the object. In particular, this allows inertial frames of reference to be defined as those frames of reference where Newton's First Law holds true.

Newton's Second Law connects dynamics and kinematics by relating the net force exerted on an object (i.e. the vector sum of the forces exerted on an object) to its acceleration and its mass:

$$\vec F^{net} = \sum_i \vec F_i = m\vec a$$

Newton's Third law states that forces always come in pairs that are exerted on different objects. If object A exerts a force on object B, then object B exerts a force that is equal in magnitude but opposite in direction on object A.

A force is a mathematical tool introduced in Newton's theory to model how different objects can influence each other. Mass can be thought of as a quantity of matter and is an intrinsic property of an object. Inertial mass refers to how that quantity of matter resists acceleration, whereas gravitational mass refers to how that quantity of mass experiences the force of gravity. As far as we can tell, inertial and gravitational mass are the same.

When applying Newton's theory, the most important part is to identify the forces that act on one object. This can be represented graphically by using a free-body diagram. The following is a common list of forces to consider when identifying the forces exerted on an object:

- Weight (is the object near the surface of a planet?).
- Normal forces (is the object in contact with any surface? There could be more than one!).
- Frictional forces (are there static or kinetic friction forces associated with the normal forces?).
- Tension forces (is something like a rope pulling on the object?).
- Drag forces (is the object moving through a fluid?).
- Spring forces (is there a spring pushing or pulling on the object?).
- Applied forces (is anything else pushing or pulling on the object?).

When applying Newton's Second Law, one needs to choose a coordinate system so that Newton's Second Law can be written out for each component. It is usually good to choose the coordinate system such that the x axis is parallel to the acceleration vector of the object.

When using Newton's Laws to model the motion of an object of mass m in a non-inertial frame of reference that is accelerating with acceleration $\vec{a}$ relative to an inertial frame of reference, an additional inertial force, $\vec{F}_I = -m\vec{a}$, must be included on the the object.

Important Equations

Newton's Second Law, in vector form, can be written as:

$$\sum \vec{F} = m\vec{a}$$

which is just a short-hand notation for the scalar equations written out for each component:

$$\sum F_x = ma_x$$
$$\sum F_y = ma_y$$
$$\sum F_z = ma_z$$

The force of gravity (or weight), $\vec{F}_g$, near the surface of the Earth is given by:

$$\vec{F}_g = m\vec{g}$$

where Earth's gravitational field has a magnitude of $g = 9.8\,\text{N/kg}$.

The force of kinetic friction exerted by one surface on another is given by::

$$f_k = \mu_k N$$

where N is the normal force between the two surfaces and μ_k is the coefficient of kinetic friction. The force of kinetic friction on a object is in the opposite direction from its motion.

The maximum value of the magnitude of the force of static friction between two surfaces with a coefficient of static friction μ_s between them, can be written as:

$$f_s \leq \mu_s N$$

The force of static friction is exerted in the direction opposite of the impeding motion.

Hookes' Law for the force exerted by a spring, is given by the following vector equation:

$$\vec{F} = -kx\hat{x}$$

where x is the distance by which the spring is compressed or extended relative to its rest length.

5.8 Thinking about the material

Reflect and research

1. What was the name of the publication in which Newton's published his three laws, and when was it published?
2. When did Galileo come up with his principle of inertia?
3. Suppose that Newton grew up in an accelerating train, with no knowledge that he is living in an accelerating train. What would Newton's first law look like in this world?
4. When you skate on ice, there is kinetic friction between your skates and the ice. Does the coefficient of kinetic friction depend on the temperature of the ice? If yes, what is the optimal temperature for skating with the least amount of friction?

To try at home

1. Place two books stacked on each other on the palm of one hand held horizontally. Use your other hand to press down (and forward) on the top book and try to move the bottom book. No matter how hard you push down (to increase the force of friction between the two books), you cannot make the bottom one move. How come?

To try in the lab

1. Propose an experiment to determine whether gravitational and inertial mass are equal.
2. Propose an experiment to measure the coefficients of static and kinetic friction between a block and a surface.

5.9 Sample problems and solutions

5.9.1 Problems

Problem 5-1:

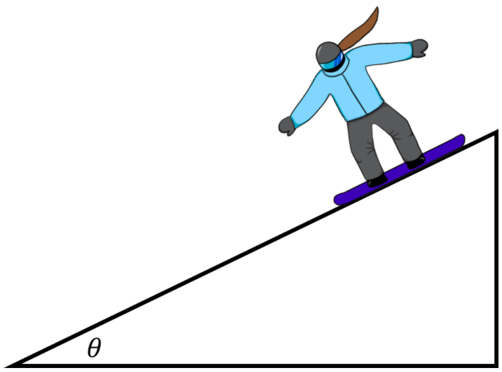


Figure 5.25: Katie snowboarding down an incline.

Katie, an amateur snowboarder, rests at the top of hill inclined by an angle of $\theta = 50°$ with respect to the horizontal, as shown in Figure 5.25. She gracefully slides down the hill until she face-plants into a large pile of snow at the bottom, 40 m from where she started. If the coefficient of kinetic friction between Katie's snowboard and the hill is $\mu_k = 0.45$, how long elapses between when she starts to glide and when she face plants? (Solution)

Problem 5-2:

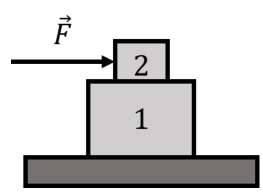


Figure 5.26: Two stacked boxes.

Two boxes with masses, m_1 and m_2, respectively, are placed on top of one another, as shown in Figure 5.26. The coefficient of static friction between the two boxes and between the boxes and the ground is $\mu_s = 0.3$. A constant force, $\vec{F}$, is exerted on box 2, as shown. Show that it is impossible for box 1 to accelerate. (Solution)

5.9.2 Solutions

Solution to problem 5-1: Before trying to solve the problem, we should think of the strategy that will allow us to model the time that it takes to arrive at the bottom. We know that Newton's Second Law relates the forces on Katie to her acceleration. If we build a model of the forces on Katie, we can then determine her acceleration. Once we know her acceleration, we can use kinematics to determine how long it takes for her to cover the distance of 40 m.

The forces exerted on Katie are:

1. $\vec{F}_g$, her weight.
2. $\vec{N}$, a normal force exerted by the slope.
3. $\vec{f}_k$, a force of kinetic friction exerted by the slope, with magnitude $f_k = \mu_k N$

This allows us to build a free-body diagram for the forces on Katie, as shown in Figure 5.27. Since Katie will glide down the slope, her acceleration will be parallel to the slope and downwards, which we showed with a thicker arrow on the free-body diagram. Our free-body diagram also shows the coordinate system that we chose, with the x axis pointing parallel to the acceleration.

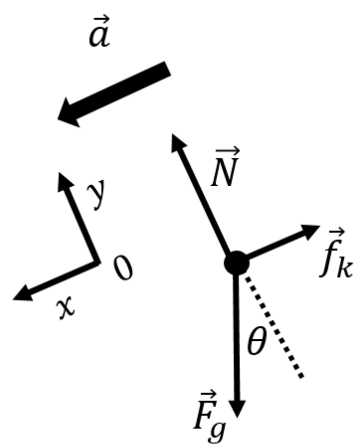


Figure 5.27: Forces acting on Katie as she snowboards.

With a free-body diagram, we can write the x and y components of Newton's Second Law. In the x direction, both the force of friction and the weight have components. The force of friction is in the negative x direction, whereas the component of gravity in the x direction is $F_g \sin\theta$. The acceleration vector is also in the x direction. Putting this altogether into Newton's Second Law:

$$\begin{aligned}\sum F_x = F_g \sin\theta - f_k = ma \\ \therefore mg \sin\theta - \mu_k N = ma\end{aligned}$$

where we used the fact that the weight is given by mg (m is Katie's mass) and the magnitude of the force of friction is given by $f_k = \mu_k N$.

Next, we write out the y component of Newton's Second Law. The normal force is in the positive y direction, whereas the component of gravity in the y direction is $-F_g \cos\theta$. The acceleration has no component in the y direction. Putting this into Newton's Second Law:

$$\begin{aligned}\sum F_y &= N - F_g \cos\theta = 0 \\ \therefore N &- mg\cos\theta = 0\end{aligned}$$

We now have two equations that describe Katie's motion:

$$\begin{aligned}mg\sin\theta - \mu_k N &= ma \\ N - mg\cos\theta &= 0\end{aligned}$$

We have three unknowns, m, N, and a, but only two equations! Hopefully, one of these will cancel out! At this point, all of the physics for the problem is done! We can now proceed to solve these equations to find the acceleration. The second equation allows us to solve for the normal force, $N = mg\cos\theta$, which we substitute into the first equation:

$$\begin{aligned}mg\sin\theta - \mu_k N &= ma \\ \therefore mg\sin\theta - \mu_k mg\cos\theta &= ma\end{aligned}$$

As you can see, the mass m can be cancelled out of this equation, and we can find the acceleration:

$$\begin{aligned}a &= g\sin\theta - \mu_k g\cos\theta \\ &= g(\sin\theta - \mu_k\cos\theta) \\ &= (9.8\,\text{N/kg})\,(\sin(50^\circ) - (0.45)\cos(50^\circ)) \\ &= 4.67\,\text{N/kg}\end{aligned}$$

At this point, we should ask ourselves if our result makes sense. In particular, we have found that the acceleration has unit of N/kg instead of m/s^2. A quick examination of Newton's Second Law shows us that these two units are equivalent:

$$\begin{aligned}F &= ma \\ a &= \frac{F}{m} \\ \therefore SI[a] &= \frac{SI[F]}{SI[m]} = \frac{\text{N}}{\text{kg}}\end{aligned}$$

Often, one write the magnitude of the Earth's gravitation field as $g = 9.8\,\text{m/s}^2$, since it has the same dimension as acceleration, and does indeed correspond to the acceleration that is felt by falling objects near the surface of the Earth. In fact, g, is usually defined as the acceleration of object near the Earth, although this is misleading, as it requires that inertial and gravitational mass be the same.

Knowing that Katie's initial velocity is $v_{0x} = 0\,\text{m/s}$, her acceleration is $a_x = a = 4.67\,\text{m/s}^2$ in the x direction (the same direction as the slope), and the distance that she must travel is $x = 40\,\text{m}$, we can find the time it takes for her to face-plant. If we set the origin of the x axis where she starts (so that her initial position along the x axis, $x_0 = 0$), the distance that she covered in the time, t, is given by:

$$
\begin{aligned}
x(t) &= x_0 + v_{0x}t + \frac{1}{2}at^2 \\
40\,\text{m} &= (0) + (0)t + \frac{1}{2}(1.31\,\text{m/s}^2)t^2 \\
\therefore t &= \sqrt{\frac{2(40\,\text{m})}{(4.67\,\text{m/s}^2)}} = 4.14\,\text{s}
\end{aligned}
$$

Katie has 4.14 s of gliding bliss before face-planting into the large pile of snow.

Solution to problem 5-2: The only way for box 1 to accelerate is if box 2 "drags" box 1 along with it through a force of friction exerted at the interface between box 1 and box 2. We need to show that the force of (static) friction exerted by the ground on box 1 will always be at least as large as the force of friction exerted by box 2 on box 1. The largest force of friction that box 2 can exert on box 1 is a force of static friction, so we model all forces between surfaces as forces of static friction.

The forces on box 2 are:

- $\vec{F}_{2g}$, its weight.
- $\vec{N}_2$, a normal exerted by box 1.
- $\vec{f}_{2s}$, a force of static friction exerted by box 1.
- $\vec{F}$, the applied force.

The forces on box 1 are:

- $\vec{F}_{1g}$, its weight.
- $-\vec{N}_2$, a normal force exerted by box 2 (downwards).
- $-\vec{f}_{2s}$, a force of static friction exerted by box 2.
- $\vec{N}_1$, a normal force exerted by the ground.
- $\vec{f}_{1s}$, a force of static friction exerted by the ground.

The are illustrated in the free-body diagram in Figure 5.28

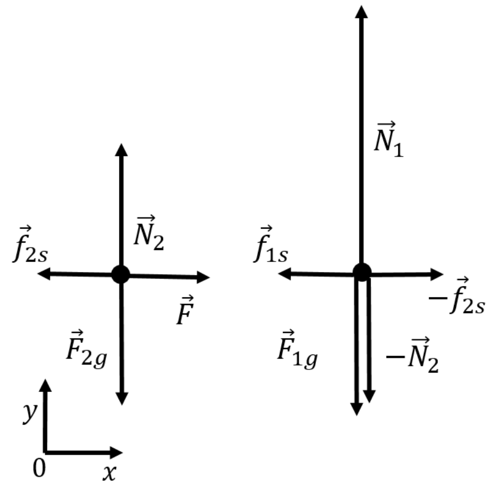


Figure 5.28: Forces on the two boxes.

Considering the y component of Newton's Second Law for box 2 (the top box), we can find the value of the normal force exerted by box 1:

$$\begin{aligned}\sum F_y &= N_2 - F_{2g} = 0\\ \therefore N_2 &= m_2 g\end{aligned}$$

The maximal magnitude of the force of static friction, f_{2s}, between the two boxes is given by:

$$f_{2s} = \mu_s N_2 = \mu_s m_2 g$$

This is the maximal magnitude of the force that can accelerate box 1. Considering the y component of Newton's Second Law applied to box 1, we can find N_1, the normal force exerted by the ground:

$$\begin{aligned}\sum F_y &= N_1 - F_{1g} - N_2 = 0\\ \therefore N_1 &= F_{1g} + N_2 = (m_1 + m_2)g\end{aligned}$$

The force of static friction exerted by the ground on box 1 will be in the opposite direction as the force of static friction exerted by box 2. The maximal magnitude of the force of static friction exerted by the ground is given by:

$$f_{1s} = \mu_s N_1 = \mu_s (m_1 + m_2)g$$

We can see that the maximal force of static friction exerted by the ground will always exceed the magnitude of the force of static friction exerted by box 2. It is thus impossible to push on box 2 to make box 1 move (as long as the force of static friction between the two boxes and the box and the ground are the same).

6 Applying Newton's Laws

In this chapter, we take a closer look at how to use Newton's Laws to build models to describe motion. Whereas the previous chapter was focused on identifying the forces that are acting on an object, this chapter focuses on using those forces to describe the motion of the object.

Newton's Laws are meant to describe "point particles", that is, objects that can be thought of as a point and thus have no orientation. A block sliding down a hill, a person on a merry-go-round, a bird flying through the air can all be modelled as point particles, as long as we do not need to model their orientation. In all of these cases, we can model the forces on the object using a free-body diagram as the location of where the forces are applied on the object do not matter. In later chapters, we will introduce the tools required to apply Newton's Second Law to objects that can rotate, where we will see that the location of where a force is exerted matters.

Learning Objectives

- Understand when an object's motion can be modelled as one dimensional (linear).
- Be able to develop models for objects undergoing linear motion.
- Be able to develop models for objects undergoing circular motion.
- Be able to develop models for objects undergoing arbitrary three dimensional motion.
- Understand the forces involved in circular motion, and understand that "centripetal" and "centrifugal" forces are not really forces.

Think About It

If a person swings on a swing where the ropes are damaged, where are the ropes most likely to break?

A) at the bottom of the trajectory, when the speed is the greatest.
B) at the top of the trajectory, when the speed is zero.
C) at the point in the trajectory where the speed is one half of its maximal value.

6.1 Statics

When using Newton's Laws to model an object, one can identify two broad categories of situations: static and dynamic. In static situations, the acceleration of the object is zero. By Newton's Second Law, this means that the vector sum of the forces (and torques, as we will see in a later chapter) exerted on an object must be zero. In dynamic situations, the acceleration of the object is non-zero.

For static problems, since the acceleration vector is zero, we can choose a coordinate system in a way that results in as many forces as possible being aligned with the axes (so that we minimize the number of forces that we need to break up into components).

Example 6-1

You push horizontally with a force $\vec F$ on a box of mass m that is resting against a vertical wall, as shown in Figure 6.1. The coefficient of static friction between the wall and the box is μ_s. What is the minimum magnitude of the force that you must exert for the box to remain stationary?

Solution

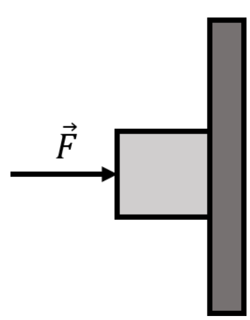


Figure 6.1: A horizontal force exerted on box that is resting against a wall.

Since the acceleration of the box is zero, the vector sum of the forces exerted on the box is zero. We start by identifying the forces exerted on the box; these are:

1. $\vec{F}$, the horizontal force that you exert on the box.
2. $\vec{F}_g$, the weight of the box, with magnitude mg.
3. $\vec{N}$, a normal force exerted by the wall on the box. The force is in the horizontal direction, in the opposite direction to $\vec{F}$.
4. $\vec{f}_s$, a vertical force of static friction between the wall and the box. The force points upwards as the "impeding motion" of the block is downwards. The force will have at most a magnitude of $f_s \leq \mu_s N$, since the force of static friction depends on the other forces exerted on the object.

The forces are shown in the free-body diagram in Figure 6.2, along with our choice of coordinate system which was chosen so that all forces are either in the x or y direction.

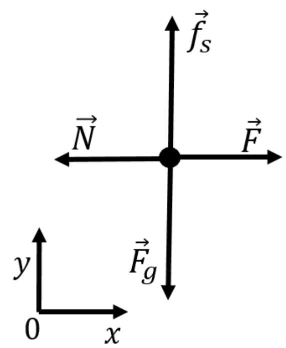


Figure 6.2: Free-body diagram of the forces exerted on the box.

The x component of Newton's Second Law is:

$$\begin{aligned}\sum F_x &= F - N = 0\\ \therefore N &= F\end{aligned}$$

which tells us that the normal force exerted by the wall has the same magnitude as the applied force, $\vec{F}$. The y component of Newton's Second Law is:

$$\begin{aligned}\sum F_y &= f_s - F_g = 0\\ \therefore f_s - mg &= 0\\ \therefore f_s &= mg\end{aligned}$$

which tells us that the force of friction must have the same magnitude as the weight. This makes sense, since they are the only forces with components in the y direction, and thus, they must cancel each other out.

The force of friction will be less than or equal to $\mu_s N$, and thus less than or equal to $\mu_s F$, since $\vec{F}$ and $\vec{N}$ have the same magnitude (from the x component of Newton's

Second Law). Furthermore, since $f_s = mg$, we can write:

$$\begin{aligned} f_s &\leq \mu_s F \\ \therefore mg &\leq \mu_s F \\ \therefore \frac{mg}{\mu_s} &\leq F \end{aligned}$$

which gives us the condition that $F \geq mg/\mu_s$, and thus the minimum magnitude of F in order to keep the box from sliding down.

Although we used the lesser than or equal to sign in the above equations, we could have used an equal sign if we were confident that the force of friction has its maximal magnitude, $f_s = \mu_s N$. The maximal magnitude of the force of friction is proportional to the force that we exert (since $N = F$); if we want to exert the least amount of force F, then we need the force of friction to be equal to its maximal magnitude which needs to be equal to the weight of the box.

Discussion: This model for the minimal required force makes sense because:

- The dimension of mg/μ_s is force.
- If the mass of the box is increased, then one needs to push harder against the box to keep it up.
- If the coefficient of static friction, μ_s, is increased, one does not need to push as hard.

6.2 Linear motion

We can describe the motion of an object whose *velocity vector does not continuously change direction* as "linear" motion. For example, an object that moves along a straight line in a particular direction, then abruptly changes direction and continues to move in a straight line can be modelled as undergoing linear motion over two different segments (which we would model individually). An object moving around a circle, with its velocity vector continuously changing direction, would not be considered to be undergoing linear motion. For example, paths of objects undergoing linear and non-linear motion are illustrated in Figure 6.3.

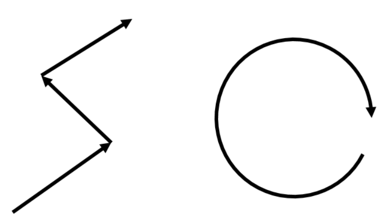


Figure 6.3: (Left:) Displacement vectors for an object undergoing three segments that can each be modelled as linear motion. (Right:) Path of an object whose velocity vector changes continuously and cannot be considered as linear motion.

When an object undergoes linear motion, we always model the motion of the object over straight segments separately. Over one such segment, the acceleration vector will be co-linear with the displacement vector of the object (parallel or anti-parallel - note that the acceleration can change direction as it would from a spring force, but will always be co-linear with the displacement).

Example 6-2

A block of mass m is placed at rest on an incline that makes an angle θ with respect to the horizontal, as shown in Figure 6.4. The block is nudged slightly so that the force of static friction is overcome and the block starts to accelerate down the incline. At the bottom of the incline, the block slides on a horizontal surface. The coefficient of kinetic friction between the block and the incline is μ_{k1}, and the coefficient of kinetic friction between the block and horizontal surface is μ_{k2}. If one assumes that the block started at rest a distance L from the bottom of the incline, how far along the horizontal surface will the block slide before stopping?

Solution

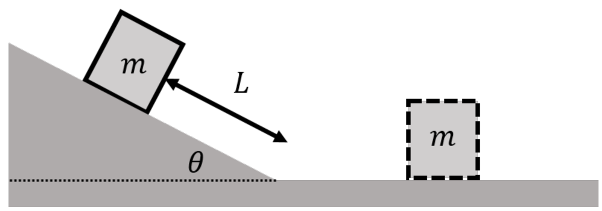


Figure 6.4: A block slides down an incline before sliding on a flat surface and stopping.

We can identify that this is linear motion that we can break up into two segments: (1) the motion down the incline, and (2), the motion along the horizontal surface. We will thus identify the forces, draw the free-body diagram for the block, and use Newton's Second Law twice, once for each segment.

It is often useful to describe the motion in words to help us identify the steps required in building a model for the block. In this case we could say that:

1. The block slides down the incline and accelerates in the direction of motion. By identifying the forces and applying Newton's Second Law, we can determine its acceleration which will be parallel to the incline.
2. The block will reach a certain speed at the bottom of the incline, which we can determine from kinematics by knowing that the block travelled a distance L, with a known acceleration and that it started at rest.
3. The block will decelerate along the horizontal surface. Again, by identifying the forces and using Newton's Second Law, we will be able to determine the

acceleration of the block.

4. The block will stop after having travelled an unknown distance, which we can find by using kinematics and knowing the acceleration of the block as well as its initial velocity at the bottom of the incline.

Our first step is thus to identify the forces on the block while it is on the incline. These are:

1. $\vec{F}_g$, its weight.
2. $\vec{N}_1$, a normal force exerted by the incline.
3. $\vec{f}_{k1}$, a force of kinetic friction exerted by the incline. The force is opposite of the direction of motion, and has a magnitude given by $f_{k1} = \mu_{k1} N_1$.

These are shown on the free-body diagram in Figure 6.5. As usual, we drew the acceleration, $\vec{a}_1$, on the free-body diagram, and chose the direction of the x axis to be parallel to the acceleration.

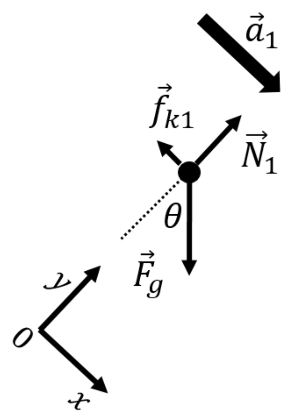


Figure 6.5: Free-body diagram for the block when it is on the incline.

Writing out the x component of Newton's Second Law, and using the fact that the acceleration is in the x direction ($\vec{a} = a_1\hat{x}$):

$$\sum F_x = F_g \sin\theta - f_{k1} = ma_1$$
$$\therefore mg\sin\theta - \mu_{k1}N_1 = ma_1$$

where we expressed the magnitude of the kinetic force of friction in terms of the normal force exerted by the plane, and the weight in terms of the mass and gravitational field, g. The y component of Newton's Second Law can be written:

$$\sum F_y = N_1 - F_g\cos\theta = 0$$
$$\therefore N_1 = mg\cos\theta$$

which we used to express the normal force in terms of the weight. We can use this expression for the normal force by substituting it into the equation we obtained from the x component to find the acceleration along the incline:

$$\begin{aligned} mg\sin\theta - \mu_{k1}N_1 &= ma_1 \\ mg\sin\theta - \mu_{k1}mg\cos\theta &= ma_1 \\ \therefore a_1 &= g(\sin\theta - \mu_{k1}\cos\theta) \end{aligned}$$

Now that we know the acceleration down the incline, we can easily find the velocity at the bottom of the incline using kinematics. We choose the origin of the x axis to be zero where the block started ($x_0 = 0$), so that the block is at position $x = L$ at the bottom of the incline. Using kinematics, we can find the speed, v, given that the initial speed, $v_0 = 0$:

$$\begin{aligned} v^2 - v_0^2 &= 2a_1(x - x_0) \\ v^2 &= 2a_1 L \\ \therefore v &= \sqrt{2a_1 L} \\ &= \sqrt{2Lg(\sin\theta - \mu_{k1}\cos\theta)} \end{aligned}$$

We can now proceed to build a model for the second segment. We first identify the forces on the block when it is on the horizontal surface; these are:

1. $\vec F_{g1}$, its weight.
2. $\vec N_2$, a normal force exerted by the horizontal surface. This is in general different than the normal force exerted when the block was on the inclined plane.
3. $\vec f_{k2}$, a force of kinetic friction exerted by the horizontal surface. The force is opposite of the direction of motion, and has a magnitude given by $f_{k2} = \mu_{k2}N_2$.

The forces are illustrated by the free-body diagram in Figure 6.6, where we showed the acceleration vector, $\vec a_2$, which we determined to be to the left since the block is decelerating. We also chose an xy coordinate system such that the x axis is anti-parallel to the acceleration, so that the motion is in the positive x direction (and the acceleration in the negative x direction).

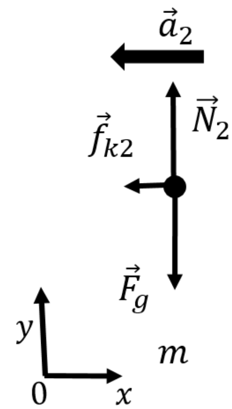


Figure 6.6: Free-body diagram for the block when it is sliding along the horizontal surface. We (arbitrarily) chose the positive x direction to be in the direction of motion and anti-parallel to the acceleration. We could easily have chosen the opposite direction.

Writing out the x component of Newton's Second Law:

$$\sum F_x = -f_{k2} = -ma_2$$
$$\therefore \mu_{k2} N_2 = ma_2$$

where we expressed the force of kinetic friction using the normal force. We have to be careful here with the sign of the acceleration; the equation that we wrote implies that a_2 is a positive number, since μ_{k2} is positive and N_2 is also positive (it is the magnitude of the normal force). a_2 is the magnitude of the acceleration, and we included the fact that the acceleration points in the negative x direction when we put a negative sign in the first line. The x component of the acceleration is $-a_2$, and the vector is given by $\vec{a}_2 = -a_2\hat{x}$.

The y component of Newton's Second Law will allow us to find the normal force:

$$\sum F_y = N_2 - F_g = 0$$
$$\therefore N_2 = mg$$

which we can substitute back into the x equation to find the magnitude of the acceleration along the horizontal surface:

$$ma_2 = \mu_{k2} N_2$$
$$\therefore a_2 = \mu_{k2} g$$

Now that we have found the acceleration along the horizontal surface, we can use kinematics to find the distance that the block travelled before stopping. We choose the origin of the x axis to be the bottom of the incline ($x_0 = 0$), the acceleration is negative $a_x = -a_2 = -mu_{k2}g$, the final speed is zero, $v = 0$, and the initial speed, v_0 is given by our model for the first segment. Using one of the kinematic equations:

$$\begin{aligned}
v^2 - v_0^2 &= 2(-a_2)(x - x_0) \\
v_0^2 &= 2a_2x \\
\therefore x &= \frac{1}{2a_2}v_0^2 \\
&= \frac{1}{2\mu_{k2}g}2Lg(\sin\theta - \mu_{k1}\cos\theta) \\
\therefore x &= \frac{(\sin\theta - \mu_{k1}\cos\theta)}{\mu_{k2}}L
\end{aligned}$$

Discussion: The model for the distance x that it takes the block to stop makes sense because:

- All of the terms in the fraction are dimensionless, so the value of x will have the same dimension as L.
- If we make L bigger, then x will be bigger (if we release the block from higher up on the incline, it will have more time to accelerate and will slide further before stopping).
- If we make μ_{k1} bigger, then x will be smaller: if we increase friction on the incline, the block will have a smaller acceleration and smaller speed at the bottom.
- If we increase the friction with the horizontal plane (increase μ_{k2}), then x will be reduced (it won't slide as far if there is more friction on the horizontal plane).
- If we increase θ, the numerator will be larger, so x will increase (the block will accelerate more down a steeper incline and end up further).

Checkpoint 6-1

A present is placed at rest on a plane that is inclined, at a distance L from the bottom of the incline, much like the box in Example 6-2 above. At the bottom of the incline, the box is determined to have a speed v. If the box is instead released from a distance of $4L$ from the bottom of the incline, what will its speed at the bottom of the incline be?

A) v
B) $2v$
C) $4v$
D) it depends on the coefficient of friction between the present and the plane.

6.2.1 Modelling situations where forces change magnitude

So far, the models that we have considered involved forces that remained constant in magnitude. In many cases, the forces exerted on an object can change magnitude and direction. For example, the force exerted by a spring changes as the spring changes length or the force of drag changes as the object changes speed. In these case, even if the object undergoes linear motion, we need to break up the motion into many small segments over which we can assume that the forces are constant. If the forces change continuously, we will need to break up the motion into an infinite number of segments and use calculus.

Consider the block of mass m that is shown in Figure 6.7, which is sliding along a frictionless horizontal surface and has a horizontal force $\vec{F}(x)$ exerted on it. The force has a different magnitude in the three segments of length Δx that are shown. If the block starts at position $x = x_0$ axis with speed v_0, we can find, for example, its speed at position $x_3 = 3\Delta x$, after the block travelled through the three segments.

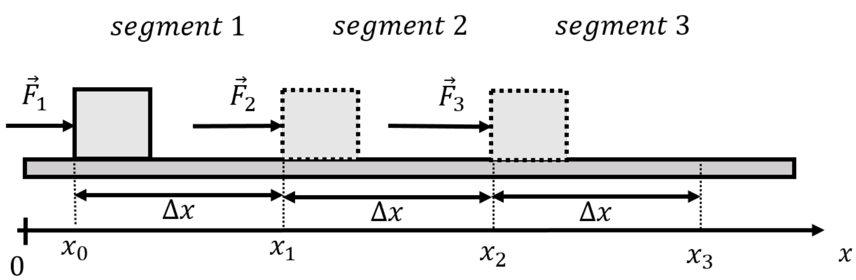


Figure 6.7: A block being pushed along a frictionless horizontal surface with a force that changes.

The horizontal force, $\vec{F}$, exerted on the block can be written as:

$$\vec{F}(x) = \begin{cases} F_1\hat{x} & x < \Delta x \quad \text{(segment 1)} \\ F_2\hat{x} & \Delta x \leq x < 2\Delta x \quad \text{(segment 2)} \\ F_3\hat{x} & 2\Delta x \leq x \quad \text{(segment 3)} \end{cases}$$

as it depends on the location of the block. To find the speed of the block at the end of the third segment, we can model each segment separately. The forces exerted on the block are the same in each segment:

1. $\vec{F}_g$, its weight, with magnitude mg.
2. $\vec{N}$, a normal force exerted by the ground.
3. $\vec{F}(x)$, an applied force that changes magnitude with position and is different in the three different segments.

The forces are illustrated in the free-body diagram show in Figure 6.8.

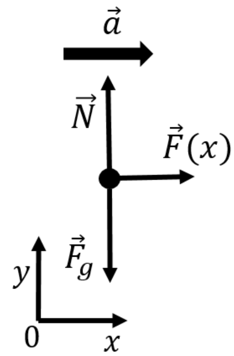


Figure 6.8: Free-body diagram for the block shown in Figure 6.7.

Newton's Second Law can be used to determine the acceleration of the block for each of the three segments, since the forces are constant within one segment. For all three segments, the y component of Newton's Second Law just tells us that the normal force exerted by the ground is equal in magnitude to the weight of the block. The x component of Newton's Second Law gives the acceleration:

$$\sum F_x = F_i = ma_i$$

where we have used the index i to indicate which segment the block is in (i can be 1, 2 or 3). The acceleration of the block in segment i is given by:

$$a_i = \frac{F_i}{m}$$

If the speed of the block is v_0 at the beginning of segment 1 ($x = x_0$), we can find its speed at the end of segment 1 ($x = x_1$), v_1, using kinematics and the fact that the acceleration in segment 1 is a_1:

$$\begin{aligned} v_1^2 - v_0^2 &= 2a_1(x_1 - x_0) \\ v_1^2 &= v_0^2 + 2a_1\Delta x \\ \therefore v_1^2 &= v_0^2 + 2\frac{F_1}{m}\Delta x \end{aligned}$$

We can now easily find the speed at the end of segment 2 ($x = x_2$), v_2, since we know the speed at the beginning of segment 2 (x_1,v_1) and the acceleration a_2:

$$\begin{aligned} v_2^2 - v_1^2 &= 2a_2(x_2 - x_1) \\ \therefore v_2^2 &= v_1^2 + 2a_2\Delta x \\ &= v_0^2 + 2\frac{F_1}{m}\Delta x + 2\frac{F_2}{m}\Delta x \end{aligned}$$

It is easy to show that the speed at the end of the third segment is:

$$v_3^2 = v_0^2 + 2\frac{F_1}{m}\Delta x + 2\frac{F_2}{m}\Delta x + 2\frac{F_3}{m}\Delta x$$

If there were N segments, with the force being different in each segment, we could use the summation notation to write:

$$v_N^2 = v_0^2 + 2\sum_{i=1}^{i=N}\frac{F_i}{m}\Delta x$$

Finally, if the magnitude of the force varied continuously as a function of x, $\vec{F}(x)$, we would model this by taking segments whose length, Δx, tends to zero (and we would need an infinite number of such segments). For example, if we wanted to know the speed of the object at position $x = X$ along the x axis, with a force that was given by $\vec{F}(x) = F(x)\hat{x}$, if the object started at position x_0 with speed v_0, we would take the following limit:

$$v^2 = v_0^2 + \lim_{\Delta x\to 0} 2\sum_{i=1}^{i=N}\frac{F(x)}{m}\Delta x$$

where $\Delta x = \frac{X}{N}$ so that as $\Delta x \to 0$, $N \to \infty$. Of course, integrals are the exact tool that allow us to evaluate the sum in this limit:

$$\lim_{\Delta x\to 0} 2\sum_{i=1}^{i=N}\frac{F_i}{m}\Delta x = 2\int_{x_0}^{X}\frac{F(x)}{m}dx$$

and the speed at position $x = X$ is given by:

$$v^2 = v_0^2 + 2\int_{x_0}^{X}\frac{F(x)}{m}dx$$

Naturally, we can find the above result starting directly from calculus. If the component of the (net) force in the x direction is given by $F(x)$, then the acceleration is given by $a(x) = \frac{F(x)}{m}$. The velocity is related to the acceleration:

$$\begin{aligned} a(x) &= \frac{dv}{dt} \\ \therefore dv &= a(x)dt \end{aligned}$$

We cannot simply integrate the last equation to find that $v = \int a(x)dt$ because the acceleration is given as a function of position, $a(x)$, and not a function of time, t. Thus, we cannot simply take the integral over t and must instead "change variables" to take the integral over x. x and t are related through velocity:

$$\begin{aligned} v &= \frac{dx}{dt} \\ \therefore dt &= \frac{1}{v}dx \end{aligned}$$

We can thus write:

$$dv = a(x)dt = a(x)\frac{1}{v}dx$$

The equation above is called a "separable differential equation", which can also be written:

$$\frac{dv}{dx} = \frac{1}{v}a(x)$$

This is called a differential equation because it relates the derivative of a function (the derivative of v with respect to x, on the left) to the function itself (v appears on the right as well). The differential equation is "separable", because we can separate out all of the quantities that depend on v and on x on different sides of the equation:

$$vdv = a(x)dx$$

This last equation says that vdv is equal to $a(x)dx$. Remember that dx is the length of a very small segment in x, and that dv is the change in velocity over that very small segment. Since the terms on the left and right are equal, if we sum (integrate) the quantity vdv over many segments, that sum must be equal to the sum (integral) of the quantity $a(x)dx$ over the same segments. Let us choose those segment such that for the beginning of the first interval the position and speed are x_0 and v_0, respectively, and the position and speed at the end of the last segment are X and V, respectively. We then must have that:

$$\begin{aligned}
\int_{v_0}^{V} vdv &= \int_{x_0}^{X} a(x)dx \\
\frac{1}{2}V^2 - \frac{1}{2}v_0^2 &= \int_{x_0}^{X} a(x)dx \\
\therefore V^2 &= v_0^2 + 2\int_{x_0}^{X} a(x)dx
\end{aligned}$$

which is the same as we found earlier. If the acceleration is constant, we recover our formula from kinematics:

$$\begin{aligned}
V^2 &= v_0^2 + 2\int_{x_0}^{X} adx \\
&= v_0^2 + 2a(X - x_0) \\
\therefore V^2 - v_0^2 &= 2a(X - x_0)
\end{aligned}$$

Example 6-3

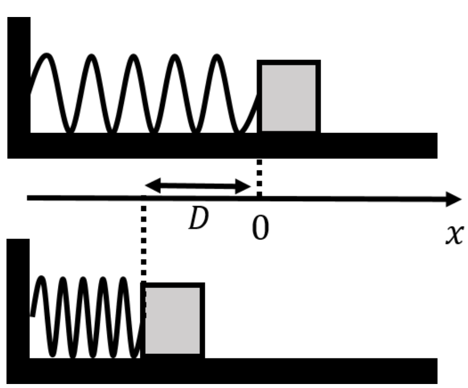


Figure 6.9: A block is launched along a frictionless surface by compressing a spring by a distance D. The top panel shows the spring when at rest, and the bottom panel shows the spring compressed by a distance D just before releasing the block.

A block of mass m can slide freely along a frictionless surface. A horizontal spring, with spring constant, k, is attached to a wall on one end, while the other end can move freely, as shown in Figure 6.9. A coordinate system is defined such that the x axis is horizontal and the free end of the spring is at $x = 0$ when the spring is at rest. The block is pushed against the spring so that the spring is compressed by a distance D. The block is then released. What speed will the block have when it leaves the spring?

Solution

As you recall, the force exerted by a spring depends on the compression or extension of the spring and is given by Hooke's Law:

$$\vec{F}(x) = -kx\hat{x}$$

where x is the position of the free end of the spring and $x = 0$ corresponds to the spring being at rest. In our case, when the edge of the block is located at $x_0 = -D$ (the spring is compressed), the force is thus in the positive x direction (since x_0 is a negative number).

The forces on the block are:

1. $\vec{F}_g$, its weight, with magnitude mg.
2. $\vec{N}$, a normal force exerted by the ground.
3. $\vec{F}(x)$, the spring force.

Since the block is not moving vertically, the magnitude of the normal force must equal the weight $N = mg$, since these are the only forces with components in the vertical direction. The x component of Newton's Second Law gives us the acceleration of the block (which depends on x):

$$\sum F_x = -kx = ma(x)$$
$$\therefore a(x) = -\frac{k}{m}x$$

Again, recall that if x is negative, then the acceleration will be in the positive direction. Since this scenario is exactly the same that we described above in the text, namely a force that varies continuously with position, we can apply the formula that we found earlier for determining the velocity after a varying force has been applied from position

$x = x_0$ to position $x = X$:

$$V^2 = v_0^2 + 2\int_{x_0}^{X} a(x)dx$$

V is the final speed that we would like to find, $v_0 = 0$ because the block starts at rest, and $x_0 = -D$ is the starting position of the block. X is the position along the x axis where the block leaves the spring.

We have to think a little about what the value of X should be: when the spring is compressed and the block accelerating, the spring is pushing the block in the positive x direction. Once the block reaches $x = 0$ the spring would want to pull the block backwards, but since it is not attached to the block, it stops exerting a force on the block at that point. The block thus leaves the spring at $x = 0$, so that the final position is $X = 0$. The speed of the block when it leaves the spring is thus:

$$\begin{aligned}
V^2 &= v_0^2 + 2\int_{x_0}^{X} a(x)dx \\
&= 0 + 2\int_{-D}^{0} a(x)dx \\
&= 2\int_{-D}^{0} -\frac{k}{m}xdx \\
&= 2\left[-\frac{k}{m}\frac{1}{2}x^2\right]_{-D}^{0} \\
&= \frac{k}{m}D^2 \\
\therefore V &= \sqrt{\frac{k}{m}}D
\end{aligned}$$

Discussion: This model for the speed of the block when it leaves the spring makes sense because:

- The dimension for the expression for V is correct (you should check this!).
- If the spring is compressed more (bigger value of D), then the speed will be higher.
- If the mass is bigger (more inertia), then the final speed will be lower.
- If the spring is stiffer (bigger value of k), then the final speed will be higher.

If you have studied physics before, you may have realized that the speed is easily found by conservation of energy:

$$\frac{1}{2}mV^2 = \frac{1}{2}kD^2$$

which gives the same value for V. As we will see in a later chapter, kinetic and potential energy are defined as they are, precisely because it makes using conservation of energy equivalent to using forces as we just did.

Example 6-4

An object of mass m is released from rest out of a helicopter. The drag (air-resistance) on the object can be modelled as having a magnitude given by bv, where v is the speed of the object and b is a constant of proportionality. How does the velocity of the object depend on time?

Solution

As the object falls through the air, the forces exerted on the object are:

1. F_g, its weight, with magnitude mg, exerted downwards.
2. F_d, the force of drag, with magnitude bv, exerted upwards.

Since the object will fall in a straight line, this is a one-dimensional problem, and we can choose the x axis to be vertical, with positive x pointing downwards, and the origin located where the object was released. The object will thus have a positive acceleration and move in the positive x direction with this choice of coordinate system. This is illustrated in the free-body diagram in Figure 6.10.

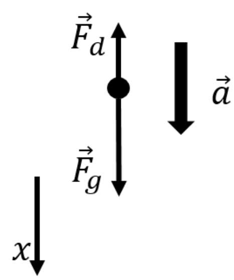


Figure 6.10: Free-body diagram for a block free-falling with drag.

Newton's Second Law for the object gives:

$$\sum F_x = F_g - F_d = ma$$
$$mg - bv = ma$$
$$\therefore a = g - \frac{b}{m}v$$

In this case, the acceleration depends explicitly on velocity rather than position, as we had before. However, we can use the same methodology to find how the velocity changes with time. First, we can note that the acceleration is zero if:

$$\begin{aligned} g - \frac{b}{m}v &= 0 \\ \therefore v &= \frac{mg}{b} \end{aligned}$$

That is, once the object reaches a speed of $v_{term} = mg/b$, it will stop accelerating, i.e. it will reach "terminal velocity". Note that this is the same condition as requiring that the drag force (bv) have the same magnitude as the weight (mg).

Writing the acceleration as $a = \frac{dv}{dt}$, we can write:

$$\frac{dv}{dt} = \left(g - \frac{b}{m}v\right)$$

which again, is a separable differential equation, in which we can write the terms that depend on v and those that depend on t on separate sides of the equal sign:

$$\begin{aligned} \frac{dv}{g - \frac{b}{m}v} &= dt \\ \frac{dv}{v - \frac{mg}{b}} &= -\frac{b}{m}dt \end{aligned}$$

where we re-arranged the equation in the second line so that it would be easier to integrate in the next step. We can find the velocity, $v(t)$, at some time, t, by stating that $v = 0$ at $t = 0$ and taking the integrals (sum) on both sides. Again, we are modelling the motion as being made up of a large number of very small segments where the quantities on both sides of the equation are the same. Thus, if we sum (integrate) those quantities over all of the same segments, the left and right hand side of the equations will still be equal to each other:

$$\begin{aligned} \int_0^{v(t)} \frac{dv}{v - \frac{mg}{b}} &= -\int_0^t \frac{b}{m}dt \\ \left[\ln\left(v - \frac{mg}{b}\right)\right]_0^{v(t)} &= -\frac{b}{m}t \\ \ln\left(v(t) - \frac{mg}{b}\right) - \ln\left(-\frac{mg}{b}\right) &= -\frac{b}{m}t \\ \ln\left(\frac{v(t) - \frac{mg}{b}}{-\frac{mg}{b}}\right) &= -\frac{b}{m}t \end{aligned}$$

where, in the last line, we used the property that $\ln(a) - \ln(b) = \ln(a/b)$. By taking the exponential on either side of the equation ($e^{\ln(x)} = x$), we can find an expression for the velocity as a function of time:

$$\begin{aligned}
\frac{v(t) - \frac{mg}{b}}{-\frac{mg}{b}} &= e^{-\frac{b}{m}t} \\
v(t) - \frac{mg}{b} &= -\frac{mg}{b}e^{-\frac{b}{m}t} \\
\therefore v(t) &= \frac{mg}{b} - \frac{mg}{b}e^{-\frac{b}{m}t} \\
&= \frac{mg}{b}\left(1 - e^{-\frac{b}{m}t}\right)
\end{aligned}$$

Discussion: This equation tells us that the velocity increases as a function of time, but the rate of increase decreases exponentially with time. At time $t = 0$, the velocity is zero, as expected. As t approaches infinity, v approaches, $\frac{mg}{b}$, which is the terminal velocity. The time dependence of the velocity is illustrated in Figure 6.11.

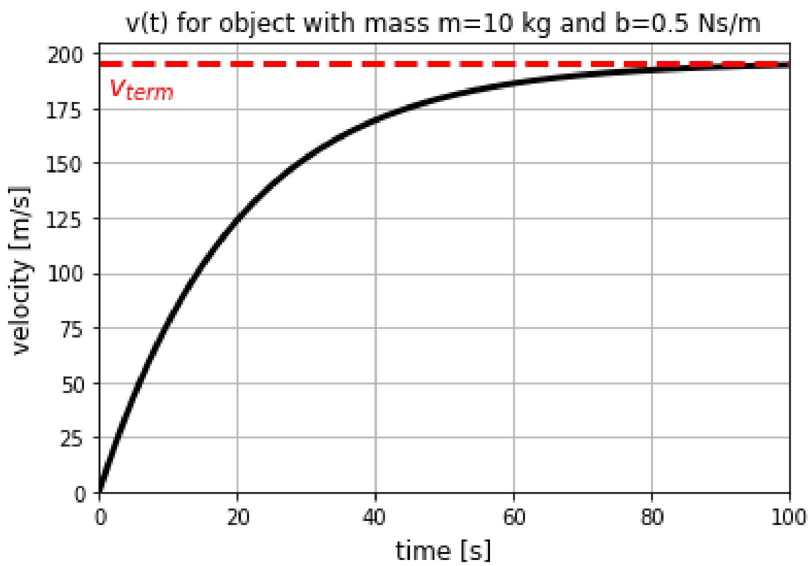


Figure 6.11: Velocity as a function of time for an object of mass $m = 10\,\mathrm{kg}$ which is free-falling from rest with a drag coefficient $b = 0.5\,\mathrm{Ns/m}$.

6.3 Uniform circular motion

As we saw in Chapter 4, "uniform circular motion" is defined to be motion along a circle with constant speed. This may be a good time to review Section 4.4 for the kinematics of motion along a circle. In particular, for the uniform circular motion of an object around a circle of radius R, you should recall that:

- The velocity vector, $\vec{v}$, is always tangent to the circle.
- The acceleration vector, $\vec{a}$, is always perpendicular to the velocity vector, because the magnitude of the velocity vector does not change.

- The acceleration vector, $\vec{a}$, always points towards the centre of the circle.
- The acceleration vector has magnitude $a = v^2/R$.
- The angular velocity, ω, is related to the magnitude of the velocity vector by $v = \omega R$ and is constant.
- The angular acceleration, α, is zero for uniform circular motion, since the angular velocity does not change.

In particular, you should recall that even if the speed is constant, the acceleration vector is always non-zero in uniform circular motion because the **velocity changes direction**. According to Newton's Second Law, this implies that there **must be a net force on the object that is directed towards the centre of the circle**[1] (parallel to the acceleration):

$$\sum \vec{F} = m\vec{a}$$

where the acceleration has a magnitude $a = v^2/R$. Because the acceleration is directed towards the centre of the circle, we sometimes call it a "radial" acceleration (parallel to the radius), a_R, or a "centripetal" acceleration (directed towards the centre), a_c.

Consider an object in uniform circular motion in a horizontal plane on a frictionless surface, as depicted in Figure 6.12.

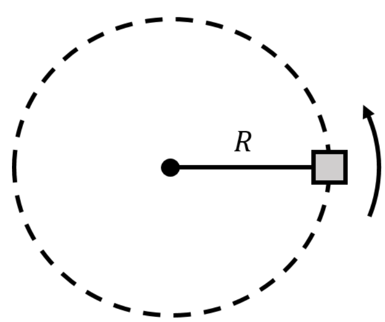


Figure 6.12: An object undergoing uniform circular motion on a frictionless surface, as seen from above.

The only way for the object to undergo uniform circular motion as depicted is if the net force on the object is directed towards the centre of the circle. One way to have a force that is directed towards the centre of the circle is to attach a string between the center of the circle and the object, as shown in Figure 6.12. If the string is under tension, the force of tension will always be towards the centre of the circle. The forces on the object are thus:

1. $\vec{F}_g$, its weight with magnitude mg.
2. $\vec{N}$, a normal forced exerted by the surface.

[1]The sum of the forces is often called the "net force" on an object, and in the specific case of uniform circular motion, that net force is sometimes called the "centripetal force" - however, it is not a force in and of itself and it is always the sum of the forces that points towards the centre of the circle.

3. $\vec{T}$, a force of tension exerted by the string.

The forces are depicted in the free-body diagram shown in Figure 6.13 (as viewed from the side), where we also drew the acceleration vector. Note that this free-body diagram is only "valid" at a particular instant in time since the acceleration vector continuously changes direction and would not always be lined up with the x axis.

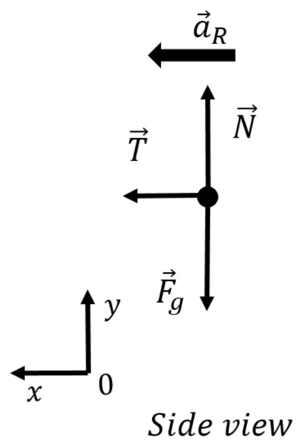


Figure 6.13: Free-body diagram (side view) for the object from Figure 6.13 undergoing uniform circular motion.

Writing out the x and y components of Newton's Second Law:

$$\sum F_x = T = ma_R$$
$$\sum F_y = N - F_g = 0$$

The y component just tells us that the normal force must have the same magnitude as the weight because the object is not accelerating in the vertical direction. The x component tells us the relation between the magnitudes of the tension in the string and the radial acceleration. Using the speed of the object, we can also write the relation between the tension and the speed:

$$T = ma_R = m\frac{v^2}{R}$$

Thus, we find that the tension in the string increases with the square of the speed, and decreases with the radius of the circle.

Checkpoint 6-2

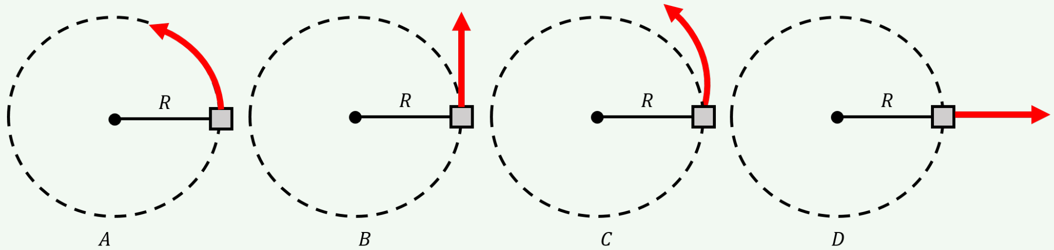


Figure 6.14: Possible trajectories (in red) that the block will follow if the string breaks.

An object is undergoing uniform circular motion in the horizontal plane, when the string connecting the object to the centre of rotation suddenly breaks. What path will the block take after the string broke?

A) A
B) B
C) C
D) D

Example 6-5

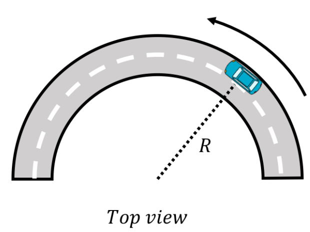


Figure 6.15: A car going around a curve that can be approximated as the arc of a circle of radius R.

A car goes around a curve which can be approximated as the arc of a circle of radius R, as shown in Figure 6.15. The coefficient of static friction between the tires of the car and the road is μ_s. What is the maximum speed with which the car can go around the curve without skidding?

Solution

If the car is going at constant speed around a circle, then the sum of the forces on the car must be directed towards the centre of the circle. The only force on the car that could be directed towards the centre of the circle is the force of friction between the tires and the road. If the road were perfectly slick (think driving in icy conditions), it would not be possible to drive around a curve since there could be no force of friction. The forces on the car are:

1. $\vec{F}_g$, its weight with magnitude mg.
2. $\vec{N}$, a normal force exerted upwards by the road.
3. $\vec{f}_s$, a force of static friction between the tires and the road. This is static friction, because the surface of the tire does not move relative to the surface of the road if the car is not skidding. The force of static friction has a magnitude that is at most $f_s \leq \mu_s N$.

The forces on the car are shown in the free-body diagram in Figure 6.16.

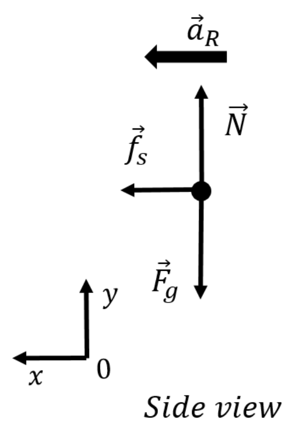


Figure 6.16: Free-body diagram for the car as seen looking at the car from the back (the centre of the curve is towards the left).

The y component of Newton's Second Law tells us that the normal force exerted by the road must equal the weight of the car:

$$\begin{aligned}\sum F_y &= N - F_g = 0\\ \therefore N &= mg\end{aligned}$$

The x component relates the force of friction to the radial acceleration (and thus to the speed):

$$\begin{aligned}\sum F_x = f_s = ma_R &= m\frac{v^2}{R}\\ \therefore f_s &= m\frac{v^2}{R}\end{aligned}$$

The force of friction must be less than or equal to $f_s \leq \mu_s N = \mu_s mg$ (since $N = mg$ from the y component of Newton's Second Law), which gives us a condition on the speed:

$$\begin{aligned} f_s = m\frac{v^2}{R} &\leq \mu_s mg \\ v^2 &\leq \mu_s gR \\ \therefore v &\leq \sqrt{\mu_s gR} \end{aligned}$$

Thus, if the speed is less than $\sqrt{\mu_s gR}$, the car will not skid and the magnitude of the force of static friction, which results in an acceleration towards the centre of the circle, will be smaller or equal to its maximal possible value.

Discussion: The model for the maximum speed that the car can travel around the curve makes sense because:

- The dimension of $\sqrt{\mu_s gR}$ is speed.
- The speed is larger if the radius of the curve is larger (one can go faster around a wider curve without skidding).
- The speed is larger if the coefficient of friction is large (if the force of friction is larger, a larger radial acceleration can be sustained).

Example 6-6

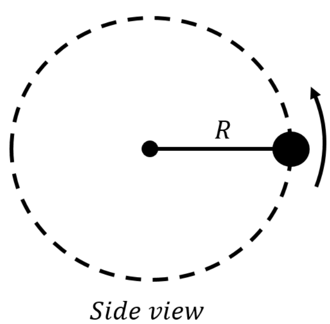


Figure 6.17: A ball attached to a string undergoing circular motion in a vertical plane.

A ball is attached to a mass-less string and executing circular motion along a circle of radius R that is in the vertical plane, as depicted in Figure 6.17. Can the speed of the ball be constant? What is the minimum speed of the ball at the top of the circle if it is able to make it around the circle?

Solution

The forces that are acting on the ball are:

1. $\vec{F}_g$, its weight with magnitude mg.
2. $\vec{T}$, a force of tension exerted by the string.

Figure 6.18 shows the free-body diagram for the forces on the ball at three different locations along the path of the circle.

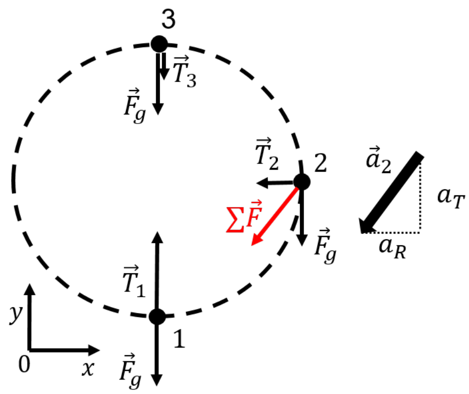


Figure 6.18: A ball attached to a string undergoing circular motion in a vertical plane.

In order for the ball to go around in a circle, there must be at least a component of the net force on the ball that is directed towards the centre of the circle at all times. In the bottom half of the circle (positions 1 and 2), only the tension can have a component directed towards the centre of the circle.

Consider in particular the position labelled 2, when the string is horizontal and the tension is equal to $\vec{T}_2$. The free-body diagram in Figure 6.18 also shows the vector sum of the weight and tension at position 2 (the red arrow labelled $\sum \vec{F}$), which points downwards and to the left. It is thus clearly impossible for the acceleration vector to point towards the centre of the circle, and the acceleration will have components that are both tangential (a_T) to the circle and radial (a_R), as shown by the vector $\vec{a}_2$ in Figure 6.18.

The radial component of the acceleration will change the direction of the velocity vector so that the ball remains on the circle, and the tangential component will reduce the magnitude of the velocity vector. According to our model, it is thus impossible for the ball to go around the circle at constant speed, and the speed must decrease as it goes from position 2 to position 3, no matter how one pulls on the string (you can convince yourself of this by drawing the free-body diagram at any point between points 2 and 3).

The minimum speed for the ball at the top of the circle is given by the condition that the tension in the string is zero just at the top of the trajectory (position 3). The ball can still go around the circle because, at position 3, gravity is towards the centre of the circle and can thus give an acceleration that is radial, even with no tension. The y component of Newton's Second Law, at position 3 gives:

$$\sum F_y = -F_g = ma_y$$
$$\therefore a_y = -g$$

The magnitude of the acceleration is the radial acceleration, and is thus related to the speed at the top of the trajectory:

$$a_R = -a_y = g = m\frac{v^2}{R}$$
$$\therefore v_{min} = \sqrt{\frac{gR}{m}}$$

which is the minimum speed at the top of the trajectory for the ball to be able to continue along the circle. The tension in the string would change as the ball moves around the circle, and will be highest at the bottom of the trajectory, since the tension has to be bigger than gravity so that the net force at the bottom of the trajectory is upwards (towards the centre of the circle).

Discussion: The model for the minimum speed of the ball at the top of the circle makes sense because:

- $\sqrt{\frac{gR}{m}}$ has the dimension of speed.
- The minimum velocity is larger if the circle has a larger radius (try this with a mass attached at the end of a string).
- The minimum velocity is larger if the mass is bigger (again, try this at home!).

Checkpoint 6-3

Consider a ball attached to a string, being spun in a vertical circle (such as the one depicted in figure 6.17). If you shortened the string, how would the minimum angular velocity (measured at the top of the trajectory) required for the ball to make it around the circle change?

A) It would decrease
B) It would stay the same
C) It would increase

6.3.1 Banked curves

As we saw in Example 6-5, there is a maximum speed with which a car can go around a curve before it starts to skid. You may have noticed that roads, highways especially, are banked where there are curves. Racetracks for cars that go around an oval (the boring kind of car races) also have banked curves. As we will see, this allows the speed of vehicles to be higher when going around the curve; or rather, it makes the curves safer as the speed at which vehicles *would* skid is higher. In Example 6-5, we saw that it was the force of static friction between the tires of the car and the road that provided the only force with a component towards the centre of the circle. The idea of using a banked curve is to change the direction of the normal force between the road and the car tires so that it, too, has a component in the direction towards the centre of the circle.

Consider the car depicted in Figure 6.19 which is seen from behind making a left turn around a curve that is banked by an angle θ with respect to the horizontal and can be modelled as an arc from a circle of radius R.

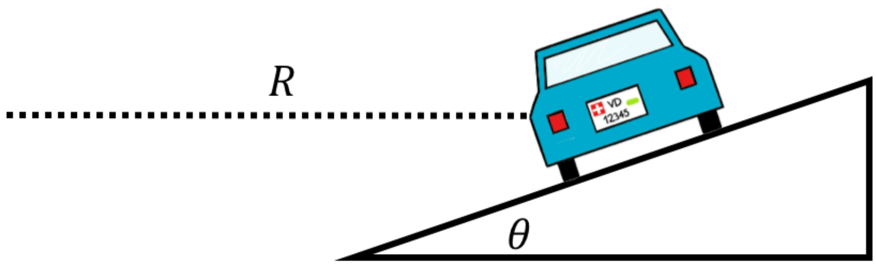


Figure 6.19: A car moving into the page and going around a banked curved so that it is turning towards the left (the centre of the circle is to the left).

The forces exerted on the car are the same as in Example 6-5, except that they point in different directions. The forces are:

1. $\vec F_g$, its weight with magnitude mg.
2. $\vec N$, a normal force exerted by the road, perpendicular to the surface of the road.
3. $\vec f_s$, a force of static friction between the tires and the road. This is static friction, because the surface of the tire does not move relative to the surface of the road if the car is not skidding. The force of static friction has a magnitude that is at most $f_s \leq \mu_s N$ and is perpendicular to the normal force. The force could be either upwards or downwards, *depending on the other forces on the car.*

A free-body diagram for the forces on the car is shown in Figure 6.20, along with the acceleration (which is in the radial direction, towards the centre of the circle), and our choice of coordinate system (choosing x parallel to the acceleration). The direction of the force of static friction is not known *a priori* and depends on the speed of the car:

- If the speed of the car is zero, the force of static friction is upwards. With a speed of zero, the radial acceleration is zero, and the sum of the forces must thus be zero. The impeding motion of the car would be to slide down the banked curve (just like a block on an incline).
- If the speed of the car is very large, the force of static friction is downwards, as the impeding motion of the car would be to slide up the bank. The natural motion of the car is to go in a straight line (Newton's First Law). If the components of the normal force and of the force of static friction directed towards the centre of the circle are too

small to allow the car to turn, then the car would slide up the bank (so the impeding motion is up the bank and the force of static friction is downwards).

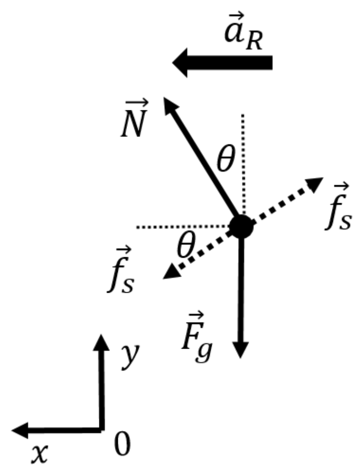


Figure 6.20: Free-body diagram for the forces on the car. The direction of the force of static friction cannot be determined, as it depends on the acceleration of the car, so it is shown twice (with dotted lines).

There is thus an "ideal speed" at which the force of static friction is precisely zero, and the x component of the normal force is responsible for the radial acceleration. At higher speeds, the force of static friction is downwards and increases in magnitude to keep the car's acceleration towards the centre of the circle. At some maximal speed, the force of friction will reach its maximal value, and no longer be able to keep the car's acceleration pointing towards the centre of the circle. At speeds lower than the ideal speed, the force of friction is directed upwards to prevent the car from sliding down the bank. If the coefficient of static friction is too low, it is possible that at low speeds, the car would start to slide down the bank (so there would be a minimum speed below which the car would start to slide down).

Let us model the situation where the force of static friction is identically zero so that we can determine the ideal speed for the banked curve. The only two forces on the car are thus its weight and the normal force. The x and y component of Newton's Second Law give:

$$\begin{aligned}\sum F_x = N\sin\theta = ma_R = m\frac{v^2}{R}\\ \therefore N\sin\theta = m\frac{v^2}{R}\end{aligned} \tag{6.1}$$

$$\begin{aligned}\sum F_y = N\cos\theta - F_g = 0\\ \therefore N\cos\theta = mg\end{aligned} \tag{6.2}$$

We can divide Equation 6.1 by Equation 6.2, noting that $\tan\theta = \sin\theta/\cos\theta$, to obtain:

$$\tan\theta = \frac{v^2}{gR}$$
$$\therefore v_{ideal} = \sqrt{gR\tan\theta}$$

At this speed, the force of static friction is zero. In practice, one would use this equation to determine which bank angle to use when designing a road, so that the ideal speed is around the speed limit or the average speed of traffic. We leave it as an exercise to determine the maximal speed that the car can go around the curve before sliding out.

6.3.2 Inertial forces in circular motion

As you sit in a car that is going around a curve, you will feel pushed outwards, away from the centre of the circle that the car is going around. This is because of your inertia (Newton's First Law), and your body would go in a straight line if the car were not exerting a net force on you towards the centre of the circle. You are not so much feeling a force that is pushing you outwards as you are feeling the effects of the car seat pushing you inwards; if you were leaning against the side of the car that is on the outside of the curve, you would feel the side of the car pushing you inwards towards the centre of the curve, even if it "feels" like you are pushing outwards against the side of the car.

If we model your motion looking at you from the ground, we would include a force of friction between the car seat (or the side of the car, or both) and you that is pointing towards the centre of the circle, so that the sum of the forces exerted on you is towards the centre of the circle. We can also model your motion from the non-inertial frame of the car. As you recall, because this is a non-inertial frame of reference, we need to include an additional inertial force, $\vec{F}_I$, that points opposite of the acceleration of the car, with magnitude $F_I = ma_R$ (if the net acceleration of the car is a_R). Inside the non-inertial frame of reference of the car, your acceleration (relative to the reference frame, i.e. the car) is zero. This is illustrated by the diagrams in Figure 6.21.

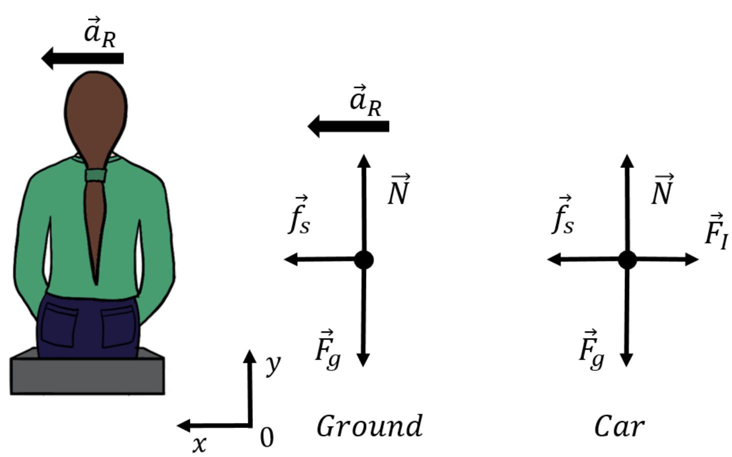


Figure 6.21: (Left:) A person sitting on a car seat in a car turning towards the left. (Centre:) Free-body diagram for the person as modelled in the inertial reference frame of the ground. (Right:) Free-body diagram for the person as modelled in the non-inertial frame of reference of the car, including an additional inertial force.

The y component of Newton's Second Law in both frames of reference is the same:

$$\begin{aligned}\sum F_y &= N - F_g = 0\\ \therefore N &= mg\end{aligned}$$

and simply tells us that the normal force is equal to the weight. In the reference frame of the ground, the x component of Newton's Second Law gives:

$$\begin{aligned}\sum F_x &= f_s = ma_R\\ \therefore f_s &= m\frac{v^2}{R}\end{aligned}$$

In the frame of reference of the car, where your acceleration is zero and an inertial force of magnitude $F_I = mv^2/R$ is exerted on you, the x component of Newton's Second Law gives:

$$\begin{aligned}\sum F_x &= f_s - F_I = 0\\ \therefore f_s - m\frac{v^2}{R} &= 0\end{aligned}$$

which of course, mathematically, is exactly equivalent. The inertial force is not a real force in the sense that it is not exerted by anything. It only comes into play because we are trying to use Newton's Laws in a non-inertial frame of reference. However, it does provide a good model for describing the sensation that we have of being pushed outwards when the car goes around a curve. Sometimes, people will refer to this force as a "centrifugal" force, which means "a force that points away from the centre". You should however remember that this is not a real force exerted on the object, but is the result of modelling motion in a non-inertial frame of reference.

Checkpoint 6-4

Jamie is driving his tricycle around a circular pond. Jamie feels a centrifugal force with magnitude F_I. If Jamie pedals twice as fast, what will be the magnitude of the centrifugal force that he experiences?

A) $\sqrt{2}F_I$
B) $\frac{1}{2}F_I$
C) $2F_I$
D) $4F_I$

6.4 Non-uniform circular motion

In non-uniform circular motion, an object's motion is along a circle, but the object's speed is not constant. In particular, the following will be true

- The object's velocity vector is always tangent to the circle.
- The speed and angular speed of the object are not constant.
- The angular acceleration of the object is not zero.
- The acceleration vector will not point towards the centre of the circle.

Since the acceleration vector does not point towards the centre of the circle, it is usually convenient to break up the acceleration vector into two components: a_R, a component that is radial (towards the centre of the circle), and a_T, a component that is tangent to the circle (and perpendicular to to the radial component). The **radial component is "responsible" for the change in direction of the velocity** such that the object goes in a circle. the magnitude of the radial acceleration is the same as it is for uniform circular motion:

$$a_R = \frac{v^2}{r}$$

where the speed is no longer constant in time. The tangential component of the acceleration is responsible for changing the magnitude of the velocity of the object:

$$a_T = \frac{dv}{dt}$$

Example 6-7

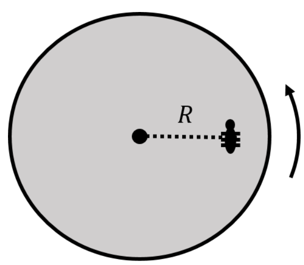


Figure 6.22: An ant on a horizontal turntable that is starting to spin, as seen from above.

A small ant is sleeping on a turntable just as the turntable starts to spin from rest, with an angular acceleration $\alpha = 1\,\mathrm{rad/s^2}$ that is small enough so that, initially, the ant remains on the turntable. The ant is a distance $R = 0.1\,\mathrm{m}$ from the centre of the turntable, as shown in Figure 6.22 and the coefficient of static friction between the ant's "feet" and the turntable is $\mu_s = 0.5$. After how much time will the ant slide off from the turntable?

Solution

As the turntable accelerates, the force of static friction between the turntable and the ant will keep the ant moving with the turntable. Once the turntable is going fast enough, the force of friction will no longer be large enough to provide the total acceleration that is required to keep the ant moving with the turntable (with a constant tangential component of the acceleration and an increasing radial component of the acceleration).

The forces on the ant are:

1. $\vec{F}_g$, its weight, with magnitude mg.
2. $\vec{N}$, a normal force exerted by the turntable on the ant.
3. $\vec{f}_s$, a force of static friction exerted by the turntable on the ant. The force of friction will be such that it has both radial and tangential components.

A free-body diagram for the forces on the ant is shown in Figure 6.23, as seen from above and from the side, for some point in time. We have chosen the point in time to be just when the ant is about to slide off of the turntable, when the force of static friction makes an unknown angle θ with the x axis. We have placed the origin of the coordinate system at the centre of the turntable and chosen the x axis such that the ant is located on the positive x axis with its velocity in the positive y direction. We used a three dimensional coordinate system where the weight and normal force are exerted in the z (vertical) direction since the acceleration vector of the ant will have both radial (x) and tangential (y) components.

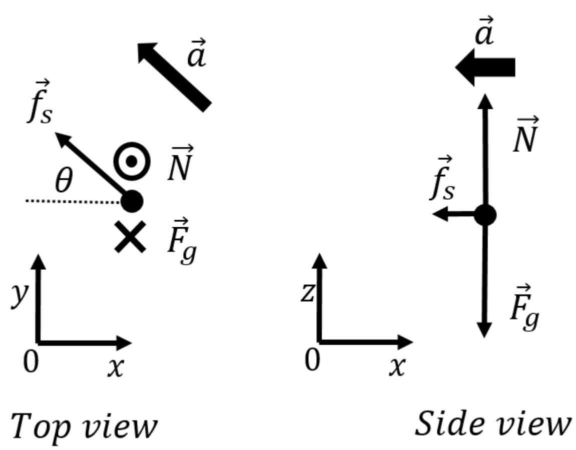


Figure 6.23: (Left:) Forces on the ant as seen from above. The normal force is out of the page ($\odot$), whereas the weight is into the page ($\times$). (Right:) Forces on the ant as seen from the side. Note that the acceleration vector and force of static friction also have components in the y direction, which is why their magnitude is shown as being smaller than in the top view.

Newton's Second Law has to be written out in three components. The z component relates the weight and normal force:

$$\sum F_z = N - F_g = 0$$
$$\therefore N = mg$$

The x component of Newton's Second Law is such that the x component of the acceleration is its radial component:

$$\sum F_x = -f_s \cos\theta = -ma_R = -m\frac{v^2}{R}$$
$$\therefore f_s \cos\theta = m\frac{v^2}{R}$$

The y component of Newton's Second relates the tangential component of the force of static friction to the tangential component of the acceleration:

$$\sum F_y = f_s \sin\theta = ma_T$$
$$\therefore f_s \sin\theta = m\alpha R$$

where we used the fact that the (linear) tangential acceleration, a_T, is related to the angular acceleration, α, by:

$$a_T = \alpha R$$

Summarizing the three equations that we obtained from the three components of Newton's Second Law:

$$f_s \cos\theta = m\frac{v^2}{R}$$
$$f_s \sin\theta = m\alpha R$$
$$N = mg$$

Also, note that the speed, $v(t)$ at some time t is given by simple kinematics:

$$v(t) = v_0 + a_T t = (0) + \alpha R t$$

The ant will start to slip when the force of friction reaches its maximal amplitude, $f_s = \mu_s N = \mu_S mg$. The x of Newton's Second Law can be used to find an expression for the time at which force of friction reaches its maximal value (in terms of the unknown angle θ):

$$f_s \cos\theta = m\frac{v^2}{R}$$
$$\mu_s g \cos\theta = R\alpha^2 t^2$$
$$\therefore t = \sqrt{\frac{\mu_s g \cos\theta}{R\alpha^2}}$$

We can use the y component to determine the angle θ:

$$\begin{aligned}
f_s \sin\theta &= m\alpha R \\
\mu_s g \sin\theta &= \alpha R \\
\therefore \sin\theta &= \frac{\alpha R}{\mu_s g} \\
\therefore \theta &= \sin^{-1}\left(\frac{\alpha R}{\mu_s g}\right) = \sin^{-1}\left(\frac{(1\,\text{rad/s}^2)(0.1\,\text{m})}{(0.5)(9.8\,\text{N/kg})}\right) \\
&= 1.17^\circ
\end{aligned}$$

The angle is very small, and we see that the force of friction is mostly directed towards the centre of the circle. The radial acceleration is thus much larger than the tangential acceleration. We can then use the angle to find the time using the expression we derived above:

$$\begin{aligned}
t &= \sqrt{\frac{\mu_s g \cos\theta}{R\alpha^2}} = \sqrt{\frac{(0.5)(9.8\,\text{N/kg})\cos(1.17^\circ)}{(0.1\,\text{m})(1\,\text{rad/s}^2)^2}} \\
&= 7.0\,\text{s}
\end{aligned}$$

6.5 Summary

Key Takeaways

When the velocity of an object does not change direction continuously ("linear motion"), we can model its motion independently over several segments in such a way that the motion is one dimensional in each segment. This allows us to choose a coordinate system in each segment where the acceleration vector is co-linear with one of the axes.

When the forces on an object changes continuously, we need to use calculus to determine the motion of the object. If the velocity vector for an object changes direction continuously, we need to model the motion in each dimension independently.

If an object undergoes uniform circular motion, the acceleration vector and the sum of the forces always point towards the centre of the circle. In the radial direction, Newton's Second Law gives

$$\sum \vec{F} = ma_R = m\frac{v^2}{R}$$

If an object's speed is changing as it moves around a circle the acceleration vector will have a component that is towards the centre of the circle (the radial component) and a component that is tangential to the circle. The tangential component is responsible for the change in speed, whereas the radial component is responsible for the change in direction of the velocity.

In a reference frame that is rotating about a circle, an inertial force, sometimes called the centrifugal force, appears to push all objects co-moving with the reference frame towards the outside of the circle.

6.6 Thinking about the material

Reflect and research

1. Is there a maximum speed with which an object can spin? (Something about the thing eventually flying apart if it rotates too fast, as the atoms can not be held together at some point - maybe there is a cool video to look up?)

To try at home

1. Spin a mass on a string in a vertical circle, what is the tension in the string when the mass is at the top for it to barely make it around?
2. Spin a mass on a string in a vertical circle, how does the minimum speed at the top of the circle to barely make it around depend on the radius of the circle or the mass?
3. Spin a mass on a string in a vertical circle, describe the motion if the mass does not have the minimum speed to make it around the circle. If it makes it to the top, does it automatically make it all the way around the circle?

To try in the lab

1. Build a conical pendulum and determine whether the opening angle of the cone is related to the speed of the bob, in the way that you expect it to be.

6.6.1 Problems and Solutions

Problem 6-1: Consider a conical pendulum with a mass m, attached to a string of length L. The mass executes uniform circular motion in the horizontal plane, about a circle of radius R, as shown in Figure 6.24. One can think of the horizontal circle and the point where the string is attached to as forming a cone. The circular motion is such that the (constant) angle between the string and the vertical is θ. (Solution)

a) Derive an expression for the tension in the string.
b) Derive an expression for the speed of the mass.
c) Derive an expression for the period of the motion.

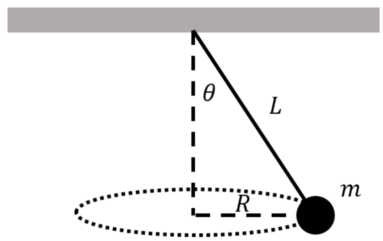


Figure 6.24: The conical pendulum.

Problem 6-2: Barb and Kenny are going to the amusement park. Barb insists on riding the giant roller coaster, but Kenny is scared that they will fall out of the roller coaster at the top of the loop. Barb reassures Kenny by asking the roller coaster technician for more information. The technician says that they will be travelling at 15 m/s when upside down, and that the roller coaster loop has a radius of 22 m. Kenny is still sceptical. Is he correct in being sceptical? (Solution)

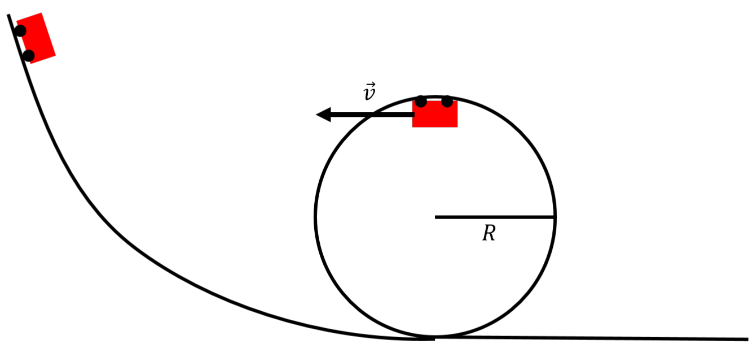


Figure 6.25: The roller coaster

6.6.2 Solutions

Solution to problem 6-1:

a) We start by identifying the forces that are acting on the mass. These are:
   - $\vec{F}_g$, its weight, with a magnitude mg.
   - $\vec{F}_T$, a force of tension exerted by the string.

   The forces are illustrated in Figure 6.26, along with our choice of coordinate system and the direction of the acceleration of the mass (towards the centre of the circle).

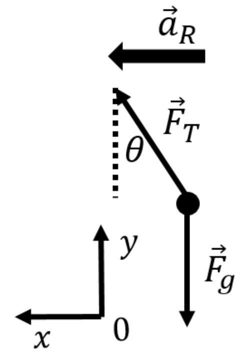


Figure 6.26: Forces acting on the conical pendulum

The y component of Newton's Second law gives the relation between the tension in the string, the weight, and the angle θ

$$\begin{aligned}\sum F_y &= 0\\ F_T\cos\theta - F_g &= 0\\ F_T\cos\theta &= mg\\ \therefore F_T &= \frac{mg}{\cos\theta}\end{aligned}$$

b) In order for the mass to move in a circle, the net force must be directed towards the centre of the circle at all times. The x component of Newton's Second Law, combined with our expression for the magnitude of the tension, F_T, allows us to determine the speed of the mass:

$$\begin{aligned}\sum F_x &= ma_r\\ F_T\sin\theta &= m\frac{v^2}{R}\\ \left(\frac{mg}{\cos\theta}\right)\sin\theta &= m\frac{v^2}{R}\\ g\tan\theta &= \frac{v^2}{R}\\ \therefore v &= \sqrt{gR\tan\theta}\end{aligned}$$

c) Now that we know the speed, we can easily find the period, T, of the motion:

$$\begin{aligned}
T &= \frac{2\pi R}{v} \\
&= \frac{2\pi R}{\sqrt{gR\tan\theta}} = 2\pi\sqrt{\frac{R}{g\tan\theta}}
\end{aligned}$$

Solution to problem 6-2: We need to determine if the speed of Barb and Kenny is large enough for them to go around the circle. The minimum speed that they must have at the top of the loop is such that their weight (the only force acting on them) provides the centripetal (net) force required to go around the loop.

Writing Newton's Second Law in the vertical direction, for the case where only the weight acts on Barb or Kenny (mass m), when they are going at speed v

$$\begin{aligned}
mg &= ma_R = m\frac{v^2}{R} \\
\therefore v &= \sqrt{gR} = \sqrt{(9.8\,\text{m/s}^2)(22\,\text{m})} = 14.68\,\text{m/s}
\end{aligned}$$

This corresponds to the minimum speed that they must have at the top of the loop to make it around. If they go faster, the normal force from their seat (downwards, since they are upside-down), would result in a larger net force towards the centre of the circle. This situation corresponds to the normal force from their seat just barely reaching 0 at the top of the loop. Since the roller coaster is quoted as having a speed of 15 m/s at the top of the loop, they will just barely make it. However, this is way too close to the minimal speed to not fall out of the roller coaster, so Kenny is correct in being sceptical! The engineers designing the roller coaster should include a much bigger safety margin!

7 Work and energy

In this chapter, we introduce a new way to build models derived from Newton's theory of classical physics. We will introduce the concepts of work and energy, which will allow us to model situations using scalar quantities, such as energy, instead of vector quantities, such as forces. It is important to remember that even when we are using energy and work, these tools are derived from Newton's Laws; that is, we may not be using Newton's Second Law explicitly, but the models that we develop are still based on the same theory of classical physics.

Learning Objectives

- Understand the concept of work and how to calculate the work done by a force.
- Understand the concept of the net work done on an object and how that relates to a change in speed of the object.
- Understand the concept of kinetic energy and where it comes from.
- Understand the concept of power.

Think About It

You are holding a heavy book with your arm extended horizontally. The book does not move as you struggle to keep it from falling to the ground. Does your arm do work on the book? If you start walking to class while holding the book, does your arm do work on the book?

7.1 Work

Review Topics

- Section A.3.3 on the scalar product.
- Section B.3 on integrals.

We introduce the concept of work as the starting point for building models using energy instead of forces. Work is a scalar quantity that is meant to represent how a force exerted on an object over a given distance results in a change in speed of that object. We will first introduce the concept of work done by a force on an object, and then look at how work can change the kinematics of the object. This is analogous to how we first defined the concept of force, and then looked at how force affects motion (by using Newton's Second Law, which connected the concept of force to the acceleration of the object).

The work done by a force, $\vec{F}$, on an object over a displacement, $\vec{d}$, is defined to be:

$$\boxed{W = \vec{F}\cdot\vec{d} = Fd\cos\theta = F_x d_x + F_y d_y + F_z d_z} \tag{7.1}$$

where θ is the angle between the vectors when they are placed tail to tail, as in Figure 7.1. The dimension of work, force times displacement, is also called "energy". The S.I. unit for energy is the Joule (abbreviated J) which is equivalent to Nm or kgm^2/s^2 in base units.

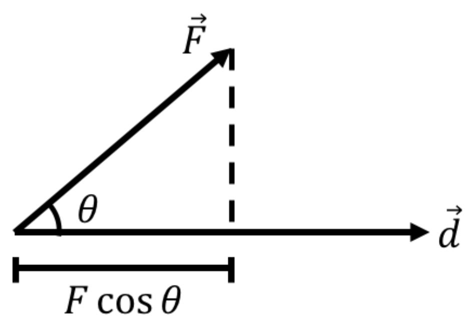


Figure 7.1: When determining the scalar product $\vec{F}\cdot\vec{d} = Fd\cos\theta$, θ is the angle between the vectors when they are placed tail to tail.

The work "done" by the force is the scalar product of the force vector and the displacement vector of the object. We say that the force "does work" if it is exerted while the object moves (has a displacement vector) and in such a way that the scalar product of the force and displacement vectors is non-zero. A force that is perpendicular to the displacement vector of an object does no work (since the scalar product of two perpendicular vectors is zero). A force exerted in the same direction as the displacement will do positive work ($\cos\theta$ positive), and a force in the opposite direction of the displacement will do negative work ($\cos\theta$ negative). As we will see, positive work corresponds to increasing the speed of the object, whereas negative work corresponds to decreasing its speed. No work corresponds to no change in speed (but could corresponds to a change in velocity).

Checkpoint 7-1

A pendulum of length R consists of a mass connected to a string (Figure 7.2). The string exerts a force of tension $\vec{F}_T$ on the mass. What is the work done by tension when the pendulum swings through an angle θ?

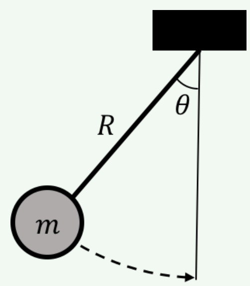


Figure 7.2: A pendulum swings through an angle θ.

A) $W = F_T R\theta$
B) $W = F_T R(1 - \cos\theta)$
C) Tension does no work on the mass.

You may be tempted to ask, "Why work? Why not something else? Why that scalar product in particular? How could we possibly have thought of that?". In general, it seems arbitrary that we introduce the quantity "work" and then find that it leads to a convenient way of building models. However, we did not just pull this quantity out of thin air! Many theorists, over many years, tried all sorts of quantities and ways to rephrase Newton's Theory that were not helpful. The quantities that make it into textbooks are the ones that turned out to be useful. You should also keep in mind that, just like force, work is a "made-up" mathematical tool that is helpful in describing the world around us. There is no such thing as work or energy; they are just useful mathematical tools.

7.1.1 Work in one dimension.

Work involves vectors, so we can first examine the concept in one dimension, before extending this to two and three dimensions. We can choose x as the coordinate in one dimension, so that all vectors only have an x component. We can write a force vector as $\vec{F} = F\hat{x}$, where F is the x component of the force (which could be positive or negative). A displacement vector can be written as $\vec{d} = d\hat{x}$, where again, d is the x component of the displacement, and can be positive or negative. In one dimension, work is thus:

$$W = \vec{F} \cdot \vec{d} = (F\hat{x}) \cdot (d\hat{x}) = Fd(\hat{x} \cdot \hat{x}) = Fd$$

where $\hat{x} \cdot \hat{x} = 1$. Consider, for example, the work done by a force, $\vec{F}$, on a box, as the box moves along the x axis from position $x = x_0$ to position $x = x_1$, as shown in Figure 7.3.

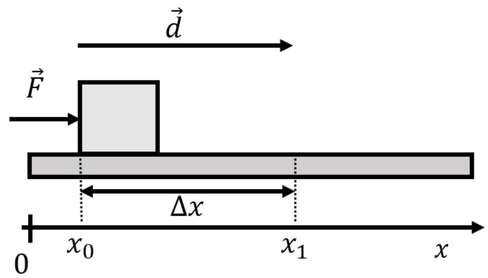


Figure 7.3: A force, $\vec{F}$, exerted on an object as it moves from position $x = x_0$ to position $x = x_1$.

We can write the length of the displacement vector as $||\vec{d}|| = d = \Delta x = x_1 - x_0$. The work done by the force is given by:

$$W = \vec{F} \cdot \vec{d} = F\hat{x} \cdot \Delta x\hat{x} = F\Delta x = F(x_1 - x_0)$$

which is a positive quantity, since $x_1 > x_0$, with our choice of coordinate system.

Checkpoint 7-2

A constant force in the positive x direction, $\vec{F}$, acts on a box, as in Figure 7.3. Consider the work done by $\vec{F}$ as the box moves from x_1 to x_0. How does it compare to the work done by $\vec{F}$ when moving from x_0 to x_1 (that we calculated above)?

A) $\vec{F}$ does no work on the box when it moves from x_0 to x_1.
B) The work has the same magnitude as before, but the work is now negative.
C) The work done by $\vec{F}$ is the same in both cases.

7.1.2 Work in one dimension - varying force

Suppose that instead of a constant force, $\vec{F}$, we have a force that changes with position, $\vec{F}(x)$, and can take on three different values between $x = x_0$ and $x = x_3$:

$$\vec{F}(x) = \begin{cases} F_1\hat{x} & x < \Delta x \\ F_2\hat{x} & \Delta x \leq x < 2\Delta x \\ F_3\hat{x} & 2\Delta x \leq x \end{cases}$$

as illustrated in Figure 7.3, which shows the force on an object as it moves from position $x = x_0$ to position $x = x_3$, along three (equal) displacement vectors, $\vec{d}_1 = \vec{d}_2 = \vec{d}_3 = \Delta x\hat{x}$.

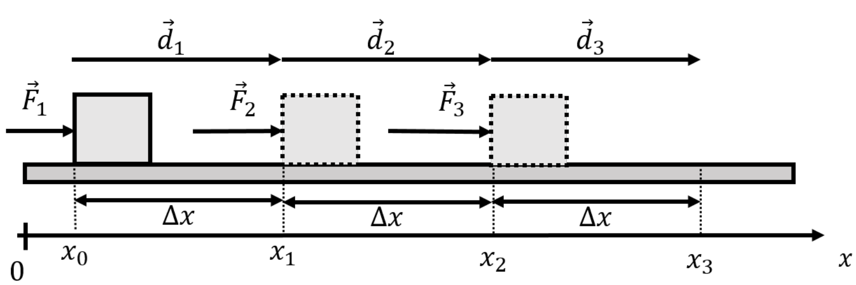


Figure 7.4: A varying force, $\vec{F}(x)$, exerted on an object as it moves from position $x = x_0$ to position $x = x_3$.

The total work done by the force over the three separate displacements is the sum of the work done over each displacement:

$$\begin{aligned} W^{tot} &= W_1 + W_2 + W_3 \\ &= \vec{F}_1 \cdot \vec{d}_2 + \vec{F}_2 \cdot \vec{d}_2 + \vec{F}_3 \cdot \vec{d}_3 \\ &= F_1 \Delta x + F_2 \Delta x + F_3 \Delta x \end{aligned}$$

If instead of 3 segments we had N segments and the x component of the force had the N corresponding values F_i in the N segments, the total work done by the force would be:

$$W^{tot} = \sum_{i=0}^{N} \vec{F}_i \cdot \Delta\vec{x}$$

where we introduced a vector $\Delta\vec{x}$ to be the vector of length Δx pointing in the positive x direction. In the limit where $\vec{F}(x)$ changes continuously as a function of position, we take the limit of an infinite number of infinitely small segments of length dx, and the sum becomes an integral:

$$\boxed{W^{tot} = \int_{x_0}^{x_f} \vec{F}(x) \cdot d\vec{x}} \tag{7.2}$$

where the work was calculated in going from $x = x_0$ to $x = x_f$, and $d\vec{x} = dx\hat{x}$ is an infinitely small displacement vector (of length dx) in the positive x direction.

Example 7-1

A block is pressed against the free end of a horizontal spring with spring constant, k, so as to compress the spring by a distance D relative to its rest length, as shown in Figure 7.5. The other end of the spring is fixed to a wall. What is the work done by the spring force on the block in going from $x = -D$ to $x = 0$? What is the work done by the block on the spring over the same displacement?

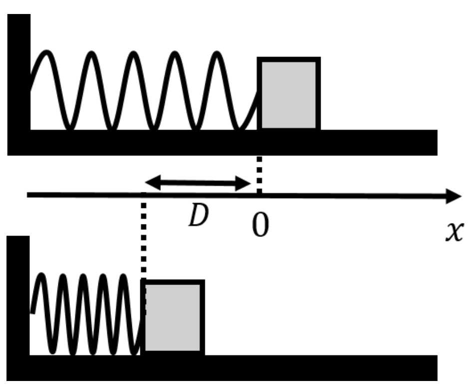


Figure 7.5: A block is pressed against a horizontal spring so as to compress the spring by a distance D relative to its rest length.

Solution

The force exerted by the spring on the block changes continuously with position, according to Hooke's law:

$$\vec{F}(x) = -kx\hat{x}$$

and points in the positive x direction when the end of the spring has a negative x position (with our coordinate choice illustrated in Figure 7.5, where the origin is located at the rest length of the spring). To calculate the work done by the force, we sum the work done by the force over many infinitesimally small displacements $d\vec{x}$ (using an integral):

$$\begin{aligned}
W &= \int_{-D}^{0} \vec{F}(x) \cdot d\vec{x} \\
&= \int_{-D}^{0} (-kx\hat{x}) \cdot (dx\hat{x}) \\
&= \int_{-D}^{0} -kxdx(\hat{x} \cdot \hat{x}) \\
&= -\int_{-D}^{0} kxdx \\
&= -\left[\frac{1}{2}kx^2\right]_{-D}^{0} \\
&= \frac{1}{2}kD^2
\end{aligned}$$

In order to determine the work that was done by the block on the spring, we need to determine the force, $\vec{F}'(x)$, exerted by the block on the spring. By Newton's Third Law, this is equal in magnitude but opposite in direction to the force exerted by the spring on the block:

$$\vec{F}'(x) = -\vec{F}(x) = kx\hat{x}$$

The work done by the block on the spring over the same displacement is:

$$\begin{aligned}
W' &= \int_{-D}^{0} \vec{F}'(x) \cdot d\vec{x} \\
&= \int_{-D}^{0} (kx\hat{x}) \cdot (dx\hat{x}) \\
&= \int_{-D}^{0} kxdx = -\frac{1}{2}kD^2
\end{aligned}$$

which is negative. This makes sense because the force exerted by the block on the spring is in the direction opposite to the direction of displacement, so the work should be negative.

7.1.3 Work in multiple dimensions

First, consider the work done by a force $\vec{F}$ in pulling a crate over a displacement $\vec{d}$, in the case where the force is directed at an angle θ above the horizontal, as shown in Figure 7.6, and the displacement is along the x axis (or rather, we chose the x axis to be parallel to the displacement).

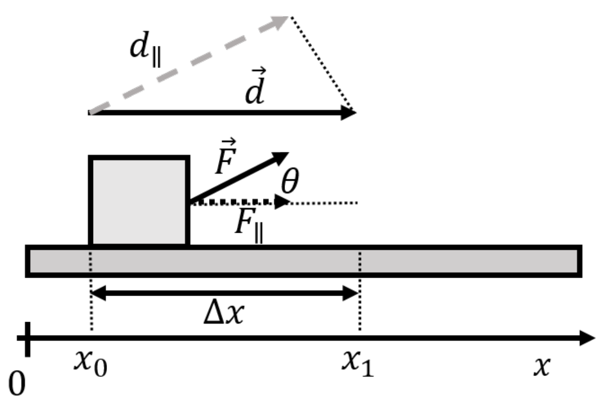


Figure 7.6: A force, $\vec{F}$, exerted on an object as it moves from position $x = x_0$ to position $x = x_1$.

The work done by the force is given by:

$$\begin{aligned}
W = \vec{F} \cdot \vec{d} &= Fd\cos\theta \\
&= F_{\parallel}d \\
&= Fd_{\parallel}
\end{aligned}$$

where we highlighted the fact that the scalar product "picks out" components of vectors that are parallel to each other. $F_{\parallel} = F\cos\theta$ is the component of $\vec{F}$ that is parallel to $\vec{d}$, and $d_{\parallel} = d\cos\theta$ is the component of $\vec{d}$ that is parallel to $\vec{F}$. These are also shown in Figure 7.6.

Checkpoint 7-3

Brent and Dean pull two crates by using ropes that make the same angle above the horizontal and with the same force. The magnitude of the crates' displacement is the same, but Dean's crate moves horizontally on the ground while Brent's crate moves up a frictionless ramp that is parallel to the rope used to pull the crate. Who did more work on the crate?

A) Dean because there is friction between his crate and the ground.
B) Brent.
C) They did the same amount of work.

In general, if an object is moving along an arbitrary path, we cannot choose the x axis to be parallel to the displacement or to the force. If the path can be sub-divided into straight segments over which the force is constant, as in Figure 7.7, we can calculate the work done by the force over each segment and add the work done in each segment together to obtain the total work done by the force. Note that, in general, the work done by a force as an object moves from one position to another depends on the particular path that was taken between the two positions, since different paths will have difference lengths.

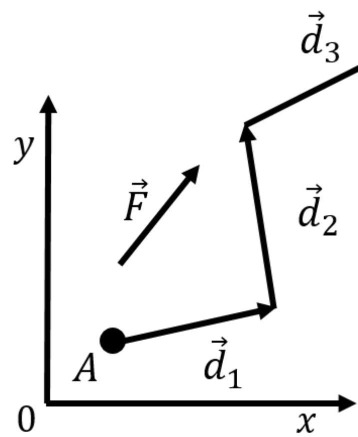


Figure 7.7: An arbitrary two dimensional path of an object from A to B broken into three straight segments.

Example 7-2

Compare the work done by the force of kinetic friction in sliding a crate along a horizontal surface from position A (coordinates x_A, y_A) to position B (coordinates x_B, y_B) using the two different paths depicted in Figure 7.8. Assume that the mass of the crate is m and that the coefficient of kinetic friction between the crate and the ground is μ_k.

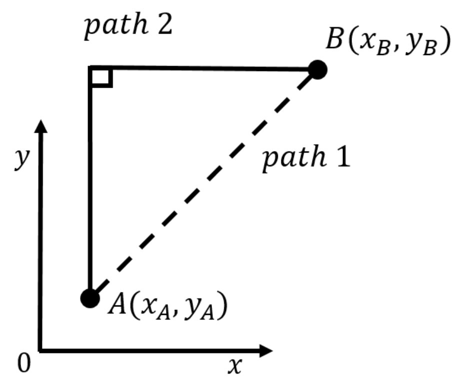


Figure 7.8: Two possible paths to slide a crate from position A to position B, as seen from above.

Solution

The force of kinetic friction is always in the direction opposite to that of motion. Thus, regardless of the path taken, the force of friction will do negative work.

Let us first calculate the work done by the force of kinetic friction along the first path (the straight line). The force of kinetic friction will have a magnitude:

$$f_k = \mu_k N = \mu_k mg$$

The normal force will have the same magnitude as the weight because the crate is not moving (accelerating) in the direction perpendicular to the xy plane. The displacement vector from A to B can be written as:

$$\vec{d} = (x_B - x_A)\hat{x} + (y_B - y_A)\hat{y}$$
$$\therefore ||\vec{d}|| = d = \sqrt{(x_B - x_A)^2 - (y_B - y_A)^2}$$

The force of kinetic friction will be in the opposite direction of the displacement vector, so the angle between the two vectors is 180° ($\cos\theta = -1$). The work done by the force of kinetic friction is thus:

$$W = \vec{f_k} \cdot \vec{d} = f_k d \cos\theta = -\mu_k mg\sqrt{(x_B - x_A)^2 - (y_B - y_A)^2}$$

and is negative, as expected.

For path 2, we break up the motion into two segments, with displacements vectors $\vec{d_1}$

(along y) and $\vec{d}_2$ (along x). We can write the two displacement vectors as:

$$\begin{aligned}
\vec{d}_1 &= 0\hat{x} + (y_B - y_A)\hat{y} \\
\therefore ||\vec{d}_1|| &= d_1 = (y_B - y_A) \\
\vec{d}_2 &= (x_B - x_A)\hat{x} + 0\hat{y} \\
\therefore ||\vec{d}_2|| &= d_2 = (x_B - x_A)
\end{aligned}$$

Along each segment, the force of kinetic friction is anti-parallel to the displacement (note that the force of friction changes direction over the two segments), but the magnitude is $f_k = \mu_k mg$. The work done along the first segment is thus:

$$W_1 = \vec{f}_k \cdot \vec{d}_1 = f_k d_1 \cos\theta = -\mu_k mg(y_B - y_A)$$

The work done along the second segment is:

$$W_2 = \vec{f}_k \cdot \vec{d}_2 = f_k d_2 \cos\theta = -\mu_k mg(x_B - x_A)$$

And the total work done by the force of kinetic friction over the second path is:

$$W^{tot} = W_1 + W_2 = -\mu_k mg\left((x_B - x_A) + (y_B - y_A)\right)$$

which is more work than was done along path 1. This makes sense because for both paths, the force of friction has the same magnitude and is always in the opposite direction of motion; thus, the longer the path, the more work will be done by the force.

Example 7-3

A box of mass m is moved from the floor onto a table using two different paths, as shown in Figure 7.9. The table is a horizontal distance L away from where the box starts and a height H above the floor. Compare the work done by the weight of the box along the two possible paths.

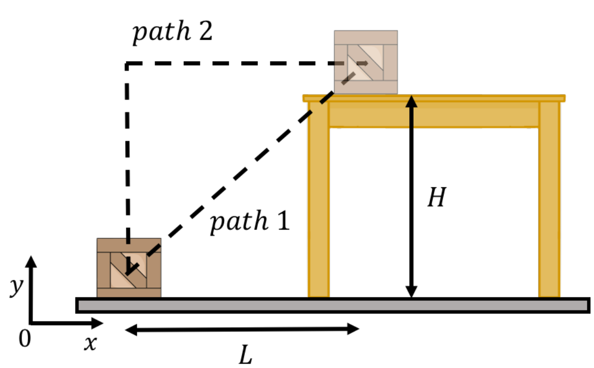


Figure 7.9: Two possible paths to move a box from the floor onto a table.

Solution

We can use a coordinate system such that the origin coincides with the initial position of the box. x is horizontal and y is vertical, as shown in Figure 7.9. The weight of the box can be written as:

$$\vec{F}_g = -mg\hat{y}$$

and points in the negative y direction with a magnitude of mg. To calculate the work done by the weight along the first path, we first determine the corresponding displacement vector, $\vec{d}$:

$$\vec{d} = L\hat{x} + H\hat{y}$$

and we can then determine the work:

$$\begin{aligned}
W &= \vec{F}_g \cdot \vec{d} = (-mg\hat{y}) \cdot (L\hat{x} + H\hat{y}) \\
&= F_x d_x + F_y d_y = (0)(L) + (-mg)(H) \\
&= -mgH
\end{aligned}$$

Along path 1, the work done by the weight is negative, and does not depend on the horizontal distance L. Let us now calculate the work done along the second path, which we break up into two segments with displacement vectors $\vec{d}_1$ (vertical) and $\vec{d}_2$ (horizontal). The displacement vectors are:

$$\begin{aligned}
\vec{d}_1 &= H\hat{y} \\
\vec{d}_2 &= L\hat{x}
\end{aligned}$$

The work done along the vertical segment is:

$$\begin{aligned}
W_1 &= \vec{F}_g \cdot \vec{d}_1 = (-mg\hat{y}) \cdot (H\hat{y}) \\
&= -mgH
\end{aligned}$$

The work done along the horizontal segment is:

$$\begin{aligned}
W_2 &= \vec{F}_g \cdot \vec{d}_2 = (-mg\hat{y}) \cdot (L\hat{x}) \\
&= 0
\end{aligned}$$

which is zero, because the force of gravity is always vertical and thus perpendicular to the displacement vector of the horizontal segment. The total work done by the weight along the second path is:

$$W^{tot} = W_1 + W_2 = -mgH$$

which is the same as the work done along path 1. As we will see, when a force is constant in magnitude and direction, the work that it does on an object in going from one position to another is independent of the path taken. This was not the case in Example 7-2, because the direction of the force of kinetic friction depends on the direction of the displacement.

Checkpoint 7-4

Clare and Amelia go down two different slides, as shown in Figure 7.10. Clare and Amelia have the same mass and the slides have the same non-zero coefficients of friction.

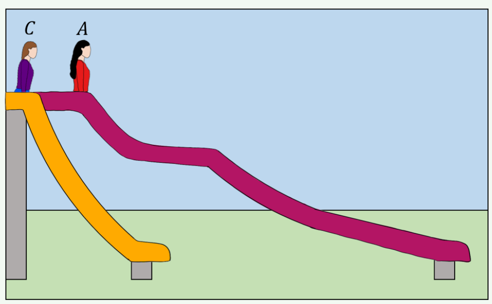


Figure 7.10: Clare (C) and Amelia (A) go down two different slides of the same height.

For each of the following forces, decide whether the force: does more work on Clare, does more work on Amelia, or does the same amount of work on both.

1. The force of gravity...
2. The force of friction...
3. The normal force from the slide...

The most general case for which we can calculate the work done by a force is the case when the force changes continuously along a path where the displacement also changes direction continuously. This is illustrated in Figure 7.11 which shows an arbitrary path between two points A and B, and a force, $\vec{F}(\vec{r})$, that depends on position ($\vec{r}$). In general, the work done by the force on an object that goes from A to B will depend on the actual path that was taken.

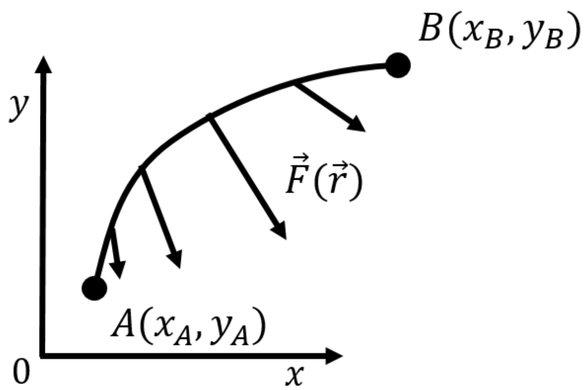


Figure 7.11: An arbitrary path between two points A and B with a force that depends on position, $\vec{F}(\vec{r})$.

The strategy for calculating the work in the general case is the same: we break up the path into small straight segments with displacement vectors $d\vec{l}$ (Figure 7.12) where we assume that the force is constant over the segment. The total work is the sum of the work over each segment:

$$\boxed{W = \int_A^B \vec{F}(\vec{r}) \cdot d\vec{l}} \tag{7.3}$$

As usual, we use the integral symbol to indicate that you need to take an infinite number of infinitely small segments $d\vec{l}$ in order to calculate the sum.

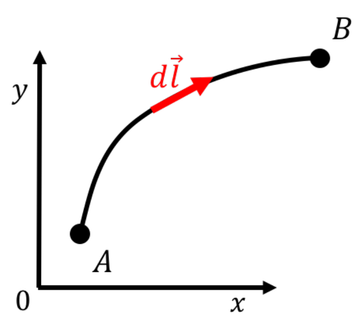


Figure 7.12: We divide the path into infinitesimally small segments with displacement vectors $d\vec{l}$.

You should note that this is not an integral like any other that we have seen so far: the integral is not over a single integration variable (usually we use x), but it is the integral (the sum!) over the specific path that we have chosen in going from A to B. This is called a "path integral", and is generally difficult to evaluate.

Example 7-4

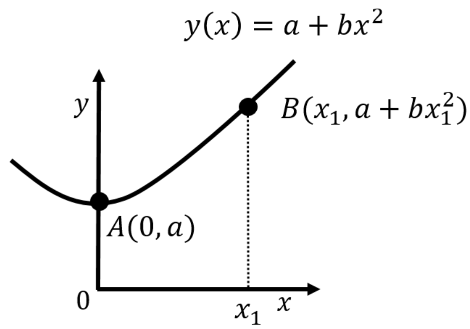


Figure 7.13: A parabolic path between A and B.

A force, $\vec{F}(\vec{r}) = \vec{F}(x, y) = F_x\hat{x} + F_y\hat{y}$, is exerted on an object. The object starts at position A and ends at position B, along a parabolic path, $y(x) = a + bx^2$, as depicted in Figure 7.13. What is the work done by the force, $\vec{F}$, along this trajectory?

Solution

In this case, the force can change with position (if F_x and F_y are not constant), and the direction of the path changes continuously. When we break up the path into small segments $d\vec{l}$, we need to incorporate the equation of the parabola to include the fact that $d\vec{l}$ must always be tangent to the parabola. Consider one small segment along the trajectory and the infinitesimal displacement vector $d\vec{l}$ at that point, as in Figure 7.14.

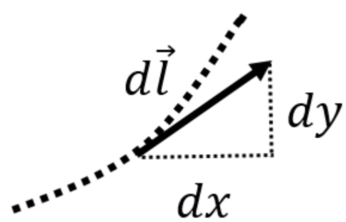


Figure 7.14: The infinitesimal displacement vector, $d\vec{l}$.

We can write the x and y components of the vector as infinitesimal distances, dx and dy, along the x and y axes, respectively. The vector $d\vec{l}$ can thus be written:

$$d\vec{l} = dx\hat{x} + dy\hat{y}$$

The total work done by the force is then:

$$\begin{aligned}
W &= \int_A^B \vec{F}(\vec{r}) \cdot d\vec{l} \\
&= \int_A^B (F_x\hat{x} + F_y\hat{y}) \cdot (dx\hat{x} + dy\hat{y}) \\
&= \int_A^B (F_x dx + F_y dy) \\
\therefore W &= \int_A^B F_x dx + \int_A^B F_y dy
\end{aligned}$$

where in the last line, we simply used the property that the integral of a sum is the sum of the corresponding integrals. At this point, we have two integrals over integration variables (x and y) that are meaningful. However, we have not yet used the fact that our path is a parabola, and in general, we expect that the shape of the path is important. By saying that we are integrating (or calculating the work) over a specific path, we are really saying that x and y are not independent; that is, if we know the value of x at some point on the path, we know the corresponding value of y ($y = a + bx^2$).

Since x and y are not independent, we can use a "substitution of variables" in order to express y in terms of x, and dy in terms of dx:

$$\begin{aligned}
y(x) &= a + bx^2 \\
\frac{dy}{dx} &= 2bx \\
\therefore dy &= 2bxdx
\end{aligned}$$

This allows us to convert the integral over y to an integral over x, which also allows us to be explicit for the limits of the integral (in our example, the integral goes from $x = 0$ to $x = x_1$):

$$\begin{aligned}
W &= \int_A^B F_x dx + \int_A^B F_y dy \\
&= \int_0^{x_1} F_x dx + \int_0^{x_1} F_y (2bxdx) \\
&= \int_0^{x_1} (F_x + 2bxF_y) dx
\end{aligned}$$

where we would need to know how F_x and F_y depends on x and y in order to actually evaluate the integral.

For example, if the force were constant (F_x and F_y constant), then the work done along the parabolic path would be:

$$\begin{aligned} W &= \int_0^{x_1} (F_x + 2bxF_y)dx \\ &= \left[F_x x + bF_y x^2\right]_0^{x_1} \\ &= F_x x_0 + bF_y x_0^2 \end{aligned}$$

As we mentioned earlier, **if the force is constant in magnitude and direction**, then the work done is independent of path. We can easily check this, using the displacement vector $\vec{d} = x_1\hat{x} + bx_1^2\hat{y}$:

$$\begin{aligned} W &= \vec{F} \cdot \vec{d} = (F_x\hat{x} + F_y\hat{y}) \cdot (x_1\hat{x} + bx_1^2\hat{y}) \\ &= F_x x_1 + bF_y x_1^2 \end{aligned}$$

as we found above.

7.1.4 Net work done

So far, we have considered the work done on an object by a single force. If more than one force is exerted on an object, then each force can do work on the object, and we can calculate the "net work" done on the object by adding together the work done by each force. We will show that this is equivalent to first calculating the net force on the object, F^{net} (i.e. the vector sum of the forces on the object), and then calculating the work done by the net force.

Suppose that three forces, $\vec{F}_1$, $\vec{F}_2$, and $\vec{F}_3$ are exerted on an object as it moves such that its displacement vector is $\vec{d}$. The net work done on the object is easily shown to be equivalent to the work done by the net force::

$$\begin{aligned} W^{net} &= W_1 + W_2 + W_3 \\ &= \vec{F}_1 \cdot \vec{d} + \vec{F}_2 \cdot \vec{d} + \vec{F}_3 \cdot \vec{d} \\ &= (F_{1x}d_x + F_{1y}d_y + F_{1z}d_z) + (F_{2x}d_x + F_{2y}d_y + F_{2z}d_z) + (F_{3x}d_x + F_{3y}d_y + F_{3z}d_z) \\ &= (F_{1x} + F_{2x} + F_{3x})d_x + (F_{1y} + F_{2y} + F_{3y})d_y + (F_{1z} + F_{2z} + F_{3z})d_z \\ &= \vec{F}^{net} \cdot \vec{d} \end{aligned}$$

where $\vec{F}^{net} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3$ is the net force. The result is easily generalized to any number of forces, including if those forces change as a function of position:

$$W^{net} = \int_A^B F^{net}(\vec{r}) \cdot d\vec{l}$$

Example 7-5

You push with an unknown horizontal force, $\vec{F}$, against a crate of mass m that is located on an inclined plane that makes an angle θ with respect to the horizontal, as shown in Figure 7.15. The coefficient of kinetic friction between the crate and the incline is μ_k. You push in such a way that that crates moves at a constant speed up the incline. What is the net work done on the crate if it moves up the incline by a distance d?

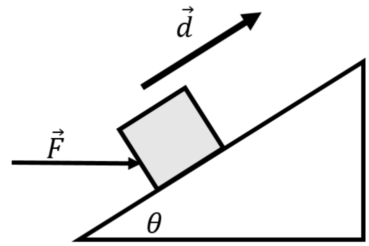


Figure 7.15: A crate being pushed up an incline.

Solution

Although the answer may be obvious, let's go the long way about it and calculate the work done by each force, and then sum them together to get the total work done. We start by identifying the forces exerted on the crate:

1. $\vec{F}$, the applied force, of unknown magnitude, $\vec{F}$.
2. $\vec{F}_g$, the weight of the crate, with magnitude mg.
3. $\vec{N}$, a normal force exerted by the incline.
4. $\vec{f}_k$, a force of kinetic friction, with magnitude $\mu_k N$, that points in the direction opposite of $\vec{d}$.

These are shown in the free-body diagram in Figure 7.16, along with our choice of coordinate system, and the displacement vector.

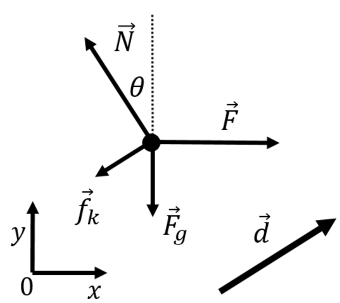


Figure 7.16: Free-body diagram for the crate on the incline.

With our choice of coordinate system, the displacement vector is given by:

$$\vec{d} = d(\cos\theta\hat{x} + \sin\theta\hat{y})$$

Before calculating the work done by each force, we need to determine the magnitude of the normal force (and thus of the force of kinetic friction). Since the crate is moving at a constant velocity, its **acceleration is zero**, so the sum of the forces must be zero. Writing out the y component of Newton's Second Law allows us to find the magnitude of the normal force:

$$\begin{aligned}\sum F_y &= N\cos\theta - F_g - f_k\sin\theta = 0\\ \therefore mg &= N\cos\theta - \mu_k N\sin\theta = N(\cos\theta - \mu_k\sin\theta)\\ \therefore N &= \frac{mg}{\cos\theta - \mu_k\sin\theta}\end{aligned}$$

Writing out the x component of Newton's Second Law allows us to find the magnitude of the unknown force F:

$$\begin{aligned}\sum F_x &= F - N\sin\theta - f_k\cos\theta = 0\\ \therefore F &= N\sin\theta + \mu_k N\cos\theta = N(\sin\theta + \mu_k\cos\theta)\\ &= mg\frac{\sin\theta + \mu_k\cos\theta}{\cos\theta - \mu_k\sin\theta}\end{aligned}$$

We now proceed to calculate the work done by each force. The work done by the normal force is identically zero, since it is perpendicular to the displacement vector. The work done by the applied force, $\vec{F} = F\hat{x}$, is:

$$\begin{aligned}W_F &= \vec{F}\cdot\vec{d} = (F\hat{x})\cdot(d(\cos\theta\hat{x} + \sin\theta\hat{y}))\\ &= Fd\cos\theta = mg\frac{\sin\theta + \mu_k\cos\theta}{\cos\theta - \mu_k\sin\theta}d\cos\theta\end{aligned}$$

The work done by the force of gravity, $\vec{F}_g = -mg\hat{y}$, is:

$$\begin{aligned}W_g &= \vec{F}_g\cdot\vec{d} = (-mg\hat{y})\cdot(d(\cos\theta\hat{x} + \sin\theta\hat{y}))\\ &= -mgd\sin\theta\end{aligned}$$

The work done by the force of friction, $\vec{f}_k$, noting that $\vec{f}_k$ and $\vec{d}$ are antiparallel:

$$\begin{aligned}W_f &= \vec{f}_k\cdot\vec{d} = -f_k d = -\mu_k N d\\ &= -\mu_k\frac{mg}{\cos\theta - \mu_k\sin\theta}d\end{aligned}$$

The net work done on the crate is thus:

$$\begin{aligned}
W^{net} &= W_F + W_g + W_f \\
&= mg\frac{\sin\theta + \mu_k\cos\theta}{\cos\theta - \mu_k\sin\theta}d\cos\theta - mgd\sin\theta - \mu_k\frac{mg}{\cos\theta - \mu_k\sin\theta}d \\
&= mgd\left(\frac{\sin\theta + \mu_k\cos\theta}{\cos\theta - \mu_k\sin\theta}\cos\theta - \sin\theta - \mu_k\frac{1}{\cos\theta - \mu_k\sin\theta}\right) \\
&= mgd\left(\frac{(\sin\theta + \mu_k\cos\theta)\cos\theta - \sin\theta(\cos\theta - \mu_k\sin\theta) - \mu_k}{\cos\theta - \mu_k\sin\theta}\right) \\
&= mgd\left(\frac{\sin\theta\cos\theta + \mu_k\cos^2\theta - \sin\theta\cos\theta + \mu_k\sin^2\theta - \mu_k}{\cos\theta - \mu_k\sin\theta}\right) \\
&= mgd\left(\frac{\mu_k(\cos^2\theta + \sin^2\theta) - \mu_k}{\cos\theta - \mu_k\sin\theta}\right) \\
&= 0
\end{aligned}$$

where we used the fact that $\cos^2\theta + \sin^2\theta = 1$. Thus we find that the net work done on the crate is zero!

Discussion: Of course, this makes sense, because the net force on the crate is zero, since it is not accelerating, so the net work done is also zero. As a consequence, or rather, by construction, we have the condition that if the net work done on an object is zero, then that object does not accelerate. We thus have a scalar quantity (work) that can tell us something about whether an object is changing speed. In the next section, we introduce a new quantity, "kinetic energy", to describe how an object's speed changes when the net work done is not zero.

Olivia's Thoughts

Pay close attention to the words "on" and "by." There are a few things about this that can be tricky:

1. In Example 7-5, we were asked to find the **net work** done **on** the crate. Sometimes, the question won't specify that it wants you to find the net work, and will just say "What is the work done **on** the crate?" When you are just asked for the work done "on" an object, the question is implicitly asking for the *net* work done on the object.

2. Just because the net work done **on** an object is zero doesn't mean that the work done **by** each of the forces is zero. This may seem obvious, but it's easy to get tripped up on a test or exam. If you are reading a question about work and it says that the object is moving at a constant speed, it's tempting to just jump ahead and say that the work must be equal to zero. However, you can only say this if it's asking you for the net work done on the object. For instance, in example 7-5, we concluded that since the crate was moving at a constant speed, the net work was equal to zero. But if the question asked you to find the work done on the crate **by gravity**, that would mean something different. The work done **by gravity** in this case is not equal to zero (it's actually negative).

3. The work done "on" an object is not the same as the net work done "by" that object. For example, say you are in a tug-of-war and you pull the other team towards you, but you yourself do not move. The net work done **on** you is zero, but the work done **by** you is not zero. So, when you are talking about work, you should always state explicitly whether the work is being done "on" the object or "by" the object.

Note: The wording won't always be like this - sometimes it will say "How much work do you do on the box?" instead of "How much work is done **by** you on the box," so always be careful. Still, looking for key words like "by" and "on" is a good place to start.

Checkpoint 7-5

A 2 kg box sits on a horizontal surface. A constant horizontal force of 6 N is applied to the box. The box moves with a constant acceleration of $2\,\mathrm{m/s^2}$. Which of the following has the greatest magnitude?

A) The work done by the applied force.
B) The work done by friction.
C) The net work done on the box.

7.2 Kinetic energy and the work energy theorem

At this point, you should be comfortable calculating the net work done on an object upon which several forces are exerted. As we saw in the previous section, the net work done on an object is connected to the object's acceleration; if the net force on the object is zero, then the net work done and acceleration are also zero. In this section, we derive a new quantity, kinetic energy, which allows us to connect the work done on an object with its change in speed. This will allow us to describe motion using only scalar quantities. Like the definition of work, the following derivation appears to "come out of thin air". Remember, though, that theorists have tried all sorts of mathematical tricks to reformulate Newton's Theory, and this is the one that worked.

Consider the most general case of an object of mass m acted upon by a net force, $\vec F^{net}(\vec r)$, which can vary in magnitude and direction. We wish to calculate the net work done on the object as it moves along an arbitrary path between two points, A and B, in space, as shown in Figure 7.17. The instantaneous acceleration of the object, $\vec a$, is shown along with an "element of the path", $d\vec l$.

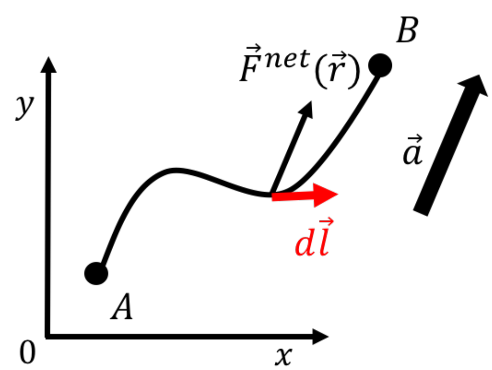


Figure 7.17: An object moving along an arbitrary path between points A and B that is acted upon by a net force $\vec F^{net}$.

The net work done on the object can be written:

$$W^{net} = \int_A^B F^{net}(\vec r)\cdot d\vec l$$

and is in general a difficult integral to evaluate for an arbitrary path. Our goal is to find a way to evaluate this integral by finding a function, K, with the property that:

$$\int_A^B F^{net}(\vec r)\cdot d\vec l = K_B - K_A$$

That is, we will only have to evaluate K at the end points of the path in order to determine the value of the integral. In this way, the function K is akin to an anti-derivative.

In order to determine the form for the function K, we start by noting that, by using Newton's

Second Law, we can write the integral for work in terms of the acceleration of the object:

$$\sum \vec{F} = \vec{F}^{net} = m\vec{a}$$
$$\therefore \int_A^B F^{net}(\vec{r}) \cdot d\vec{l} = \int_A^B m\vec{a} \cdot d\vec{l} = m \int_A^B \vec{a} \cdot d\vec{l}$$

where we assumed that the mass of the object does not change along the path and can thus be factored out of the integral. Consider the scalar product of the acceleration, $\vec{a}$, and the path element, $d\vec{l} = dx\hat{x} + dy\hat{y} + dz\hat{z}$, written in terms of the velocity vector:

$$\begin{aligned}
\vec{a} &= \frac{d\vec{v}}{dt} \\
\therefore \vec{a} \cdot d\vec{l} &= \frac{d\vec{v}}{dt} \cdot d\vec{l} \\
&= \left(\frac{dv_x}{dt}\hat{x} + \frac{dv_y}{dt}\hat{y} + \frac{dv_z}{dt}\hat{z}\right) \cdot (dx\hat{x} + dy\hat{y} + dz\hat{z}) \\
&= \frac{dv_x}{dt}dx + \frac{dv_y}{dt}dy + \frac{dv_z}{dt}dz
\end{aligned}$$

Any of the terms in the sum can be re-arranged so that the time derivative acts on the element of path (dx, dy, or dz) instead of the velocity, for example:

$$\frac{dv_x}{dt}dx = \frac{dx}{dt}dv_x$$

where we recognize that $\frac{dx}{dt} = v_x$. We can thus write the scalar product between the acceleration vector and the path element as:

$$\begin{aligned}
\vec{a} \cdot d\vec{l} &= \frac{dv_x}{dt}dx + \frac{dv_y}{dt}dy + \frac{dv_z}{dt}dz \\
&= \frac{dx}{dt}dv_x + \frac{dy}{dt}dv_y + \frac{dz}{dt}dv_z \\
&= v_x dv_x + v_y dv_y + v_z dv_z
\end{aligned}$$

The integral for the net work done can be written as:

$$\begin{aligned}
W^{net} &= \int_A^B F^{net}(\vec{r}) \cdot d\vec{l} = m \int_A^B (v_x dv_x + v_y dv_y + v_z dv_z) \\
&= m \int_A^B v_x dv_x + m \int_A^B v_y dv_y + m \int_A^B v_z dv_z
\end{aligned}$$

which corresponds to the sum of three integrals over the three independent components of the velocity vector. The components of the velocity vector are functions that change over the path and have fixed values at either end of the path. Let the velocity vector of the object at point A be $\vec{v}_A = (v_{Ax}, v_{Ay}, v_{Az})$ and the velocity vector at point B be $\vec{v}_B = (v_{Bx}, v_{By}, v_{Bz})$. The integral over, say, the x component of velocity is then:

$$\begin{aligned}
m \int_A^B v_x dv_x &= m \int_{v_{Ax}}^{v_{Bx}} v_x dv_x = m \left[\frac{1}{2}v_x^2\right]_{v_{Ax}}^{v_{Bx}} \\
&= \frac{1}{2}m(v_{Bx}^2 - v_{Ax}^2)
\end{aligned}$$

We can thus write the net work integral as:

$$\begin{aligned}
W^{net} &= m\int_A^B v_x dv_x + m\int_A^B v_y dv_y + m\int_A^B v_z dv_z \\
&= \frac{1}{2}m(v_{Bx}^2 - v_{Ax}^2) + \frac{1}{2}m(v_{By}^2 - v_{Ay}^2) + \frac{1}{2}m(v_{Bz}^2 - v_{Az}^2) \\
&= \frac{1}{2}m(v_{Bx}^2 + v_{By}^2 + v_{Bz}^2) - \frac{1}{2}m(v_{Ax}^2 + v_{Ay}^2 + v_{Az}^2) \\
&= \frac{1}{2}mv_B^2 - \frac{1}{2}mv_A^2
\end{aligned}$$

where we recognized that the magnitude (squared) of the velocity is given by $v_A^2 = v_{Ax}^2 + v_{Ay}^2 + v_{Az}^2$. We have thus arrived at our desired result; namely, we have found a function of speed, $K(v)$, that when evaluated at the endpoints of the path allows us to calculate the net work done on the object over that path:

$$\boxed{K(v) = \frac{1}{2}mv^2} \tag{7.4}$$

That is, if you know the speed at the start of the path, v_A, and the speed at the end of the path, v_B, then the net work done on the object along the path between A and B is given by:

$$\boxed{W^{net} = \Delta K = K(v_B) - K(v_a)} \tag{7.5}$$

We call $K(v)$ the "kinetic energy" of the object. We can say that the net work done on an object in going from A to B is equal to its change in kinetic energy (final kinetic energy minus initial kinetic energy). It is important to note that we defined kinetic energy in a way that it is equal to the net work done. You may have already seen kinetic energy from past introductions to physics as a quantity that is just given; here, we instead derived a function that has the desired property of being equal to the net work done and called it "kinetic energy".

The relation between the net work done and the change in kinetic energy is called the "Work-Energy Theorem" (or Work-Energy Principle). It is the connection that we were looking for between the dynamics (the forces from which we calculate work) and the kinematics (the change in kinetic energy). Unlike Newton's Second Law, which relates two vector quantities (the vector sum of the forces and the acceleration vector), the Work-Energy Theorem relates two scalar quantities to each other (work and kinetic energy). Although we introduced the kinetic energy as a way to calculate the integral for the net work, if you know the value of the net work done on an object, then the Work-Energy Theorem can be used to calculate the change in speed of the object.

Most importantly, the Work-Energy theorem introduces the concept of "energy". As we will see in later chapters, there are other forms of energy in addition to work and kinetic energy. The Work-Energy Theorem is the starting point for the idea that you can convert one form of energy into another. The Work-Energy Theorem tells us how a force, by doing work, can provide kinetic energy to an object or remove kinetic energy from an object.

Example 7-6

A net work of W was done on an object of mass m that started at rest. What is the speed of the object after the work has been done on the object?

Solution

Using the Work-Energy Theorem:

$$W = \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2$$

where v_i is the initial speed of the object and v_f is its final speed. Since the initial speed is zero, we can easily find the final speed:

$$v_f = \sqrt{\frac{2W}{m}}$$

Example 7-7

A block is pressed against the free end of a horizontal spring with spring constant, k, so as to compress the spring by a distance D relative to its rest length, as shown in Figure 7.18. The other end of the spring is fixed to a wall.

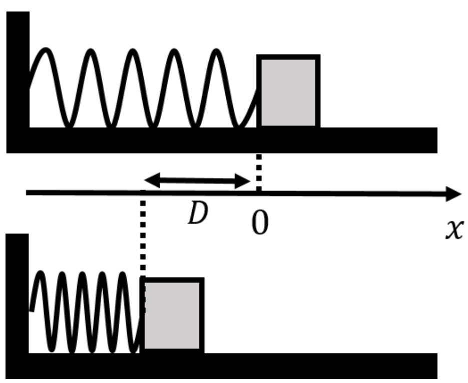


Figure 7.18: A block is pressed against a horizontal spring so as to compress the spring by a distance D relative to its rest length.

If the block is released from rest and there is no friction between the block and the horizontal surface, what is the speed of the block when it leaves the spring?

Solution

This is the same problem that we presented in Chapter 6 in Example 6-3, where we solved a differential equation to find the speed.

Our first step is to calculate the net work done on the object in going from $x=-D$ to $x=0$ (which corresponds to when the object leaves the spring, as discussed in Example 6-3). The forces on the object are:

1. $\vec{F}_g$, its weight, with magnitude mg.
2. $\vec{N}$, the normal force exerted by the ground.
3. $\vec{F}(x)$, the force from the spring, with magnitude kx.

Both the normal force and weight are perpendicular to the displacement, so they will do no work. The net work done is thus the work done by the spring, which we calculated in Example 7-1 to be:

$$W^{net}=W_F=\frac{1}{2}kD^2$$

By the Work-Energy Theorem, this is equal to the change in kinetic energy. Noting that the object started at rest ($v_i=0$), the final speed v_f is found to be:

$$\begin{aligned}
W^{net}&=\frac{1}{2}mv_f^2-\frac{1}{2}mv_i^2=\frac{1}{2}mv_f^2-0\\
\frac{1}{2}kD^2&=\frac{1}{2}mv_f^2\\
\therefore v_f&=\sqrt{\frac{kD^2}{m}}
\end{aligned}$$

Example 7-8

A block is pressed against the free end of a horizontal spring with spring constant, k, so as to compress the spring by a distance D relative to its rest length, as shown in Figure 7.19. The other end of the spring is fixed to a wall.

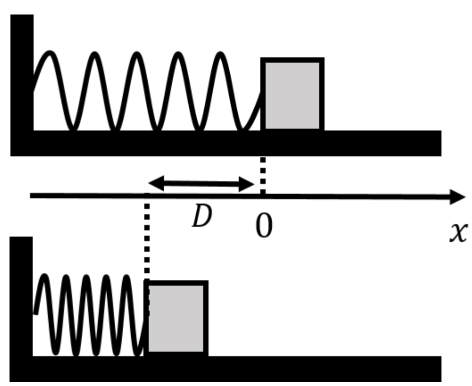


Figure 7.19: A block is pressed against a horizontal spring so as to compress the spring by a distance D relative to its rest length.

If the block is released from rest and the coefficient of kinetic friction between the block and the horizontal surface is μ_k, what is the speed of the block when it leaves the spring?

Solution

This is the same example as the previous one, but with kinetic friction. The forces on the block are:

1. $\vec{F}_g$, its weight, with magnitude mg.
2. $\vec{N}$, the normal force exerted by the ground on the block.
3. $\vec{F}(x)$, the force from the spring, with magnitude kx.
4. $\vec{f}_k$, the force of kinetic friction, with magnitude $\mu_k N$.

Both the normal force and weight are perpendicular to the displacement, so they will do no work. Furthermore, since the acceleration in the vertical direction is zero, the normal force will have the same magnitude as the weight ($N = mg$). The magnitude of the force of kinetic friction is thus $f_k = \mu_k mg$. The net work done will be the sum of the work done by the spring, W_F, and the work done by friction, W_f:

$$W^{net} = W_F + W_f$$

We have already determined the work done by the spring:

$$W_F = \frac{1}{2}kD^2$$

The work done by the force of kinetic friction will be negative (since it is in the direction opposite of the motion) and is given by:

$$W_f = \vec{f}_k \cdot \vec{d} = -f_k D = -\mu_k mgD$$

Applying the work energy theorem, and noting that the block started at rest ($v_i = 0$), the final speed v_f is found to be:

$$\begin{aligned}
W^{net} = W_F + W_f &= \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2 \\
\frac{1}{2}kD^2 - \mu_k mgD &= \frac{1}{2}mv_f^2 \\
\therefore v_f &= \sqrt{\frac{kD^2}{m} - 2\mu_k gD}
\end{aligned}$$

Discussion: We can think of this in terms of the concept of energy. The spring does positive work on the block, and so it increases its kinetic energy. Friction does negative work on the block, decreasing its kinetic energy. Only the spring is "introducing" energy into the block, as friction is removing that energy by doing negative work. Another way to think about it is that the spring is inputting energy; some of that energy goes into increasing the kinetic energy of the block, and some of it is lost by friction.

The energy that is lost to friction can be thought of as "thermal energy" (heat) that goes up into heating the block and the surface. Indeed, if you rub your hand against the table, you will notice that it gets warmer; you are losing some of the energy introduced to your hand by the work done by your arm into heating up the table and your hand! This shows that we can think about modelling friction using thermal energy rather than a force.

7.3 Power

We finish the chapter by introducing the concept of "power", which is the rate at which work is done on an object, or more generally, the rate at which energy is being converted from one form to another. If an amount of work, ΔW, was done in a period of time Δt, then the work was done at a rate of:

$$\boxed{P = \frac{\Delta W}{\Delta t}} \tag{7.6}$$

where P is called the power. The SI unit for power is the "Watt", abbreviated W, which corresponds to $\text{J/s} = \text{kgm}^2/\text{s}^3$ in base SI units. If the rate at which work is being done changes with time, then the instantaneous power is defined as:

$$\boxed{P = \frac{dW}{dt}} \tag{7.7}$$

You have probably already encountered power in your everyday life. For example, your 1000 W hair dryer consumes "electrical energy" at a rate of 1000 J per second and converts it into the kinetic energy of the fan as well as the thermal energy to heat up the air. Horsepower (hp) is an imperial unit of power that is often used for vehicles, the conversion being $1\,\text{hp} = 746\,\text{W}$. A 100 hp car thus has an engine that consumes the chemical energy released by burning gasoline at a rate of 7.46×10^4 J per second and converts it into work done on the car as well as into heat.

Checkpoint 7-6

Two cranes lift two identical boxes off of the ground. One crane is twice as powerful as the other. Both cranes do the same amount of work on the boxes and operate at full power. Which of the following statements is true of the boxes, once the cranes have done work on them?

A) One box has been lifted twice as high as the other.
B) The boxes are lifted to the same height in the same amount of time.
C) The boxes are lifted to the same height, but it takes one of the boxes twice as long to get there.
D) One box is lifted twice as high as the other, but it takes the same amount of time to get there.

Example 7-9

If a car engine can do work on the car with a power of P, what will be the speed of the car at some time t if the car was at rest at time $t = 0$?

Solution

First, we need to calculate how much total work was done on the car:

$$W = Pt$$

Then, using the Work-Energy Theorem, we can find the speed of the car at some time t:

$$\begin{aligned}
W &= \frac{1}{2}mv_f^2 - \frac{1}{2}mv_i^2 \\
Pt &= \frac{1}{2}mv_f^2 \\
\therefore v_f &= \sqrt{\frac{2Pt}{m}}
\end{aligned}$$

Discussion: The model for the final speed of the car makes sense because:

- The dimension of the expression for v_f is speed (you should check this!).
- The speed is greater if either the time or power are greater (so the speed is larger if more work is done on the car).
- The speed is smaller if the mass of the car is greater (the acceleration of the car will be less if the mass of the car is larger).

Example 7-10

You are pushing a crate along a horizontal surface at constant speed, v. You find that you need to exert a force of $\vec{F}$ on the crate in order to overcome the friction between the crate and the ground. How much power are you expending by pushing on the crate?

Solution

We need to calculate the rate at which the force, $\vec{F}$, that you exert on the crate does work. If the crate is moving at constant speed, v, then in a time Δt, it will cover a distance, $d = v\Delta t$. Since you exert a force in the same direction as the motion of the crate, the work done over that distance d is:

$$\Delta W = \vec{F} \cdot \vec{d} = Fd\cos(0) = Fv\Delta t$$

The power corresponding to the work done in that period of time is thus:

$$P = \frac{\Delta W}{\Delta t} = Fv$$

This is quite a general result for the rate at which a force does work when it is exerted on an object moving at constant speed.

Olivia's Thoughts

Example 7-10 ties into what I brought up earlier. If you think to yourself: "The velocity is constant, so the work must be zero", the formula,

$$P = \frac{\Delta W}{\Delta t} = Fv$$

wouldn't make any sense. Since v is a constant velocity, the power would always be equal to zero, which of course isn't right. Again, remember that when the velocity is constant, it is only the **net work** that is equal to zero. In Example 7-10, it's asking for the power that **you** are expending by pushing on the crate (which is the same as asking for the rate of the work done **by** you **on** the crate). So, the formula does indeed make sense.

7.4 Summary

Key Takeaways

The work, W, done on an object by a force, $\vec{F}$, while the object has moved through a displacement, $\vec{d}$, is defined as the scalar product:

$$\begin{aligned} W = \vec{F} \cdot \vec{d} &= Fd\cos\theta \\ &= F_x d_x + F_y d_y + F_z d_z \end{aligned}$$

If the force changes with position and/or the object moves along an arbitrary path in space, the work done by that force over the path is given by:

$$W = \int_A^B \vec{F}(\vec{r}) \cdot d\vec{l}$$

If multiple forces are exerted on an object, then one can calculate the net force on the object (the vector sum of the forces), and the net work done on the object will be equal to the work done by the net force:

$$W^{net} = \int_A^B \vec{F}^{net}(\vec{r}) \cdot d\vec{l}$$

If the net work done on an object is zero, that object does not accelerate.

We can define the kinetic energy, $K(v)$ of an object of mass m that has speed v as:

$$K(v) = \frac{1}{2}mv^2$$

The Work-Energy Theorem states that the net work done on an object in going from position A to position B is equal to the object's change in kinetic energy:

$$W^{net} = \Delta K = \frac{1}{2}mv_B^2 - \frac{1}{2}mv_A^2$$

where v_A and v_B are the speed of the object at positions A and B, respectively.

The rate at which work is being done is called power and is defined as:

$$P = \frac{dW}{dt}$$

If a constant force $\vec{F}$ is exerted on an object that has a constant velocity $\vec{v}$, then the power that corresponds to the work being done by that force is:

$$\begin{aligned} P &= \frac{d}{dt}W = \frac{d}{dt}(\vec{F}\cdot\vec{d}) \\ &= \vec{F}\cdot\frac{d}{dt}\vec{d} = \vec{F}\cdot\vec{v} \end{aligned}$$

Important Equations

Work:

$$\begin{aligned} W &= \vec{F}\cdot\vec{d} = Fd\cos\theta \\ W &= F_x d_x + F_y d_y + F_z d_z \\ W &= \int_A^B \vec{F}(\vec{r})\cdot d\vec{l} \\ W^{net} &= \int_A^B \vec{F}^{net}(\vec{r})\cdot d\vec{l} \end{aligned}$$

Kinetic Energy:

$$K(v) = \frac{1}{2}mv^2$$

Work-Energy Theorem:

$$W^{net} = \Delta K = \frac{1}{2}mv_B^2 - \frac{1}{2}mv_A^2$$

Power:

$$\begin{aligned} P &= \frac{dW}{dt} \\ P &= \vec{F}\cdot\vec{v} \end{aligned}$$

7.5 Thinking about the material

Reflect and research

1. When was the concept of work first introduced?
2. To construct the pyramids, the ancient Egyptians used simple machines, like levers, to accomplish tasks that would not be possible otherwise. Apply what we know about work to find out how levers help people lift incredibly heavy objects.
3. After an accident, investigators use skid marks to figure out how fast the cars were going before the crash. Use your knowledge of work, figure out how they do this.
4. The Tesla Model S can accelerate from 0-100 km/h in as little as 2.7 seconds. Calculate the power of the car in horsepower. Why is it unusual for a 7 seat sedan, like the Model S, to have such a short acceleration time? Investigate how it's possible for the Tesla to accelerate so quickly.

To try at home

1. Measure the power that you can output with your legs, and describe how you made the measurement.

To try in the lab

1. Propose an experiment to measure the thermal energy associated with a force of kinetic friction.
2. Propose an experiment to test the Work-Energy Theorem.

7.6 Sample problems and solutions

7.6.1 Problems

Problem 7-1: A ski jump can is modelled as a ramp of height $h = 5\,\text{m}$, as shown in Figure 7.20. The landing area is at the same height as the bottom of the ramp. A skier of mass $m = 80\,\text{kg}$ is moving at a speed $v_i = 15\,\text{m/s}$ when they reach the bottom of the ramp. When the skier lands the jump, their speed is measured to be $v_f = 12\,\text{m/s}$. Ignore air resistance.

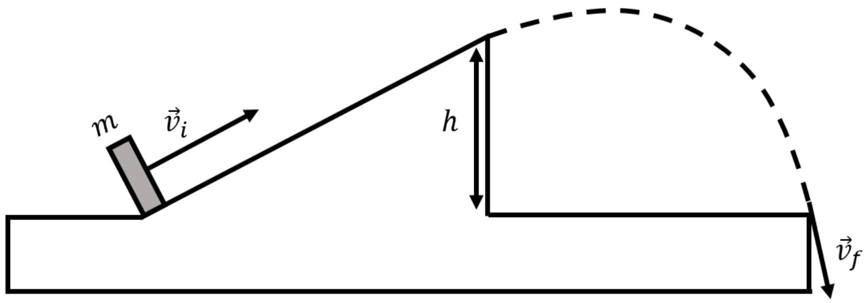


Figure 7.20: A person of mass m goes off a ski jump of height h.

(Solution)

a) What is the speed of the skier the instant they leave the ski jump, at the top of the ramp?
b) Use the answer from part (a) to find the work done by friction the friction between the ramp and the skier.

Problem 7-2: A child of mass m sits on a swing of length L, as in Figure 7.21. You push the child with a horizontal force $\vec{F}$. You apply the force in such a way that the child moves at a constant speed (note that $\vec{F}$ will not have a constant magnitude). (Solution)

a) How much work do you do to move the child from $\theta = 0$ to $\theta = \theta_1$?
b) Use a detailed diagram to show that the work done by $\vec{F}$ is equal to mgh, where h is the change in height of the child.

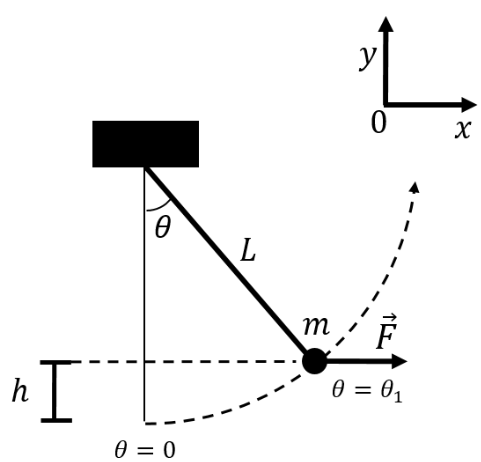


Figure 7.21: A child on a swing is pushed from $\theta = 0$ to $\theta = \theta_1$ at constant speed with a horizontal force, $\vec{F}$.

7.6.2 Solutions

Solution to problem 7-1:

a) We start by defining a coordinate system. We choose the x axis to be horizontal and positive in the direction of motion, and we choose the y axis to be vertical and the positive direction upwards.
We will determine the speed at the top of the ramp, v_t, using the Work-Energy Theorem:

$$W^{net} = \frac{1}{2}mv_f^2 - \frac{1}{2}mv_t^2$$

where W^{net} is the net work done on the skier as they "fly" through the air. While the skier is in the air, the only force acting on them is gravity, $\vec F = -mg\hat y$. The path of the skier is a parabola, so that the displacement vector changes direction continuously. The work done by gravity is given by:

$$W = \int \vec F_g \cdot d\vec l$$

where $d\vec l$ is an infinitesimal displacement along the trajectory, as shown in Figure 7.22.

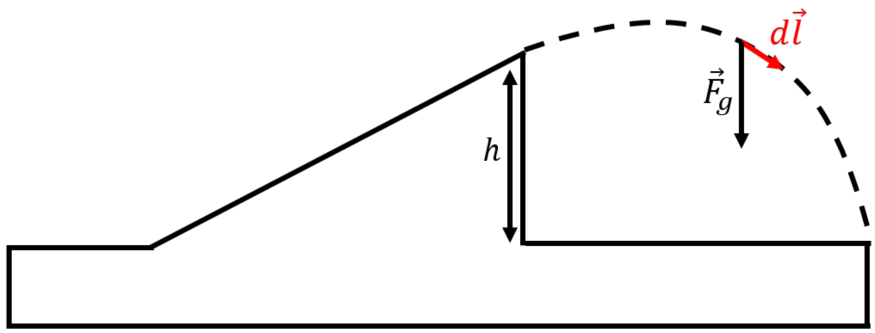


Figure 7.22: Infinitesimal displacement along the trajectory of the jump.

The displacement vector will have x and y components:

$$d\vec l = dx\hat x + dy\hat y$$

The scalar product with the force of gravity is thus:

$$\vec F_g \cdot d\vec l = (-mg\hat y) \cdot (dx\hat x + dy\hat y) = -mgdy$$

The work done by gravity can thus be converted into an integral over y (for which we know the start and end values), and is given by:

$$W = \int \vec F_g \cdot d\vec l = \int_h^0 -mgdy = [-mgy]_h^0 = mgh$$

The work done by gravity is positive, which makes sense, since the force of gravity is generally in the same direction as the net displacement (downwards). We did not

need to take into account the specific shape of the trajectory, because the force was constant in magnitude and direction (see Example 7-4).
We can now find the speed of the skier when they leave the jump using the Work-Energy theorem:

$$\begin{aligned}
W^{net} &= \frac{1}{2}mv_f^2 - \frac{1}{2}mv_t^2 \\
mgh &= \frac{1}{2}mv_f^2 - \frac{1}{2}mv_t^2 \\
\therefore v_t &= \sqrt{v_f^2 - 2gh} = \sqrt{(12\,\text{m/s})^2 - 2(9.8\,\text{m/s}^2)(5\,\text{m})} = 6.8\,\text{m/s}
\end{aligned}$$

b) We can again use the Work-Energy Theorem to determine the work done by friction as the skier slides up the ramp. We know that the speed of the skier at the bottom of the ramp is v_i, and we just found that the speed of the skier at the top of the ramp is $v_t = \sqrt{v_f^2 - 2gh}$. The net work done on the skier going up the ramp is equal to:

$$\begin{aligned}
W^{net} &= \frac{1}{2}mv_t^2 - \frac{1}{2}mv_i^2 \\
&= \frac{1}{2}m(v_t^2 - v_i^2) = \frac{1}{2}m(v_f^2 - 2gh - v_i^2) \\
&= \frac{1}{2}m(v_f^2 - v_i^2) - mgh
\end{aligned}$$

The net work done is also the sum of the work done by each of the forces acting on the skier as they slide up the ramp. The forces on the skier are the force of gravity, the force of friction, and the normal force. The normal force does no work, since it is always perpendicular to the displacement. The net work is thus the sum of the work done by the force gravity, W_g, and the work done by the force of friction, W_f, over the displacement corresponding to the length of the ramp:

$$W^{net} = W_g + W_f$$

The work done by gravity is:

$$W_g = \vec{F}_g \cdot \vec{d} = (-mg\hat{y}) \cdot (d_x\hat{x} + h\hat{y}) = -mgh$$

where $\vec{d}$ is the displacement vector up the ramp (unknown horizontal distance, d_x, and vertical distance, h). We can now determine the work done by the force of friction:

$$\begin{aligned}
W^{net} &= W_g + W_f \\
\frac{1}{2}m(v_f^2 - v_i^2) - mgh &= -mgh + W_f \\
\therefore W_f &= \frac{1}{2}m(v_f^2 - v_i^2) = \frac{1}{2}(80\,\text{kg})((12\,\text{m/s})^2 - (15\,\text{m/s})^2) = -3240\,\text{J}
\end{aligned}$$

And we find that the force of friction did negative work (it reduced the kinetic energy of the skier).

Discussion: Over the course of the jump, the skier started at the bottom of the ramp with a given kinetic energy, then lost some of that energy going up the ramp (in the form of loss to friction and negative work done by gravity). During the airborne phase, gravity did positive work and the skier gained back some of the kinetic energy that they had lost going up the ramp. Thus the net work done by the force of friction is the difference in kinetic energies between the final landing point and the beginning of the ramp, because friction is the only force that did a net amount of (negative) work over the whole trajectory (gravity did no net work over the whole trajectory). This example shows how we can start to think about energy as something that is "conserved", which we will explore in more detail in the next chapter.

Solution to problem 7-2:

a) We want to find the work done by the applied force $\vec F$. We first need to find an expression for the magnitude of $\vec F$, based on the fact that the child is not accelerating. The forces on the child are:
   - $\vec F_g$, their weight, with magnitude mg.
   - $\vec F_T$, the tension in the rope, which changes with the angle, θ.
   - $\vec F$, the applied force, which change in magnitude as the angle, θ, changes.

   The forces are illustrated in Figure 7.23.

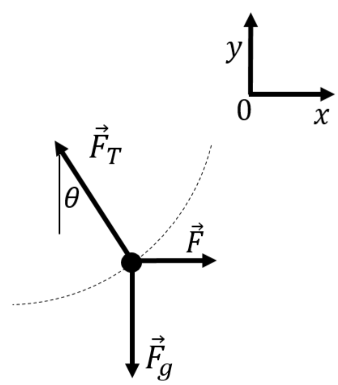


Figure 7.23: A free-body diagram of the forces exerted on the child.

The child is moving at a constant speed, so the net force is equal to zero. The sum of the x and y components of the forces are equal to zero (Newton's Second Law):

$$\sum F_x = F - F_T \sin\theta = 0$$
$$\sum F_y = F_T \cos\theta - mg = 0$$

Rearranging these equations gives:

$$F = F_T \sin\theta$$
$$mg = F_T \cos\theta$$

We want an expression for F that does not depend on F_T (since F_T is unknown), so we can divide one equation by the other:

$$\frac{F}{mg} = \frac{F_T \sin\theta}{F_T \cos\theta} = \tan\theta$$
$$\therefore F(\theta) = mg\tan\theta$$

where we indicated that the force $\vec{F}(\theta)$ depends on the angle θ. The work done by the force, $\vec{F}$, is given by:

$$W_F = \int_A^B \vec{F}(\theta)\cdot d\vec{l}$$

$d\vec{l}$ is the "path element" along part of the arc of circle over which the child moves, as illustrated in Figure 7.24. We have an expression for how $\vec{F}$ changes in magnitude as a function of the angle θ, and it would thus be convenient to perform the integral over the angle θ.

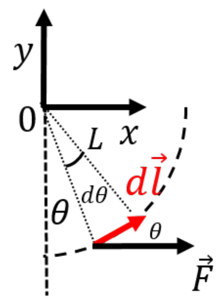


Figure 7.24: A path element along the circular trajectory of the swing.

We can use polar coordinate, (r, θ), instead of cartesian coordinates to describe the displacement vector, $d\vec{l}$. If the vector subtends an arc on the circle that makes an infinitesimal angle, $d\theta$, as illustrated, then the length of the vector $d\vec{l}$ is given by:

$$dl = Ld\theta$$

where L is the radius of the circle. The vector $d\vec{l}$ makes an angle θ with the horizontal, and thus with the vector, $\vec{F}$. The dot product between $\vec{F}$ and $d\vec{l}$ can thus be written as:

$$\vec{F}(\theta)\cdot d\vec{l} = Fdl\cos\theta = (mg\tan\theta)(Ld\theta)\cos\theta = mgL\sin\theta d\theta$$

We can now write the integral for the work using limit that are based on the angle θ, from $\theta = 0$ to $\theta = \theta_1$:

$$\begin{aligned} W &= \int_0^{\theta_1} mgL\sin\theta d\theta \\ &= mgL[-\cos\theta]_0^{\theta_1} = mgL(1-\cos\theta_1) \end{aligned}$$

b) We know that the work done by $\vec{F}$ is $W = mgL(1-\cos\theta_1)$. So, we want to prove that $L(1-\cos\theta_1)$ is equal to h. Expanding $L(1-\cos\theta_1)$ gives:

$$L(1-\cos\theta_1) = L - L\cos\theta_1$$

This can be illustrated on a diagram, as in Figure 7.25, which shows that h is equal to $L - L\cos\theta$.

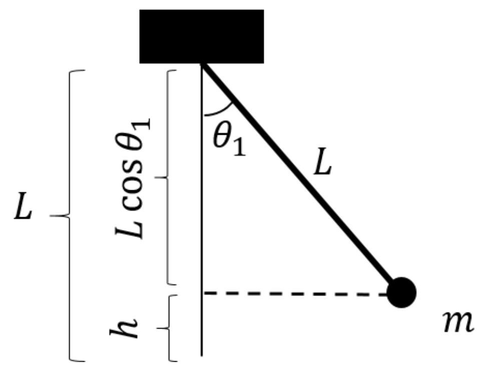


Figure 7.25: A diagram showing the geometry of the problem

Discussion: The net force acting on the mass is equal to zero, so the net work must be equal to zero. The two forces that do work on the mass are the applied force $\vec{F}$, and gravity. The work done by the applied force if mgh, so the work done by gravity must be $-mgh$.

8 Potential Energy and Conservation of Energy

In this chapter, we continue to develop the concept of energy in order to introduce a different formulation for Classical Physics that does not use forces. Although we can describe many phenomena using energy instead of forces, this method is completely equivalent to using Newton's Three Laws. As such, this method can be derived from Newton's formulation, as we will see. Because energy is a scalar quantity, for many problems, it leads to models that are much easier to develop mathematically than if one had used forces. The chapter will conclude with a presentation of the more modern approach, using "Lagrangian Mechanics", that is currently preferred in physics and forms the basis for extending our description of physics to the microscopic world (e.g. quantum mechanics).

Learning Objectives

- Understand the difference between conservative and non-conservative forces.
- Understand how to define potential energy for a conservative force.
- Understand how to use potential energy to calculate work.
- Understand the definition of mechanical energy.
- Understand how to use conservation of mechanical energy.
- Understand how to apply the Lagrangian formulation in a simple case.

Think About It

Three roller coaster carts start at position $x = 0$, where they are all at the same height (Figure 8.1). All of the carts start with the same velocity. At x_1, which roller coaster cart will be moving the fastest?

All of the roller coasters end at ground level, at x_2. Which roller coaster cart will be moving the fastest at x_2? Will all of them make it to x_2? Who will get there first? Assume that the roller coaster track is frictionless.

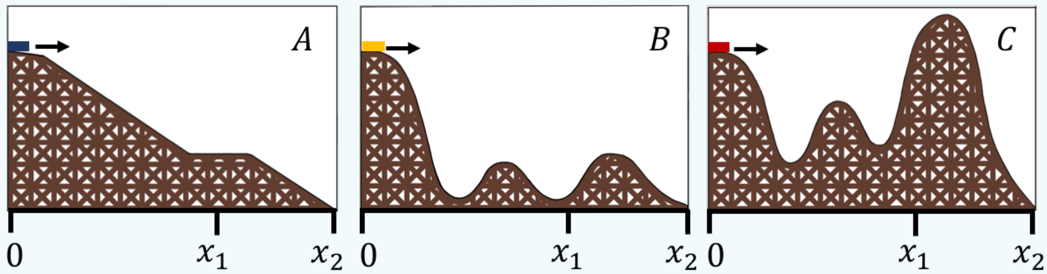


Figure 8.1: Three roller coasters that start at the same height and end at the same height.

8.1 Conservative forces

In Chapter 7, we introduced the concept of work, W, done by a force, $\vec{F}(\vec{r})$, acting on an object as it moves along a path from position A to position B:

$$W = \int_A^B \vec{F}(\vec{r}) \cdot d\vec{l} \tag{8.1}$$

where $\vec{F}(\vec{r})$ is a force vector that, in general, is different at different positions in space ($\vec{r}$). We can also say that $\vec{F}$ depends on position by writing $\vec{F}(\vec{r}) = \vec{F}(x, y, z)$, since the position vector, $\vec{r}$, is simply the vector $\vec{r} = x\hat{x} + y\hat{y} + z\hat{z}$. That is, $\vec{F}(\vec{r})$ is just a short hand notation for $\vec{F}(x, y, z)$, and $d\vec{l}$ is a (very) small segment along the particular path over which one calculates the work.

The above integral is, in general, difficult to evaluate, as it depends on the specific path over which the object moved. In Example 7-2 of Chapter 7, we calculated the work done by friction on a crate that was slid across the floor along two different paths and indeed found that the work depended on the path that was taken. In Example 7-3 of the same chapter, we saw that the work done by the force of gravity when moving a box along two different paths did not depend on the path chosen[1].

We call "conservative forces" those forces for which the work done only depends on the initial and final positions and not on the path taken between those two positions. "Non-conservative" forces are those for which the work done does depend on the path taken. The

[1] At least for those two paths that we tried in the example.

force of gravity is an example of a conservative force, whereas friction is an example of a non-conservative force.

This means that the work done by a conservative force on a "closed path" is zero; that is, **the work done by a conservative force on an object is zero if the object moves along a path that brings it back to its starting position.** Indeed, since the work done by a conservative force only depends on the location of the initial and final positions, and not the path taken between them, the work has to be zero if the object ends in the same place as where it started (a possible path is for the object to not move at all).

Consider the work done by gravity in raising (displacement $\vec{d_1}$) and lowering (displacement $\vec{d_2} = -\vec{d_1}$) an object back to its starting position along a vertical path, as depicted in Figure 8.2.

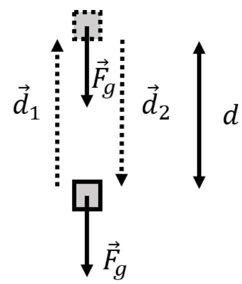


Figure 8.2: An object that has moved up and back down.

The total work done by gravity on this particular closed path is easily shown to be zero, as the work can be broken up into the negative work done as the object moves up (displacement vector $\vec{d_1}$) and the positive work done as the object moves down (displacement vector $\vec{d_2}$):

$$W^{tot} = \vec{F}_g \cdot \vec{d_1} + \vec{F}_g \cdot \vec{d_2} = -mgd + mgd = 0$$

In order to write the path integral of the force over a closed path, we introduce a new notation to indicate that the starting and ending position are the same:

$$\int_A^A \vec{F}(\vec{r}) \cdot d\vec{l} = \oint \vec{F}(\vec{r}) \cdot d\vec{l}$$

The condition for a force to be conservative is thus:

$$\boxed{\oint \vec{F}(\vec{r}) \cdot d\vec{l} = 0} \tag{8.2}$$

since this means that the work done over a closed path is zero. The condition for this integral to be zero can be found by Stokes' Theorem:

$$\oint \vec{F}(\vec{r}) \cdot d\vec{l} = \int_S \left[\left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) \hat{x} + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) \hat{y} + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) \hat{z} \right] \cdot d\vec{A}$$

where the integral on the right is called a "surface integral" over the surface, S, enclosed by the closed path over which the work is being calculated. Don't worry, it is way beyond the scope of this text to understand this integral or Stokes' Theorem in detail! It is however useful in that it gives us the following conditions on the components of a force for that force to be conservative (by requiring the terms in parentheses to be zero):

$$\begin{aligned}\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} &= 0\\ \frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} &= 0\\ \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} &= 0 \end{aligned} \tag{8.3}$$

In general:

1. A force can be conservative if it only depends on position in space, and not speed, time, or any other quantity.
2. A force is conservative if it is constant in magnitude and direction.

Checkpoint 8-1

You push a crate from point A to point B along a horizontal surface. Is the force you exert a conservative force?

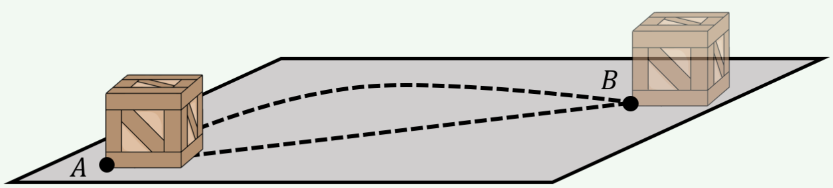


Figure 8.3: You push a crate from A to B along any path.

A) Yes
B) No
C) Not enough information

Example 8-1

Is the force of gravity on an object of mass m, near the surface of the Earth, given by:

$$\vec{F}(x,y,z) = 0\hat{x} + 0\hat{y} - mg\hat{z}$$

conservative? Note that we have defined the z axis to be vertical and positive upwards.

Solution

The force is expected to be conservative since it is constant in magnitude and direction.

We can verify this using the conditions in Equation 8.3:

$$\begin{aligned}
\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} &= \frac{\partial}{\partial y}(-mg) - 0 && = 0\\
\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} &= 0 - \frac{\partial}{\partial x}(-mg) && = 0\\
\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} &= 0 - 0 && = 0
\end{aligned}$$

and the force is indeed conservative since all three conditions are zero.

Example 8-2

Is the force given by:

$$\vec{F}(x,y,z) = \frac{-k}{r^3}\vec{r} = \frac{-kx}{(x^2+y^2+z^2)^{\frac{3}{2}}}\hat{x} + \frac{-ky}{(x^2+y^2+z^2)^{\frac{3}{2}}}\hat{y} + \frac{-kz}{(x^2+y^2+z^2)^{\frac{3}{2}}}\hat{z}$$

conservative?

Solution

Since the force only depends on position, it *could* be conservative, so we must check using the conditions from Equation 8.3:

$$\begin{aligned}
\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} &= \frac{\partial}{\partial y}\left(\frac{-kz}{(x^2+y^2+z^2)^{\frac{3}{2}}}\right) - \frac{\partial}{\partial z}\left(\frac{-ky}{(x^2+y^2+z^2)^{\frac{3}{2}}}\right)\\
&= \frac{3kz(2y)}{2(x^2+y^2+z^2)^{\frac{5}{2}}} - \frac{3ky(2z)}{2(x^2+y^2+z^2)^{\frac{5}{2}}} = 0
\end{aligned}$$

$$\begin{aligned}
\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} &= \frac{\partial}{\partial z}\left(\frac{-kx}{(x^2+y^2+z^2)^{\frac{3}{2}}}\right) - \frac{\partial}{\partial x}\left(\frac{-kz}{(x^2+y^2+z^2)^{\frac{3}{2}}}\right)\\
&= \frac{3kx(2z)}{2(x^2+y^2+z^2)^{\frac{5}{2}}} - \frac{3kz(2x)}{2(x^2+y^2+z^2)^{\frac{5}{2}}} = 0
\end{aligned}$$

$$\begin{aligned}
\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} &= \frac{\partial}{\partial x}\left(\frac{-ky}{(x^2+y^2+z^2)^{\frac{3}{2}}}\right) - \frac{\partial}{\partial y}\left(\frac{-kx}{(x^2+y^2+z^2)^{\frac{3}{2}}}\right)\\
&= \frac{3ky(2x)}{2(x^2+y^2+z^2)^{\frac{5}{2}}} - \frac{3kx(2y)}{2(x^2+y^2+z^2)^{\frac{5}{2}}} = 0
\end{aligned}$$

where we used the Chain Rule to take the derivatives. Since all of the conditions are zero, the force is conservative. As we will see, the force represented here is similar mathematically to both the force that Newton introduced in his Universal Theory of Gravity, and the force introduced by Coulomb as the electric force, which are both conservative.

8.2 Potential energy

In this section, we introduce the concept of "potential energy". Potential energy is a scalar function of position that can be defined for any conservative force in a way to make it easy to calculate the work done by that force over any path. Since the work done by a conservative force in going from position A to position B does not depend on the particular path taken, but only on the end points, we can write the work done by a conservative force in terms of a "potential energy function", $U(\vec{r})$, that can be evaluated at the end points:

$$\boxed{-W = -\int_A^B \vec{F}(\vec{r}) \cdot d\vec{l} = U(\vec{r}_B) - U(\vec{r}_A) = \Delta U} \tag{8.4}$$

where we have have chosen to define the function $U(\vec{r})$ so that it relates to the **negative** of the work done for reasons that will be apparent in the next section. Figure 8.4 shows an example of an arbitrary path between two points A and B in two dimensions for which one could calculate the work done by a conservative force using a potential energy function.

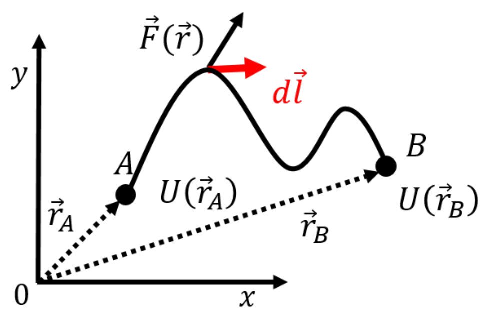


Figure 8.4: Illustration of calculating the work done by a conservative function along an arbitrary path by taking the difference in potential energy evaluated at the two endpoints, $-W = U(\vec{r}_B) - U(\vec{r}_A)$.

Once we know the function for the potential energy, $U(\vec{r})$, we can calculate the work done by the associated force along any path. In order to determine the function, $U(\vec{r})$, we can calculate the work that is done along a path over which the integral for work is easy (usually, a straight line).

For example, near the surface of the Earth, the force of gravity on an object of mass, m, is given by:

$$\vec{F}_g = -mg\hat{z}$$

where we have defined the z axis to be vertical and positive upwards. We already showed in Example 8-1 that this force is conservative and that we can thus define a potential energy function. To do so, we can calculate the work done by the force of gravity over a straight vertical path, from position A to position B, as shown in Figure 8.5.

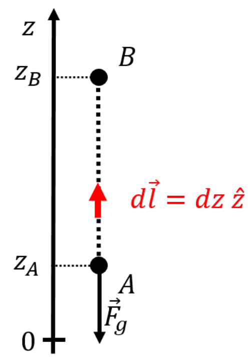


Figure 8.5: A vertical path for calculating the work done by gravity.

The work done by gravity from position A to position B is:

$$\begin{aligned}
W &= \int_A^B \vec{F}(\vec{r}) \cdot d\vec{l} \\
&= \int_{z_A}^{z_B} (-mg\hat{z}) \cdot (dz\hat{z}) \\
&= -mg \int_{z_A}^{z_B} dz \\
&= -mg(z_B - z_A)
\end{aligned}$$

By inspection, we can now identify the functional form for the potential energy function, $U(\vec{r})$. We require that:

$$-W = U(\vec{r}_B) - U(\vec{r}_A) = U(z_B) - U(z_A)$$

where we replaced the position vector, $\vec{r}$, with the z coordinate, since this is a one dimensional situation. Therefore:

$$\begin{gathered}
-W = mg(z_B - z_A) = U(z_B) - U(z_A) \\
\therefore U(z) = mgz + C
\end{gathered}$$

and we have found that, for the force of gravity near the surface of the Earth, one can define a potential energy function (by inspection), $U(z) = mgz + C$.

It is important to note that, since it is only the **difference** in potential energy that matters when calculating the work done, the potential energy function can have an arbitrary constant, C, added to it. Thus, **the value of the potential energy function is meaningless, and only differences in potential energy are meaningful and related to the work done on an object**. In other words, it does not matter where the potential energy is equal to zero, and by choosing C, we can therefore choose a convenient location where the potential energy is zero.

Checkpoint 8-2

When we found the work done by gravity, we defined positive z to be upwards. If we instead chose positive z to be downwards, how would the potential energy function be defined?

A) The potential energy function would be the same, $U(z) = mgz + C$.
B) The potential energy function would be the same but negative, $U(z) = -mgz+C$

Checkpoint 8-3

Can an object have a negative potential energy?

A) Yes
B) No

Example 8-3

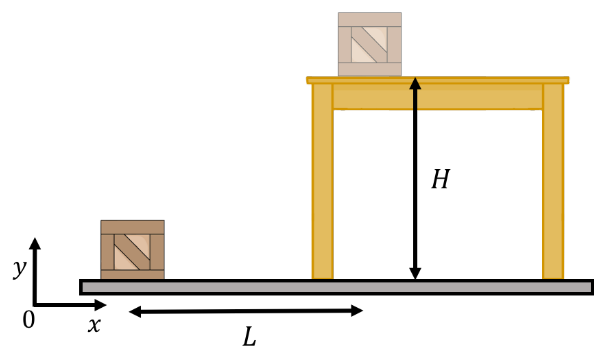


Figure 8.6: A box moved from the ground up onto a table.

Calculate the work done **by the force of gravity** when a box of mass, m, is moved from the ground up onto a table that is a distance L away horizontally and H vertically, as illustrated in Figure 8.6. How much work must be done by a person moving the box?

Solution

Since the force of gravity is conservative, we can use the potential energy function given by:

$$U(z) = mgz + C$$

to calculate the work done by the force of gravity when the box is moved. The work done by gravity will only depend on the change in height, H, as the potential energy function only depends on the z coordinate of an object. We can choose the origin of our coordinate system to be the ground and choose the constant $C = 0$, so that the potential energy function at the starting position of the box is:

$$U(z_A = 0) = mg(0) = 0$$

The potential energy function when the box is on the table, with $z = H$, is given by:

$$U(z_B = H) = mgH$$

The change in potential energy, $\Delta U = U(z_B) - U(z_A)$ is equal to the negative of the work done by gravity. The work done by gravity, W_g, is thus:

$$\begin{aligned}
-W_g &= U(z_B) - U(z_A) = mgH - 0 \\
\therefore W_g &= -mgH
\end{aligned}$$

which is the same as what we found in Example 7-3 of Chapter 7. The work done by gravity is negative, as we found previously. This makes sense because gravity has a component opposite to the direction of motion.

The work done by a person, W_p, to move the box can easily be found by considering the net work done on the box. While the box is moving, only the person and gravity are exerting forces on the box, so those are the only two forces performing work. Since the box starts and ends at rest, the net work done on the box must be zero (no change in kinetic energy, recall the Work-Energy Theorem):

$$\begin{aligned}
W^{net} &= 0 = W_g + W_p \\
\therefore W_p &= -W_g = mgH
\end{aligned}$$

Discussion: We find that the person had to do positive work, which makes sense, since they had to exert a force with a component in the direction of motion (upwards). It is also interesting to note that it does not matter if the person exerted a constant force or whether they varied the force that they exerted on the box as they moved it: the amount of work done by the person is fixed to be the negative of the work done by gravity.

Example 8-4

The force exerted by a spring that is extended or compressed by a distance, x, is given by Hookes' Law:

$$\vec{F}(x) = -kx\hat{x}$$

where the x axis is defined to be co-linear with the spring and the origin is located at the rest position of the spring. Show that the force exerted by the spring onto an object is conservative and determine the corresponding potential energy function.

Solution

Since the force depends on position, it could be conservative, which we can check with the conditions from Equation 8.3:

$$\begin{aligned}
\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} &= 0 - 0 &= 0\\
\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} &= \frac{\partial}{\partial z}(-kx)) - 0 &= 0\\
\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} &= 0 - \frac{\partial}{\partial y}(-kx)) &= 0
\end{aligned}$$

and the force is indeed conservative. To determine the potential energy function, let us calculate the work done by the spring from position x_A to position x_B:

$$\begin{aligned}
W &= \int_A^B \vec{F}(\vec{r}) \cdot d\vec{l}\\
&= \int_{x_A}^{x_B} (-kx\hat{x}) \cdot dx\hat{x}\\
&= \int_{x_A}^{x_B} (-kx)dx = \left[-\frac{1}{2}kx^2\right]_{x_A}^{x_B}\\
&= -\left(\frac{1}{2}kx_B^2 - \frac{1}{2}kx_A^2\right)
\end{aligned}$$

Again, comparing with:

$$-W = U(\vec{r}_B) - U(\vec{r}_A) = U(x_B) - U(x_A)$$

We can identify the potential energy for a spring:

$$U(x) = \frac{1}{2}kx^2 + C$$

where, in general, the constant C can take any value. If we choose $C = 0$, then the potential energy is zero when the spring is at rest, although it is not important what choice is made. Note that in one dimension, the potential energy function is the negative of the anti-derivative of the function that gives the x component of the force.

Checkpoint 8-4

A conservative force acts on an object that is initially at rest. No other forces act on the object. Does the object move in a way that increases its potential energy or decreases its potential energy?

A) Increases.
B) Decreases.
C) It depends on the choice of C for the corresponding potential energy.

8.2.1 Recovering the force from potential energy

Given a (scalar) potential energy function, $U(\vec r)$, it is possible to determine the (vector) force that is associated with it. Take, for example, the potential energy from a spring (Example 8-4):

$$U(x) = \frac{1}{2}kx^2 + C$$

As you recall from Example 8-4, to find this function (in one dimension), we took the x component of the spring force and (effectively) found the negative of its anti-derivative, which we defined as the potential energy function:

$$\begin{aligned}
F(x) &= -kx \\
U(x) &= -\int F(x)dx = \int (kx)dx = \frac{1}{2}kx^2 + C \\
\therefore F(x) &= -\frac{d}{dx}U(x)
\end{aligned}$$

Thus, the force can be obtained from the negative of the potential energy function, by taking its derivative with respect to position.

In three dimensions, the situation is similar, although the potential energy function (and the components of the force vector) will generally depend on all three position coordinates, x, y, and z. In three dimensions, the the three components of the force vector are given by taking the gradient of the negative of the potential energy function[2]:

$$\begin{aligned}
\vec{F}(\vec{r}) &= -\vec{\nabla}U(\vec{r}) = -\vec{\nabla}U(x,y,z) \\
\therefore F_x(x,y,z) &= -\frac{\partial}{\partial x}U(x,y,z) \\
\therefore F_y(x,y,z) &= -\frac{\partial}{\partial y}U(x,y,z) \\
\therefore F_z(x,y,z) &= -\frac{\partial}{\partial z}U(x,y,z)
\end{aligned} \tag{8.5}$$

8.3 Mechanical energy and conservation of energy

Recall the Work-Energy Theorem, which relates the net work done on an object to its change in kinetic energy, along a path from point A to point B:

$$W^{net} = \Delta K = K_B - K_A$$

where K_A is the object's initial kinetic energy and K_B is its final kinetic energy. Generally, the net work done is the sum of the work done by conservative forces, W^C, and the work done by none conservative forces, W^{NC}:

$$W^{net} = W^C + W^{NC}$$

The work done by conservative forces can be expressed in terms of changes in potential energy functions. For example, suppose that two conservative forces, $\vec{F}_1$ and $\vec{F}_2$, are exerted on the object. The work done by those two forces is given by:

$$\begin{aligned}
W_1 &= -\Delta U_1 \\
W_2 &= -\Delta U_2
\end{aligned}$$

where U_1 and U_2 are the changes in potential energy associated with forces $\vec{F}_1$ and $\vec{F}_2$, respectively. We can re-arrange the Work-Energy Theorem as follows[3]:

$$\begin{aligned}
W^{net} = W^C + W^{NC} &= -\Delta U_1 - \Delta U_2 + W^{NC} = \Delta K \\
\therefore W^{NC} &= \Delta U_1 + \Delta U_2 + \Delta K
\end{aligned}$$

That is, the work done by non-conservative forces is equal to the sum of the changes in potential and kinetic energies. In general, we can use ΔU to represent the change in the total potential energy of the object. The total potential energy is the sum of the potential energies associated with each of the conservative forces acting on the object ($\Delta U = \Delta U_1 + \Delta U_2$ above). The above expression can thus be written in a more general form:

$$\boxed{W^{NC} = \Delta U + \Delta K} \tag{8.6}$$

[2]As you may recall from Appendix B, the gradient is a vector that points towards the direction of maximal increase in a multi-variate function.

[3]This is why we defined potential energy as negative of the work; it becomes a positive term when we move it to the same side of the equation as the kinetic energy!

In particular, note that if there are no non-conservative forces doing work on the object:

$$\boxed{\Delta K + \Delta U = 0} \quad \text{if no non-conservative forces} \tag{8.7}$$
$$-\Delta U = \Delta K$$

That is, the sum of the changes in potential and kinetic energies of the object is always zero. This means that if the potential energy of the object increases, then the kinetic energy of the object must decrease by the same amount.

We can introduce the "mechanical energy", E, of an object as the sum of the potential and kinetic energies of the object:

$$\boxed{E = U + K} \tag{8.8}$$

If the object started at position A, with potential energy U_A and kinetic energy K_A, and ended up at position B with potential energy U_B and kinetic energy K_B, then we can write the mechanical energy at both positions and its change ΔE, as:

$$\begin{aligned}
E_A &= U_A + K_A \\
E_B &= U_B + K_B \\
\Delta E &= E_B - E_A \\
&= U_B + K_B - U_A - K_A \\
\therefore \Delta E &= \Delta U + \Delta K
\end{aligned}$$

Thus, the change in mechanical energy of the object is equal to the work done by non-conservative forces:

$$W^{NC} = \Delta U + \Delta K = \Delta E$$

and if there is no work done by non-conservative forces on the object, then the mechanical energy of the object does not change:

$$\begin{aligned}
\Delta E &= 0 \quad \text{if no non-conservative forces} \\
\therefore E &= \text{constant}
\end{aligned}$$

This is what we generally call the "conservation of mechanical energy". If there are no non-conservative forces doing work on an object, its mechanical energy is conserved (i.e. constant).

The introduction of mechanical energy gives us a completely different way to think about mechanics. We can now think of an object as having "energy" (potential and/or kinetic), and we can think of forces as changing the energy of the object.

Checkpoint 8-5

Is the value of an an object's mechanical energy meaningful, or is it only the difference in mechanical energy that is meaningful?

A) Yes, the value of the mechanical energy is meaningful. At any given time, an object will have a quantifiable amount of mechanical energy.
B) No, the value is not meaningful because the value of potential energy is arbitrary. Only differences in mechanical energy are meaningful.
C) No, the value is not meaningful because both the potential and kinetic energies are arbitrary. Their values will change depending on where you set the energy to be zero.
D) It depends on which conservative forces act on the object (and therefore what "kind" of potential energy the object has).

We can also think of the work done by non-conservative forces as a type of change in energy. For example, the work done by friction can be thought of as a change in thermal energy (feel the burn as you rub your hand vigorously on a table!). If we can model the work done by non-conservative forces as a type of "other" energy, $-W^{NC} = \Delta E^{other}$, then we can state that:

$$\Delta E^{other} + \Delta U + \Delta K = 0$$

which is what we usually refer to as "conservation of energy". That is, the total energy in a system, including kinetic, potential and any other form (e.g. thermal, electrical, etc.) is constant unless some external agent is acting on the system.

We can always include that external agent in the system so that the total energy of the system is constant. The largest system that we can have is the Universe itself. Thus, the total energy in the Universe is constant and can only transform from one type into another, but no energy can ever be added or removed from the Universe.

Olivia's Thoughts

Here's an example that may help you understand the concept of external agents and energy conservation. Say we have a mass that hangs from a spring, so that the mass oscillates up and down like a yo-yo. If we define our system to include the spring, the mass, and gravity, energy will be conserved (the energy is transformed from potential energy to kinetic energy and back again).

Now, what if someone is holding the end of the spring and they start walking so that the whole system accelerates? Energy is not conserved because the system is gaining kinetic energy, seemingly out of nowhere. The system is being acted on by an *external agent* (the person). If we expand our system so that it includes the spring, the mass, gravity, *and the person*, energy is conserved. Instead of the kinetic energy "coming out of nowhere", we can see that it is actually coming from the person converting chemical energy in their body in order to move their muscles.

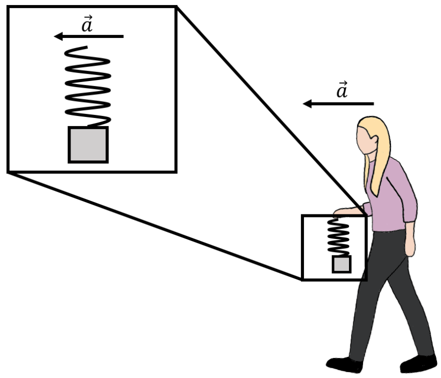


Figure 8.7: A person accelerates a mass and spring by walking. If the system does not include the person, energy is not conserved. If it does include the person, energy is conserved.

But what if there's an external agent acting on our new system? We can keep "zooming out" to include more and more external sources in the definition of our system. If you kept zooming out, eventually you would reach the point where the whole Universe was included in your system. At this point, you can't zoom out any more. This means that, if the Universe is your system, energy must always be conserved because there can't be any external agents acting on the system.

Example 8-5

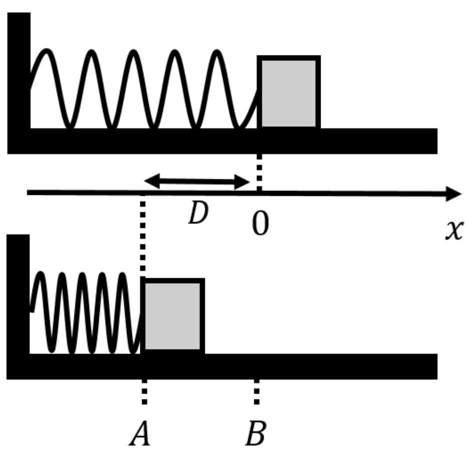


Figure 8.8: A block is launched along a frictionless surface by compressing a spring by a distance D. The top panel shows the spring when at rest, and the bottom panel shows the spring compressed by a distance D just before releasing the block.

A block of mass m can slide along a horizontal frictionless surface. A horizontal spring, with spring constant, k, is attached to a wall on one end, while the other end can move, as shown in Figure 8.8. A coordinate system is defined such that the x axis is horizontal and the free end of the spring is at $x = 0$ when the spring is at rest. The block is pushed against the spring so that the spring is compressed by a distance D. The block is then released. What speed will the block have when it leaves the spring?

Solution

This is again the same example that we saw in Chapters 6 and 7. We will show here that it is solved very easily using conservation of energy. The forces acting on the block are:

1. Weight, which does no work since it is perpendicular to the block's displacement.
2. The normal force, which does no work since it is perpendicular to the block's displacement.
3. The force from the spring, which is conservative and can be modelled with a potential energy $U(x) = \frac{1}{2}kx^2$, where x is the position of the end of the spring.

The block starts at rest at position A ($x = -D$), where the spring is compressed by a distance D, and leaves the spring at position B ($x = 0$), where the spring is at its rest position.

At position A, the kinetic energy of the block is $K_A = 0$ since the block is at rest, and the potential energy from the spring force of the block is $U_A = \frac{1}{2}kD^2$. The mechanical energy of the block at position A is thus:

$$\begin{aligned}
K_A &= 0 \\
U_A &= \frac{1}{2}kD^2 \\
\therefore E_A &= U_A + K_A = \frac{1}{2}kD^2
\end{aligned}$$

At position B, the spring potential energy of the block is zero (since the spring is at rest), and all of the energy is kinetic:

$$\begin{aligned}
K_B &= \frac{1}{2}mv_B^2 \\
U_B &= 0 \\
\therefore E_B &= U_B + K_B = \frac{1}{2}mv_B^2
\end{aligned}$$

Since there are no non-conservative forces doing work on the block, the mechanical energies at A and B are the same:

$$\begin{aligned}
W^{NC} &= \Delta E = E_B - E_A = 0 \\
\therefore E_B &= E_A \\
\frac{1}{2}mv_B^2 &= \frac{1}{2}kD^2 \\
v_B &= \sqrt{\frac{kD^2}{m}}
\end{aligned}$$

as we found previously.

Example 8-6

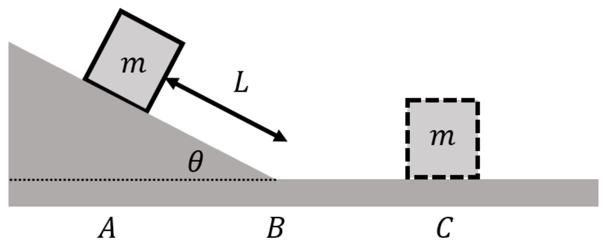


Figure 8.9: A block slides down an incline before sliding on a flat surface and stopping.

A block of mass m is placed at rest on an incline that makes an angle θ with respect to the horizontal, as shown in Figure 8.9. The block is nudged slightly so that the force of static friction is overcome and the block starts to accelerate down the incline. At the bottom of the incline, the block slides on a horizontal surface. The coefficient of kinetic friction between the block and the incline is μ_{k1}, and the coefficient of kinetic friction between the block and horizontal surface is μ_{k2}. If one assumes that the block started at rest a distance L from the bottom of the incline, how far along the horizontal surface will the block slide before stopping?

Solution

This is the same problem we solved in Example 6-2. In that case, we solved for the acceleration of the block using Newton's Second Law and then used kinematics to find how far the block went. We can solve this problem in a much easier way using conservation of energy.

It is still a good idea to think about what forces are applied on the object in order to determine if there are non-conservative forces doing work. In this case, the forces on the block are:

1. The normal force, which does no work, as it is always perpendicular to the motion.
2. Weight, which does work when the height of the object changes, which we can model with a potential energy function.
3. Friction, which is a non-conservative force, whose work we must determine.

Let us divide the motion into two segments: (1) a segment along the incline (positions A to B in Figure 8.9), where gravitational potential energy changes, and (2), the horizontal segment from positions B to position C on the figure. We can then apply conservation of energy for each segment.

Starting with the first segment, we can choose the gravitational potential energy to be zero when the block is at the bottom of the incline. The block starts at a height $h = L\sin\theta$ above the bottom of the incline. The gravitational potential energy for the beginning and end of the first segment are thus:

$$\begin{aligned}U_A &= mgL\sin\theta\\ U_B &= 0\end{aligned}$$

Since the block starts at rest, its kinetic energy is zero at position A, and if the speed of the box is v_B at position B, we can write its kinetic energy at both positions as:

$$\begin{aligned}K_A &= 0\\ K_B &= \frac{1}{2}mv_B^2\end{aligned}$$

The mechanical energy of the object at positions A and B is thus:

$$\begin{aligned}E_A &= U_A + K_A = mgL\sin\theta\\ E_B &= U_B + K_B = \frac{1}{2}mv_B^2\\ \Delta E &= E_B - E_A = \frac{1}{2}mv_B^2 - mgL\sin\theta\end{aligned}$$

Finally, since we have a non-conservative force, the force of kinetic friction, acting on the first segment, we need to calculate the work done by that force. We found in Example 6-2 that the force of friction had magnitude $f_k = \mu_{k1}N = \mu_{k1}mg\cos\theta$. Since the force of friction is anti-parallel to the displacement vector, which points down the incline and has length L, the work done by friction is:

$$W^{NC} = W_f = -f_kL = -\mu_{k1}mg\cos\theta L$$

Applying conservation of energy along the first segment, we have:

$$\begin{aligned}
W^{NC} &= \Delta E \\
-\mu_{k1} mg\cos\theta L &= \frac{1}{2}mv_B^2 - mgL\sin\theta \\
\therefore \frac{1}{2}mv_B^2 &= mgL\sin\theta - \mu_{k1} mg\cos\theta L
\end{aligned}$$

Note that the above equation, in words, could be read as, "the change in kinetic energy ($\frac{1}{2}mv_B^2$) is equal to the negative change in potential energy ($mgL\sin\theta$) minus the work done by friction ($\mu_{k1} mg\cos\theta L$)". In other words, the block had potential energy, which was converted into kinetic energy and heat (the work done by friction can be thought of as thermal energy).

We now proceed in an analogous way for the second segment, from position B to position C. The only force that can do work along this segment (of length x) is the force of kinetic friction, since both the weight and normal force are perpendicular to the displacement. There are no conservative forces doing work, so there is no change in potential energy. The initial kinetic energy is K_B (from above), and the final kinetic energy, K_C, is zero. The change in mechanical energy is thus:

$$\begin{aligned}
\Delta E = E_C - E_B &= K_C - K_B = -K_B \\
&= -\frac{1}{2}mv_B^2 \\
&= -mgL\sin\theta + \mu_{k1} mg\cos\theta L
\end{aligned}$$

where, in the last line, we used the result from the first segment. The work done by the force of friction along the horizontal segment of (undetermined) length x is:

$$W^{NC} = W_f = -f_k x = -\mu_{k2} N x = -\mu_{k2} mgx$$

Finally, we can find x by setting the work done by non-conservative forces equal to the change in mechanical energy:

$$\begin{aligned}
W^{NC} &= \Delta E \\
-\mu_{k2} mgx &= -mgL\sin\theta + \mu_{k1} mg\cos\theta L \\
\therefore x &= L\frac{1}{\mu_{k2}}(\sin\theta - \mu_{k1}\cos\theta)
\end{aligned}$$

which is the same result that we obtained in Example 6-2.

Discussion: By using conservation of energy, we were able to model the motion of the block down the incline in a way that was much easier than what was done in Example 6-2. Furthermore, although we modelled friction as a non-conservative force doing work, we gained some insight into the idea that this could be thought of as an energy loss. In terms of energy, we would say that the block initially had gravitational potential energy, which was then converted into kinetic energy as well as thermal energy (in the heat generated by friction).

8.4 Energy diagrams and equilibria

We can write the mechanical energy of an object as:

$$E = K + U$$

which will be a constant if there are no non-conservative forces doing work on the object. This means that if the potential energy of the object increases, then its kinetic energy must decrease by the same amount, and vice-versa.

Consider a block that can slide on a frictionless horizontal surface and that is attached to a spring, as is shown in Figure 8.10 (left side), where $x = 0$ is chosen as the position corresponding to the rest length of the spring. If you push on the block so as to compress the spring by a distance D and then release it, the block will initially accelerate because of the spring force in the positive x direction until the block reaches the rest position of the spring ($x = 0$ on the diagram). When it passes that point, the spring will exert a force in the opposite direction. The block will continue in the same direction and decelerate until it stops and turns around. It will then accelerate again towards the rest position of the spring, and then decelerate once the spring starts being compressed again, until the block stops and the motion repeats. We say that the block "oscillates" back and forth about the rest position of the spring.

We can describe the motion of the block in terms of its total mechanical energy, E. Its potential energy is given by:

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

On the right of Figure 8.10 is an "Energy Diagram" for the block, which allows us to examine how the total energy, E, of the block is divided between kinetic and potential energy depending on the position of the block. The vertical axis corresponds to energy and the horizontal axis corresponds to the position of the block.

The total mechanical energy, $E = 25\,\text{J}$, is shown by the horizontal red line. Also illustrated are the potential energy function ($U(x)$ in blue), and the kinetic energy, ($K = E - U(x)$, in dotted black).

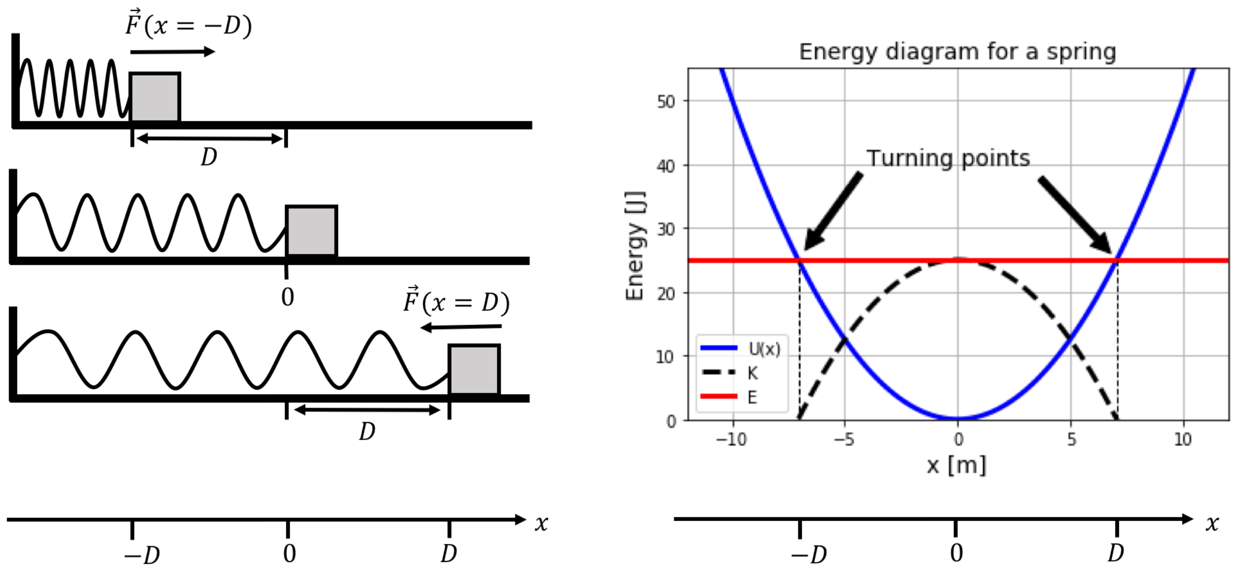


Figure 8.10: Left: The block oscillates about the rest position of the spring, between $x=-D$ and $x=D$. Right: The energy diagram for the block. This diagram is for a spring with spring constant $k=1\,\mathrm{N/m}$.

The energy diagram allows us to describe the motion of the object attached to the spring in terms of energy. A few things to note:

1. At $x=\pm D$, the potential energy is equal to E, so the kinetic energy is zero. The block is thus instantaneously at rest at those positions.
2. At $x=0$, the potential energy is zero, and the kinetic energy is maximal. This corresponds to where the block has the highest speed.
3. The kinetic energy of the block can never be negative[4], thus, the block cannot be located outside the range $[-D,+D]$, and we would say that the motion of the block is "bound". The points between which the motion is bound are called "turning points".

An analysis of the energy diagram tells us that the block is bound between the two turning points, which themselves are equidistant from the origin. When we initially compress the spring, we are "giving" the block "spring potential energy". As the block starts to move, the potential energy of the block is converted into kinetic energy as it accelerates and then back into potential energy as it decelerates.

Checkpoint 8-6

Calculate the positions of the turning points for the situation shown in Figure 8.10. The total energy is 25 J and the spring constant is $k=1\,\mathrm{N/m}$.

By looking at only the potential energy function, without knowing that it is related to a

[4]Remember, the kinetic energy is given by $K=\frac{1}{2}mv^2$. Since neither mass nor the value of v^2 can be negative, the kinetic energy of an object can never be negative.

spring, we can come to the same conclusions; namely that the motion is bound as long as the total mechanical energy is not infinite. We call the point $x = 0$ a "stable equilibrium", because it is a local minimum of the potential energy function. If the object is displaced from the equilibrium point, it will want to move back towards that point. This can also be understood in terms of the force associated with the potential energy function:

$$F = -\frac{d}{dx}U(x)$$

The local minimum occurs where the derivative of the potential function is equal to zero. Thus, the **equilibrium point is given by the condition that the force associated with the potential is zero** ($x = 0$ in the case of the potential energy from a spring). The equilibrium is a stable equilibrium because the force associated with the potential energy function ($F(x) = -kx$ for the spring) points towards the equilibrium point.

The potential energy function for an object with total mechanical energy, E, can be thought of as a little "roller coaster", on which you place a marble and watch it "roll down" the potential energy function. You can think of placing a marble where $U(x) = E$ and releasing it. The marble would then roll down the potential energy function, just as an actual marble would roll down a real slope, mimicking the motion of the object along the x axis. This is illustrated in Figure 8.11 which shows an arbitrary potential energy function and a marble being placed at a location where the potential energy is equal to E.

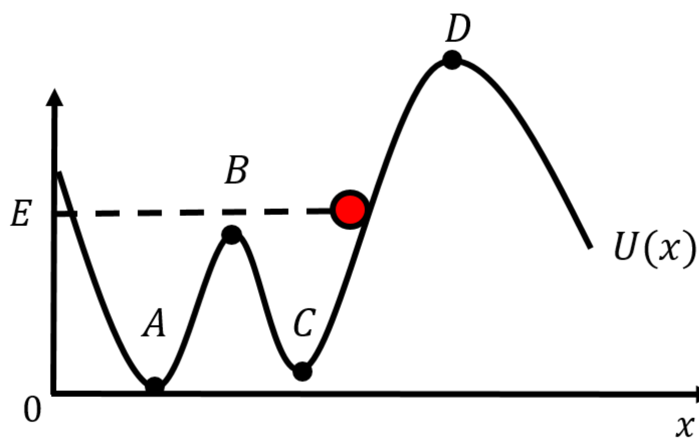


Figure 8.11: Arbitrary potential energy function and illustration of visualizing a marble rolling down the function by placing the marble on the potential energy function at a point where $U(x) = E$.

The motion of the marble will be bound between the two points where the potential energy function is equal to E. When the marble is placed as shown, it will roll towards the left, just as if it were a real marble on a track. Since the potential energy is increasing as a function of x at the point where we placed the marble, the force is in the negative x direction (remember, the force is the negative of the derivative of the potential energy function). With the given energy, the marble would never be able to make it to point D, as it does not have enough energy to "climb up the hill". It would roll down, through point C, up to point B, down to point A, and then turn around where $U(x) = E$ and return to where it started.

Locations A and C on the diagram are stable equilibria, because if a marble is placed in one of those locations and nudged slightly, it will come back to the equilibrium point (or oscillate about that point). Points B and D are "unstable equilibria", because if the marble is placed there and nudged, it will not immediately come back to those points. Note that if the marble were placed at point D and nudged towards the right, the motion of the marble would be unbound on the right, and it would keep going in that direction.

Now, say an object's potential energy is described by the function in Figure 8.11, and the object has total energy E. The object's motion along the x axis will be exactly the same as the projection of the marble's motion on the x axis.

Checkpoint 8-7

A force, $F(x)$, acts on an object. The potential energy function, $U(x)$, associated with the force is given by $U(x) = a(x-6)^2(x-1)(x-3)+20\,\text{J}$, where a is a positive constant. $U(x)$ is plotted in Figure 8.12. Use the "marble" method to determine the direction of the force at $x = 5$. Confirm your answer by finding the value of the force , $F(x)$, at $x = 5$.

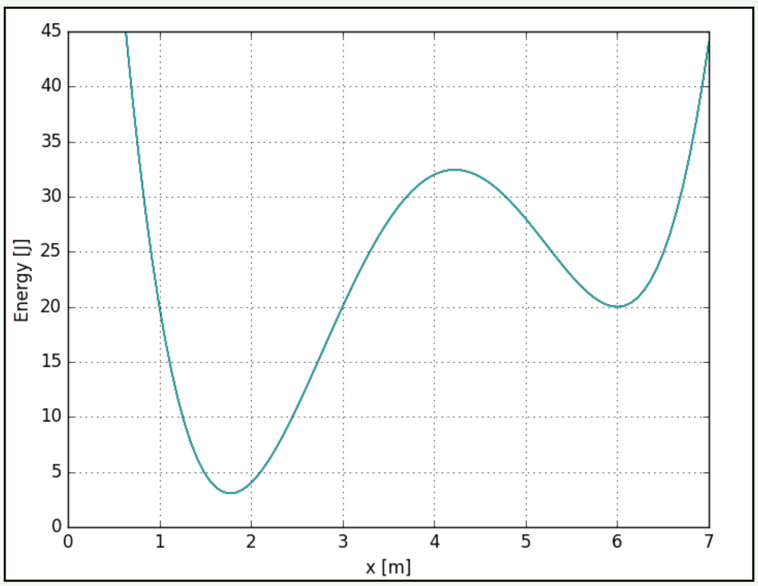


Figure 8.12: A potential energy function $U(x)$. The x-axis represents the x position and the y-axis represents the energy.

A) $F(x=5) = -10a$
B) $F(x=5) = 10a$
C) $F(x=5) = 20a$
D) $F(x=5) = -20a$

8.5 Advanced Topic: The Lagrangian formulation of classical physics

So far, we have seen that, based on Newton's Laws, one can formulate a description of motion that is based solely on the concept of energy. A lot of research was done in the eighteenth century to reformulate a theory of mechanics that would be equivalent to Newton's Theory but whose starting point is the concept of energy instead of the concept of force. This "modern" approach to classical mechanics is primarily based on the research by Lagrange and Hamilton.

Although it is beyond the scope of this text to go into the details of this formulation, it is worth taking a quick look in order to get a better sense of how physicists seek to generalize theories. It is also worth noting that the Lagrangian formulation is the method by which theories are developed for quantum mechanics and modern physics.

The Lagrangian description of a "system" is based on a quantity, L, called the "Lagrangian", which is defined as:

$$\boxed{L = K - U} \tag{8.9}$$

where K is the kinetic energy of the system, and U is its potential energy. A "system" can be a rather complex collection of objects, although we will illustrate how the Lagrangian formulation is implemented for a single object of mass m moving in one dimension under the influence of gravity. Let x be the direction of motion (which is vertical) such that the potential and kinetic energies of the object are given by:

$$\begin{aligned} U(x) &= mgx \\ K(v_x) &= \frac{1}{2}mv_x^2 \\ \therefore L(x, v_x) &= \frac{1}{2}mv_x^2 - mgx \end{aligned}$$

where we chose the potential energy to be zero at $x = 0$, and v_x is the velocity of the object.

In the modern formulation of classical mechanics, the motion of the system will be such that the following integral is minimized:

$$S = \int L dt$$

where L can depend on time explicitly or implicitly (through the fact that position and velocity, on which the Lagrangian depends, are themselves time-dependent). The requirement that the above integral be minimized is called the "Principle of Least Action"[5], and is thought to be the fundamental principle that describes all of the laws of physics. The condition for the action to be minimized is given by the Euler-Lagrange equation:

$$\boxed{\frac{d}{dt}\left(\frac{\partial L}{\partial v_x}\right) - \frac{\partial L}{\partial x} = 0} \tag{8.10}$$

[5]The integral, S, is called the "action" of the system.

Thus, in the Lagrangian formulation, one first writes down the Lagrangian for the system, and then uses the Euler-Lagrange equation to obtain the "equations of motion" for the system (i.e. equation that give the kinematic quantities, such as acceleration, for the system).

Given the Lagrangian that we found above for a particle moving in one dimension under the influence of gravity, we can determine each term in the Euler-Lagrange equation:

$$\begin{aligned}
\frac{\partial L}{\partial v_x} &= \frac{\partial}{\partial v_x}\left(\frac{1}{2}mv_x^2 - mgx\right) = mv_x \\
\therefore \frac{d}{dt}\left(\frac{\partial L}{\partial v_x}\right) &= \frac{d}{dt}(mv_x) = ma_x \\
\frac{\partial L}{\partial x} &= \frac{\partial}{\partial x}\left(\frac{1}{2}mv_x^2 - mgx\right) = -mg
\end{aligned}$$

Putting these into the Euler-Lagrange equation:

$$\begin{aligned}
\frac{d}{dt}\left(\frac{\partial L}{\partial v_x}\right) - \frac{\partial L}{\partial x} &= 0 \\
(ma_x) - (-mg) &= 0 \\
ma_x &= -mg \\
\therefore a_x &= -g
\end{aligned}$$

which is exactly equivalent to using Newton's Second Law (the second last step is equivalent to $F = ma$). In the Lagrangian formulation, we do not need the concept of force. Instead, we describe possible "interactions" by a potential energy function. That is why you may sometimes hear of physicists talking about the "Weak interaction" instead of the "Weak force" when they are talking about one of the four fundamental interactions (forces) of Nature. This is because, in the modern formulation of physics, one does not use the concept of force, and instead thinks of potential energy functions to model what we would call a force in the Newtonian approach.

Emmy Noether, a mathematician in the early twentieth century, proved a theorem that makes the Lagrangian formulation particularly aesthetic. Noether's theorem states that for any symmetry in the Lagrangian, there exists a quantity that is conserved. For example, if the Lagrangian does not depend explicitly on time, then a quantity, which we call energy, is conserved[6].

The Lagrangian that we had above for a particle moving under the influence of gravity did not depend on time explicitly, and thus energy is conserved (gravitational potential energy is converted into kinetic energy and there are no non-conservative forces). If the

[6] If the Lagrangian does not depend on time, then we can shift the system in time and the equations of motion would be unaffected. We say that the Lagrangian is symmetric, or unaffected, by changes in time.

Lagrangian did not depend on position, then a quantity that we call "momentum"[7] would be conserved. In this case, momentum in the x direction was not conserved because the Lagrangian depended on x through the potential energy.

Olivia's Thoughts

We saw in this chapter that describing systems in terms of energy is often easier than describing them in terms of forces. The Lagrangian gives us a way to get the same information we would get from Newton's laws (like the acceleration, etc.), but using energy as a starting point. The Lagrangian method is really useful when we are looking at motion in multiple dimensions, or when we are describing complicated systems. Using the Lagrangian is actually really simple, and just like with forces, you can pretty much approach every problem the same way. Here are the basic steps to follow:

1. Find two expressions for your system: one for the potential energy (U) and one for the kinetic energy (K). This often ends up being the hardest step.
2. Write down the Lagrangian, $L = K - U$, using the expressions you just found.
3. Pick a coordinate. (In one dimension, this is trivial, but it will be important once you start working in multiple dimensions). The Euler-Lagrange equation was given to you as:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v_x}\right) - \frac{\partial L}{\partial x} = 0$$

   because we are working in one dimension. You can actually pick whichever coordinate you are interested in. For example, if you were interested in the motion of your object in the y direction, you would pick y as your coordinate and write:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v_y}\right) - \frac{\partial L}{\partial y} = 0$$

4. Now you just have to do what the equation above tells you to do, which is to start with your Lagrangian (your $L = K - U$ equation) and take a bunch of derivatives. If you try to just plug L into the Euler-Lagrange equation and do all the derivatives at once, it can get confusing. I recommend finding the components separately. I like to start by taking the partial derivative with respect to velocity, $\frac{\partial L}{\partial v_y}$, then taking its derivative with respect to time. Next, I find $\frac{\partial L}{\partial y}$ and then put it all together.
5. That's it! When you've taken the derivatives (and simplified a bit), you'll have an "equation of motion" that gives you information about the motion of the object. You can then use this equation however you want!

[7] See chapter 10

8.6 Summary

Key Takeaways

A force is conservative if the work done by that force on a closed path is zero:

$$\oint \vec{F}(\vec{r}) \cdot d\vec{l} = 0$$

Equivalently, the force is conservative if the work done by the force on an object moving from position A to position B does not depend on the particular path between the two points. The conditions for a force to be conservative are given by:

$$\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} = 0$$
$$\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} = 0$$
$$\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} = 0$$

In particular, a force that is constant in magnitude and direction will be conservative. A force that depends on quantities other than position (e.g. speed, time) will not be conservative. The force exerted by gravity and the force exerted by a spring are conservative.

For any conservative force, $\vec{F}(\vec{r})$, we can define a potential energy function, $U(\vec{r})$, that can be used to calculate the work done by the force along any path between position A and position B:

$$-W = -\int_A^B \vec{F}(\vec{r}) \cdot d\vec{l} = U(\vec{r}_B) - U(\vec{r}_A) = \Delta U$$

where the change in potential energy function in going from A to B is equal to the negative of the work done in going from point A to point B. We can determine the function $U(\vec{r})$ by calculating the work integral over an "easy" path (e.g. a straight line that is co-linear with the direction of the force).

It is important to note that an arbitrary constant can be added to the potential energy function, because only differences in potential energy are meaningful. In other words, we are free to choose the location in space where the potential energy function is defined to be zero.

We can break up the net work done on an object as the sum of the work done by conservative (W^C) and non-conservative forces (W^{NC}):

$$W^{net} = W^{NC} + W^C = W^{NC} - \Delta U$$

where ΔU is the difference in the total potential energy of the object (the sum of the potential energies for each conservative force acting on the object).

The Work-Energy Theorem states that the net work done on an object in going from position A to position B is equal to the object's change in kinetic energy:

$$W^{net} = \frac{1}{2}mv_B^2 - \frac{1}{2}mv_A^2 = \Delta K$$

We can thus write that the total work done by non conservative forces is equal to the change in potential and kinetic energies:

$$W^{NC} = \Delta K + \Delta U$$

In particular, if no non-conservative forces do work on an object, then the change in total potential energy is equal to the negative of the change in kinetic energy of the object:

$$-\Delta U = \Delta K$$

We can introduce the mechanical energy, E, of an object as:

$$E = U + K$$

The net work done by non-conservative forces is then equal to the change in the object's mechanical energy:

$$W^{NC} = \Delta E$$

In particular, if no net work is done on the object by non-conservative forces, then the mechanical energy of the object does not change ($\Delta E = 0$). In this case, we say that the mechanical energy of the object is conserved.

The Lagrangian description of classical mechanics is based on the Lagrangian, L:

$$L = K - U$$

which is the difference between the kinetic energy, K, and the potential energy, U, of the object. The equations of motion are given by the Principle of Least Action, which leads to the Euler-Lagrange equation (written here for the case of a particle moving in one dimension):

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v_x}\right) - \frac{\partial L}{\partial x} = 0$$

Important Equations

Conditions for a force to be conservative:

$$\oint \vec F(\vec r)\cdot d\vec l = 0$$

$$\begin{aligned}\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} &= 0\\ \frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} &= 0\\ \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} &= 0\end{aligned}$$

Potential energy for a conservative force:

$$\Delta U = -W$$
$$U(\vec r_B) - U(\vec r_A) = -\int_A^B \vec F(\vec r)\cdot d\vec l$$

Work-energy theorem:

$$W^{net} = \frac{1}{2}mv_B^2 - \frac{1}{2}mv_A^2 = \Delta K$$

Work:

$$\begin{aligned}W^{net} &= W^{NC} + W^C = W^{NC} - \Delta U\\ W^{NC} &= \Delta K + \Delta U\end{aligned}$$

Energy:

$$\begin{aligned}E &= U + K\\ W^{NC} &= \Delta E\end{aligned}$$

Lagrange:

$$L = K - U$$
$$\frac{d}{dt}\left(\frac{\partial L}{\partial v_x}\right) - \frac{\partial L}{\partial x} = 0$$

8.7 Thinking about the material

Reflect and research

1. When did Lagrange publish his theory of classical mechanics, and what was the name of the publication?
2. What is D'Alembert's contribution to the field of classical mechanics?
3. Who first proposed the Principle of Least Action, and when?
4. What is an example of a situation not already covered that you can describe where mechanical energy is conserved?
5. Under what symmetry is angular momentum conserved?
6. Think of three renewable energy sources and describe how they use conservation of energy to produce electricity.
7. What is a Rube Goldberg machine? Look up some videos of Rube Goldberg machines, and find the coolest one!

To try at home

1. Design a small catapult or slingshot that you can build using materials found at home. Describe how these machines work using conservation of energy.

To try in the lab

1. Propose an experiment to test that energy is conserved in a system where only gravity acts.

8.8 Sample problems and solutions

8.8.1 Problems

Problem 8-1: A ball of mass m is dropped onto a vertical spring with spring constant k. The spring will compress until the ball comes to rest. How much will it compress if the ball is dropped from a height h above the spring? (Solution)

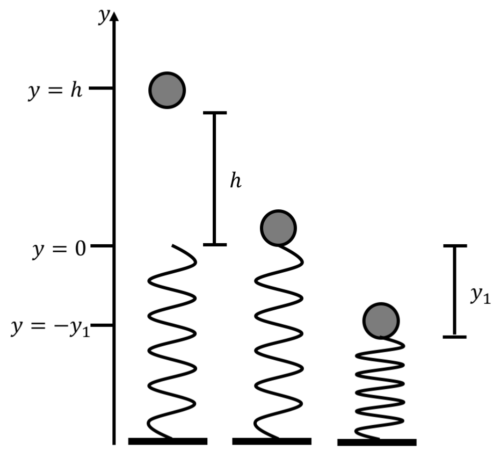


Figure 8.13: A ball is dropped from rest onto a vertical spring.

Problem 8-2: A simple pendulum consists of a mass m connected to a string of length L. The pendulum is released from an angle θ_0 from the vertical. Use conservation of energy to find an expression for the velocity of the mass as a function of the angle. (Solution)

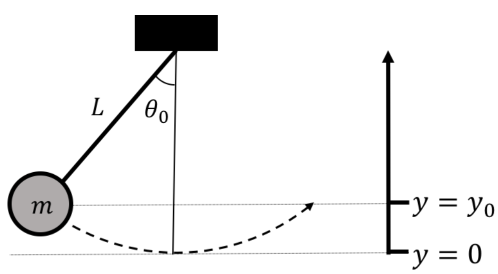


Figure 8.14: A pendulum is released from rest an angle θ_0 from the vertical.

Problem 8-3: A block of mass m sits on a frictionless horizontal surface. It is attached to a wall by a spring with a spring constant k. The mass is pushed so as to compress the spring and then it is released (Figure 8.15). Use the Lagrangian formalism to find an equation of motion for the mass/spring system (i.e. use the Lagrangian to determine the acceleration of the mass). (Solution)

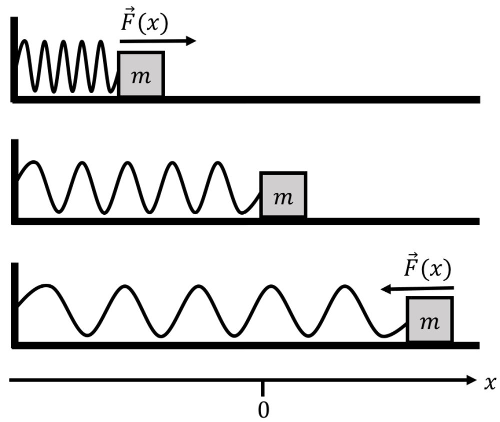


Figure 8.15: A mass attached to a spring oscillates about the rest position of the spring.

8.8.2 Solutions

Solution to problem 8-1: The two forces acting on the ball are gravity and the spring force. Both are conservative, so we can use conservation of mechanical energy. We will find the energy of the ball when it is at a height h above the spring, and the energy of the ball when the spring is fully compressed. Then, we will use conservation of mechanical energy to determine the compression of the spring.

Remember that the total mechanical energy is the sum of the total potential energy and the kinetic energy, $E = U + K$. Let's call the initial position of the ball A and the final position of the ball B. You will notice that we set up our coordinate system so that y is positive upwards, with $y = 0$ at the point where the ball comes into contact with the spring. We choose to define both the gravitational potential energy and spring potential energy so that they are zero at $y = 0$.

Since the ball starts from rest, its kinetic energy is zero at position A. At this point, the ball is not touching the spring, so the potential energy from the spring force is zero. The mechanical energy of the ball at position A is simply equal to its gravitational potential energy:

$$\begin{aligned}
E_A &= U_A + K_A \\
E_A &= mgh
\end{aligned}$$

At position B, the ball is again at rest, so the kinetic energy of the ball is zero. Now that the ball is in contact with the spring, it will experience a force from the spring that can be modelled with a potential energy $U(y) = \frac{1}{2}ky_1^2$, where y_1 is the distance between the rest position of the spring and its compressed length. At point B ($y = -y_1$), the ball will have both spring and gravitational potential energy, so its mechanical energy at position B is given by:

$$\begin{aligned}
E_B &= U_B + K_B = U_B \\
U_B &= mg(-y_1) + \frac{1}{2}ky_1^2 \\
E_B &= -mgy_1 + \frac{1}{2}ky_1^2
\end{aligned}$$

Since mechanical energy is conserved in this system (no non-conservative forces are doing work), we can now set $E_A = E_B$ and solve for y_1:

$$\begin{aligned}
E_A &= E_B \\
mgh &= -mgy_1 + \frac{1}{2}ky_1^2 \\
0 &= \frac{1}{2}ky_1^2 - mgy_1 - mgh
\end{aligned}$$

where in the last line we rewrote the expression as a quadratic equation. We can solve for

y_1 with the quadratic formula:

$$y_1 = \frac{mg \pm \sqrt{(mg)^2 - 4(1/2k)(-mgh)}}{k}$$
$$y_1 = \frac{mg \pm \sqrt{mg(mg + 2kh)}}{k}$$

We now have an expression for the amount the spring is compressed, y_1, in terms of our known values.

Solution to problem 8-2: We are going to find a general expression for the energy of the system, and then use this expression to find the velocity at any point. There are two forces acting on the mass:

1. The force of tension (from the string). This force is perpendicular to the direction of motion at any point, so it does no work on the mass.
2. The force of gravity, which has a potential energy function given by $U(y) = mgy$. We choose the gravitational potential energy to be zero when the pendulum hangs vertically (when $\theta = 0$ and $y = 0$).

The mechanical energy of the mass is conserved, and at any point is given by the sum of its kinetic and its gravitational potential energies:

$$E = mgy + \frac{1}{2}mv^2$$

We want to find the velocity as a function of θ, so we need to write y in terms of θ. As you may recall from Problem 7-2, we saw that from the geometry of the problem, we can express the height of the mass as $y = L - L\cos\theta$, or $L(1 - \cos\theta)$, where y is the height as measured from the bottom point of the motion. You can refer to Figure 7.25 to refresh your memory. The energy at any point is then:

$$E = mgL(1 - \cos\theta) + \frac{1}{2}mv^2$$

Conservation of energy tells us that the total energy at any point must be the same as the initial energy. So, we can use our initial conditions to find the total energy of the system. The mass starts from rest (initial kinetic energy is zero) an angle θ_0 above the vertical:

$$E = mgL(1 - \cos\theta) + \frac{1}{2}mv^2$$
$$E_{initial} = mgL(1 - \cos\theta_0)$$

Now that we have found the total energy of the system, we can write our general expression for the energy of the system at any point:

$$E = mgL(1 - \cos\theta) + \frac{1}{2}mv^2$$
$$mgL(1 - \cos\theta_0) = mgL(1 - \cos\theta) + \frac{1}{2}mv^2$$

All that's left to do is simplify the expression and rearrange for v:

$$\begin{aligned}
mgL(1-\cos\theta_0) &= mgL(1-\cos\theta)+\frac{1}{2}mv^2 \\
gL(1-\cos\theta_0)-gL(1-\cos\theta) &= \frac{1}{2}v^2 \\
gL-gL\cos\theta_0-gL+gL\cos\theta &= \frac{1}{2}v^2 \\
gL(\cos\theta-\cos\theta_0) &= \frac{1}{2}v^2 \\
\therefore v &= \sqrt{2gl(\cos\theta-\cos\theta_0)}
\end{aligned}$$

Discussion: We can see from this expression that the speed will be maximized when $\cos\theta$ is maximized, which will occur when $\theta = 0$ (when the pendulum is vertical). This is as we expected. We can also see that we will get an imaginary number if the magnitude of θ is greater than θ_0, showing that the motion is constrained between $-\theta_0$ and θ_0. Finally, we showed that the velocity of the pendulum does not depend on the mass!

Solution to problem 8-3: We are going to find an equation of motion of the system using the Lagrangian method. We choose to use a one dimension coordinate system, with the x axis defined to be co-linear with the spring, positive in the direction where the spring is extended, and set the origin to be located at the rest position of the spring. The kinetic energy and potential energy of the mass are given by

$$\begin{aligned}
K &= \frac{1}{2}mv_x^2 \\
U &= \frac{1}{2}kx^2
\end{aligned}$$

since the only force exerted on the mass that can do work is the force from the spring. We have chosen the potential energy to be zero at $x = 0$. The Lagrangian for this system is:

$$\begin{aligned}
L &= K-U \\
L &= \frac{1}{2}mv_x^2-\frac{1}{2}kx^2
\end{aligned}$$

The Euler-Lagrange equation in one dimension is:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v_x}\right)-\frac{\partial L}{\partial x}=0$$

We can calculate the terms of the Euler-Lagrange equation:

$$\begin{aligned}
\frac{\partial L}{\partial v_x} &= \frac{\partial}{\partial v_x}\left(\frac{1}{2}mv_x^2 - \frac{1}{2}kx^2\right) \\
&= mv_x \\
\therefore \frac{d}{dt}\left(\frac{\partial L}{\partial v_x}\right) &= \frac{d}{dt}(mv_x) \\
&= ma_x \\
\text{and} \qquad \frac{\partial L}{\partial x} &= \left(\frac{1}{2}mv_x^2 - \frac{1}{2}kx^2\right) \\
&= -kx
\end{aligned}$$

and then put them together to get:

$$\begin{aligned}
\frac{d}{dt}\left(\frac{\partial L}{\partial v_x}\right) - \frac{\partial L}{\partial x} &= 0 \\
\therefore ma_x &= -kx
\end{aligned}$$

We can see that this equation of motion is equivalent to Newton's Second Law.

9 Gravity

In previous chapters, we have so far learned about Newton's Theory of Classical Mechanics, which allowed us to model the motion of an object based on the forces acting on the object. In this chapter, we present the theories that describe the force of gravity itself. We will see several theories of gravity and focus primarily on Newton's Universal Theory of Gravity.

Learning Objectives

- Understand Kepler's Laws.
- Understand Newton's Universal Theory of Gravity.
- Understand Gauss' Law and the gravitational field.
- Understand how to use energy to describe orbits.
- Understand how Einstein's General Theory of Relativity differs from Newton's theory of gravity.

Think About It

A person stands on a scale at the top of Mount Logan, the tallest mountain in Canada. How will her measured weight compare to her weight at sea level?

A) It will be slightly less than her weight at sea level.
B) It will be equal to her weight at sea level.
C) It will be slightly more than her weight at sea level.

9.1 Kepler's Laws

Although humans have long been fascinated by the motion of objects in the sky, it was Johannes Kepler, in the early seventeenth century, that was the first to write down quantitative rules that described the motion of planets around the Sun. His theory was based on the extensive and detailed observations recorded by Tycho Brahe in the late sixteenth century.

Kepler proposed three laws that describe all of the data that Tycho Brahe had collected about planetary motion:

1. The path of a planet around the Sun is described by an ellipse with the Sun at once of its foci.
2. All planets move in such a way that the area swept by a line connecting the planet and the Sun in a given period of time is constant.
3. The ratio between the orbital periods, T, of two planets squared is equal to the ratio of the semi-major axes, s, of their orbits cubed:

$$\left(\frac{T_1}{T_2}\right)^2 = \left(\frac{s_1}{s_2}\right)^3$$

We examine these three laws in more detail in the sections that follow. It should also be noted that, even though Kepler's Laws were derived for planets orbiting the Sun, they apply to any body that is orbiting any other body under the influence of gravity[1].

9.1.1 Kepler's First Law

Kepler noticed that the motion of all planets followed the path of an ellipse with the Sun located at one of its foci. Figure 9.1 shows a diagram of an ellipse, along with its two foci, its semi-major axis, s, its semi-minor axis, b, and its eccentricity, e. The eccentricity is a measure of how far a focus is from the centre of the ellipse. A larger eccentricity thus corresponds to a "flatter" ellipse. Note that a circle is just a special case of an ellipse, with both foci located at the centre of the circle.

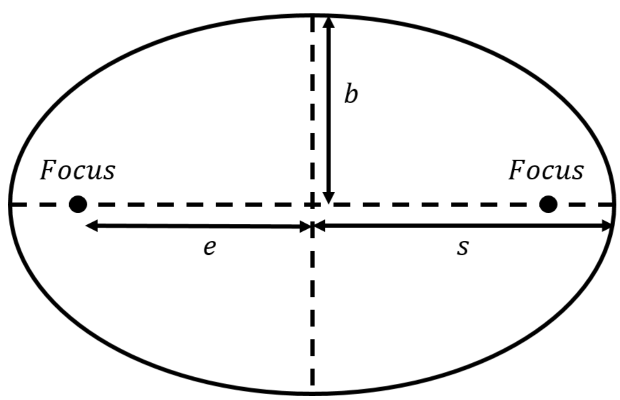


Figure 9.1: A ellipse, showing its two foci, its semi-major axis, s, its semi-minor axis, b, and its eccentricity, e.

The sun is located at one of the foci. The point of closest approach to the Sun is called the "perihelion" of the orbit (or "perigee" if the orbit is not around the Sun), and the point furthest from the Sun is called the "aphelion" of the orbit (or "apogee" if the orbit is not around the Sun), as shown in Figure 9.2.

[1]In fact, they apply for any two bodies orbiting each other if the force between them is an "inverse-square" law, such as the gravitational and electric forces.

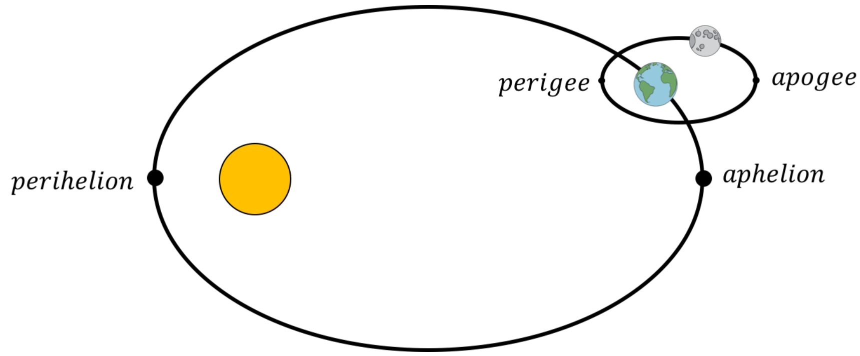


Figure 9.2: The orbit of the Earth around the Sun, showing the perihelion and aphelion, and the orbit of the Moon around the Earth, showing the perigee and the apogee. (Not to scale.)

Checkpoint 9-1

Order the ellipses from smallest eccentricity to largest eccentricity.

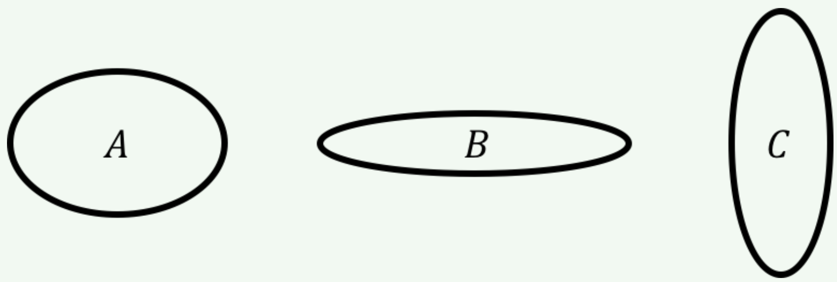


Figure 9.3: Three ellipses, each with a different eccentricity.

9.1.2 Kepler's Second Law

Kepler's Second Law is really a statement about the speed of a planet in an elliptical orbit. It states that the area swept by a line connecting the planet and the Sun in a given period of time is fixed. This is illustrated in Figure 9.4, which shows the elliptical orbit of a planet around the Sun located at one of the foci, and the area swept out when the planet goes from A to B and from C to D.

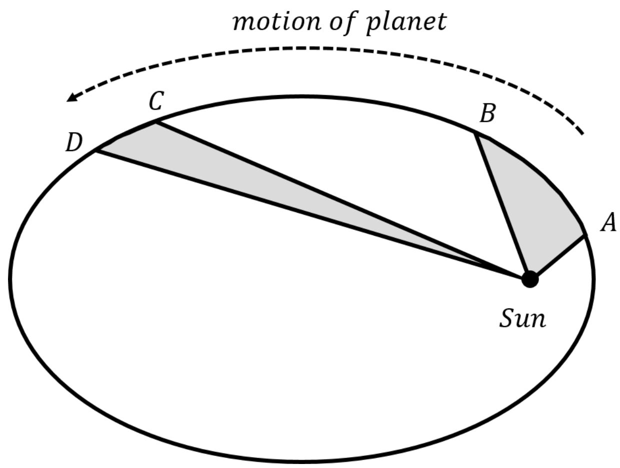


Figure 9.4: Illustration of Kepler's Second Law, showing the area that is "swept" by a planet in a fixed period of time.

Kepler's Second Law states that the two areas that are shown by the greyed out sections in the figure are the same if the planet took the same amount of time to travel between points A and B as it did to travel between points C and D. Because the points C and D are further away from the Sun than points A and B, the distance between points C and D must be smaller than the distance between points A and B for the two areas to be the same. This, in turn, implies that the planet must be moving slower between C and D than between points A and B. The speed of a planet is thus greatest at the perihelion and smallest at the aphelion. As we will see in a later chapter, Kepler's Second Law is equivalent to the statement that the angular momentum of the planet about the Sun is conserved.

Checkpoint 9-2

Based on Kepler's second law, what can you say about the speed of a planet in a **circular** orbit?

9.1.3 Kepler's Third Law

Kepler's Third Law is quantitative and relates the orbital periods (T) and the semi-major axes (s) between any two planets in orbit around the Sun:

$$\left(\frac{T_1}{T_2}\right)^2 = \left(\frac{s_1}{s_2}\right)^3$$

We can re-arrange this relation so that all of the quantities related to one planet are on the same side of the equal sign:

$$\frac{T_1^2}{s_1^3} = \frac{T_2^2}{s_2^3} = \text{constant}$$

In other words, the ratio between the orbital period squared and the semi-major axis cubed is a constant, independent of the particular planet. In Example 9-2, we will use Newton's Universal Theory of Gravity to evaluate the constant.

Checkpoint 9-3

An object is in a circular orbit with radius r and has an orbital speed v. If you double the radius of the circular orbit, what will be the value of the orbital speed?

A) $2v$
B) $8v$
C) $\sqrt{8}v$.
D) $\frac{1}{\sqrt{2}}v$

9.2 Newton's Universal Theory of Gravity

Newton supposedly gained insight into the gravitational force by observing an apple falling from a tree and concluding that if it is the same force that makes apples fall at sea level and at the top of a mountain, perhaps that force can be exerted all the way up to the moon. It is rather remarkable that Newton was able to make the connection between falling apples and the motion of the moon around the Earth to find a single theory to describe both situations.

We should be clear that the theory of gravity is a different theory than Newton's "Laws of Motion" (Newton's Three Laws). The Laws of Motion introduce the concepts of force and inertial mass, and prescribe how to use those concepts in order to model motion using kinematics. Newton's Universal Theory of Gravity is a theory that describes the force of gravity that two bodies with (gravitational) mass exert on each other.

Newton's Universal Theory of Gravity states that if two bodies, with masses M_1 and M_2, located at positions $\vec{r}_1$ and $\vec{r}_2$, respectively, are separated by a distance, r, then M_2 will exert an attractive force on M_1, $\vec{F}_{12}$, given by:

$$\vec{F}_{12} = -G\frac{M_1 M_2}{r^2}\hat{r}_{21} \tag{9.1}$$

where $\hat{r}_{21}$ is the unit vector from M_2 to M_1:

$$\begin{aligned}\vec{r}_{21} &= \vec{r}_2 - \vec{r}_1\\ \hat{r}_{21} &= \frac{1}{r}\vec{r}_{21}\end{aligned}$$

as shown in Figure 9.5. $\vec{F}_{12}$ should be read as "the force on body 1 from body 2". $G = 6.67 \times 10^{-11}\,\text{Nm}^2/\text{kg}^2$ is Newton's Universal Constant of Gravity. It should be noted that Newton's theory for the force of gravity written in this form only applies to either point masses (separated by a distance r) or spherical bodies whose centres are separated by a distance r that is larger than the radius of either sphere.

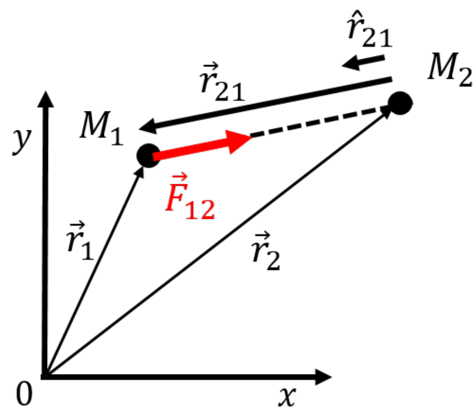


Figure 9.5: Illustration of the vectors involved in Newton's Universal Theory of Gravity.

Originally, Newton argued that the force of gravity would be proportional to the masses of the bodies, and inversely proportional to the square of the distance between them:

$$F_{12} \propto \frac{M_1 M_2}{r^2}$$

and G is simply the constant of proportionality.

Because of Newton's Third Law, body 1 exerts a force on body 2 that is equal in magnitude but opposite in direction:

$$\vec{F}_{12} = -\vec{F}_{21}$$

Example 9-1

Calculate the magnitude of the force of gravity between yourself and a person standing 50 cm from you and compare that to your weight at the surface of the Earth (the force of gravity exerted by the Earth on you).

Solution

If we assume that the two people have a mass of 50 kg, the force of gravity exerted by one on the other, if they are separated by 50 cm, is given by:

$$F = G\frac{M_1 M_2}{r^2} = (6.67 \times 10^{-11}\,\mathrm{Nm^2/kg^2})\frac{(50\,\mathrm{kg})(50\,\mathrm{kg})}{(0.5\,\mathrm{m})^2} = 6.67 \times 10^{-7}\,\mathrm{N}$$

This is a very small force, compared to their weight, F_g:

$$F_g = M_1 g = (50\,\mathrm{kg})(9.8\,\mathrm{N/kg}) = 490\,\mathrm{N}$$

which is approximately 700 million times bigger.

Discussion: The force of gravity is a very weak force when compared to other forces in Nature, such as the electric force between two charges. Newton's Universal Constant of Gravity is very small, so the force of gravity between two objects is very small unless either of the masses involved are very large, or the distance between them is very small. In general, when modelling the motion of objects on the Earth, it is generally safe to ignore the forces of gravity between objects and only include their weight (the force of gravity from the Earth).

Checkpoint 9-4

The radius of the Earth is 6371 km. What is the order of magnitude of the Earth's mass?

A) 10^{24}kg
B) 10^{18}kg
C) 10^{19}kg
D) 10^{21}kg
E) Not enough information.

Example 9-2

Determine the constant in Kepler's Third Law for planets orbiting the Sun, namely the value of the ratio:

$$\frac{s^3}{T^2}$$

where s is the semi-major axis and T is the orbital period.

Solution

Since Kepler's Third Law holds for any body orbiting the Sun, we can easily determine the ratio by considering a planet of mass m in a circular orbit of radius R centred about the Sun. The semi-major axis of the orbit is equal to the radius of the orbit for a circular orbit ($s = R$).

If the planet is in a circular orbit about the Sun, its speed, v, will be constant, by Kepler's Second Law, and it will thus be executing uniform circular motion. The only force exerted on the planet is the force of gravity exerted by the Sun. Thus the force of gravity must be equal to the mass of the planet times its radial (centripetal) acceleration, a_R, which is given by:

$$a_R = \frac{v^2}{R}$$

Newton's Second Law for the planet can be written as:

$$\begin{aligned}\sum F = F_g &= ma_R \\ G\frac{Mm}{R^2} &= m\frac{v^2}{R} \\ G\frac{M}{R} &= v^2\end{aligned}$$

where M is the mass of the Sun. The speed of the planet is given by the circumference of the orbit divided by the orbital period T, since it is constant:

$$v = \frac{2\pi R}{T}$$

Re-arranging the equation from Newton's Second Law:

$$\begin{aligned}G\frac{M}{R} &= v^2 \\ G\frac{M}{R} &= \frac{4\pi^2 R^2}{T^2} \\ \therefore \frac{R^3}{T^2} &= G\frac{M}{4\pi^2}\end{aligned}$$

Thus, we find that the ratio of the cube of the orbital radius to the period squared is a constant that depends only on the mass of the Sun, which will then be the same for all planets (as it does not depend on, say, the mass of the planet that we considered).

Discussion: The relation above can, for example, be used to determine the mass of the Sun, since we can use geometry to determine the semi-major axis for the orbit of a planet, as Kepler did with the data from Tycho Brahe.

Example 9-3

The acceleration due to Earth's gravity depends on the force that the Earth exerts on an object. Using the Earth's mass and radius, determine the acceleration of falling objects due to Earth's gravity at the surface of the Earth. Also, determine the altitude where the acceleration due to Earth's gravity is half of that at the Earth's surface.

Solution

We can find the acceleration due to Earth's gravity by determining the acceleration of a mass m upon which gravity is the only acting force. In other words, we model an object that is in free-fall a distance R away from the centre of the Earth. Newton's Second Law can be used in one dimension (corresponding to the direction that connects the falling mass to the centre of the Earth):

$$\sum F = G\frac{Mm}{R^2} = ma$$
$$\therefore a = G\frac{M}{R^2}$$

where $M = 5.97 \times 10^{24}\,\text{kg}$ is the mass of the Earth. At the surface of the Earth, $R = R_\oplus = 6.371 \times 10^6\,\text{m}$:

$$\begin{aligned} a &= G\frac{M}{R_\oplus^2} = (6.67 \times 10^{-11}\,\text{Nm}^2/\text{kg}^2)\frac{(5.97 \times 10^{24}\,\text{kg})}{(6.371 \times 10^6\,\text{m})^2} \\ &= 9.81\,\text{m/s}^2 \end{aligned}$$

which, of course, is the value of g that we have been using so far for the acceleration due to gravity near the surface of the Earth. To find the altitude at which this is reduced by half, we first find the value of R that corresponds to this acceleration:

$$\frac{1}{2}G\frac{M}{R_\oplus^2} = G\frac{M}{R^2}$$
$$\therefore R = \sqrt{2}R_\oplus = 9.0 \times 10^6\,\text{m}$$

which corresponds to an altitude of $h = R - R_\oplus = 2640\,\text{km}$, well above the Earth's atmosphere.

Discussion: The acceleration of falling objects decreases as one moves further from the centre of the Earth. It is thus an approximation to assume that g is a constant, although in most cases this is a very good approximation. In practice, the value of g will depend both on the distance from the centre of the Earth and the composition (density) of the material in the Earth's crust below where g is being measured. Precise measurements of g have bee used to determine the composition of the Earth's crust and to search for mineral and oil deposits.

9.2.1 Weight and apparent weight

You have probably seen images of astronauts floating around the International Space Station (ISS) or other orbiting vessels, and heard of the term "weightlessness" to describe their motion. The ISS is in orbit at an altitude of approximately 400 km, where the force of Earth's gravity is far from negligible (in Example 9-3 we showed that one needs to go to an altitude of 2640 km for the force to be reduced by half of that at the surface of the Earth). The contradiction between being weightless and the fact that weight is not zero is resolved by understanding that the popular term "weightless" is imprecise from a physics perspective.

The correct term to use from a physics perspective is to say that the *apparent weight* of the astronauts is zero when they are floating around. Weight is the magnitude of the force of gravity exerted by the Earth. Apparent weight is, for example, the force that one measures when standing on a spring scale, which is equal to the normal force exerted by the spring scale on the person. Apparent weight could also be determined by the tension in a string from which a person is suspended. The apparent weight is the sum of the forces exerted on a person minus the force of gravity. If gravity is the only force exerted on a person (or object), that person's apparent weight is zero, which is what is popularly called being weightless.

One way to feel weightless is when you are in free-fall (e.g. the first few seconds of a parachute jump from an airplane). One can think of being in orbit as continuously falling towards the centre of the Earth, but with an initial velocity in a direction such that you never actually collide with the Earth. The feeling of weightlessness will occur any time that the only force exerted on you is the force of gravity. If you are in a spacecraft in any orbit and the only force on the spacecraft is from gravity (i.e. no rockets or wings are exerting any forces), then everything in the spacecraft will have the same acceleration, since gravity is the only force acting on anything in the spacecraft, and it will appear that everything is just floating. To an outside observer, it would be clear that the spacecraft and its contents are all accelerating.

Effects of Earth's rotation

Earth's rotation affects the apparent weight of objects near the Earth's surface. Consider a person standing on a spring scale at the North pole (top free-body diagram in Figure 9.6). The only two forces exerted on the person are their weight, $\vec{F}_g$, and the normal force, $\vec{N}$, exerted by the spring scale. Since the person is not accelerating, the normal force and the weight have the same magnitude and opposite directions. The scale will thus read the actual weight of the person[2].

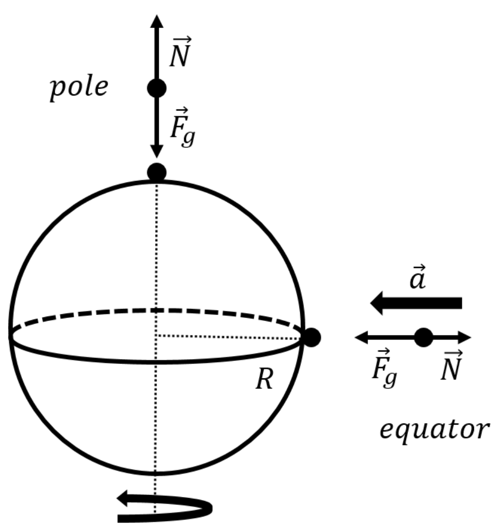


Figure 9.6: The apparent weight, given by the normal force, is different at the Earth's equator because a person's acceleration is non-zero as they rotate with the Earth.

[2]The weight that is displayed on the scale is equal in magnitude to the normal force exerted by the scale on the person. It is the reaction force to that normal force.

Consider, instead, a person standing on a spring scale at the equator (Figure 9.6). That person is accelerating because they are undergoing uniform circular motion as they rotate along with the Earth. Again, the only forces acting on the person are their weight and the normal force exerted by the scale. The sum of the forces must now be equal to the person's mass, m, times the radial acceleration, a_r, that is necessary for them to follow the surface of the Earth as the Earth rotates about its axis. Newton's Second Law allows us to find the magnitude of the normal force acting on the person:

$$\begin{aligned}\sum F &= F_g - N = ma_r = m\frac{v^2}{R}\\ \therefore N &= F_g - m\frac{v^2}{R}\\ &= G\frac{Mm}{R^2} - m\frac{v^2}{R}\\ &= m\left(G\frac{M}{R^2} - \frac{v^2}{R}\right)\\ &= m\left(g - \frac{v^2}{R}\right)\end{aligned}$$

where M is the mass of the Earth, R is the radius of the Earth, and v is the speed at the surface of the Earth due to the Earth's rotation. In the last line, we used the result from Example 9-3 where we determined the value of g in terms of the mass and radius of the Earth.

We see that the normal force is reduced compared to what it would be if the Earth were not rotating ($v = 0$) or if one is standing at one of the poles. Your apparent weight, which you can measure by standing on a spring scale, is thus smaller at the equator than it is at the poles. The quantity in parenthesis can be thought of as a modified or "effective" value of g at the equator.

Checkpoint 9-5

What is the effective value of g at the equator?

A) $9.80\,\text{m/s}^2$
B) $9.78\,\text{m/s}^2$
C) $9.70\,\text{m/s}^2$
D) $9.51\,\text{m/s}^2$

If you are circling the Earth a distance R from the centre of the Earth at a constant speed v, it is possible for your apparent weight to be zero. Imagine standing on a scale in an aircraft that is circling the Earth and measuring your apparent weight with the spring scale. As the speed of the aircraft increases, your apparent weight, N, decreases according to the formula that we just found:

$$N = m\left(G\frac{M}{R^2} - \frac{v^2}{R}\right)$$

At a certain speed, v, your apparent weight is zero and you feel weightless:

$$G\frac{M}{R^2} = \frac{v^2}{R}$$
$$\therefore v = \sqrt{G\frac{M}{R}}$$

This speed corresponds to a centripetal acceleration that is exactly equal to that due to gravity. This makes sense, since gravity is the only force that is acting on you (the normal force is now zero), which is exactly what we call being in orbit.

Earth's rotation has some interesting consequences for stationary objects. At any position on Earth that is not at the equator or the poles, the sum of the forces on any stationary object (meaning stationary relative to the Earth's surface) cannot be zero. This is because the object must rotate along with the Earth, so the net force on the object must point toward the centre of the circle about which that location on Earth is rotating.

Take, for example, a plumb line, which consists of a mass hanging from a string. The two forces acting on the mass are gravity and tension. The force of gravity must point towards the centre of the Earth. We would expect that the force of tension, and therefore the string, would point directly away from the centre of the Earth. However, we find that if the plumb line is located at some angle θ from the equator (but not at the equator or poles), as in Figure 9.7, then the string will point slightly away from the centre of the Earth. In order for the mass to remain stationary relative to the ground, it must rotate along with the Earth (radius R) around a circle of radius $R\cos\theta$. Thus, the tension from the string cannot point away from the centre of the Earth, because the net force must point towards the centre of the circle of radius $R\cos\theta$.

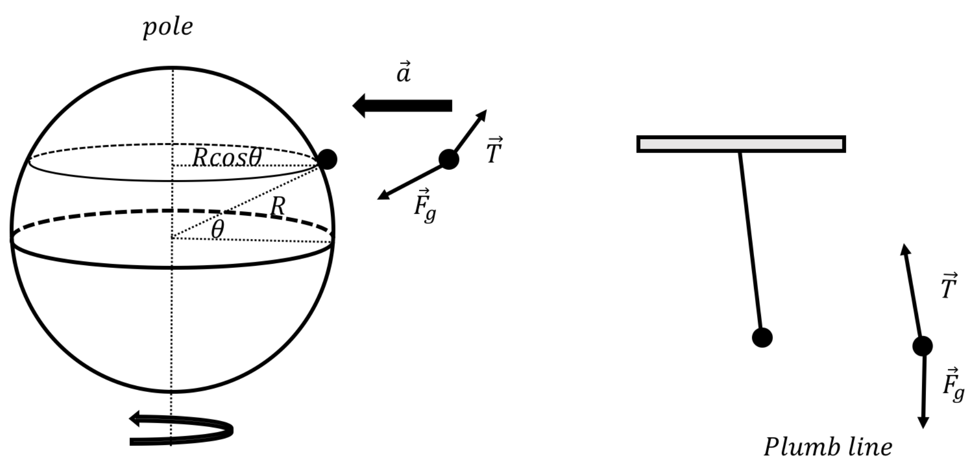


Figure 9.7: Away from the equator and poles, a plumb line (right) does not point towards the centre of the Earth, because the net force on the mass must provide the acceleration towards the centre of the circle (of radius $R\cos\theta$) about which the plumb line rotates due to the Earth's rotation. Note that the deflection of the plumb line is highly exaggerated.

Checkpoint 9-6

You cut the string of the plumb line. Where does the mass land relative to its initial latitude (the angle θ in Figure 9.7)?

A) At the same latitude.
B) Closer to the nearest pole.
C) Closer to the equator.

9.2.2 The gravitational field

The gravitational force exerted on a mass m by a mass M can be written as:

$$\vec F(\vec r) = -G\frac{Mm}{r^2}\hat r$$

if we define a coordinate system with the origin located at the centre of mass M so that $\vec r$ is the position of m relative to M. We can define the "gravitational field", $\vec g(\vec r)$, at position, $\vec r$, due to the presence of mass M as the gravitational force per unit mass exerted by M:

$$\boxed{\vec g(\vec r) = \frac{\vec F(\vec r)}{m} = -\frac{GM}{r^2}\hat r} \tag{9.2}$$

The word "field" is just a mathematical term for a function that depends on position. Since $\vec g(\vec r)$ is a vector, we would refer to it as a "vector field".

Defining the gravitational field makes it easy to calculate the force of gravity from M on any mass m:

$$\vec F_g = m\vec g(\vec r)$$

At the surface of the Earth, the magnitude of the gravitational field is given by:

$$g(R_\oplus) = \frac{GM}{R_\oplus^2} = 9.81\,\text{N/kg}$$

where $R_\oplus$ is the radius of the Earth. Of course, this also corresponds to the acceleration of objects in free-fall near the surface of the Earth, which we can find from Newton's Second Law:

$$\begin{aligned}\sum \vec F = \vec F_g &= m\vec a\\ m\vec g(R_\oplus) &= m\vec a\\ \therefore \vec a &= \vec g(R_\oplus)\end{aligned}$$

but we see here why it more precise to refer to g as the "magnitude of the gravitational field at the surface of the Earth" rather than "the acceleration due to Earth's gravity". It is also worth noting that the two are only equal if the gravitational mass (on the left of the equation in the second line) is the same as the inertial mass (on the right of the equation). The gravitational mass is the mass that appears in the gravitational force defined by Newton, whereas the inertial mass is the mass that appears with the acceleration in Newton's Second Law.

Checkpoint 9-7

Two small objects with different masses, m_1 and m_2, are located a distance r from a nearby star. What can you say about the magnitude of the gravitational field and the magnitude of the gravitational force on m_1 and m_2?

A) The field is different and the forces are different.
B) The field is different but the forces are the same.
C) The field is the same but the forces are different.
D) The field is the same and the forces are the same.

Suppose that there are two large mass bodies, M_1 and M_2, and a smaller mass body, m. We can calculate the net gravitational force on m by summing the gravitational force vectors from M_1 and M_2 separately. If the gravitational fields from M_1 and M_2 are given by $\vec{g}_1(\vec{r})$ and $\vec{g}_2(\vec{r})$, respectively, then the total gravitational force on m is given by:

$$\begin{aligned}\vec{F} &= m\vec{g}_1(\vec{r}) + m\vec{g}_2(\vec{r}) = m(\vec{g}_1(\vec{r}) + \vec{g}_2(\vec{r}))\\ &= m\vec{g}(\vec{r})\end{aligned}$$

where we have introduced the total gravitational field:

$$\vec{g}(\vec{r}) = \vec{g}_1(\vec{r}) + \vec{g}_2(\vec{r})$$

In other words, if there are multiple bodies that result in a non-negligible force of gravity, we can calculate their gravitational fields independently and sum them together to define a net gravitational field, $\vec{g}(\vec{r})$, that models the net force of gravity from all of the bodies. The net gravitational force on a new body of mass m' is then simply given by $m'\vec{g}(\vec{r})$, and we do not need to add any more vectors together. For example, when calculating the motion of satellites that can be influenced by the force of gravity from both the Earth and the Moon, we simply need to calculate the net gravitational field from the Earth and Moon, and the motion of any satellite can then be modelled using that net gravitational field.

Checkpoint 9-8

There are two planets with equal mass located a distance d apart. Position A is located midway between the two planets. Position B is located a distance $d/2$ from one of the planets, in the position shown in Figure 9.8. How does the field at A compare to the field at B?

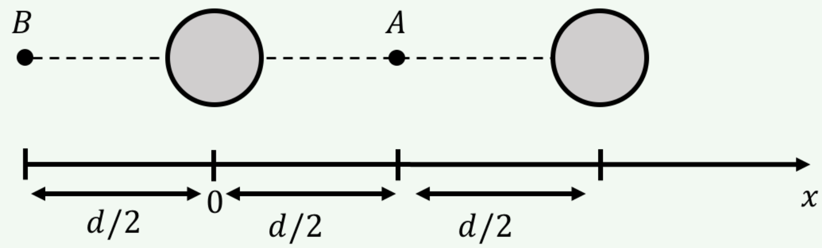


Figure 9.8: Two planets are a distance d apart. We are considering the gravitational field at two positions, A and B, located near the planets.

A) The magnitude of the field is the same at A and B.
B) The magnitude of the field is greater at A than at B.
C) The magnitude of the field is greater at B than at A.

9.2.3 Gauss' Law

Newton's Universal Theory of Gravity postulates that the force of gravity between two bodies decreases as the squared of the distance between those two bodies. Using the terminology of a field, we would say that the strength of the gravitational field from an object decreases as the inverse of the square of the distance to that object. This is an example of a what we generally call an "inverse-square law". The electric force between two charges is also given by an inverse-square law, and we now understand that these forces behave as if they were "transmitted" by waves or particles.

If a force is given by an inverse-square law, then Gauss' Law gives a way to determine the strength of the field that is associated with that force. In the case of gravity, Gauss' Law states that:

$$\oint \vec{g}(\vec{r}) \cdot d\vec{A} = 4\pi G M^{enc}$$

where the integral on the left is an integral over a "closed surface" that completely surrounds the mass for which we want to determine the gravitational field. To evaluate the integral, imagine taking a closed surface, S, that surrounds the mass. The vector $d\vec{A}$ is defined as the vector that goes with a small element of that surface and points outwards from the closed surface. The magnitude of the vector is equal to the area of that small surface (dA), as illustrated in Figure 9.9. You can then take the scalar product of $d\vec{A}$ with the gravitational field, $\vec{g}(\vec{r})$, at that point on the surface. If you sum all of those scalar products, you get the value of the integral on the left. Gauss' Law states that the value of that integral is equal $4\pi G$ times the total mass that is enclosed by the surface.

Olivia's Thoughts

If you want to know if a surface is closed, ask yourself if you could put water inside the surface and not be worried about it spilling out. For example, if you put water in a sphere or a cube , the water would not spill out even if you shook it around, so they are closed surfaces. A flat square is an open surface because there is no "inside" to even put the water in. A bowl is an open surface because, though you can put water in it, the water could spill out.

We will go into more detail about Gauss' Law when we cover electromagnetism, but it is worth seeing how to use it in a simple scenario. Figure 9.9 shows a spherical body of mass M and radius R for which we would like to determine the value of the gravitational field at a distance r from the centre of the body.

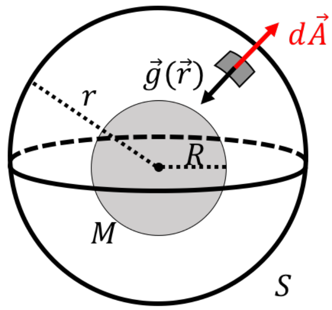


Figure 9.9: Example of a spherical Gaussian surface, S, of radius r centred about a body of mass M and radius R. An element of the surface, $d\vec{A}$ is also shown along with the gravitational field, $\vec{g}(\vec{r})$, at that point.

To do so, we draw a "Gaussian surface", S, that is a sphere with a radius r, and centred about the body. At any point on the surface, the area element vector $d\vec{A}$ points away from the centre of the spherical surface. The gravitational field vector, $\vec{g}(\vec{r})$ will always point towards the centre of the spherical surface, as illustrated. Furthermore, by symmetry, the magnitude of $\vec{g}(\vec{r})$ is constant along the whole Gaussian surface. Thus, the scalar product $\vec{g}(\vec{r}) \cdot d\vec{A} = -g(r)dA$ everywhere along the surface (it is negative because the two vectors are anti-parallel). The integral is thus given by:

$$\oint \vec{g}(\vec{r}) \cdot d\vec{A} = -g(r) \oint dA$$

where we factored $g(r)$ out of the integral, since the magnitude of $\vec{g}(\vec{r})$ is constant for all of the area elements dA on the sphere. Remember that an integral is a sum. The integral $\oint dA$ thus means "sum all of the area elements dA over the entire surface S". Thus, the integral is the total area of the spherical surface S[3]:

$$\oint \vec{g}(\vec{r}) \cdot d\vec{A} = -g(r) \oint dA = -g(r)(4\pi r^2)$$

[3] The surface area of a sphere of radius r is $4\pi r^2$.

Inserting this into Gauss' Law, we find:

$$\oint \vec{g}(\vec{r}) \cdot d\vec{A} = 4\pi G M^{enc}$$
$$-g(r)(4\pi r^2) = 4\pi G M^{enc}$$
$$\therefore g(r) = -\frac{GM}{r^2}$$

where $M^{enc} = M$ is the total mass enclosed by the Gaussian surface (in this case, the entire mass M is enclosed). This is of course the result that we expected and obtained earlier from Newton's formulation. Note that Gauss' Law is only easy to use if the system is highly symmetric (e.g. spherically symmetric), and that it does not give the direction of the field vector, which must be obtained from symmetry arguments.

Olivia's Thoughts

Here's an analogy to describe Gauss's Law for gravity: A famous celebrity is doing an event, and they attract a certain number of fans who want to get as close to the celebrity as possible. You put up a barricade around the celebrity. The gravitational field is represented by how crowded it is somewhere along the barricade. If a second celebrity is at the event, they will attract their own fans, so there will be more people around the barricade. The number of celebrities is kind of like the enclosed mass M^{enc}.

A photographer is coming to the event, and you told him to stand at some location that is a distance r from the celebrities. The photographer wants to know how crowded it will be when he is standing behind the barricade at that location. Gauss's law gives us a way to figure this out. If you know which celebrities are at the event (M^{enc}), you can determine how many people will be there (this is like finding $4\pi G M^{enc}$). Then, if you can build a barricade such that the fans are evenly distributed around it, and you know how long that barricade is ($\oint dA$), you can easily calculate how crowded it will be at some point along the barricade (you can just divide the number of people by the length of the barricade).

The barricade represents our Gaussian surface and, like a Gaussian surface, it can be whatever shape we want as long as it encloses the celebrities and passes through the point we are interested in. If we want to make sure the people are spread out evenly, the shape of the barricade is going to depend on the specific case. Let's take the example of our single spherical body. This is analagous to having one celebrity at the event. Figure 9.10 shows two possible barricades we could build. Although we can technically build the barricade on the left, it doesn't help us because the areas closer to the celebrity will be more crowded. Instead, we want to build the barricade on the right, which is a circle of radius r, because the fans are evenly spread out. This is why we use a spherical Gaussian surface when we're considering the field due to a spherical body - at any point a distance r from the body, the field will be the same. (Note: Remember that, unlike the barricade, the Gaussian surface isn't a physical thing, so it won't affect the gravitational field. It is just a mathematical tool that allows us to take advantage of what the field already looks like.)

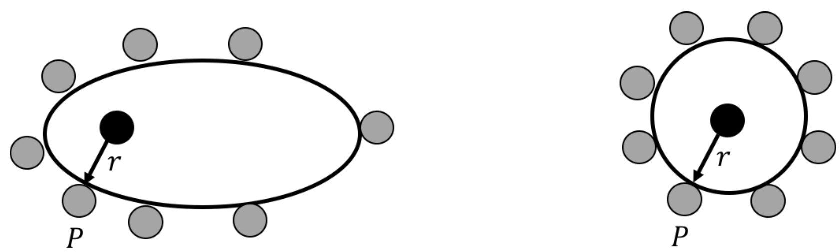


Figure 9.10: A celebrity (black dot) attracts fans (grey dots). A photographer (dot labelled "P") stands behind the barricade a distance r away. This shows two possible barricades we could build around the celebrity. The density of the fans is not uniform for the barricade on the left, so we would not choose that shape to evaluate the Gaussian integral.

We can also use Gauss' Law to determine the gravitational field **inside** of the body of mass M and radius R. This is illustrated in Figure 9.11, which shows a spherical Gaussian surface of radius r that is **inside** of the body of mass M.

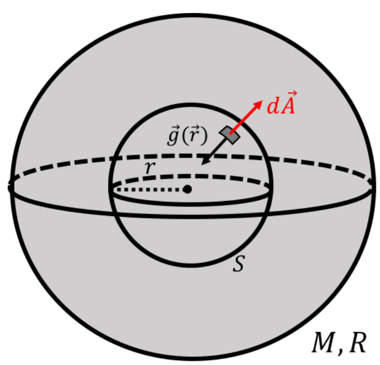


Figure 9.11: Example of a spherical Gaussian surface, S, of radius r centred inside a body of mass M and radius R.

The gravitational field inside of the body of mass M is also symmetric and constant in magnitude across the whole surface, so that the integral is the same as before:

$$\oint \vec{g}(\vec{r}) \cdot d\vec{A} = -g(r)(4\pi r^2)$$

However, in order to use Gauss' Law, we need to determine the mass of the body that is enclosed within the spherical surface, which will be less than M. If we assume that the mass density, ρ, of the object is constant (the body is made of a uniform material), then the density is simply the mass of the object over its volume:

$$\rho = \frac{M}{\frac{4}{3}\pi R^3}$$

The amount of mass enclosed by the spherical surface of radius r is the density multiplied by the volume of a sphere of radius r:

$$M^{enc} = \rho\frac{4}{3}\pi r^3 = M\frac{r^3}{R^3}$$

Applying Gauss' Law, we can now find the magnitude of the gravitational field inside of the spherical body at a distance r from the centre:

$$\oint \vec{g}(\vec{r}) \cdot d\vec{A} = 4\pi G M^{enc}$$
$$-g(r)(4\pi r^2) = 4\pi G M\frac{r^3}{R^3}$$
$$\therefore g(r) = -\frac{GM}{R^3}r$$

And we find that, inside a uniform spherical body of mass M, the gravitational field increases linearly with radius as one moves out from the centre. At the centre of the body, the gravitational field is zero.

Checkpoint 9-9

What can you say about the magnitude of the gravitational field inside a spherical **shell** of mass M?

A) It increases as you move out from the centre of the spherical shell.
B) It decreases as you move out from the centre of the spherical shell.
C) It is equal to zero.
D) It is nonzero and constant in magnitude.

9.3 Gravitational potential energy

Consider a large spherical body of mass M with a coordinate system whose origin coincides with the centre of the spherical body (for example, the large body could be the Earth). The force, $\vec{F}(\vec{r})$ on a body of mass m (for example, a satellite), located at a position $\vec{r}$ is then given by:

$$\vec{F}(\vec{r}) = -G\frac{Mm}{r^2}\hat{r} = -G\frac{Mm}{r^3}\vec{r}$$

where in the second equality, we use the fact that the unit vector in the direction of $\vec{r}$ is simply the vector $\vec{r}$ divided by its magnitude. We can write the force out in Cartesian coordinates:

$$\begin{aligned}\vec{r} &= x\hat{x} + y\hat{y} + z\hat{z}\\ r &= \sqrt{x^2+y^2+z^2} = (x^2+y^2+z^2)^{\frac{1}{2}}\\ \therefore \vec{F}(x,y,z) &= -G\frac{Mm}{(x^2+y^2+z^2)^{\frac{3}{2}}}(x\hat{x} + y\hat{y} + z\hat{z})\end{aligned}$$

Mathematically, this is equivalent to the force that we considered in Example 8-2 of Chapter 8, which we showed was a conservative force. The force of gravity in Newton's theory is thus a conservative force, for which we can determine a potential energy function.

In order to determine the gravitational potential energy function for the mass m in the presence of a mass M, we calculate the work done by the force of gravity on the mass m over a path where the integral for work will be "easy" to evaluate, namely a straight line. Figure 9.12 shows such a path in the radial direction, r, over which it will be easy to calculate the work done by the force of gravity from mass M when mass m moves from being a distance r_A to a distance r_B from the centre of mass M.

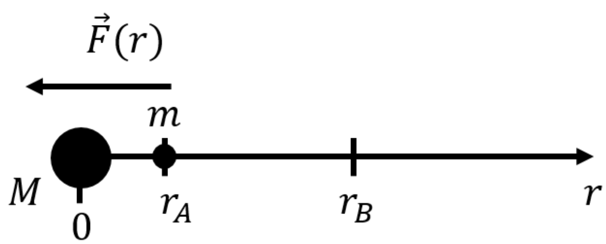


Figure 9.12: Calculating the work done on a mass m by the force of gravity exerted by mass M when mass m moves from a distance r_A to a distance r_B from the center of mass M.

The work done by the force of gravity on m in going from r_A to r_B is given by:

$$\begin{aligned}W &= \int_{r_A}^{r_B} \vec{F}(r)\cdot d\vec{r} = \int_{r_A}^{r_B}\left(-G\frac{Mm}{r^2}\hat{r}\right)\cdot d\vec{r} = \int_{r_A}^{r_B} -G\frac{Mm}{r^2}dr\\ &= \left[G\frac{Mm}{r}\right]_{r_A}^{r_B} = G\frac{Mm}{r_B} - G\frac{Mm}{r_A}\end{aligned}$$

The difference in potential energy in going from position A to position B is given by the negative of the work done by the force:

$$\Delta U = U(r_B) - U(r_A) = -W = G\frac{Mm}{r_A} - G\frac{Mm}{r_B}$$

By inspection, we can identify the potential energy function for gravity:

$$\boxed{U(r) = -G\frac{Mm}{r} + C} \tag{9.3}$$

which is determined only up to a constant, C.

A particularly useful choice of constant is $C = 0$. This corresponds to choosing the potential energy to be zero only when r goes to infinity. That is, the potential energy of mass m is zero only when it is infinitely far away from mass M. The choice of constant C corresponds to the (arbitrary) value of the potential energy when mass m is infinitely far from mass M. When mass m is not infinitely far away, it has **negative** potential energy (if $C = 0$). This is not a problem! Remember, the only thing that is meaningful is a difference in potential energy, so the specific value of the potential energy has no meaning. The kinetic energy of an object, on the other hand, has to be positive.

Recall that if there are no other forces acting on an object, that object will move in such a way to reduce its potential energy. If the object of mass m is located at some distance r from the object of mass M, the force of gravity will attract m so that r decreases. As r decreases in magnitude, the potential energy becomes more negative (larger in magnitude, but further away from zero), and the potential energy of m will indeed decrease as it accelerates due to the force of gravity.

9.3.1 Mechanical energy with gravity

Unless noted otherwise, we will continue our discussion of gravitational potential energy with the particular choice of constant $C = 0$:

$$\boxed{U(r) = -G\frac{Mm}{r}} \tag{9.4}$$

Furthermore, we will assume that M is a large body, such as the Earth, which we can consider as fixed, and focus our discussion on describing the motion of mass m (e.g. a satellite). If M is much bigger than m, they will both experience a force of gravity from each other of the same magnitude (Newton's Third Law), but because M is so much larger, its acceleration will be much smaller (Newton's Second Law). Thus, it is a good approximation to assume that M is stationary and that only m moves when $M >> m$.

We can define the total mechanical energy of mass m when it has a speed v (relative to M) and is located at a distance r from the centre of mass M:

$$E = U + K = -G\frac{Mm}{r} + \frac{1}{2}mv^2$$

where the kinetic energy term is always positive. If gravity is the only force exerted on mass m, then the mechanical energy, E, as defined above, will be a constant. The mechanical energy of an object can give us insight into the possible motion of the object.

Imagine launching a rocket straight upwards from the surface of the Earth; once all of the fuel has burnt up, the rocket's mechanical energy becomes constant as the rocket engine stops doing work on the rocket. As soon as the engine stops providing thrust, the rocket will start to slow down as the force of gravity attracts the rocket back to Earth. If the rocket is going fast enough, it will be able to completely escape the Earth's gravitational pull and travel to infinity (we assume that there are no other planets or the Sun, just the Earth exists!). If, on the other hand, the rocket's speed is too low, it will eventually stop and fall back to Earth. This is the same thing that happens to you when you try to jump vertically. If you could jump hard enough, you would be able to escape the Earth's gravitational pull!

In terms of mechanical energy, we can ask ourselves if the mechanical energy of the rocket is large enough to escape the Earth's gravitational pull. Specifically, we can ask ourselves what the value of the rocket's kinetic energy would be when it reaches infinity. The kinetic energy of the rocket is given by:

$$K = E - U$$

If the rocket is infinitely far from the Earth, then its potential energy is zero, and the kinetic energy is equal to E.

If the mechanical energy, E, is negative, it is not possible for the rocket to ever make it to infinity because its kinetic energy would have to be negative. In other words, if the mechanical energy is negative, then the object of mass m can never escape the gravitational pull of object M. We say that m is "gravitationally bound" to M.

If the mechanical energy, E, is exactly zero, then the object's kinetic energy will become zero just as it reaches infinity. In other words, it will just barely be able to escape the gravitational pull from mass M. The condition for this to happen is:

$$\begin{aligned}
E &= 0 \\
K &= -U \\
\frac{1}{2}mv^2 &= G\frac{Mm}{r} \\
\therefore v_{esc} &= \sqrt{\frac{2GM}{r}}
\end{aligned}$$

which we can interpret as a condition for the speed of the rocket. If at some distance r from M, the rocket has the speed given by the condition above, then it will have enough kinetic energy to escape the gravitational pull of M. We call this speed the "escape velocity".

Finally, if the mechanical energy is greater than zero, then the rocket will have enough energy to escape the gravitational pull of M and have a non-zero speed when it reaches infinity.

Checkpoint 9-10

What is the escape velocity from the surface of the Earth?

A) $4.29 \times 10^6\,\text{km/s}$
B) $1.25 \times 10^5\,\text{km/s}$
C) $11.2\,\text{km/s}$
D) $9.81\,\text{km/s}$

Example 9-4

Show that an object of mass m in a circular orbit of radius r around a body of mass M has half of the kinetic energy required to escape the gravitational pull of M.

Solution

The only force acting on the object is gravity, so it has a mechanical energy given by:

$$E = U + K$$
$$E = -G\frac{Mm}{r} + \frac{1}{2}mv^2$$

In order for the object to just escape the gravitational pull of M, it's mechanical energy must be equal to zero:

$$E = 0$$
$$\therefore K_{esc} = -U$$

Since the object is in a circular orbit, we can use Newton's Second Law to find an expression for v^2:

$$F_{net} = \frac{mv^2}{r}$$
$$\frac{GMm}{r^2} = \frac{mv^2}{r}$$
$$\frac{GM}{r} = v^2$$

where in the second line we used the fact that F_{net} is equal to the force of gravity exerted by M on the object. The kinetic energy of the object is thus:

$$K = \frac{1}{2}mv^2$$
$$K = \frac{1}{2}\frac{GMm}{r}$$

You will notice that this is very similar to our expression for U. In fact, we have:

$$K = -\frac{1}{2}U$$
$$\therefore K = \frac{1}{2}K_{esc}$$

Note: We can also see that the velocity of an object in a circular orbit is equal to $\sqrt{GM/r}$, which is half the escape velocity, $v_{esc} = \sqrt{2GM/r}$

Types of orbits

The mechanical energy of a body of mass m determines whether it is gravitationally bound to (i.e. cannot escape) the body of mass M. The path (orbit) that m will take depends on its velocity with respect to M. Clearly, if the velocity of m is directed at the centre of M, then m will just collide with M. In all other cases, the orbit that m will take depends on the mechanical energy of m as well as the speed of m at the point of closest approach to M (see Figure 9.13). The velocity of m at the point of closest approach will always be perpendicular to the line joining the centres of m and M. The different possible orbits are:

1. A **circular orbit** of radius R (where R is the distance of closest approach) if the **mechanical energy is negative** (i.e. it is bound) and the speed is exactly equal to the value necessary for the gravitational force to provide the required centripetal acceleration for uniform circular motion:

$$\sum F = G\frac{Mm}{R^2} = m\frac{v^2}{R}$$
$$\therefore v_{circ} = \sqrt{\frac{GM}{R}}$$

2. An **elliptical orbit** if the **mechanical energy is negative** and the speed at the point of closest approach is different than that required for a circular orbit.
3. A **parabolic orbit** if the **mechanical energy is exactly zero**.
4. A **hyperbolic orbit** if the **mechanical energy is bigger than zero**.

The possible orbits are illustrated in Figure 9.13, and are curves in the family of "conic sections", as they can be found by the intersection of a plane and a cone. All conic sections have at least one "focus" point (ellipses have two) that corresponds to the location of M.

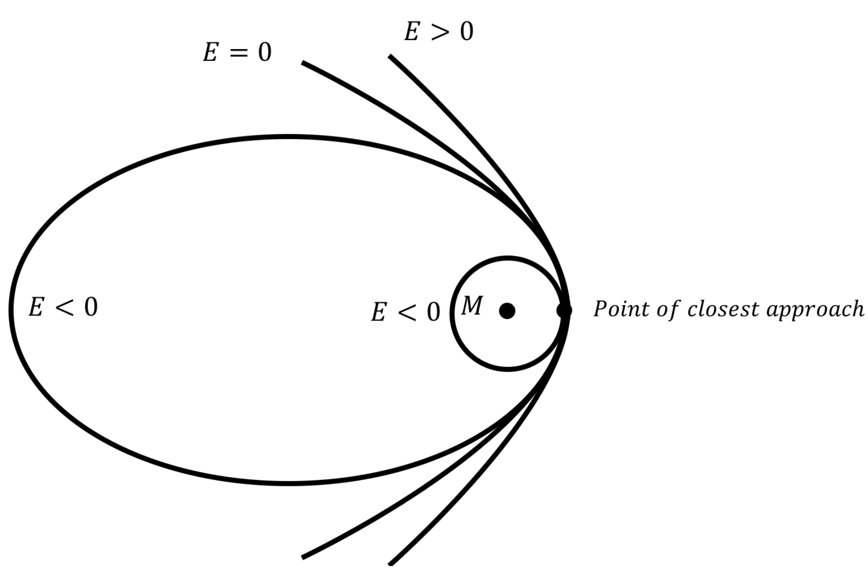


Figure 9.13: The different possible orbits of m due to the gravitational force of M depend on the mechanical energy, E, of m. The orbits are drawn in a frame of reference where M is at rest.

9.4 Einstein's Theory of General Relativity

Newton's Universal Theory of Gravity was extremely successful at describing the motion of planets in the solar system, and allowed for high precision astronomy. For example, precision measurements of Uranus's orbit showed that it appeared to be inconsistent with Newton's theory, unless the gravitational influence of another planet was included in the model. This led to the discovery of the planet Neptune.

However, some issues with Newton's theory were uncovered. The orbit of Mercury was shown to be different than what Newton's theory could describe, but searches for another planet (Vulcan) were unsuccessful. In addition, Albert Einstein's theory of Special Relativity, published in 1905, was found to be incompatible with Newton's theory of gravity. One of the consequences of Special Relativity is that nothing can propagate faster than the speed of light. Newton's Universal Theory of Gravity implies that the gravitational force is transmitted instantaneously. In Newton's theory, if the Sun suddenly disappeared, Earth would immediately "fall out" of its orbit, and we would immediately know that the Sun has disappeared. This would violate Special Relativity because there cannot be a mechanism that would allow us to know that the Sun has disappeared faster than it would take light to propagate from the Sun. In other words, for the 8 min that are required for light to travel from the Sun to the Earth, we cannot know that the Sun has disappeared: only when we literally see the Sun disappear would the Earth be "allowed" to fall out of its orbit.

Einstein's Theory of General Relativity is a theory developed by Einstein in order to describe gravity in a way that is consistent with Special Relativity and the propagation of light. Einstein was famous for his "thought experiments," which allow us to think about some of the implications of a theory, even if the experiments would be very difficult to carry out in practice. One such thought experiment is to consider what someone would observe in an accelerating frame of reference.

Consider an observer in an elevator, as illustrated in Figure 9.14. If the elevator is stationary at the surface of the Earth (left panel), and the observer is standing on a scale, they could measure their weight, mg, on the scale. The two forces on the observer are their weight and the normal force, which would be equal in magnitude since the observer is not accelerating. The normal force, read out by the scale, would thus correspond to their weight. To be more precise, the normal force would be equal to $m_G g$, where m_G is the gravitational mass of the observer (that mass which is related to the force of gravity experienced by a mass).

If the elevator was instead placed in empty space, and the elevator was accelerating upwards with an acceleration of g (right panel), the observer would still be able to measure their weight by stepping on the scale. The only force on the observer is the normal force from the scale, which must be equal to its mass times their acceleration $N = ma = mg$, since the observer is accelerating with the elevator. In this case, it is the inertial mass of the observer, m_I, that comes into play, so the normal force read on the scale is $m_I g$.

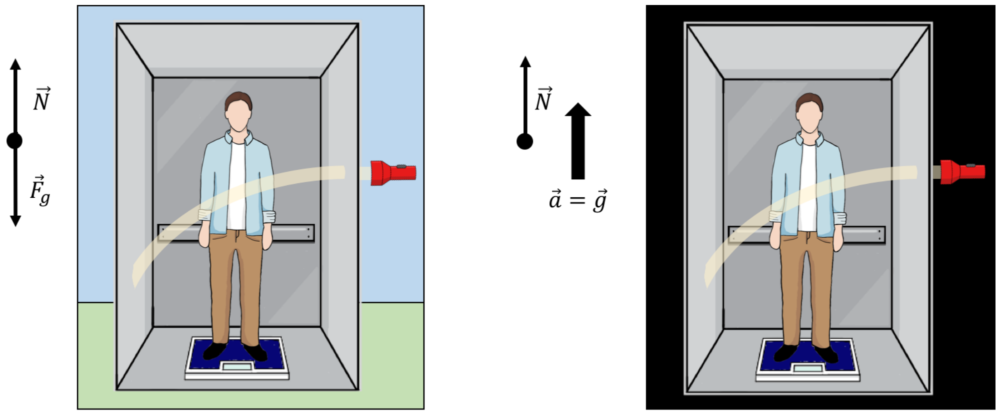


Figure 9.14: Left: A person standing on a scale in an elevator at rest at the surface of the Earth. Right: A person in an elevator that is accelerating in empty space with the same acceleration as that due to gravity at the Earth's surface. The curvature of the light beam is exaggerated.

Einstein postulated that it would be impossible for the observer to distinguish whether they are at rest on the surface of the Earth, or in empty space accelerating with an acceleration of g. In other words, he postulated that the inertial and gravitational masses are exactly equivalent. This is what is called the "Equivalence Principle".

This simple statement has dramatic implications. Special Relativity requires that light will travel in a straight line in empty space. If a beam of light enters and then exits the elevator, the observer on Earth and the one accelerating in empty space must observe the same thing, since they cannot distinguish between being on Earth or accelerating in space. The observer in space, who is accelerating, will observe that the beam of light bends as it crosses the elevator (the beam travels in a straight line as observed in an inertial reference frame, so the person in the accelerating elevator would see it follow a parabolic path). The observer on Earth must thus observe the same thing, namely that light will follow a curved path in the presence of a gravitational field.

But...light must travel in a straight line in empty space. That means that if the path of a beam of light is curved near Earth, it must be because space itself is curved in the presence of a gravitational field! In other words, Einstein's Theory of General Relativity describes how the presence of mass (or energy) results in a curvature of space (and time).

Imagine a ladybug on the side of a basketball. If the ladybug starts moving in what it believes to be a straight line, it will actually move in a curved path along the surface of the ball, as in Figure 9.15. This is like the curved path of light that we observe. If we didn't know the ball was there, we would just think that the bug was moving along a curved path. In the same way, if an observer is not aware of the curvature of spacetime, it appears that light follows a curved path.

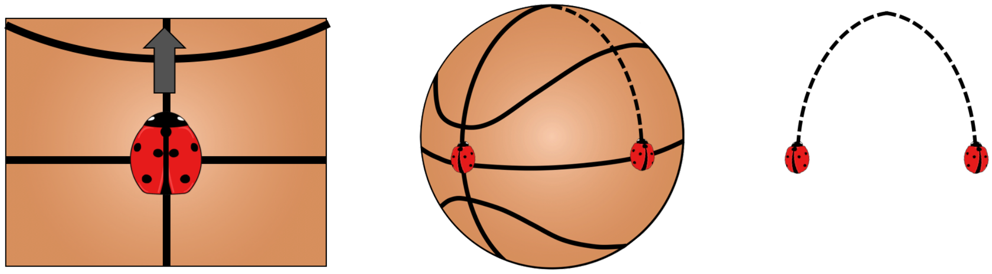


Figure 9.15: Left: A ladybug perceives itself to be moving in a straight line. Center: The basketball is curved, so the ladybugs follow curved paths. Right: What an observer would see if they didn't know the basketball was there.

Now imagine there's a second ladybug. Both bugs start at the middle of the ball and start moving towards the top of the ball in what they think is a straight line (as shown in the center panel of Figure 9.15). When the bugs start moving, they are parallel to each other, so if the ball was not curved, the ladybugs would never meet. However, because it is curved, the ladybugs will eventually cross paths. If you were not aware that the ball was there, you would have to conclude that there was some force attracting the bugs to each other, just like if you were unaware that spacetime was curved, you would conclude that massive bodies moving towards each other are attracted by a gravitational force.

Objects that are moving in a gravitational field are actually following Newton's First Law (they are moving at constant velocity in a straight line and no force is exerted on them). It is strange and unexpected, but high precision measurements confirm that this correctly describes everything that we have measured!

Einstein's theory was able to describe the orbit of Mercury, and the prediction that gravity leads to light following a curved path was confirmed by Eddington within five years of Einstein's theory being published. Another implication of the theory is that time goes by slower in the presence of a gravitational field. Clocks on Earth run slower than clocks in orbit (where the gravitational field is weaker). This effect is taken into account when using GPS to determine your position on Earth, since this is based on comparing the time that it takes signals to arrive to your position on Earth from different satellites. This is also somewhat reasonably well described in the movie "Interstallar", where time is seen to pass much slower for a set of astronauts in the vicinity of a black hole, where the gravitational field is strong.

9.5 Summary

Key Takeaways

Kepler was the first to synthesize a large amount of data to quantitatively describe gravity with his three laws:

1. The path of a planet around the Sun is described by an ellipse with the Sun at once of its foci.
2. Planets move in such a way that the area swept by a line connecting the planet and the Sun in a given period of time is constant, independent of the location of the planet.
3. The ratio between the orbital periods, T, squared of two planets is equal to the ratio of the semi-major axes, s, of their orbits cubed:

$$\left(\frac{T_1}{T_2}\right)^2 = \left(\frac{s_1}{s_2}\right)^3$$

Newton described the attractive force of gravity exerted between two bodies of mass M_1 and M_2 (which must be point masses) as:

$$\vec{F}_{12} = -G\frac{M_1 M_2}{r^2}\hat{r}_{21}$$

where $\vec{F}_{12}$ is the force on body 1 from body 2, r is the distance between the two bodies, and $\vec{r}_{21}$ is the vector from body 2 to body 1. The motion of a body under the influence of only this force will satisfy all of Kepler's Laws, if the body is gravitationally bound.

The gravitational field, $\vec{g}(\vec{r})$, from a body of mass M, is defined as the gravitational force that another body would experience per unit mass:

$$\vec{g}(\vec{r}) = \frac{\vec{F}(\vec{r})}{m} = -G\frac{M}{r^2}\hat{r}$$

The field can be used to determine the corresponding gravitational force, $\vec{F}_g$, that a body of mass m would experience if located at a position $\vec{r}$ relative to the body of mass M:

$$F_g = m\vec{g}(\vec{r})$$

When describing the motion of objects near the surface of the Earth, it is thus more precise to refer to $g = 9.8\,\mathrm{N/kg}$ as the magnitude of the Earth's gravitational field at the surface of the Earth, then to refer to $g = 9.8\,\mathrm{m/s^2}$ as the acceleration due to Earth's gravity. The two are only equal if gravitational mass (the m in the above equation) and inertial mass (the m in Newton's Second Law) are the same.

Gauss' Law, which applies to all inverse-square force laws, can be used to determine the magnitude of the gravitational field from a body of mass M, even it if it not a point mass:

$$\oint \vec{g}(\vec{r}) \cdot d\vec{A} = 4\pi G M^{enc}$$

Since the force described by Newton's theory is conservative, we can define a potential energy function. The gravitational potential energy of a mass m located a distance r away from a mass M is:

$$U(r) = -G\frac{Mm}{r} + C$$

A convenient choice of the constant is $C = 0$, as this corresponds to the gravitational potential energy being equal to zero when m is infinitely far away from M.

The mechanical energy, E, of an object of mass m that is located at a distance r from an object of mass M, if gravity is the only conservative force exerted on m, is given by:

$$E = K + U = \frac{1}{2}mv^2 - G\frac{Mm}{r}$$

where we have explicitly chosen $C = 0$, and v is the speed of m relative M (considered to be at rest). Furthermore, if no non-conservative forces do work on the body of mass m, the mechanical energy, E, is constant.

If the mechanical energy of m is negative, it is gravitationally bound to M. Depending on the mechanical energy of m and its velocity at the point of closest approach to M, the orbit of m will be described by one of four conic sections (circle, ellipse, parabola, hyperbola).

Einstein's Theory of General Relativity describes gravitation as the bending of space and time caused by the presence of mass and energy. In Einstein's theory, objects follow straight (inertial) paths and do not feel a force of gravity. The curvature of space is what results in their apparent motion not being a straight line. Einstein's theory is based on the Equivalence Principle (inertial and gravitational mass are exactly equal) and the properties of how light propagates according to the Theory of Special Relativity.

Important Equations

Kepler's Third Law:

$$\left(\frac{T_1}{T_2}\right)^2 = \left(\frac{s_1}{s_2}\right)^3$$

Gravitational force and gravitational field:

$$\vec{F}_{12} = -G\frac{M_1 M_2}{r^2}\hat{r}_{21}$$
$$\vec{g}(\vec{r}) = -G\frac{M}{r^2}\hat{r}$$
$$F_g = m\vec{g}(\vec{r})$$

Gauss's Law:

$$\oint \vec{g}(\vec{r}) \cdot d\vec{A} = 4\pi G M^{enc}$$

Gravitational potential energy and mechanical energy:

$$U(r) = -G\frac{Mm}{r} + C$$
$$E = K + U = \frac{1}{2}mv^2 - G\frac{Mm}{r}$$

9.6 Thinking about the material

Reflect and research

1. When you look at the night sky, how can you tell the difference between a planet and a star?
2. What was the relationship between Tycho Brahe and Johannes Kepler?
3. How did Tycho Brahe collect all the data that Kepler used?
4. How much time elapsed between Kepler publishing his laws and Newton publishing his Universal Theory of Gravity?
5. What was Kepler's original intention when he synthesized Tycho Brahe's observations? What was he hoping to show?
6. What was Ptolemy's theory of gravity based upon?
7. Who was the first to suggest that planets revolved around the Sun instead of the Earth?
8. Explain how the force of gravity from the moon results in tides on both sides of the Earth.
9. Explain what an L1 Lagrange point is, and how it does not violate Kepler's Third Law.
10. How did Eddington confirm that light follows a curved path in a gravitational field?

To try in the lab

1. Theory project: Prove, based on Newton's Universal Theory of Gravity, that the motion of orbiting bodies is given by a conic section.
2. Write a computer simulation to plot the orbit of two bodies, and explore how the total mechanical energy of one object affects its motion. If the two bodies have the same mass, and both move, where is the focus of the conical section describing their respective paths?

9.7 Sample problems and solutions

9.7.1 Problems

Problem 9-1: Geosynchronous satellites are satellites that are placed in a circular orbit around the Earth in such a way that their orbital period is synchronized with the 24 h rotation period of the Earth. The advantage of geosynchronous satellites is that they are always above the same point on Earth, which makes them useful for establishing communication networks. At what altitude must geosynchronous satellites be placed? (Solution)

Problem 9-2: How much energy must be expended in order to place a satellite of mass $m = 1000\,\mathrm{kg}$ in a geosynchronous circular orbit around the Earth, if the satellite is launched from the North Pole of the Earth? How much energy is this per kilogram of satellite placed in orbit? (Solution)

Problem 9-3: Find an expression for the gravitational field due to a thin uniform rod of mass M at point P, which is a distance h above the midsection of the rod (Figure 9.16).

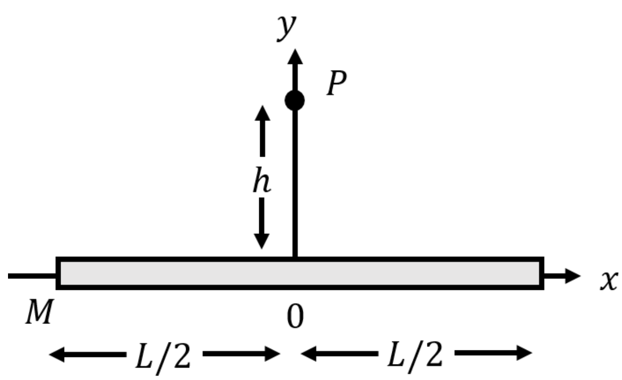


Figure 9.16: A thin rod of mass M and length L produces a gravitational field at a point P located above the midsection of the rod.

(Solution)

9.7.2 Solutions

Solution to problem 9-1: When a satellite orbits the Earth, the only force on the satellite is the force of gravity from the Earth. Since the satellite is in a circular orbit, that force of gravity must point towards the centre of the Earth in such a way that the satellite has the correct radial acceleration, a_R, to stay in uniform circular motion:

$$a_r = \frac{v^2}{R}$$

where v is the speed of the satellite, and R is the distance between the satellite and the centre of the Earth (i.e. the centre of the circular orbit). The magnitude of the force of gravity on the satellite of mass m is given by:

$$F = G\frac{Mm}{R^2}$$

where M is the mass of the Earth. Newton's Second Law applied to the satellite is:

$$\begin{aligned}\sum F_r = F &= ma_r \\ \therefore G\frac{Mm}{R^2} &= m\frac{v^2}{R}\end{aligned}$$

The speed of the satellite can be found from the fact that it must travel a distance of $2\pi R$ (the circumference of the orbit) in a period $T = 24\,\text{h}$:

$$v = \frac{2\pi R}{T}$$

which we can substitute into the equation from Newton's Second Law to find the distance R (i.e. the radius of the circular orbit):

$$\begin{aligned}G\frac{Mm}{R^2} &= m\frac{v^2}{R} \\ G\frac{M}{R^2} &= \frac{(2\pi R)^2}{T^2 R} \\ G\frac{M}{R^2} &= \frac{4\pi^2 R}{T^2} \\ \therefore R &= \sqrt[3]{G\frac{MT^2}{4\pi^2}} \\ &= \sqrt[3]{(6.67\times 10^{-11}\,\text{Nm}^2/\text{kg}^2)\frac{(5.97\times 10^{24}\,\text{kg})(86\,400\,\text{s})^2}{4\pi^2}} \\ &= 42.2\times 10^6\,\text{m}\end{aligned}$$

which corresponds to the distance between the satellite and the centre of the Earth. To obtain the "altitude", h, namely the distance from the surface of the Earth to the satellite, we must subtract the radius of the Earth, $R_\oplus = 6.371\times 10^6\,\text{m}$ from this distance:

$$h = R - R_\oplus = 35.9\times 10^6\,\text{m}$$

Thus, geosynchronous satellites are located at an altitude of approximately 36 000 km.

Discussion: Note that we could have also easily used Kepler's Third Law to determine the radius of the orbit, since we already know the period (24 h), and we know the value of the constant for Kepler's Third Law from Example 9-2.

Solution to problem 9-2: We need to calculate how much work must be done for the satellite to go from being at rest at the surface of the Earth to being in a geosynchronous orbit. That work will be done by a non-conservative force (a rocket engine). The work done by the non-conservative force, W, is equal to the satellite's change in mechanical energy:

$$W = \Delta E = E_B - E_A$$

The initial mechanical energy of the satellite, E_A, is given by its gravitational potential energy (it has no kinetic energy at the surface of the Earth when at the North Pole - on the equator, it would have kinetic energy due to the Earth's rotation):

$$E_A = K + U = 0 - G\frac{Mm}{R_\oplus}$$

where $M = 5.97 \times 10^{24}$ kg is the mass of the Earth, and $R_\oplus = 6.731 \times 10^6$ m is the radius of the Earth.

In orbit, the energy of the rocket, E_B, is given by:

$$E_B = K + U = \frac{1}{2}mv^2 - G\frac{Mm}{R}$$

where $R = 42.2 \times 10^6$ m is the radius of the geosynchronous orbit (Problem 9-1) and v is the speed of the satellite in orbit. The speed is given by:

$$v = \frac{2\pi R}{T}$$

where $T = 24$ h is the orbital period. The net work that must be done to place the satellite in orbit is thus given by:

$$\begin{aligned}
W = E_B - E_A &= \frac{1}{2}mv^2 - G\frac{Mm}{R} - \left(-G\frac{Mm}{R_\oplus}\right)\\
&= \frac{1}{2}m\frac{4\pi^2R^2}{T^2} + GMm\left(\frac{1}{R_\oplus} - \frac{1}{R}\right)\\
&= \frac{1}{2}(1000\,\text{kg})\frac{4\pi^2(42.2\times10^6\,\text{m})^2}{(86\,400\,\text{s})^2}\\
&+ (6.67\times10^{-11}\,\text{Nm}^2/\text{kg}^2)(5.97\times10^{24}\,\text{kg})(1000\,\text{kg})\left(\frac{1}{(6.731\times10^6\,\text{m})} - \frac{1}{(42.2\times10^6\,\text{m})}\right)\\
&= 5.78\times10^{10}\,\text{J}
\end{aligned}$$

This corresponds to the energy that must be imparted to a 1000 kg satellite for it to end up in a geosynchronous orbit. This corresponds to 5.78×10^7 J/kg as the energy required per kilogram of payload placed in geosynchronous orbit. Although we calculated work as if it were work done by a force, we can think of this work coming from stored chemical potential energy in the fuel of the rocket carrying the satellite.

Discussion: The energy that we found above is the minimum energy that one must provide to the satellite. In practice, in order to place a satellite in orbit, one will also need to provide enough energy to accelerate the rocket that carries the satellite up into orbit, which is typically much heavier than the satellite. If the satellite were instead launched from the equator of the Earth, the satellite would already have some initial kinetic energy due to the rotation of the Earth, and one would need to provide less energy to place it in orbit. This is the reason that most rockets are launched from near the equator (think French Guyana, Florida, Kazakhstan) in a direction that is roughly parallel with the Earth's rotation.

Solution to problem 9-3: We cannot use Gauss's law to determine the magnitude of the field because the gravitational field lacks symmetry (i.e. the field will be different at the ends of the rod than along the length of the rod). The gravitational field due to a body of mass M is given by:

$$\vec{g}(\vec{r}) = -\frac{GM}{r^2}\hat{r}$$

Our strategy will be to break the rod into very small segments of length dx. Each segment, of mass dM, will make a small contribution, $d\vec{g}$, to the gravitational field, as shown in Figure 9.17. We will then take the sum of all these contributions to find the net field.

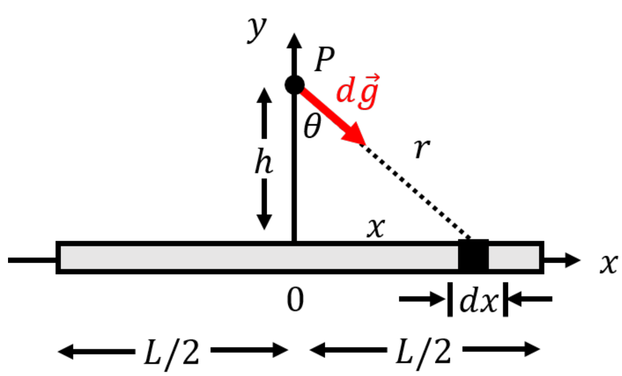


Figure 9.17: A thin rod of mass M and length L produces a gravitational field at a point P located above the midsection of the rod. Each segment of the rod dx will contribute to the gravitational field.

The gravitational field due to each segment is given by:

$$d\vec{g} = -\frac{GdM}{r^2}\hat{r}$$

The element of the field, $d\vec{g}$, will point in a different direction for each segment dx. You can conclude from Figure 9.17 that, due to symmetry, the x components of the field from each segment will cancel out (for the segment dx shown in the diagram, there will be an identical segment on the other side of the rod). The net field will point in the $-\hat{y}$ direction, so we are only interested in the vertical component of $d\vec{g}$. Using our diagram, this means that we want to find the magnitude of $dg\cos\theta$:

$$dg\cos\theta = \frac{GdM}{r^2}\cos\theta$$

The magnitude of the gravitational field at point P is thus given by:

$$g = \int dg\cos\theta = \int \frac{GdM}{r^2}\cos\theta$$

The integral is written over dM, where both r, and θ are different for each different mass element, dM. We need to express any variable that changes for different mass elements in terms of a single variable of integration. We will choose θ as the variable of integration, and thus need to express r and dM in terms of θ, $d\theta$, and other constants.

The distance, r, between P and a mass element dM located at angle θ is easily found to be:

$$\begin{aligned} r &= \frac{h}{\cos\theta} \\ \therefore \frac{1}{r^2} &= \frac{\cos^2\theta}{h^2} \end{aligned}$$

dM can easily be expressed in term of dx (the length of the mass element in the x direction) and λ, the mass per unit length of the rod:

$$dM = \lambda dx = \frac{M}{L}dx$$

We now need to express dx in terms of $d\theta$. This can be found as follows, by first expressing x in terms of θ, and then taking the derivative of x with respect to θ

$$\begin{aligned} x &= h\tan\theta \\ \therefore \frac{dx}{d\theta} &= \frac{h}{\cos^2\theta} \\ \therefore dx &= \frac{h}{\cos^2\theta}d\theta \end{aligned}$$

Now that we have found the small change in x that results from a small change in θ, we can write the mass element, dM, in terms of the $d\theta$:

$$dM = \frac{M}{L}dx = \frac{M}{L}\frac{h}{\cos^2\theta}d\theta$$

We can now write the integral in terms of θ:

$$\begin{aligned}
g = \int \frac{GdM}{r^2}\cos\theta = G\int\frac{1}{r^2}\cos\theta dM \\
= G\int\left(\frac{\cos^2\theta}{h^2}\right)\cos\theta\left(\frac{M}{L}\frac{h}{\cos^2\theta}\right) \\
= \frac{GM}{Lh}\int\cos\theta d\theta
\end{aligned}$$

Now that we have the integral over θ, we need to set the limits to correspond to the values of θ at each end of the rod. The angle will have the same magnitude for each end of the rod, θ_0, given by:

$$\sin\theta_0 = \frac{L}{2\sqrt{h^2+\frac{L^2}{4}}}$$

The magnitude of the field is thus given by:

$$\begin{aligned}
g = \frac{GM}{Lh}\int_{-\theta_0}^{\theta_0}\cos\theta d\theta = \frac{GM}{Lh}[\sin\theta]_{-\theta_0}^{\theta_0} \\
= \frac{2GM}{Lh}\sin\theta_0 = \frac{2GM}{Lh}\frac{L}{2\sqrt{h^2+\frac{L^2}{4}}}
\end{aligned}$$

The gravitational field at point P is thus given by:

$$\vec{g} = -\frac{2GM}{Lh}\frac{L}{2\sqrt{h^2+\frac{L^2}{4}}}\hat{y}$$

10 Linear momentum and the centre of mass

In this chapter, we introduce the concepts of linear momentum and of centre of mass. Momentum is a quantity that, like energy, can be defined from Newton's Second Law, to facilitate building models. Since momentum is often a conserved quantity within a system, it can make calculations much easier than using forces. The concepts of momentum and of centre of mass will also allow us to apply Newton's Second Law to systems comprised of multiple particles including solid objects.

Learning Objectives

- Understand how to calculate linear momentum.
- Understand how to calculate impulse and that it corresponds to a change in momentum.
- Understand when and how to apply conservation of linear momentum to model situations.
- Understand the difference between elastic and inelastic collisions, and when mechanical energy is conserved.
- Understand how to calculate the centre of mass of an object.

Think About It

You hit a pool ball square on with the cue ball. If both balls have the same mass, and you can neglect any "english" on the cue ball, what happens to the cue ball?

A) It stops.
B) It continues, with half of its original speed.
C) It continues, with its original speed.
D) It rebounds, with its original speed.

10.1 Momentum

10.1.1 Momentum of a point particle

We can define the momentum, $\vec{p}$, of a particle of mass m and velocity $\vec{v}$ as the vector quantity:

$$\boxed{\vec{p} = m\vec{v}} \tag{10.1}$$

Since this is a vector equation, it corresponds to three equations, one for each component of the momentum vector. It should be noted that the numerical value for the momentum of a particle is arbitrary, as it depends in which frame of reference the velocity of the particle is defined. For example, your velocity with respect to the surface of the Earth is zero, so your momentum relative to the surface of the Earth is zero. However, relative to the surface of the Sun, your velocity, and momentum, are not zero. As we will see, forces are related to a *changes* in momentum, just as they are related to a change in velocity (acceleration).

If the particle has a constant mass, then the time derivative of its momentum is given by:

$$\frac{d}{dt}\vec{p} = \frac{d}{dt}m\vec{v} = m\frac{d}{dt}\vec{v} = m\vec{a}$$

and we can write this as Newton's Second Law, since $m\vec{a}$ must be equal to the vector sum of the forces on the particle of mass m:

$$\boxed{\frac{d}{dt}\vec{p} = \sum \vec{F} = \vec{F}^{net}} \tag{10.2}$$

The equation above is the original form in which Newton first developed his theory. It says that the net force on an object is equal to the rate of change of its momentum. **If the net force on the object is zero, then its momentum is constant** (as is its velocity). In terms of components, Newton's Second Law written for the rate of change of momentum is given by:

$$\begin{aligned}
\frac{dp_x}{dt} &= \sum F_x \\
\frac{dp_y}{dt} &= \sum F_y \\
\frac{dp_z}{dt} &= \sum F_z
\end{aligned}$$

Example 10-1

A particle of mass m is released from rest and allowed to fall freely under the influence of gravity near the Earth's surface (assume that drag is negligible). Is the mechanical energy of the particle conserved? Is the momentum of the particle conserved? If momentum is not conserved, how does momentum change with time? Do your answers change if the force of drag cannot be ignored?

Solution

First, we model the falling particle assuming that there is no force of drag. The only force exerted on the particle is thus its weight.

The mechanical energy of the particle will be conserved only if there are no non-conservative forces doing work on the particle. Since the force of gravity is the only force acting on the particle, its mechanical energy is conserved.

The total momentum of the particle is not conserved, because the sum of the forces on the particle is not zero. Choosing the z axis to be vertical and positive upwards, Newton's Second Law in the z direction is given by:

$$\sum F_z = -mg = \frac{dp_z}{dt}$$

Note that the x and y components of momentum are conserved, since there are no forces with components in that direction. We can find how the z component of the momentum changes with time by taking the anti-derivative of the force with respect to time (from $t = 0$ to $t = T$):

$$\begin{aligned}
\frac{dp_z}{dt} &= -mg \\
\int dp_z &= \int_0^T (-mg)dt \\
p_z(T) - p_z(0) &= -mgT \\
\therefore p_z(T) &= p_z(0) - mgT
\end{aligned}$$

where the z component of momentum, $p_z(T)$ at some time T, is given by its value at time $t = 0$ plus $-mgT$. If the object started at rest ($\vec{v} = 0$), then the magnitude of the momentum, as a function of time, is given by:

$$p(t) = p_z(t) = -mgt$$

and indeed changes with time.

If the force of drag were not negligible, there would be a non-conservative force acting on the particle, so its mechanical energy would no longer be conserved. The particle will accelerate until it reaches terminal velocity. During that phase of acceleration, the net force on the particle is not zero (it is accelerating), so its momentum is not conserved. Once the particle reaches terminal velocity, the net force on the particle is zero, and its momentum is conserved from then on.

Discussion: This simple example highlights the fact that mechanical energy and momentum are conserved under different conditions. Just because one is conserved does not mean that the other is conserved. It also shows that Newton's Second Law is a statement about change in momentum, not momentum itself (just like it is a statement about acceleration, change in velocity, not velocity).

10.1.2 Impulse

When we introduced the concept of energy, we started by calculating the "work", W, done by a force exerted on an object over a specific path between two points:

$$W = \int_A^B \vec F \cdot d\vec l$$

We then introduced kinetic energy, K, to be that quantity whose change is equal to the net work done on the particle

$$W^{net} = \int_A^B \vec F^{net} \cdot d\vec l = \Delta K$$

where the net force, $\vec F^{net}$, is the vector sum of the forces on the particle.

We can do the same thing, but instead of integrating the force over distance, we can integrate it over time. We thus introduce the concept of "impulse", $\vec J$, of a force, as that force integrated from an initial time, t_A, to a final time, t_B:

$$\vec J = \int_{t_A}^{t_B} \vec F dt \tag{10.3}$$

where it should be clear that impulse is a vector quantity (and the above vector equation thus corresponds to one integral per component). Impulse is, in general, defined as an integral because the force, $\vec F$, could change with time. If the force is constant in time (magnitude and direction), then we can define the impulse without using an integral:

$$\vec J = \vec F \Delta t$$

where Δt is the amount of time over which the force was exerted. Although the force might never be constant, we can sometimes use the above formula to calculate impulse using an average value of the force.

Checkpoint 10-1

What is the SI unit for impulse?

A) $\text{kg} \cdot \text{m/s}^2$
B) $\text{kg} \cdot \text{s}^2$
C) $\text{kg} \cdot \text{m/s}$
D) $\text{kg} \cdot \text{m/s}^3$

Example 10-2

Estimate the impulse that is given to someone's head when they are slapped in the face.

Solution

When we slap someone's face with our hand, our hand exerts a force on their face during the period of time, Δt, over which our hand is in contact with their face. During that period of time, the force on their face goes from being 0, to some unpleasantly high value, and then back to zero, so the force cannot be considered constant.

Let us estimate the average magnitude of the slapping force by considering the deceleration of our slapping hand and modelling the motion as one-dimensional. Let us assume that our slapping hand has a mass $m = 1\,\text{kg}$ and that it is has a speed of $2\,\text{m/s}$ just before it makes contact. Furthermore, let us assume that it is in contact with the face for a period of time Δt. This allows us to find the average acceleration of our hand and thus the average force exerted by the face on our hand to stop it:

$$\begin{aligned} a &= \frac{\Delta v}{\Delta t} \\ \therefore F &= ma = m\frac{\Delta v}{\Delta t} \end{aligned}$$

By Newton's Third Law, the force decelerating our hand has the same magnitude as the force that our hand exerts on the face, allowing us to calculate the impulse given to the person's head:

$$\begin{aligned} J = F\Delta t &= \left(m\frac{\Delta v}{\Delta t}\right)\Delta t = m\Delta v \\ &= (1\,\text{kg})(2\,\text{m/s}) = 2\,\text{kg}\cdot\text{m/s} \end{aligned}$$

Discussion: Note that the impulse given to the head corresponds exactly to the change in momentum of the hand ($\Delta p = m\Delta v$).

So far, we calculated the impulse that is given by a single force. We can also consider the net impulse given to an object by the net force exerted on the object:

$$\vec J^{net} = \int_{t_A}^{t_B} \vec F^{net} dt$$

Compare this to Newton's Second Law written out using momentum:

$$\begin{aligned} \frac{d}{dt}\vec p &= \vec F^{net} \\ \int_{\vec p_A}^{\vec p_B} d\vec p &= \int_{t_A}^{t_B} \vec F^{net} dt \\ \vec p_B - \vec p_A &= \int_{t_A}^{t_B} \vec F^{net} dt \\ \therefore \Delta\vec p &= \int_{t_A}^{t_B} \vec F^{net} dt \end{aligned}$$

and we find that the net impulse received by a particle is precisely equal to its change in momentum:

$$\boxed{\Delta\vec{p} = \vec{J}^{net}} \tag{10.4}$$

This is similar to the statement that the net work done on an object corresponds to its change in kinetic energy, although one should keep in mind that momentum is a vector quantity, unlike kinetic energy.

Example 10-3

A car moving with a speed of 100 km/h collides with a building and comes to a complete stop. The driver and passenger each have a mass of 80 kg. The driver wore a seat belt that extended during the collision, so that the force exerted by the seatbelt on the driver acted for about 2.5 s. The passenger did not wear a seat belt and instead was slowed down by the force exerted by the dashboard, over a much smaller amount of time, 0.2 s. Compare the average decelerating force experienced by the driver and the passenger.

Solution

We can calculate the change in momentum of both people, which will be equal to the impulse they received as they collided with the seatbelt or with the dashboard. Since we know the duration in time that the forces were exerted, we can calculate the average force involved in order to give the required impulse. We can assume that this all happens in one dimension, so we use scalar quantities instead of vectors.

The change in momentum along the direction of motion for either the driver or passenger is given by:

$$\Delta p = p_B - p_A = (0) - p_A = -mv_A$$

where v_A is the initial speed of the car, and the final momentum of either person is zero.

The change in momentum is equal to the impulse received by either person during a period of time Δt, which is related to the force that was exerted on them:

$$J = F\Delta t = \Delta p = -mv_A$$
$$F = -m\frac{v_A}{\Delta t}$$

For the driver, this corresponds:

$$F = (80\,\text{kg})\frac{(27.8\,\text{m/s})}{(2.5\,\text{s})} = 890\,\text{N}$$

and for the passenger:

$$F = (80\,\text{kg})\frac{(27.8\,\text{m/s})}{(0.2\,\text{s})} = 11\,120\,\text{N}$$

The force on the driver is thus comparable to their weight, whereas the passenger experiences an average force that is more than 10 times their weight.

Discussion: Any mechanism that results in a longer collision time will help to reduce the forces that are involved. This is why cars are designed to crumple in head-on collisions. We can understand this in terms of the crumpling of the car absorbing some of the kinetic energy of the car, as well as lengthening the time of the collision so that the forces involved are smaller. You may also hear people that look at modern cars that are all crumpled up after a crash and say something along the lines of "They sure don't make cars the way they used to". But of course, that is by design; it is safer if the car crumples up (and cars are designed to crumple up in specific areas, not the passenger cabin).

Note that we did not need to use impulse to calculate the average force, since we could have just used kinematics to determine the acceleration and Newton's Second Law to calculate the corresponding force. Using impulse is equivalent by construction, but sometimes, it is easier mathematically.

10.1.3 Systems of particles: internal and external forces

So far, we have only used Newton's Second Law to describe the motion of a single point mass particle or to describe the motion of an object whose orientation we did not need to describe (e.g. a block sliding down a hill). In this section, we consider what happens when there are multiple point particles that form a "system".

In physics, we loosely define a system as the ensemble of objects/particles that we wish to describe. So far, we have only described systems made of one particle, so describing the motion of the system was equivalent to describing the motion of that single particle. A system of two particles could be, for example, two billiard balls on a pool table. To describe that system, we would need to provide functions that describe the positions, velocities, and forces exerted on both balls. We can also define functions/quantities that describe the system as a whole, rather than the details. For example, we can define the total kinetic energy of the system, K, corresponding to the sum of kinetic energies of the two balls. We can also define the total momentum of the system, $\vec{P}$, given by the vector sum of the momenta of the two balls.

When considering a system of multiple particles, we distinguish between **internal** and **external** forces. Internal forces are those forces that the particles in the system exert on each other. For example, if the two billiard balls in the system collide with each other, they will each exert a force on the other during the collision; those forces are internal. External forces are all other forces exerted on the particles of the system. For example, the force of gravity and the normal force from the pool table are both external forces exerted on the balls in the system (exerted by the Earth, or by the pool table, neither of which we considered to be part of the system). The force exerted by a person hitting one of the balls with a pool queue is similarly an external force. What we consider to be a system is arbitrary; we could consider the pool table and the Earth to be part of the system along with the two balls; in that case, the normal force and the weight of the balls would become internal forces. The classification of whether a force is internal or external to a system of course depends on what is considered part of the system.

Checkpoint 10-2

Two pool balls crash against each other. Is this force of gravity exerted by one ball on the other an internal or external force?

A) Internal.
B) External.

The key property of internal forces is that **the vector sum of the internal forces in a system is zero**. Indeed, Newton's Third Law states that for every force exerted by object A on object B, there is a force that is equal in magnitude and opposite in direction exerted by object B on object A. If we consider both objects to be in the same system, then the sum of the internal forces between objects A and B must sum to zero. It is important to note that this is quite different than what we have discussed so far about summing forces. The forces that sum to zero are exerted on *different* objects. Thus far, we had only ever considered summing forces that are exerted on the same object in order to apply Newton's Second Law. We have never encountered a situation where "action" and "reaction" forces are summed together, because they act on different objects.

Emma's Thoughts

Internal vs. External forces - what is the "system" and what forces should we consider?

As discussed above, internal and external forces can only be considered in the context of a specific system. So, how do we define this "system"? How far do we go when defining the system?

For example, let's say that you kick a soccer ball, and it hits a nearby lawn chair, knocking it down. You want to determine what will happen to the soccer ball after it hits the lawn chair. What is defined to be the system here, and how should the forces be classified? Is the force exerted by the soccer ball on the lawn chair an external force? Should we consider the friction between the first foot particle that touches the first soccer ball particle?

The best way to approach "defining the system" is to pin down exactly what you're trying to model. Here, specifically, you are trying to determine the velocity of the ball after it hits the lawn chair. In this situation, thinking about the friction between individual foot and soccer ball particles wouldn't help us to figure out the final velocity of the soccer ball. Rather, thinking of the soccer ball and lawn chair as two giant, continuous particles, colliding and exchanging energy would be helpful. In this situation, it would be useful to consider the "system" to be the soccer ball and lawn chair only.

The force exerted by the soccer ball on the lawn chair would be an internal force, as this gives us information as to the final velocity of the soccer ball and is a force exchanged between the particles within the system. The force that gravity exerts on the lawn chair, normal force on the person's foot and the force exerted by the foot on the soccer ball are all forces that we would consider "external".

Remember - "internal" and "external" are not magical properties of a specific type of force. These definitions are made by us in the quest of building useful models.

10.1.4 Conservation of momentum

Consider a system of two particles with momenta $\vec{p}_1$ and $\vec{p}_2$. Newton's Second Law must hold for each particle:

$$\frac{d\vec{p}_1}{dt} = \sum_k \vec{F}_{1k}$$
$$\frac{d\vec{p}_2}{dt} = \sum_k \vec{F}_{2k}$$

where F_{ik} is the k-th force that is acting on particle i. We can sum these two equations together:

$$\frac{d\vec{p}_1}{dt} + \frac{d\vec{p}_2}{dt} = \sum_k \vec{F}_{1k} + \sum_k \vec{F}_{2k}$$

The quantity on the right is the sum of the forces exerted on particle 1 plus the sum of the forces exerted on particle 2. In other words, it is the sum of all of the forces exerted on all of the particles in the system, which we can write as a single sum. On the left hand side, we have the sum of the two time derivatives of the momenta, which is equal to the time-derivative of the sum of the momenta. We can thus re-write the equation as:

$$\frac{d}{dt}(\vec{p}_1 + \vec{p}_2) = \sum \vec{F}$$

where, again, the sum on the right is the sum over all of the forces exerted on the system. Some of those forces are external (e.g. gravity exerted by Earth on the particles), whereas some of the forces are internal (e.g. a contact force between the two particles). We can separate the sum into a sum over all external forces ($\vec{F}^{ext}$) and a sum over internal forces ($\vec{F}^{int}$):

$$\sum \vec{F} = \sum \vec{F}^{ext} + \sum \vec{F}^{int}$$

The sum of the internal forces is zero:

$$\sum \vec{F}^{int} = 0$$

because for every force that particle 1 exerts on particle 2, there will be an equal and opposite force exerted by particle 2 on particle 1. We thus have:

$$\frac{d}{dt}(\vec{p}_1 + \vec{p}_2) = \sum \vec{F}^{ext}$$

Furthermore, if we introduce the "total momentum of the system", $\vec{P} = \vec{p}_1 + \vec{p}_2$, as the sum of the momenta of the individual particles, we find:

$$\frac{d\vec{P}}{dt} = \sum \vec{F}^{ext}$$

which is the equivalent of Newton's Second Law for a system where, $\vec{P}$, is the total momentum of the system, and the sum of the forces is only over external forces to the system.

Note that the derivation above easily extends to any number, N, of particles, even though we only did it with $N = 2$. In general, for the "ith particle", with momentum $\vec{p}_i$, we can write Newton's Second Law:

$$\frac{d\vec{p}_i}{dt} = \sum_k \vec{F}_{ik}$$

where the sum is over only those forces exerted on particle i. Summing the above equation for all N particles in the system:

$$\frac{d}{dt}\sum_i \vec{p}_i = \sum \vec{F}^{ext} + \sum \vec{F}^{int}$$

where the sum over internal forces will vanish for the same reason as above. Introducing the total momentum of the system, $\vec{P}$:

$$\vec{P} = \sum_i \vec{p}_i$$

We can write an equation for the time-derivative of the total momentum of the system:

$$\boxed{\frac{d\vec{P}}{dt} = \sum \vec{F}^{ext}} \tag{10.5}$$

where the sum of the forces is the sum over all forces external to the system. Thus, **if there are no external forces on a system, then the total momentum of that system is conserved** (if the time-derivative of a quantity is zero then that quantity is constant).

We already argued in the previous section that we can make all forces internal if we choose our system to be large enough. If we make the system be the Universe, then there are no forces external to the Universe, and the total momentum of the Universe must be constant:

$$\begin{aligned} \frac{d\vec{P}^{Universe}}{dt} &= \sum_{Universe} \vec{F}^{ext} = 0 \\ \therefore \vec{P}^{Universe} &= \text{constant} \end{aligned}$$

In summary, we saw that:

- If no forces are exerted on a single particle, then the momentum of that particle is constant (conserved).
- In a system of particles, the total momentum of the system is conserved if there are no external forces on the system.
- If there are no non-conservative forces exerted on a particle, then that particle's mechanical energy is constant (conserved).
- In a system of multiple particles, the total mechanical energy of the system will be conserved if there are no non-conservative forces exerted on the system.

When we refer to a force being "exerted on a system", we mean exerted on one or more of the particles in the system. In particular, the sum of the work done by internal forces is not necessarily zero, so **energy and momentum are thus conserved under different conditions**.

Example 10-4

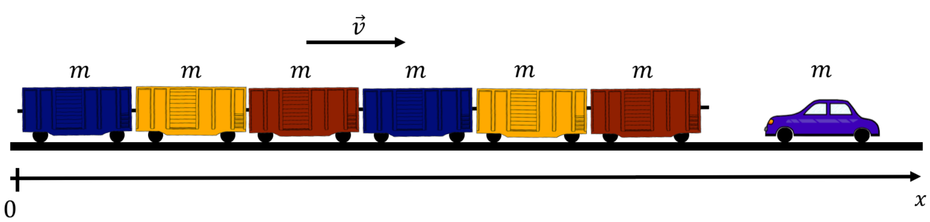


Figure 10.1: A train with N cars of mass m about to collide with a car of mass m that is at rest on the track.

Consider a train made of N cars of equal mass m that is travelling at constant speed v along a straight piece of track where friction and drag are negligible, as depicted in Figure 10.1. An empty car of mass m was left at rest on the track in front of the train. The train collides with the empty car which stays attached to the front of the train. What is the speed of the train after the collision? Is the total mechanical energy of the system conserved?

Solution

When the train collides with the car, it will exert a "collision" force on the car, and the car will exert an opposite force on the train. If we consider both of the train and the car as being part of the same system, then those collision forces will be internal, and the momentum of the system (train + car) will be conserved. The train and car both experience external forces from Earth's gravity and the normal force from the train tracks. However, those two sets of forces cancel each other out, since neither the train nor the car have any acceleration in the vertical direction (the sum of the forces on each object has no net vertical component). Thus, there are no net external forces on the car+train system, and the total momentum of the system is conserved through the collision.

We can model this system in one dimension (along the track), defining our x axis. We choose the ground as a frame of reference, the positive direction parallel to the initial velocity of the train, and the origin to be located where the car initially starts. Before the collision, the x component of the momenta of the train (mass Nm) and car (mass m) are:

$$\begin{aligned}p_{train} &= Nmv\\ p_{car} &= 0\end{aligned}$$

After the collision, the car is attached to the train (and thus has the same speed, v'), so the momenta of the train and car after the collision are:

$$\begin{aligned}p'_{train} &= Nmv'\\ p'_{car} &= mv'\end{aligned}$$

where the primes $'$ denote quantities after the collision. Applying conservation of momentum to the system, the total momentum before and after the collision must be equal:

$$\begin{aligned}p_{train} + p_{car} &= p'_{train} + p'_{car}\\ \therefore Nmv &= Nmv' + mv'\\ \therefore v' &= \frac{N}{N+1}v\end{aligned}$$

and the speed of the train with the additional car attached is reduced by a factor $N/(N+1)$ compared to what it was before the collision.

We can check to see if the mechanical energy of the system is conserved, since we know the speeds of the train and car before and after the collision. Since all of the motion is horizontal, gravity and the normal force do no work on either the train or car, so their mechanical energy can be taken as their kinetic energy (their gravitational potential energy does not change after the collision). The total mechanical energy of the system,

E, before the collision is the kinetic energy of the train:

$$E = \frac{1}{2}Nmv^2$$

The total mechanical energy of the system, E', after the collision is:

$$\begin{aligned} E' &= \frac{1}{2}Nmv'^2 + \frac{1}{2}mv'^2 = \frac{1}{2}(N+1)mv'^2 \\ &= \frac{1}{2}(N+1)m\left(\frac{N}{N+1}v\right)^2 \\ &= \frac{1}{2}m\frac{N^2}{N+1}v^2 \end{aligned}$$

and we see that $E' < E$, and thus that the total mechanical energy of the system is not conserved (it is reduced after the collision).

Discussion: We could have solved this problem by carefully modelling the force exerted by the car on the train during the collision, which would have allowed us to find the speed of the train after the collision using its acceleration. This would have required a detailed model for that force, which we do not have. However, by realizing that the train and car could be considered as a system with no net external forces exert on it, we were able to easily find the speed of the train after the collision using conservation of momentum.

We also found that mechanical energy was not conserved. This makes physical sense because, for the car to remain attached to the train, there presumably had to be some significant forces in play that "crushed" the car into the train. Some of the initial kinetic energy of the train was used to deform the train and the car during the collision. We can also think of deforming a material as giving it energy. Sometimes that energy is recoverable (e.g. compressing a spring), sometimes, it is not (e.g. crushing a car).

If the car and train were equipped with large springs to absorb the energy of the impact, the collision could have conserved mechanical energy, as the springs compress and then expand back. The speed of the car and train would then be different after the collision in this case (see example 10-7). It is a feature of collisions where the two bodies remain attached to each other that mechanical energy is not conserved.

10.2 Collisions

In this section we go through a few examples of applying conservation of momentum to model collisions. Collisions can loosely be defined as events where the momenta of individual particles in a system are different before and after the event.

We distinguish between two types of collisions: **elastic** and **inelastic** collisions. Elastic collisions are those for which the total mechanical energy of the system is conserved during the collision (i.e. it is the same before and after the collision). Inelastic collisions are those for which the total mechanical energy of the system is not conserved. In either case, to model the system, one chooses to define the system such that there are no external forces on the system so that total momentum is conserved.

10.2.1 Inelastic collisions

In this section, we give a few examples of modelling inelastic collisions. Inelastic collisions are usually easier to handle mathematically, because one only needs to consider conservation of momentum and does not use conservation of energy (which usually involves equations that are quadratic in the speeds because of the kinetic energy term).

Example 10-5

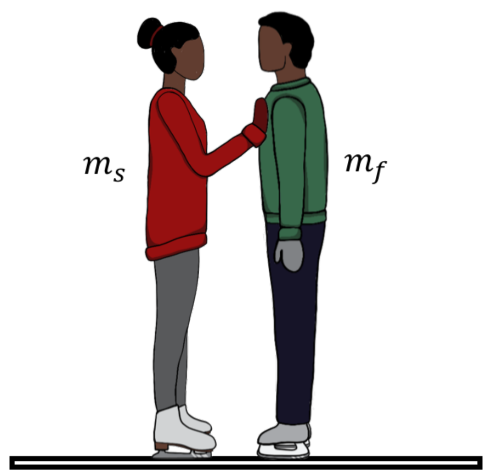


Figure 10.2: One skater pushing another on a frictionless horizontal surface.

You (mass m_s) and your friend (mass m_f) face each other on ice skates on an ice surface that is slippery enough that friction can be considered negligible, as shown in Figure 10.2. You shove your friend away from you so that he moves with velocity $\vec{v}_f$ away from you (the velocity is measured relative to the ice). Is the collision elastic? What is your speed relative to the ice after you shoved your friend?

Solution

We can consider the system as being comprised of you and your friend. There are no net external forces on the system (gravity and normal forces cancel each other), so the momentum of the system will be conserved.

The mechanical energy will not be conserved. You had to use chemical potential energy stored in your muscles to shove your friend. Thus, external energy (i.e. not mechanical energy from you or your friend) was injected into the system, and we should expect the total mechanical energy to be larger after the collision.

Before the collision, both you and your friend have zero speed, and thus zero kinetic energy and zero momentum. After the collision, your friend has a velocity $\vec{v}_f$. We can use conservation of total momentum, $\vec{P}$, to determine your velocity, $\vec{v}_s$, after the collision.

$$\begin{aligned}\vec{P} &= \vec{P}'\\ 0 &= m_s\vec{v}_s + m_f\vec{v}_f\\ \therefore \vec{v}_s &= -\frac{m_f}{m_s}\vec{v}_f\end{aligned}$$

where primes ($'$) denote a quantity after the collision. We find that your velocity is in the opposite direction from that of your friend. Before the collision, the mechanical energy, E, of the system is zero (we can ignore gravitational potential energy, since everything is in the horizontal plane). After the collision, the mechanical energy, E', is:

$$E' = \frac{1}{2}m_s v_s^2 + \frac{1}{2}m_f v_f^2$$

which is clearly bigger than the mechanical energy before the collision (i.e. 0), as we suspected it would be.

Discussion: We find that you recoil in the opposite direction, which makes sense. If you push your friend in one direction, Newton's Third Law says that your friend pushes you in the opposite direction. Your speed furthermore depends on the ratio of your friend's mass to yours. This also makes sense, because if you both feel the same force, the person with the smallest mass will have the highest speed; if your mass is higher than your friend's, then your speed after the collision will be smaller than your friend's.

We also saw that mechanical energy was not conserved. In terms of energy, we can explain this by saying that you burned up chemical potential energy stored in your muscles in order to shove your friend. Because we included both you and your friend in the system, the shove was an internal force and momentum is conserved. Of course, if we had considered only you as the system, then your momentum would not have been conserved during the collision.

The type of collision that we described here is also sometimes called an "explosion". You can imagine all of the parts that make up a bomb as small particles. When the bomb explodes, chemical potential energy is converted into the kinetic energy of the bomb fragments. If you consider all of the particles/fragments of the bomb as a system, then the total momentum of all of the bomb fragments is conserved (and equal to zero if the bomb was initially at rest). Again, mechanical energy would not be conserved (and would increase) as the chemical potential energy is converted into mechanical energy.

Example 10-6

A proton of mass m_p and initial velocity $\vec{v}_p$ collides inelastically with a nucleus of mass m_N at rest, as shown in Figure 10.3. A coordinate system is set up as shown, such that the initial velocity of the proton is in the x direction. After the collision, the proton's speed is measured to be v'_p and its velocity vector is found to make an angle θ with the x axis as shown. What is the velocity vector of the nucleus after the collision? Assume that the collision takes place in vacuum.

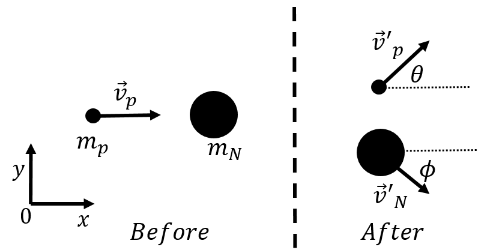


Figure 10.3: A proton of mass m_p colliding inelastically with a nucleus of mass m_N.

Solution

As a system, we consider the proton and the nucleus together, so that the total momentum of the system is conserved during the collision, as no other external forces are exerted on the two particles (since they are in vacuum). Because momentum is a vector, each component of the total momentum, $\vec{P}$, is conserved during the collision:

$$\begin{aligned}\vec{P} &= \vec{P}'\\ \therefore P_x &= P'_x\\ \therefore P_y &= P'_y\end{aligned}$$

where, as usual, primes ($'$) denote quantities after the collision. After the collision, both particles will have velocity vectors that have x and y components. Let the velocity vector of the nucleus after the collision be $\vec{v}'_N$ and let ϕ be the angle that it makes with the x axis, as shown in Figure 10.3.

We can start by considering the conservation of the x component of the total momentum. The initial and final momenta in the x direction are given by:

$$\begin{aligned}P_x &= m_p v_p\\ P'_x &= m_p v'_p \cos\theta + m_N v'_N \cos\phi\\ \therefore m_p v_p &= m_p v'_p \cos\theta + m_N v'_N \cos\phi\end{aligned}$$

which gives us a first equation to determine the final velocity of the nucleus.

The y component of the total momentum before the collision is zero since we chose the coordinate system such that the initial velocity of the proton is in the x direction. The initial and final momenta in the y direction are given by:

$$\begin{aligned}
P_y &= 0\\
P'_y &= m_p v'_p \sin\theta - m_N v'_N \sin\phi\\
\therefore m_p v'_p \sin\theta &= m_N v'_N \sin\phi
\end{aligned}$$

which gives us a second equation to solve for the velocity of the nucleus. With the two equations from momentum conservation, we can solve for the magnitude and direction of the velocity of the nucleus. From the y component of momentum conservation, we can find an expression for the speed of the nucleus:

$$\begin{aligned}
m_p v'_p \sin\theta &= m_N v'_N \sin\phi\\
\therefore v'_N &= \frac{m_p}{m_N} v'_p \sin\theta \frac{1}{\sin\phi}
\end{aligned}$$

which we can substitute into the x equation for momentum conservation to solve for the angle ϕ:

$$\begin{aligned}
m_p v_p &= m_p v'_p \cos\theta + m_N v'_N \cos\phi\\
m_p v_p &= m_p v'_p \cos\theta + m_N \frac{m_p}{m_N} v'_p \sin\theta \frac{\cos\phi}{\sin\phi}\\
v_p &= v'_p \cos\theta + v'_p \sin\theta \frac{1}{\tan\phi}\\
\therefore \tan\phi &= \frac{v'_p \sin\theta}{v_p - v'_p \cos\theta}
\end{aligned}$$

If we were given numbers for the initial and final speed of the proton, as well as the angle θ, we would be able to find a value for the angle ϕ, which we could then use to determine the final speed of the nucleus:

$$v'_N = \frac{m_p}{m_N} v'_p \sin\theta \frac{1}{\sin\phi}$$

Discussion: By using the conservation of momentum equation and writing out the x and y components, we were able to find two equations to determine the magnitude and direction of the nucleus' velocity after the collision. In the limit where $m_N >> m_p$, the final speed of the nucleus would be very small (close to zero).

10.2.2 Elastic collisions

In this section, we give a few examples of modelling elastic collisions. Even though it is mechanical energy that is conserved in an elastic collision, one can almost always simplify this to only kinetic energy being conserved. If a collision takes place in a well localized position in space (i.e. before and after the collision are the same point in space), then the potential energies of the objects involved will not change, thus any change in their mechanical energy is due to a change in kinetic energy.

Example 10-7

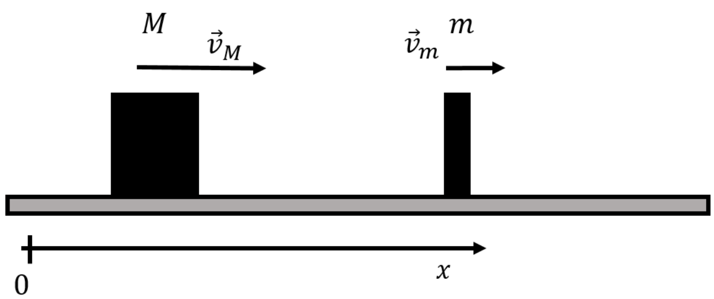


Figure 10.4: Two blocks about to collide elastically.

A block of mass M moves with velocity $\vec{v}_M$ in the x direction, as shown in Figure 10.4. A block of mass m is moving with velocity $\vec{v}_m$ also in the x direction and collides elastically with block M. Both blocks slide with no friction on the horizontal surface. What are the velocities of the two blocks after the collision?

Solution

Because this is an elastic collision, both the total momentum and total mechanical energy are conserved. Equating the total momentum before and after the collision, and considering only the x component gives the following equation:

$$\vec{P} = \vec{P}'$$
$$Mv_M + mv_m = Mv'_M + mv'_m$$

where the primes ($'$) correspond to the quantities after the collision. Note that, in principle, the x components of the velocities (v_M, v'_M, v_m, v'_m) could be negative numbers if the corresponding block is moving in the negative x direction.

For the mechanical energy of the two blocks, we only need to consider their kinetic energy since their gravitational potential energies are the same before and after the collision on the horizontal surface. The total mechanical energy of the system, before and after the collision is given by:

$$\begin{aligned}
E &= E' \\
\frac{1}{2}Mv_M^2 + \frac{1}{2}mv_m^2 &= \frac{1}{2}Mv_M'^2 + \frac{1}{2}mv_m'^2 \\
\therefore Mv_M^2 + mv_m^2 &= Mv_M'^2 + mv_m'^2
\end{aligned}$$

where we cancelled the factor of one half in the last line. This gives two equations (conservation of energy and momentum) and two unknowns (the two speeds after the collision). This is not a linear system of equations, because the equation from conservation of energy is quadratic in the speeds.

The following method allows many models for elastic collisions between two particles to be solved easily by converting the quadratic equation from energy conservation into an equation that is linear in the speeds. First, write both equations so that the quantities related to each particle are on opposite sides of the equation. For momentum, this gives:

$$\begin{aligned}
Mv_M + mv_m &= Mv_M' + mv_m' \\
\therefore M(v_M - v_M') &= m(v'm - v_m)
\end{aligned} \tag{10.6}$$

For conservation of energy, this gives:

$$\begin{aligned}
Mv_M^2 + mv_m^2 &= Mv_M'^2 + mv_m'^2 \\
\therefore M(v_M^2 - v_M'^2) &= M(v_m'^2 - v_m^2)
\end{aligned} \tag{10.7}$$

which we can re-write as:

$$\begin{aligned}
M(v_M^2 - v_M'^2) &= M(v_m'^2 - v_m^2) \\
M(v_M - v_M')(v_M + v_M') &= M(v_m' - v_m)(v_m' + v_m)
\end{aligned}$$

We can then divide Equation 10.7 by Equation 10.6:

$$\begin{aligned}
\frac{M(v_M - v_M')(v_M + v_M')}{M(v_M - v_M')} &= \frac{M(v_m' - v_m)(v_m' + v_m)}{m(v'm - v_m)} \\
\therefore v_M + v_M' &= v_m' + v_m
\end{aligned}$$

which gives us an equation that is much easier to work with, since it is linear in the speeds. If we re-arrange this last equation back so that quantities before and after the collision are on different sides of the equality:

$$\boxed{v_M - v_m = -(v_M' - v_m')}$$

we can see that the relative speed between M and m is the same before and after the collision. That is, if block M "saw" block m approaching with a speed of 3 m/s before the collision, it would "see" block m moving *away* with speed 3 m/s after the collision, regardless of the actual directions and velocities of the block, if the collision was elastic.

By using this equation with the original conservation of momentum equation, we now have two equations and two unknowns that are easy to solve:

$$\begin{aligned} v_M - v_m &= -(v'_M - v'_m) \\ Mv_M + mv_m &= Mv'_M + mv'_m \end{aligned}$$

Solving for v'_m in both equations gives:

$$\begin{aligned} v_M - v_m &= -(v'_M - v'_m) \\ \therefore v'_m &= v_M + v'_M - v_m \\ Mv_M + mv_m &= Mv'_M + mv'_m \\ \therefore v'_m &= \frac{1}{m}(Mv_M + mv_m - Mv'_M) \end{aligned}$$

Equating the two expressions for v'_m allows us to solve for v'_M:

$$\begin{aligned} \frac{1}{m}(Mv_M + mv_m - Mv'_M) &= v_M + v'_M - v_m \\ Mv_M + mv_m - Mv'_M &= mv_M + mv'_M - mv_m \\ (M-m)v_M + 2mv_m &= (M+m)v'_M \\ \therefore v'_M &= \frac{M-m}{M+m}v_M + \frac{2m}{M+m}v_m \end{aligned}$$

One can easily solve for the other speed, v'_m:

$$\therefore v'_m = \frac{m-M}{M+m}v_m + \frac{2M}{M+m}v_M$$

And writing these together:

$$\begin{aligned} v'_M &= \frac{M-m}{M+m}v_M + \frac{2m}{M+m}v_m \\ v'_m &= \frac{m-M}{M+m}v_m + \frac{2M}{M+m}v_M \end{aligned}$$

Discussion: The formulas that we obtained above are valid for any one dimensional elastic collision.

Checkpoint 10-3

Two trains of equal masses collide elastically on a track. If train A had a speed v and train B was at rest, what are the speeds of the trains after the collision?

A) Both trains A and B travel away from each other with speeds $\frac{1}{2}v$.
B) Train A will be at rest and train B will move away with a speed v.
C) Both trains A and B will stick together and move at a speed of v.
D) Train B will be at rest and train A will move away at a speed of v.

Example 10-8

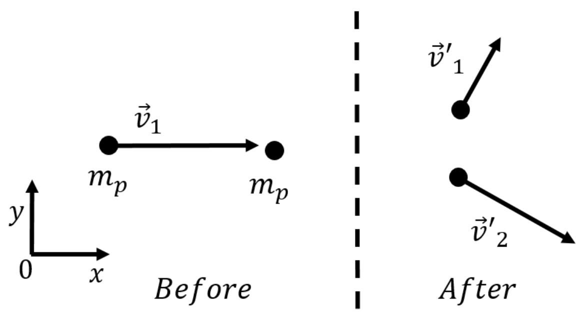


Figure 10.5: A proton elastically collides with a proton at rest.

A proton of mass m and initial velocity $\vec{v}_1$ collides elastically with a second proton that is at rest. After the collision, the two protons have velocities $\vec{v}'_1$ and $\vec{v}'_2$, as shown in Figure 10.5. Show that the velocity vectors of the two protons are perpendicular after the collision.

Solution

This example highlights a particular feature of elastic collisions when the two objects have the same mass and one of the objects is initially at rest. The conservation of momentum for the system comprised of the two protons can be written as:

$$m\vec{v}_1 = m\vec{v}'_1 + m\vec{v}'_2$$
$$\vec{v}_1 = \vec{v}'_1 + \vec{v}'_2$$

where the left hand side corresponds to the initial total momentum and the right hand side to the total momentum after the collision. In the second line, we cancelled out the mass, and obtained a vector relation between the velocity vectors. We can graphically illustrate the vector relation as in Figure 10.6 which shows the triangle that is formed by adding the two outgoing velocity vectors to obtain the initial velocity vector.

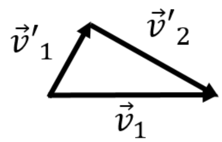


Figure 10.6: Graphical illustration of the relation between the initial and final velocity vectors as a vector sum.

Conservation of kinetic energy for the collision can be written as:

$$\frac{1}{2}mv_1^2 = \frac{1}{2}mv_1'^2 + \frac{1}{2}mv_2'^2$$
$$v_1^2 = v_1'^2 + v_2'^2$$

where the left hand side corresponds to the initial kinetic energy and the right hand side to the final kinetic energy. We cancelled the mass and factor of one half in the second line. This last equation gives a relation between the magnitudes of the velocity vectors. By comparing the equation above to Pythagoras' theorem, and by inspecting the triangle in Figure 10.6, it is clear that the triangle must be a right angle triangle, and thus that $\vec{v}'_1$ and $\vec{v}'_2$ must be perpendicular.

10.2.3 Frames of reference

Review Topics

Before proceeding, you may wish to review Sections 3.4 and 4.1.2 on expressing velocities in different frames of reference.

Because the momentum of a particle is defined using the velocity of the particle, its value depends on the reference frame in which we chose to measure that velocity. In some cases, it is useful to apply momentum conservation in a frame of reference where the total momentum of the system is zero. For example, consider two particles of mass m_1 and m_2, moving towards each other with velocities $\vec{v}_1$ and $\vec{v}_2$, respectively, as measured in a frame of reference S, as illustrated in Figure 10.7.

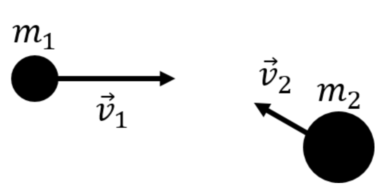


Figure 10.7: Two particles moving towards each other.

In the frame of reference S, the total momentum, $\vec{P}$, of the two particles can be written:

$$\vec{P} = m_1\vec{v}_1 + m_2\vec{v}_2$$

Consider a frame of reference, S', that is moving with velocity, $\vec{v}_{CM}$, relative to the frame of reference S. In that frame of reference, the velocities of the two particles are different and given by:

$$\vec{v}'_1 = \vec{v}_1 - \vec{v}_{CM}$$
$$\vec{v}'_2 = \vec{v}_2 - \vec{v}_{CM}$$

The total momentum, $\vec{P}'$, in the frame of reference S' is then given by[1]:

$$\begin{aligned}\vec{P}' &= m_1\vec{v}'_1 + m_2\vec{v}'_2\\ &= m_1(\vec{v}_1 - \vec{v}_{CM}) + m_2(\vec{v}_2 - \vec{v}_{CM})\\ &= m_1\vec{v}_1 + m_2\vec{v}_2 - (m_1 + m_2)\vec{v}_{CM}\end{aligned}$$

We can choose the velocity of the frame S', $\vec{v}_{CM}$, such that the total momentum in that frame of reference is zero:

$$\begin{aligned}\vec{P}' &= 0\\ m_1\vec{v}_1 + m_2\vec{v}_2 - (m_1 + m_2)\vec{v}_{CM} &= 0\\ \therefore \vec{v}_{CM} &= \frac{m_1\vec{v}_1 + m_2\vec{v}_2}{m_1 + m_2}\end{aligned}$$

This "special" frame of reference, in which the total momentum of the system is zero, is called the "centre of mass frame of reference". The velocity of centre of mass frame of reference can easily be obtained if there are N particles involved instead of two:

$$\boxed{\therefore \vec{v}_{CM} = \frac{m_1\vec{v}_1 + m_2\vec{v}_2 + m_3\vec{v}_3 + \dots}{m_1 + m_2 + m_3 + \dots} = \frac{\sum m_i\vec{v}_i}{\sum m_i}} \tag{10.8}$$

Again, you should note that because the above equation is a vector equation, it represents one equation per component of the vectors. For example, the x component of the velocity of the centre of mass frame of reference is given by:

$$\therefore v_{CMx} = \frac{m_1 v_{1x} + m_2 v_{2x} + m_3 v_{3x} + \dots}{m_1 + m_2 + m_3 + \dots} = \frac{\sum m_i v_{ix}}{\sum m_i}$$

Example 10-9

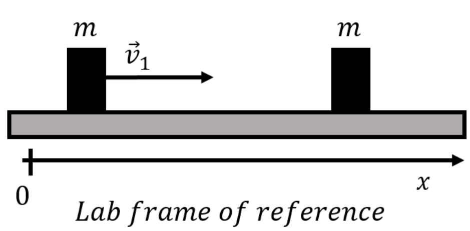


Figure 10.8: One block approaching another identical block at rest, as seen in the lab frame of reference.

[1] Note that we are using primes ($'$) to denote quantities in a different reference frame, not after a collision.

In the frame of reference of a lab, a block of mass m has a velocity $\vec{v}_1$ directed along the positive x axis and is approaching a second block of mass m that is at rest ($\vec{v}_2 = 0$), as shown in Figure 10.8. What is the velocity of the centre of mass frame? What is the velocity of each block in the centre of mass frame? Verify that the total momentum is zero in the centre of mass frame.

Solution

Since this is a one dimensional situation, we only need to evaluate the x component of the velocity of the centre of mass:

$$\begin{aligned}
\vec{v}_{CM} &= \frac{m_1\vec{v}_1 + m_2\vec{v}_2}{m_1 + m_2} \\
\therefore v_{CMx} &= \frac{m_1 v_{1x} + m_2 v_{2x}}{m_1 + m_2} \\
&= \frac{mv_1 + m(0)}{m + m} \\
&= \frac{1}{2}v_1
\end{aligned}$$

The centre of mass frame of reference is thus also moving along the positive direction of the x axis, but with a speed that is half of that of the moving block. In the centre of mass frame of reference, it appears that the block on the left is slower than in the lab frame and that the block on the right is moving in the negative x direction. The velocities of the two blocks in the centre of mass frame of reference are given by:

$$\begin{aligned}
v_1' &= v_1 - v_{CMx} = \frac{1}{2}v_1 \\
v_2' &= (0) - v_{CMx} = -\frac{1}{2}v_1
\end{aligned}$$

Thus, in the reference frame of the centre of mass, the two block are approaching each other with the same speed ($v_1/2$), which is only the case because the two blocks have the same mass. The blocks, as viewed in the centre of mass frame of reference, are shown in Figure 10.9.

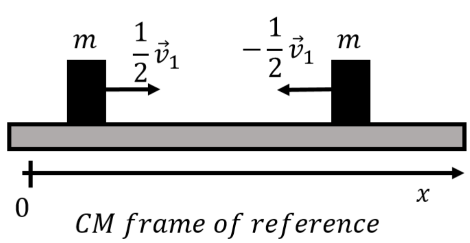


Figure 10.9: In the centre of mass frame of reference, the block approach each other with the same speed, because they have the same mass.

Clearly, the total momentum is zero in the centre of mass frame of reference:

$$\vec{P}' = m\vec{v}'_1 + m\vec{v}'_2 = m\left(\frac{1}{2}\vec{v}_1 - \frac{1}{2}\vec{v}_1\right) = 0$$

Discussion: As we have seen, in the centre of mass frame of reference the total momentum is zero. If there are only two particles, and they have the same mass, then, in the centre of mass frame of reference, they both have the same speed and move either towards or away from each other.

10.3 The centre of mass

In this section, we show how to generalize Newton's Second Law so that it may describe the motion of an object that is not a point particle. Any object can be described as being made up of point particles; for example, those particles could be the atoms that make up regular matter. We can thus use the same terminology as in the previous sections to describe a complicated object as a "system" comprised of many point particles, themselves described by Newton's Second Law. A system could be a rigid object where the point particles cannot move relative to each other, such as atoms in a solid[2]. Or, the system could be a gas, made of many atoms moving around, or it could be a combination of many solid objects moving around.

In the previous section, we saw how the total momentum and the total mechanical energy of the system could be used to describe the system as a whole. In this section, we will define the centre of mass which will allow us to describe the position of the system as a whole.

Consider a system comprised of N point particles. Each point particle i, of mass m_i, can be described by a position vector, $\vec{r}_i$, a velocity vector, $\vec{v}_i$, and an acceleration vector, $\vec{a}_i$, relative to some coordinate system in an inertial frame of reference. Newton's Second Law can be applied to any one of the particles in the system:

$$\sum_k \vec{F}_{ik} = m_i\vec{a}_i$$

where $\vec{F}_{ik}$ is the k-th force exerted on particle i. We can write Newton's Second Law once for each of the N particles, and we can sum those N equations together:

$$\sum_k \vec{F}_{1k} + \sum_k \vec{F}_{2k} + \sum_k \vec{F}_{3k} + \dots = m_1\vec{a}_1 + m_2\vec{a}_2 + m_3\vec{a}_3 + \dots$$
$$\sum \vec{F} = \sum_i m_i\vec{a}_i$$

where the sum on the left is the sum of all of the forces exerted on all of the particles in the system[3] and the sum over i on the right is over all of the N particles in the system. As

[2] In reality, even atoms in a solid can move relative to each other, but they do not move by large amounts compared to the object.

[3] Again, we are summing together forces that are acting on **different** particles

we have already seen, the sum of all of the forces exerted on the system can be divided into separate sums over external and internal forces:

$$\sum \vec{F} = \sum \vec{F}^{ext} + \sum \vec{F}^{int}$$

and the sum over the internal forces is zero[4]. We can thus write that the sum of the external forces exerted on the system is given by:

$$\sum \vec{F}^{ext} = \sum_i m_i \vec{a}_i \tag{10.9}$$

We would like this equation to resemble Newton's Second Law, but for the system as a whole. Suppose that the system has a total mass, M:

$$M = m_1 + m_2 + m_3 + \cdots = \sum_i m_i$$

we would like to have an equation of the form:

$$\sum \vec{F}^{ext} = M\vec{a}_{CM} \tag{10.10}$$

to describe the system as a whole. However, it is not (yet) clear what is accelerating with acceleration, $\vec{a}_{CM}$, since the particles in the system could all be moving in different directions. Suppose that there is a point in the system, whose position is given by the vector, $\vec{r}_{CM}$, in such a way that the acceleration above is the second time-derivative of that position vector:

$$\vec{a}_{CM} = \frac{d^2}{dt^2}\vec{r}_{CM}$$

We can compare Equations 10.9 and 10.10 to determine what the position vector $\vec{r}_{CM}$ corresponds to:

$$\begin{aligned}
\sum \vec{F}^{ext} &= \sum_i m_i \vec{a}_i = \sum_i m_i \frac{d^2}{dt^2}\vec{r}_i \\
\sum \vec{F}^{ext} &= M\vec{a}_{CM} = M\frac{d^2}{dt^2}\vec{r}_{CM} \\
\therefore M\frac{d^2}{dt^2}\vec{r}_{CM} &= \sum_i m_i \frac{d^2}{dt^2}\vec{r}_i
\end{aligned}$$

Re-arranging, and noting that the masses are constant in time, and so they can be factored into the derivatives:

$$\begin{aligned}
\frac{d^2}{dt^2}\vec{r}_{CM} &= \frac{1}{M}\sum_i m_i \frac{d^2}{dt^2}\vec{r}_i \\
\frac{d^2}{dt^2}\vec{r}_{CM} &= \frac{d^2}{dt^2}\left(\frac{1}{M}\sum_i m_i \vec{r}_i\right) \\
\therefore \vec{r}_{CM} &= \frac{1}{M}\sum_i m_i \vec{r}_i
\end{aligned}$$

[4]Recall, the internal forces are those forces that particles in the system are exerting on one another. Because of Newton's Third Law, these will sum to zero.

where in the last line we set the quantities that have the same time derivative equal to each other[5]. $\vec{r}_{CM}$ is the vector that describes the position of the "centre of mass" (CM). The position of the centre of mass is described by Newton's Second Law applied to the system as a whole:

$$\boxed{\sum \vec{F}^{ext} = M\vec{a}_{CM}} \tag{10.11}$$

where M is the total mass of the system, and the sum of the forces is the sum over only external forces on the system.

Although we have formally derived Newton's Second Law for a system of particles, we really have been using this result throughout the text. For example, when we modelled a block sliding down an incline, we never worried that the block was made of many atoms all interacting with each other and the surroundings. Instead, we only considered the external forces on the block, namely, the normal force from the incline, any frictional forces, and the total weight of the object (the force exerted by gravity). Technically, the force of gravity is not exerted on the block as a whole, but on each of the atoms. However, when we sum the force of gravity exerted on each atom:

$$m_1\vec{g} + m_2\vec{g} + m_3\vec{g} + \dots = (m_1 + m_2 + m_3 + \dots)\vec{g} = M\vec{g}$$

we find that it can be modelled by considering the block as a single particle of mass M upon which gravity is exerted. The centre of mass is sometimes described as the "centre of gravity", because it **corresponds to the location where we can model the total force of gravity, $M\vec{g}$, as being exerted**. When we applied Newton's Second Law to the block, we then described the motion of the block as a whole (and not the motion of the individual atoms). Specifically, we modelled the motion of the centre of mass of the block.

The position of the centre of mass is a vector equation that is true for each coordinate:

$$\begin{aligned}
\vec{r}_{CM} &= \frac{1}{M}\sum_i m_i\vec{r}_i \\
\therefore x_{CM} &= \frac{1}{M}\sum_i m_i x_i \\
\therefore y_{CM} &= \frac{1}{M}\sum_i m_i y_i \\
\therefore z_{CM} &= \frac{1}{M}\sum_i m_i z_i
\end{aligned} \tag{10.12}$$

The centre of mass is that **position in a system that is described by Newton's Second Law when it is applied to the system as a whole**. The centre of mass can be thought of as an average position for the system (it is the average of the positions of the particles in the system, weighted by their mass). By describing the position of the centre of mass,

[5]Technically, the terms in the derivatives are only equal to within two constants of integration, $\vec{r}_{CM} = \frac{1}{M}\sum_i m_i\vec{r}_i + at + b$, which we can set to zero

we are not worried about the detailed positions of the all of the particles in the system, but rather only the average position of the system as a whole. In other words, this is equivalent to viewing the whole system as a single particle of mass M located at the position of the centre of mass.

Consider, for example, a person throwing a dumbbell that is made from two spherical masses connected by a rod, as illustrated in Figure 10.10. The dumbbell will rotate in a complex manner as it moves through the air. However, the centre of mass of the dumbbell will travel along a parabolic trajectory (projectile motion), because the only external force exerted on the dumbbell during its trajectory is gravity.

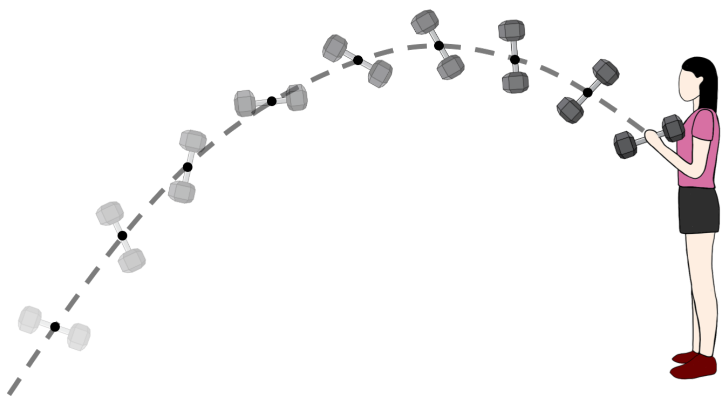


Figure 10.10: The motion of the centre of mass of a dumbbell is described by Newton's Second Law, even if the motion of the rotating dumbbell is more complex.

If we take the derivative with respect to time of the centre of mass position, we obtain the velocity of the centre of mass, and its components, which allow us to describe how the system is moving as a whole:

$$\begin{aligned}\vec{v}_{CM} &= \frac{d}{dt}\vec{r}_{CM} = \frac{1}{M}\sum_i m_i \frac{d}{dt}\vec{r}_i = \frac{1}{M}\sum_i m_i\vec{v}_i\\ \therefore v_{CMx} &= \frac{1}{M}\sum_i m_i v_{ix}\\ \therefore v_{CMy} &= \frac{1}{M}\sum_i m_i v_{iy}\\ \therefore v_{CMz} &= \frac{1}{M}\sum_i m_i v_{iz}\end{aligned} \tag{10.13}$$

Note that this is the same velocity that we found earlier for the velocity of the centre of mass frame of reference. In the centre of mass frame of reference, the total momentum of the system is zero. This makes sense, because the centre of mass represents the average position of the system; if we move "with the system", then the system appears to have zero momentum.

We can also define the total momentum of the system, $\vec{P}$, in terms of the total mass, M, of the system and the velocity of the centre of mass:

$$\begin{aligned}\vec{P} &= \sum m_i \vec{v}_i = \frac{M}{M} \sum m_i \vec{v}_i \\ &= M \vec{v}_{CM}\end{aligned}$$

which we can also use in Newton's Second Law:

$$\frac{d}{dt}\vec{P} = \sum \vec{F}^{ext}$$

and again, we see that the total momentum of the system is conserved if the net external force on the system is zero. In other words, the centre of mass of the system will move with constant velocity when momentum is conserved.

Finally, we can also define the acceleration of the centre of mass by taking the time derivative of the velocity:

$$\begin{aligned}\vec{a}_{CM} &= \frac{d}{dt}\vec{v}_{CM} = \frac{1}{M}\sum_i m_i \frac{d}{dt}\vec{v}_i = \frac{1}{M}\sum_i m_i \vec{a}_i \\ \therefore a_{CMx} &= \frac{1}{M}\sum_i m_i a_{ix} \\ \therefore a_{CMy} &= \frac{1}{M}\sum_i m_i a_{iy} \\ \therefore a_{CMz} &= \frac{1}{M}\sum_i m_i a_{iz}\end{aligned} \tag{10.14}$$

Example 10-10

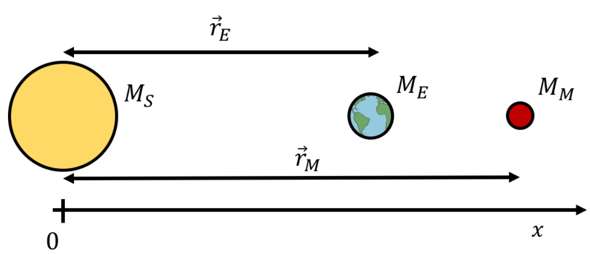


Figure 10.11: A syzygy between the Sun, Earth, and Mars.

In astronomy, a syzygy is defined as the event in which three bodies are all lined up along a straight line. For example, a syzygy occurs when the Sun (mass $M_S = 2.00 \times 10^{30}$ kg), Earth (mass $M_E = 5.97 \times 10^{24}$ kg), and Mars (mass $M_M = 6.39 \times 10^{23}$ kg) are all lined up, as in Figure 10.11. How far from the centre of the Sun is the centre of mass of the Sun, Earth, Mars system during a syzygy?

Solution

Since this is a one-dimensional problem, we can define an x axis that is co-linear with the three bodies, and find only the x coordinate of the position of the centre of mass. We are free to choose the origin of the coordinate system, so we choose the origin to be located at the centre of the Sun. This way, the position of the centre of mass along the x axis will directly correspond to its distance from the centre of the Sun.

The Sun, Earth, and Mars are not point particles. However, because they are spherically symmetric, their centres of mass correspond to their geometric centres. We can thus model them as point particles with the mass of the body located at the corresponding geometric centre. If $r_E = 1.50 \times 10^{11}\,\text{m}$ ($r_M = 2.28 \times 10^{11}\,\text{m}$) is the distance from the centre of the Earth (Mars) to the centre of the Sun, then the position of the centre of mass is given by:

$$\begin{aligned}
x_{CM} &= \frac{1}{M}\sum_i m_i x_i \\
&= \frac{M_S(0) + M_E r_E + M_M r_M}{M_S + M_E + M_M} \\
&= \frac{(2.00 \times 10^{30}\,\text{kg})(0) + (5.97 \times 10^{24}\,\text{kg})(1.50 \times 10^{11}\,\text{m}) + (6.39 \times 10^{23}\,\text{kg})(2.28 \times 10^{11}\,\text{m})}{(2.00 \times 10^{30}\,\text{kg}) + (5.97 \times 10^{24}\,\text{kg}) + (6.39 \times 10^{23}\,\text{kg})} \\
&= 5.21 \times 10^{5}\,\text{m}
\end{aligned}$$

The centre of mass of the Sun-Earth-Mars system during a syzygy is located approximately 500 km from the centre of the Sun.

Discussion: The radius of the Sun is approximately 700 000 km, so the centre of mass of the system is well inside of the Sun. The Sun is so much more massive than either of the Earth or Mars, that the two planets hardly contribute to shifting the centre of mass away from the centre of the Sun. We would generally consider the masses of the two planets to be negligible if one wanted to model how the solar system itself moves around the Milky Way galaxy.

Example 10-11

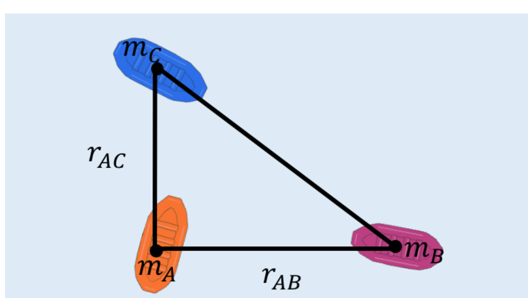


Figure 10.12: Three people on rafts on a lake.

Alice (mass m_A), Brice (mass m_B), and Chloë (mass m_C) are stranded on individual rafts of negligible mass on a lake, off of the coast of Nyon. The rafts are located at the corners of a right-angle triangle, as illustrated in Figure 10.12, and are connected by ropes. The distance between Alice and Brice is r_{AB} and the distance between Alice and Chloë is r_{AC}, as illustrated. Alice decides to pull on the rope that connects her to Chloë , while Brice decide to pull on the rope that connects him to Alice. Where will the three rafts meet?

Solution

We consider the system comprised of the three people and their rafts and model each person and their raft as a point particle with the mass concentrated at the centre of the raft. The forces exerted by pulling on the ropes are internal forces (one particle on the other), and will thus have no impact on the motion of the centre of mass of the system. There are no net external forces exerted on the system (the forces of gravity are balanced out by the forces of buoyancy from the rafts). The centre of mass of the system does not move when the people are pulling on the ropes, so they must ultimately meet at the centre of mass.

We can define a coordinate system such that the origin is located where Alice is initially located, the x axis is in the direction from Alice to Brice, and the y axis is in the direction from Alice to Chloë. The initial positions of Alice, Brice, and Chloë are thus:

$$\begin{aligned}\vec{r}_A &= 0\hat{x} + 0\hat{y}\\ \vec{r}_B &= r_{AB}\hat{x} + 0\hat{y}\\ \vec{r}_C &= 0\hat{x} + r_{AC}\hat{y}\end{aligned}$$

respectively. The x and y coordinates of the centre of mass are thus:

$$\begin{aligned}x_{CM} &= \frac{1}{M}\sum_i m_i x_i = \frac{m_A(0) + m_B r_{AB} + m_C(0)}{m_A + m_B + m_C} = \left(\frac{m_B}{m_A + m_B + m_C}\right) r_{AB}\\ y_{CM} &= \frac{1}{M}\sum_i m_i y_i = \frac{m_A(0) + m_B(0) + m_C r_{AC}}{m_A + m_B + m_C} = \left(\frac{m_C}{m_A + m_B + m_C}\right) r_{AC}\end{aligned}$$

which corresponds to the position where the three rafts will meet, relative to the initial position of Alice.

Discussion: By using the centre of mass, we easily found where the three rafts would meet. If we had used Newton's Second Law on the three rafts individually, the model would have been complicated by the fact that the forces exerted by Alice and Brice on the ropes change direction as the rafts begin to move, which would have required the use of integrals to determine the motion of each person.

10.3.1 The centre of mass for a continuous object

So far, we have considered the centre of mass for a system made of point particles. In this section, we show how one can determine the centre of mass for a "continuous object"[6]. We previously argued that if an object is uniform and symmetric, its centre of mass will be located at the centre of the object. Let us show this explicitly for a uniform rod of total mass M and length L, as depicted in Figure 10.13.

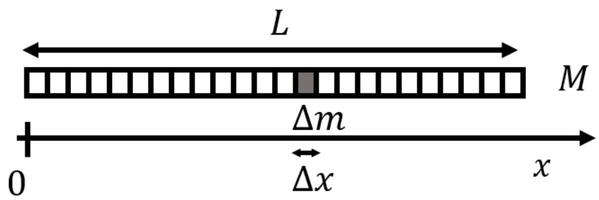


Figure 10.13: A rod of length L and mass M.

In order to determine the centre of mass of the rod, we first model the rod as being made of N small "mass elements" each of equal mass, Δm, and of length Δx, as shown in Figure 10.13. If we choose those mass elements to be small enough, we can model them as point particles, and use the same formulas as above to determine the centre of mass of the rod.

We define the x axis to be co-linear with the rod, such that the origin is located at one end of the rod. We can define the "linear mass density" of the rod, λ, as the mass per unit length of the rod:

$$\lambda = \frac{M}{L}.$$

A small mass element of length Δx, will thus have a mass, Δm, given by:

$$\Delta m = \lambda \Delta x$$

If there are N mass elements that make up the rod, the x position of the centre of mass of the rod is given by:

$$\begin{aligned} x_{CM} &= \frac{1}{M}\sum_i^N m_i x_i = \frac{1}{M}\sum_i^N \Delta m x_i \\ &= \frac{1}{M}\sum_i^N \lambda \Delta x x_i \end{aligned}$$

[6]In reality, there are of course no continuous objects since, at the atomic level, everything is made of particles.

where x_i is the x coordinate of the i-th mass element. Of course, we can take the limit over which the length, Δx, of each mass element goes to zero to obtain an integral:

$$x_{CM} = \lim_{\Delta x \to 0} \frac{1}{M} \sum_i^N \lambda \Delta x x_i = \frac{1}{M} \int_0^L \lambda x dx$$

where the discrete variable x_i became the continuous variable x, and Δx was replaced by dx (which is the same, but indicates that we are taking the limit of $\Delta x \to 0$). The integral is easily found:

$$\begin{aligned} x_{CM} &= \frac{1}{M} \int_0^L \lambda x dx = \frac{1}{M} \lambda \left[\frac{1}{2} x^2 \right]_0^L \\ &= \frac{1}{M} \lambda \frac{1}{2} L^2 = \frac{1}{M} \left(\frac{M}{L} \right) \frac{1}{2} L^2 \\ &= \frac{1}{2} L \end{aligned}$$

where we substituted the definition of λ back in to find, as expected, that the centre of mass of the rod is half its length away from one of the ends.

Suppose that the rod was instead not uniform and that its linear density depended on the position x along the rod:

$$\lambda(x) = 2a + 3bx$$

We can still find the centre of mass by considering an infinitesimally small mass element of mass dm, and length dx. In terms of the linear mass density and length of the mass element, dx, the mass dm is given by:

$$dm = \lambda(x) dx$$

The x position of the centre of mass is thus found the same way as before, except that the linear mass density is now a function of x:

$$\begin{aligned} x_{CM} &= \frac{1}{M} \int_0^L \lambda(x) x dx = \frac{1}{M} \int_0^L (2a + 3bx) x dx = \frac{1}{M} \int_0^L (2ax + 3bx^2) dx \\ &= \frac{1}{M} \left[ax^2 + bx^3 \right]_0^L \\ &= \frac{1}{M} (aL^2 + bL^3) \end{aligned}$$

In general, for a continuous object, the position of the centre of mass is given by:

$$\vec{r}_{CM} = \frac{1}{M}\int \vec{r}dm$$
$$\therefore x_{CM} = \frac{1}{M}\int xdm$$
$$\therefore y_{CM} = \frac{1}{M}\int ydm$$
$$\therefore z_{CM} = \frac{1}{M}\int zdm \tag{10.15}$$
$$\tag{10.16}$$

where in general, one will need to write dm in terms of something that depends on position (or a constant) so that the integrals can be evaluated over the spatial coordinates (x,y,z) over the range that describe the object. In the above, we wrote $dm = \lambda dx$ to express the mass element in terms of spatial coordinates.

Example 10-12

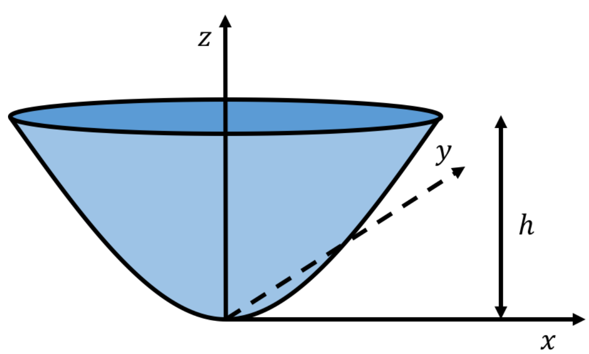


Figure 10.14: A symmetric bowl with parabolic sides is completely filled with water. The bowl has a height h.

A bowl of height h has parabolic sides and a circular cross-section, as illustrated in Figure 10.14. The bowl is filled with water. The bowl itself has a negligible mass and thickness, so that the mass of the full bowl is dominated by the mass of the water. Where is the centre of mass of the full bowl?

Solution

We can define a coordinate system such that the origin is located at the bottom of the bowl and the z axis corresponds to the axis of symmetry of the bowl. Because the bowl is full of water, and the bowl itself has negligible mass, we can model the full bowl as a uniform body of water with the same shape as the bowl and (volume) mass density ρ equal to the density of water. Furthermore, by symmetry, the centre of mass of the bowl will be on the z axis.

Because the bowl has a circular cross-section, we can divide it up into disk-shaped mass elements, dm, that have an infinitesimally small height dz, and a radius $r(z)$, that depends on their z coordinate (Figure 10.14).

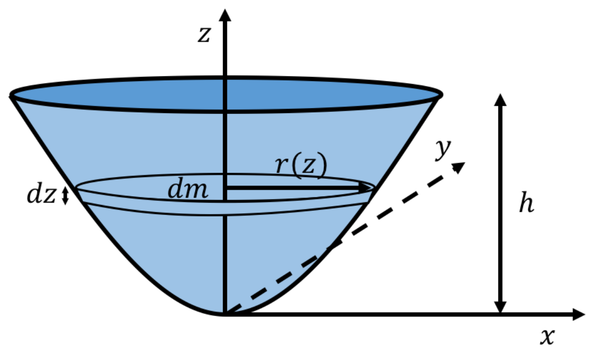


Figure 10.15: The parabolic bowl divided up into disk-shaped mass elements, dm, that have an infinitesimally small height dz, and a radius $r(z)$, that depends on their z coordinate.

The centre of mass of each disk-shaped mass element will be located where the corresponding disk intersects the z axis. The mass of one disk element is given by:

$$dm = \rho dV = \rho\pi r^2(z)dz$$

where $dV = \pi r(z)^2 dz$ is the volume of the disk with radius $r(z)$ and thickness dz. The radius of the infinitesimal disk depends on its z position, since the radii of the different disks must describe a parabola:

$$\begin{aligned}
z(r) &= \frac{1}{a^2}r^2 \\
r(z) &= a\sqrt{z} \\
\therefore dm &= \rho\pi r^2(z)dz = \rho\pi a^2 z dz
\end{aligned}$$

where we introduced the constant a so that the dimensions are correct. The constant a determines how "steep" the parabolic sides are. The z coordinate of the centre of mass is thus given by:

$$\begin{aligned}
z_{CM} &= \frac{1}{M}\int z dm = \frac{1}{M}\int_0^h z(\rho\pi a^2 z dz) = \frac{\rho\pi a^2}{M}\int_0^h z^2 dz \\
&= \frac{\rho\pi a^2}{M}\left[\frac{1}{3}z^3\right]_0^h \\
&= \frac{\rho\pi a^2}{3M}h^3
\end{aligned}$$

However, we are not quite done, since we do not know the total mass, M, of the water. To find the total mass of water, M, we proceed in an analogous way, and determine the value of the sum (integral) of all of the mass elements:

$$M = \int dm = \int_0^h \rho\pi a^2 z dz = \rho\pi a^2 \left[\frac{1}{2}z^2\right]_0^h = \frac{1}{2}\rho\pi a^2 h^2$$

Substituting this value for M, we can determine the z coordinate of the centre of mass of the full bowl:

$$z_{CM} = \frac{\rho\pi a^2}{3M}h^3 = \frac{2\rho\pi a^2}{3\rho\pi a^2 h^2}h^3 = \frac{2}{3}h$$

Regardless of the actual shape of the parabola (the parameter a), the centre of mass will always be two thirds of the way up from the bottom of the bowl.

Discussion: In determining the centre of mass of a three dimensional object, we used symmetry to argue that the x and y coordinates would be zero. We then found the z position of the centre of mass by dividing up the bowl into infinitesimally small mass elements (disks) along the direction in which we needed to find the centre of mass coordinate.

Checkpoint 10-4

True or False: The centre of mass of a continuous object is always located within the object.

A) True
B) False

10.4 Summary

Key Takeaways

The momentum vector, $\vec{p}$, of a point particle of mass, m, with velocity, $\vec{v}$, is defined as:

$$\vec{p} = m\vec{v}$$

We can write Newton's Second Law for a point particle in term of its momentum:

$$\frac{d}{dt}\vec{p} = \sum \vec{F} = \vec{F}^{net}$$

where the net force on the particle determines the rate of change of its momentum. In particular, if there is no net force on a particle, its momentum will not change.

The net impulse vector, $\vec{J}^{net}$, is defined as the net force exerted on a particle integrated from a time t_A to a time t_B:

$$\vec{J}^{net} = \int_{t_A}^{tB} \vec{F}^{net} dt$$

The net impulse vector is also equal to the change in momentum of the particle in that same period of time:

$$\vec{J}^{net} = \Delta\vec{p} = \vec{p}_B - \vec{p}_A$$

When we define a system of particles, we can distinguish between internal and external forces. Internal forces are those forces exerted by the particles in the system on each other. External forces are those forces on the particles in the system that are not exerted by the particles on each other. The sum over all of the forces on all of the particles in the system will be equal to the sum over the external forces, because the sum over all internal forces on a system is always zero (Newton's Third Law).

The total momentum of a system, $\vec{P}$, is the sum of the momenta, $\vec{p}_i$, of all of the particles in the system:

$$\vec{P} = \sum \vec{p}_i$$

The rate of change of the momentum of a system is equal to the sum of the external forces exerted on the system:

$$\frac{d}{dt}\vec{P} = \sum \vec{F}^{ext}$$

which can be thought of as an equivalent description as Newton's Second Law, but for the system as a whole. If the net (external) force on a system is zero, then the total momentum of the system is conserved.

Collisions are those events when the particles in a system interact (e.g. by colliding) and change their momenta. When modelling collisions, it is usually beneficial to first define a system for which the total momentum is conserved before and after the collision.

Collisions can be elastic or inelastic. Elastic collisions are those where, in addition to the total momentum, the total mechanical energy of the system is conserved. The total mechanical energy can usually be taken as the sum of the kinetic energies of the particles in the system.

Inelastic collisions are those in which the total mechanical energy of the system is not conserved. One can usually identify if mechanical energy was introduced or removed from the system and determine if the collision is elastic. It is important to identify when momentum and mechanical energy are conserved. Momentum is conserved if no net force is exerted on the system, whereas mechanical energy is conserved if no net work was done on the system by non-conservative forces (internal or external) or by external conservative forces.

We can always choose in which frame of reference to model a collision. In some cases, it is convenient to use the frame of reference of the centre of mass of the system, because in that frame of reference, the total momentum of the system is zero.

If a system has a total mass M, then one can use Newton's Second Law to describe its motion:

$$\begin{aligned}\sum \vec F^{ext} &= M\vec a_{CM}\\ \sum \vec F^{ext} &= \frac{d}{dt}\vec P\end{aligned}$$

where the sum of the forces is over all of the external forces on the system. The acceleration vector, $\vec a_{CM}$, describes the motion of the "centre of mass" of the system. $\vec P = M\vec v_{CM}$ is the total momentum of the system.

The centre of mass of a system is a mass-weighted average of the positions of all of the particles of mass m_i and position $\vec r_i$ that comprise the system:

$$\vec r_{CM} = \frac{1}{M}\sum_i m_i \vec r_i$$

The vector equation can be broken into components to find the x, y, and z component of the position of the centre of mass. Similarly, one can also define the velocity of the centre of mass of the system, in terms of the individual velocities, $\vec v_i$, of the particles in the system:

$$\vec{v}_{CM} = \frac{1}{M}\sum_i m_i \vec{v}_i$$

Finally, one can define the acceleration of the centre of mass of the system, in terms of the individual accelerations, $\vec{a}_i$, of the particles in the system:

$$\vec{a}_{CM} = \frac{1}{M}\sum_i m_i \vec{a}_i$$

If the system is a continuous object, we can find its centre of mass using a sum (integral) of infinitesimally small mass elements, dm, weighted by their position:

$$\begin{aligned}
\vec{r}_{CM} &= \frac{1}{M}\int \vec{r}dm \\
\therefore x_{CM} &= \frac{1}{M}\int xdm \\
\therefore y_{CM} &= \frac{1}{M}\int ydm \\
\therefore z_{CM} &= \frac{1}{M}\int zdm
\end{aligned}$$

The strategy to set up the integrals above is usually to express the mass element, dm, in terms of the position and density of the material of which the object is made. One can then integrate over position in the range defined by the dimensions of the object.

Important Equations

Momentum of a point particle:

$$\vec{p} = m\vec{v}$$
$$\frac{d}{dt}\vec{p} = \sum \vec{F} = \vec{F}^{net}$$

Impulse:

$$\vec{J}^{net} = \int_{t_A}^{tB} \vec{F}^{net} dt$$
$$\vec{J}^{net} = \Delta\vec{p} = \vec{p}_B - \vec{p}_A$$

Momentum of a system:

$$\vec{P} = \sum \vec{p}_i$$
$$\frac{d}{dt}\vec{P} = \sum \vec{F}^{ext}$$

Newton's Second Law for a system:

$$\sum \vec{F}^{ext} = M\vec{a}_{CM}$$
$$\sum \vec{F}^{ext} = \frac{d}{dt}\vec{P}$$

Position of the Centre of Mass of a system:

$$\vec{r}_{CM} = \frac{1}{M}\sum_i m_i\vec{r}_i$$

Velocity of the Centre of Mass of a system:

$$\vec{v}_{CM} = \frac{1}{M}\sum_i m_i\vec{v}_i$$

Acceleration of the Centre of Mass of a system:

$$\vec{a}_{CM} = \frac{1}{M}\sum_i m_i\vec{a}_i$$

Position of the Centre of Mass for a continuous object:

$$\vec{r}_{CM} = \frac{1}{M}\int \vec{r}dm$$
$$\therefore x_{CM} = \frac{1}{M}\int xdm$$
$$\therefore y_{CM} = \frac{1}{M}\int ydm$$
$$\therefore z_{CM} = \frac{1}{M}\int zdm$$

10.5 Thinking about the material

Reflect and research

1. Explain how Newton's Cradle illustrates the conservation of momentum. Are the collisions in Newton's Cradle elastic? Explain!
2. Gymnasts have specially engineered "crash mats" for landing after doing spins and flips in the air. Why do these crash mats have to be specially engineered, and why can't the gymnast just use a big pile of blankets?
3. Give 2 examples where the centre of mass of an object is not located inside of the object.
4. The Volvo XC60 is supposedly the safest car in the world that money can buy. Why is this?
5. In the boxing world, boxers try to "ride the punch". Research and explain how this method helps boxers to reduce injuries.

To try at home

1. Grab two or three of your friends and ask them to hold a bed sheet. Throw an egg at full speed onto the bed sheet. What happens to the egg, and why?
2. Verify that in a 1 one-dimensional elastic collision between two objects of the same mass, if one object starts at rest, the other object will end at rest after the collision (look up Newton's Cradle to get an idea).

To try in the lab

1. Propose an experiment to test whether a collision is elastic.
2. Propose an experiment to test whether momentum is conserved in a two dimensional collision.

10.6 Sample problems and solutions

10.6.1 Problems

Problem 10-1:

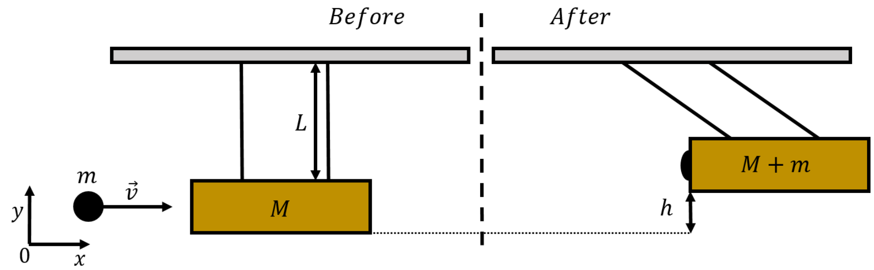


Figure 10.16: A bullet of mass m strikes and embeds itself into a ballistic pendulum of mass M.

A ballistic pendulum is a device that can be built to measure the speed of a projectile. The pendulum is constructed such that the projectile is fired at the bob of the pendulum (typically a block of wood) which then swings as illustrated in Figure 10.16, with the projectile embedded within. By measuring the height that is reached by the pendulum's bob, one can determine the speed of the projectile before it collided with the pendulum. If a ballistic pendulum with a mass M suspended at the end of strings of length L is observed to rise by a height h after being struck by a bullet of mass m, how fast was the bullet moving? (Solution)

Problem 10-2:

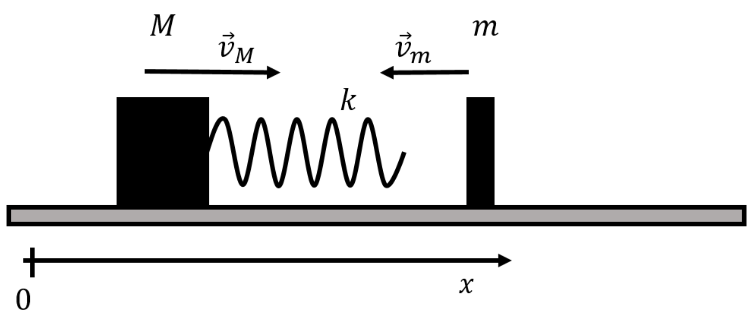


Figure 10.17: One block attached to a spring about to collide with another block.

A block of mass M with a spring of spring constant k attached to it is sliding on a frictionless surface with velocity $\vec{v}_M$ in the x direction. A second block of mass m has velocity $\vec{v}_m$ also in the x direction (shown above in the negative x direction, but let us assume that we do not necessarily know the direction, only that the two blocks will collide). During the collision between the blocks, what is the maximum amount by which the spring is compressed? (Solution)

Problem 10-3: A uniform wire is bent into a semi-circle of radius R. Where is the centre of mass of the wire? (Solution)

10.6.2 Solutions

Solution to problem 10-1: We can model this situation by dividing it into three phases:

1. Before the bullet collides with the pendulum, only the bullet has momentum in the x direction.
2. Immediately after the **inelastic** collision, the bullet and pendulum form a combined object of mass $M+m$ that has the same momentum as the bullet, in the x direction, before the pendulum starts to swing upwards.
3. The pendulum with the embedded bullet swings upwards until its kinetic energy is zero.

The collision between the bullet and pendulum is inelastic, because some of the kinetic energy of the bullet is used to deform the bullet and the pendulum. In general, any collision where two objects end up "stuck together" is inelastic.

In order to model the pendulum's motion we first apply conservation of momentum to determine the speed, v', of the pendulum and embedded bullet just after the collision. Applying conservation of momentum in the x direction to the system formed by the pendulum and the bullet, just before and after the collision, we have:

$$\begin{aligned}
P &= mv \\
P' &= (M+m)v' \\
\therefore mv &= (M+m)v' \\
\therefore v' &= \frac{m}{m+M}v
\end{aligned}$$

where P and P' are the initial and final momenta of the system, respectively. The pendulum with the bullet embedded in it will thus have a speed of v' at the bottom of the pendulum's motion, before it swings upwards.

We can now use conservation of energy to model the swinging motion since, at that point, only tension and gravity act on the pendulum, and there are no non-conservative forces. If we choose the origin to be the location of the pendulum at the bottom of its trajectory, its initial gravitational potential energy is zero and its initial mechanical energy, E, is given by:

$$E = \frac{1}{2}(m+M)v'^2$$

At the top of the trajectory, the pendulum with the embedded bullet will stop and have no kinetic energy. The mechanical energy at the top of the trajectory, E', is thus equal to the gravitational potential energy of the pendulum at a height h above the origin:

$$E' = (m+M)gh$$

Applying conservation of mechanical energy allows us to find the initial speed of the bullet:

$$\begin{aligned}
E &= E' \\
\frac{1}{2}(m+M)v'^2 &= (m+M)gh \\
v'^2 &= 2gh \\
\left(\frac{m}{m+M}v\right)^2 &= 2gh \\
\therefore v &= \frac{m+M}{m}\sqrt{2gh}
\end{aligned}$$

where is the second last line we used the expression for v' that we obtained from conservation of momentum.

Discussion: This example showed a situation in which momentum and energy were both conserved, but not at the same time. This example also highlighted how, by using conservation laws, one can derive models that are much easier to solve mathematically than if one had to model all of the forces involved.

Solution to problem 10-2: The collision is elastic because the energy used to compress the spring is "given back" when the spring extends again, since the spring force is conservative.

They key to modelling the compression of the spring is to identify the condition under which the spring is maximally compressed. This will occur at the point during the collision where the two masses will have exactly the same velocity, momentarily moving in unison as the spring is maximally compressed. Because, instantaneously, the masses have the same velocity, there is a frame of reference in which the two masses are at rest, and the momentum is zero. Of course, that frame of reference is the centre of mass frame of reference.

Because the collision is one-dimensional, we can calculate the velocity of the centre of mass as:

$$v_{CM} = \frac{Mv_M + mv_m}{m+M}$$

where we note that v_m is a negative number, since the block of mass m is moving in the negative x direction. The total momentum, $\vec{P}^{CM}$, in the centre of mass frame of reference must be zero. Writing this out for the x component and transforming the velocities of the two blocks into the centre of mass frame of reference:

$$\begin{aligned}
P_x^{CM} = M(v_M - v_{CM}) + m(v_m - v_{CM}) = 0 \\
\therefore (v_m - v_{CM}) = -\frac{M}{m}(v_M - v_{CM})
\end{aligned}$$

Also note that we can write the velocity difference $v_M - v_{CM}$ without using the centre of mass velocity:

$$\begin{aligned}
v_M - v_{CM} &= v_M - \frac{Mv_M + mv_m}{m+M} = \frac{1}{m+M}(v_M(m+M) - Mv_M - mv_m) \\
&= \frac{m}{m+M}(v_M - v_m)
\end{aligned}$$

We can then use conservation of energy in the centre of mass frame to determine the maximal compression of the spring. Before the collision, the total mechanical energy in the system, E, is the sum of the kinetic energies of the two blocks (as the spring is not compressed):

$$\begin{aligned}
E &= \frac{1}{2}m(v_m - v_{CM})^2 + \frac{1}{2}M(v_M - v_{CM})^2 \\
&= \frac{1}{2}\frac{M^2}{m}(v_M - v_{CM})^2 + \frac{1}{2}M(v_M - v_{CM})^2 \\
&= \frac{1}{2}M\left(1 + \frac{M}{m}\right)(v_M - v_{CM})^2 \\
&= \frac{1}{2}M\left(\frac{m+M}{m}\right)(v_M - v_{CM})^2 \\
&= \frac{1}{2}M\left(\frac{m+M}{m}\right)\left(\frac{m}{m+M}(v_M - v_m)\right)^2 \\
&= \frac{1}{2}\left(\frac{mM}{m+M}\right)(v_M - v_m)^2
\end{aligned}$$

where we used our expressions above to simplify the expression. When the spring is maximally compressed, the two blocks are at rest and the mechanical energy of the system, E', is all "stored" as spring potential energy:

$$E' = \frac{1}{2}kx^2$$

where x is the distance by which the spring is compressed. Equating the two allows us to determine the maximal compression of the spring:

$$\begin{aligned}
E &= E' \\
\frac{1}{2}\left(\frac{mM}{m+M}\right)(v_M - v_m)^2 &= \frac{1}{2}kx^2 \\
\therefore x &= \sqrt{\frac{1}{k}\left(\frac{mM}{m+M}\right)}(v_M - v_m)
\end{aligned}$$

Discussion: By modelling the collision in the centre of mass frame of reference, we were easily able to determine the maximal compression of the spring. This would have been more difficult in the lab frame of reference, because the two blocks would still be moving when the spring is maximally compressed, so we would have needed to determine their speeds to determine the total mechanical energy when the spring is compressed.

When we calculated the initial kinetic energy, we found that it was given by:

$$E = \frac{1}{2}\left(\frac{mM}{m+M}\right)(v_M - v_m)^2 = \frac{1}{2}M_{red}(v_M - v_m)^2$$

The combination of masses in parentheses is called the "reduced mass" of the system, and is a sort of effective mass that can be used to model the system as a whole.

Solution to problem 10-3: The curved wire is illustrated in Figure 10.18, along with a small mass element, dm, on the wire, and our choice of coordinate system (centred at the centre of the semi-circle). By symmetry, the position of the centre of mass will be located at $x = 0$, so we only need to determine the y position.

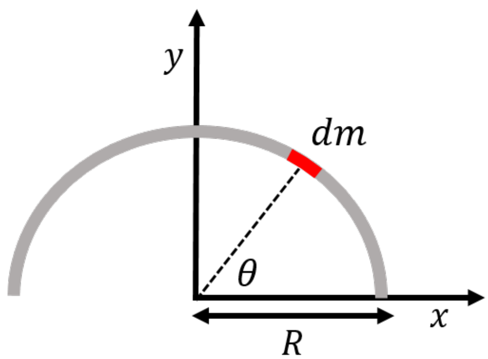


Figure 10.18: A uniform wire bent into a semi circle of radius R, and a small mass element, dm, on the wire.

The y position of the centre of mass is given by:

$$y_{CM} = \frac{1}{M}\int y dm$$

where M is the total mass of the wire. We can define the mass per unit length, λ, for the wire as:

$$\lambda = \frac{M}{\pi R}$$

We will choose to integrate the equation for the y position of the centre of mass over θ (from 0 to π), instead of over y, as it will make the integral easier (it is easier to express dm in terms of $d\theta$ than dy because the wire is curved). θ is the angle at which the mass element is located. The mass element forms an arc on the wire of length ds that subtends an angle $d\theta$. The two are related by:

$$ds = Rd\theta$$

The mass element, dm, can then be expressed in terms of the mass per unit length of the wire and the length, $Rd\theta$, of the mass element:

$$dm = \lambda ds = \lambda R d\theta$$

We also need to express the y position of the mass element using θ:

$$y = R\sin\theta$$

Now that we have expressed dm and y in terms of θ, we can determine the y position of the centre of mass:

$$\begin{aligned} y_{CM} &= \frac{1}{M}\int ydm = \frac{1}{M}\int_0^{\pi} R\sin\theta\lambda Rd\theta \\ &= \frac{R^2\lambda}{M}\int_0^{\pi}\sin\theta d\theta = \frac{R^2\lambda}{M}\Big[-\cos\theta\Big]_0^{\pi} \\ &= \frac{2R^2\lambda}{M} = \frac{2R}{\pi} \end{aligned}$$

where in the last equality, we used the expression for the mass per unit length, λ, obtained above.

11 Rotational dynamics

In this Chapter, we use Newton's Second Law to develop a formalism to describe how objects rotate. In particular, we will introduce the concept of torque which plays a similar role to that of force in non-rotational dynamics. We will also introduce the concept of moment of inertia to describe how objects resist rotational motion.

Learning Objectives

- Understand how to use vector quantities for describing the kinematics of rotations.
- Understand how to use torque to determine the angular acceleration of an object.
- Understand conditions for static and dynamic equilibrium.
- Understand how to determine the moment of inertia of an object.

Think About It

A construction worker would like to lift one end of a heavy block from the ground using a bar propped against a rock on the ground as a lever. Should he place the rock close or far from the block to make it easier to lift the block?

A) It will be easier to lift the block if the rock is close to the block.
B) It will be easier to lift the block if the rock is far from the block.
C) It does not matter where he places the rock, as long as the bar is short.
D) It does not matter where he places the rock, as long as the bar is long.

11.1 Rotational kinematic vectors

Review Topics

Before proceeding, you may wish to review:

- Section 4.4 on kinematics for circular motion.
- Section A.3.4 on the vector product.
- Section A.4.3 on axial vectors and their use in defining rotational quantities.

11.1.1 Scalar rotational kinematic quantities

Recall that we can describe the motion of a particle along a circle of radius, R, by using its angular position, θ, its angular velocity, ω, and its angular acceleration, α. With a suitable choice of coordinate systen, the angular position can be defined as the angle made by the position vector of the particles, $\vec{r}$, and the x axis of a coordinate system whose origin is the centre of the circle, as shown in Figure 11.1.

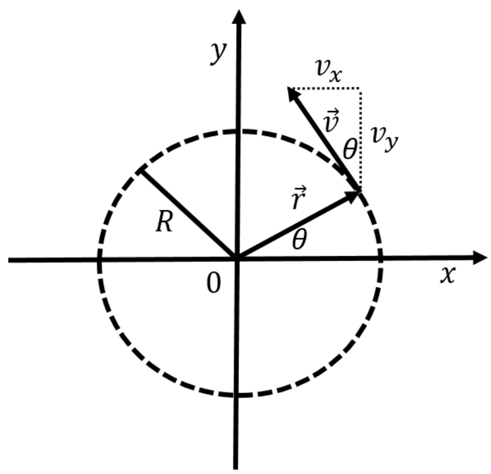


Figure 11.1: Angular position for a particle moving around the z axis (out of the page), along a circle of radius R with a centre at the origin.

The angular velocity, ω, is the rate of the change of the angular position, and the angular acceleration, α, is the rate of change of the angular velocity:

$$\omega = \frac{d}{dt}\theta$$
$$\alpha = \frac{d}{dt}\omega$$

If the angular acceleration is constant, then angular velocity and position as a function of time are given by:

$$\omega(t) = \omega_0 + \alpha t$$
$$\theta(t) = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2$$

where θ_0 and ω_0 are the angular position and velocity, respectively, at $t = 0$.

We can also describe the motion of the particle in terms of "linear" quantities (as opposed to "angular" quantities) along a one-dimensional axis that is curved along the circle. If s is the distance along the circumference of the circle, measured counter-clockwise from where the circle intersects the x axis, then it is related to the angular displacement:

$$s = R\theta$$

if θ is expressed in radians. Similarly, the linear velocity along the s axis, v_s, and the corresponding acceleration, a_s, are given by:

$$\begin{aligned} v_s &= \frac{ds}{dt} = \frac{d}{dt}R\theta = R\omega \\ a_s &= \frac{dv}{dt} = \frac{d}{dt}R\omega = R\alpha \end{aligned}$$

where the radius of the circle, R, is a constant that can be taken out of the time derivatives. For motion along a circle, the velocity vector, $\vec{v}$, of the particle is always tangent to the circle (Figure 11.1), so v_s corresponds to the speed of the particle. The acceleration vector, $\vec{a}$, is in general not tangent to the circle; a_s represents the component of the acceleration vector that is tangent to the circle. If $a_s = 0$, then $\alpha = 0$, and the particle is moving with a constant speed (uniform circular motion), and the acceleration vector points towards the centre of the circle.

Checkpoint 11-1

Which of the following statements correctly describes the speeds at points A and B on the disk rotating about an axis through its centre, as illustrated in Figure 11.2?

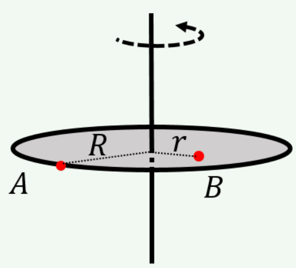


Figure 11.2: Two points at different radii on a rotating disk.

A) Both points A and B have the same angular and linear speeds.
B) Both points A and B have the same linear speed but they have different angular speeds.
C) Both points A and B have the same angular speed but they have different linear speeds.

11.1.2 Vector rotational kinematic quantities

In the previous section, we defined angular quantities to describe the motion of a particle about the z axis along a circle of radius R that lies in the xy plane. By using vectors, we can define the angular quantities for rotation about an **axis that can point in any direction**. Given an axis of rotation, the path of any particle rotating about that axis can be described by a circle that lies in the plane perpendicular to that axis of rotation, as illustrated in Figure 11.3.

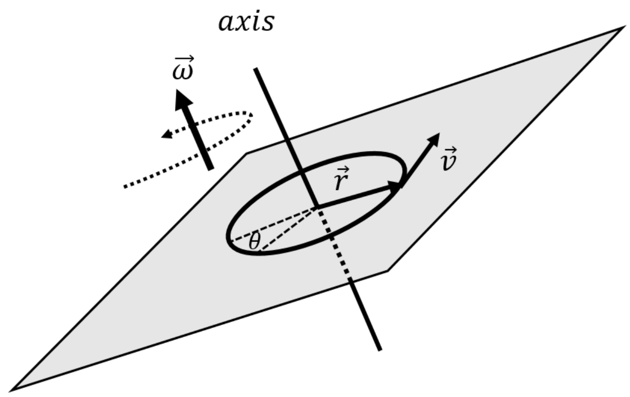


Figure 11.3: Defining the vector $\vec{r}$ and the angular velocity, $\vec{\omega}$ for a particle with velocity $\vec{v}$ rotating about an axis in a general direction.

We define the vector, $\vec{r}$, for a particle to be the vector that goes from the axis of rotation to the particle and is in a plane perpendicular to the axis of rotation, as in Figure 11.3. Given the velocity vector of the particle, $\vec{v}$, we define its angular velocity vector, $\vec{\omega}$, **about the axis of rotation**, as:

$$\boxed{\vec{\omega} = \frac{1}{r^2}\vec{r}\times\vec{v}} \tag{11.1}$$

The angular velocity vector is perpendicular to both the velocity vector and the vector $\vec{r}$, since it is defined as their cross-product. Thus, the **angular velocity vector is co-linear with the axis of rotation**. By using the angular velocity vector, we can specify **the direction of the axis of rotation as well as the direction in which the particle is rotating about that axis**. The direction of rotation is given by the right hand rule for axial vectors: when you point your thumb in the same direction as the angular velocity vector, the direction of rotation is the direction that your fingers point when you curl them, as illustrated in Figure. 11.4.

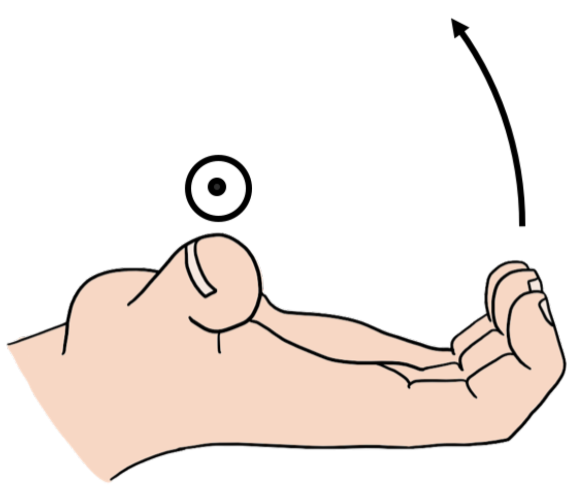


Figure 11.4: Using the right hand rule for axial vectors. In this case, the direction of rotation is counter clockwise when looking at the page (the direction that the fingers curl), so the rotation vector points out of the page (the direction of the thumb).

This definition of the angular velocity is consistent with the description from the previous section for motion about a circle of radius R that lies in the xy plane, as in Figure 11.1. In that case, the magnitude of the angular velocity is given by:

$$\begin{aligned}\omega &= \frac{1}{r^2}||\vec{r}\times\vec{v}|| = \frac{1}{r^2}rv\sin\phi = \frac{v}{R}\\ \therefore v &= R\omega\end{aligned}$$

where ϕ is the angle between the vectors $\vec{r}$ and $\vec{v}$ (90° for motion around a circle). The direction of the angular velocity in Figure 11.1 is in the positive z direction, which corresponds to counter-clockwise rotation about the z axis.

Checkpoint 11-2

You push on the right-hand side of a door to open it, as the door's hinges are on the left. The angular velocity vector of the door is:

A) Upwards
B) Downwards
C) Forwards
D) Backwards

One can always define an angular velocity vector **relative to a point of rotation**, even if the particle is not moving along a circle. If we define the vector $\vec{r}$ to be the vector from the point of rotation to the particle, then the angular velocity vector describes the motion of the particle as if it were instantaneously moving in a circle centred at the point of rotation, in a plane given by the vectors $\vec{r}$ and $\vec{v}$.

Consider, for example, the particle in Figure 11.5 which is moving in a straight line with a velocity vector in the xy plane at a position $\vec{r}$ relative to the origin. We can define its angular velocity vector relative to the origin, which will be in the positive z direction.

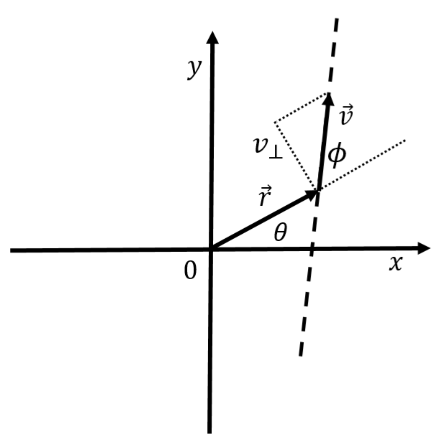


Figure 11.5: Angular position for a particle moving in a straight line.

The angular velocity describes the motion of the particle as if it were **instantaneously moving along a circle of radius r centred about the origin**. The angular velocity is related to the component of $\vec{v}$, $v_\perp$, that is perpendicular to $\vec{r}$ (which is the component tangent to the circle of radius r, in Figure 11.5):

$$||\vec{\omega}|| = \frac{1}{r^2}||\vec{r}\times\vec{v}|| = \frac{v\sin\phi}{r} = \frac{v_\perp}{r} \tag{11.2}$$

where ϕ is the angle between $\vec{r}$ and $\vec{v}$.

Similarly, we can define the angular acceleration vector, $\vec{\alpha}$, about an axis of rotation:

$$\vec{\alpha} = \frac{1}{r^2}\vec{r}\times\vec{a} \tag{11.3}$$

where $\vec{a}$ is the particle's acceleration vector, and $\vec{r}$ is the vector from the axis of rotation to the particle. The direction of the angular acceleration is co-linear with the axis of rotation and the right-hand rule gives the rotational direction of the angular acceleration. We can also define the angular acceleration about a point; in that case, the direction of the vector will define an instantaneous axis of rotation about a circle of radius r centred at the point as well as the direction of the angular acceleration about that axis.

Finally, we can define an angular displacement vector, $\vec{\theta}$, relative to an axis of rotation. The direction of the angular displacement vector will be co-linear with the axis of rotation, its direction will indicate the direction of rotation about that axis, and its magnitude (in radians) will correspond to the angular displacement (as shown in Figure 11.3). We can only relate the angular displacement vector to an infinitesimal linear displacement vector, $d\vec{s}$, since the position vector $\vec{r}$ from the axis of rotation will be different at each end of the displacement vector if the displacement is large. The infinitesimal angular displacement vector that corresponds to an infinitesimal displacement vector, $d\vec{s}$, is defined as:

$$d\vec{\theta} = \frac{1}{r^2}\vec{r}\times d\vec{s}$$

Checkpoint 11-3

Which statement is correct regarding an ant on a disk that is rotating slower and slower as illustrated?

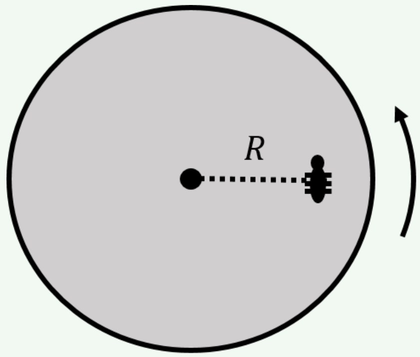


Figure 11.6: An ant on a disk.

A) The angular velocity points into the page and the angular acceleration points out of the page.
B) Both the angular velocity and acceleration point into the page.
C) Both the angular velocity and acceleration point out of the page.
D) The angular acceleration points into the page and the angular velocity points out of the page.

The instantaneous angular velocity vector is the rate of change of the angular displacement vector:

$$\vec{\omega} = \frac{d\vec{\theta}}{dt} = \frac{d}{dt}\frac{1}{r^2}\vec{r} \times d\vec{s} = \frac{1}{r^2}\vec{r} \times \vec{v}_s$$

where $\vec{v}_s$ is the (instantaneous) tangential velocity around the circle (i.e. the component of the velocity $\vec{v}$ that is perpendicular to $\vec{r}$). The angular acceleration vector is the rate of change of the angular velocity vector:

$$\vec{\alpha} = \frac{d}{dt}\vec{\omega}$$

Given the angular kinematic quantities, the related linear quantities at a position $\vec{r}$ from the axis of rotation are given by:

$$\begin{aligned} d\vec{s} &= d\vec{\theta} \times \vec{r} \\ \vec{v}_s &= \vec{\omega} \times \vec{r} \\ \vec{a}_s &= \vec{\alpha} \times \vec{r} \end{aligned} \tag{11.4}$$

where the linear quantities are always in the direction perpendicular to $\vec{r}$ (tangent to the circle, for motion around a circle). In other words, one cannot, say, take the acceleration vector, obtain the angular acceleration vector, and then get back the original acceleration vector - one will only get back the component of the acceleration vector that is perpendicular to $\vec{r}$.

Checkpoint 11-4

A particle has an angular velocity in the negative z direction. In which way is the particle's velocity vector at a point in its trajectory when it is on the positive y axis?

A) Positive z direction
B) Negative y direction
C) Positive x direction
D) Negative x direction

11.2 Rotational dynamics for a single particle

Suppose that a single force, $\vec{F}$, is acting on a particle of mass m. Newton's Second Law for the particle is then given by:

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

We can define a point of rotation such that $\vec{r}$ is the position of the particle relative to that point. We can take the cross-product of $\vec{r}$ with both sides of the equation in Newton's Second Law:

$$\vec{r} \times \vec{F} = m\vec{r} \times \vec{a}$$

The left hand-side of the equation is called "the torque of $\vec{F}$ relative to the point of rotation", and is usually denoted by $\vec{\tau}$:

$$\boxed{\vec{\tau} = \vec{r} \times \vec{F}} \tag{11.5}$$

The right-hand side of the equation is related to the angular acceleration vector, $\vec{\alpha}$, about that point of rotation:

$$m\vec{r} \times \vec{a} = mr^2\vec{\alpha}$$

Putting this altogether, we get:

$$\vec{\tau} = mr^2\vec{\alpha}$$

If more than one force is exerted on the particle, it is easy to show that the **net torque** from the net force on the particle **is equal to the sum of the torques on the particle**:

$$\begin{aligned}\vec{r} \times (\vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \dots) &= (\vec{r} \times \vec{F}_1 + \vec{r} \times \vec{F}_2 + \vec{r} \times \vec{F}_3 + \dots)\\ \therefore \vec{r} \times \sum \vec{F} &= \sum \vec{\tau} = \vec{\tau}^{net}\end{aligned}$$

We can write "Newton's Second Law for the rotational dynamics of a particle":

$$\boxed{\sum \vec{\tau} = \vec{\tau}^{net} = mr^2\vec{\alpha}} \tag{11.6}$$

This equation provides us an alternate formulation to Newton's Second Law that is useful for describing the motion of a particle that is rotating. The left-hand side of the equation corresponds to the "causes of motion" (much like the sum of the forces in Newton's Second Law), and the right-hand side of the equation to the inertia and the kinematics. A few things to note when comparing to Newton's Second Law:

1. The rotational quantities, torque and angular acceleration, **are only defined with respect to a point or axis of rotation** (as this determines the vector $\vec{r}$). If one chooses a different point of rotation, then the torque and angular acceleration will be different.
2. The angular acceleration of a particle is proportional to the net torque exerted on it, much like the linear acceleration is proportional to the net force exerted on the particle.
3. Torque about a centre of rotation can be thought of as the equivalent of a force that causes things rotate about an axis that goes through the point of rotation and that is parallel to the torque/angular acceleration vectors.
4. Instead of mass, it is mass times r^2 that plays the role of inertia and determines how large of an angular acceleration a particle will experience for a given net torque.

Example 11-1

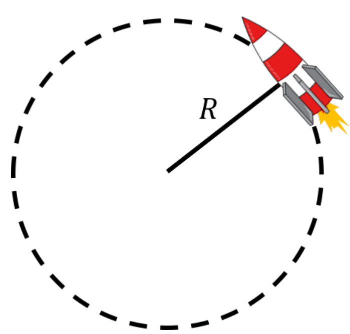


Figure 11.7: A toy rocket accelerating around a circle of radius R, as seen from above.

A toy rocket is attached to a string on a horizontal frictionless table, as shown in Figure 11.7. The rocket has a mass m and produces a constant force of thrust with a magnitude F that accelerates the rocket along a circle of radius R (the length of the string). If the rocket starts at rest, what distance along the circumference of the circle will the rocket have travelled after a time, t?

Solution

We can model the rocket as a point particle of mass m with the following forces exerted on it:

1. $\vec{F}$, the thrust of the rocket, always acting tangent to the circle.
2. $\vec{T}$, the force of tension in the string, always acting towards the centre of the circle.
3. $\vec{F}_g$, the rocket's weight, acting into the page, with magnitude mg.
4. $\vec{N}$, a normal force exerted by the table, out of the page, with magnitude mg.

Because the normal force and the weight are equal in magnitude and opposite in direction, the net force will be the sum of the force of thrust and the force of tension, which are always perpendicular to each other. Thinking about this with Newton's Second Law, we could model the force of thrust as increasing the speed of the particle, while the force of tension keeps the rocket moving in a circle (it can do no work to increase the speed, since it is always perpendicular to the motion).

We can also think about this in terms of torques and angular acceleration about the centre of the circle. The thrust will result in a net torque about the centre of rotation, which will lead to the rocket having an angular acceleration. By determining the angular acceleration, we can then model the displacement at some time, t, using kinematics. The force of tension will create no torque about the centre of the circle because the force of tension is always co-linear with the position vector, $\vec{r}$ (the cross-product of co-linear vectors is always zero).

We introduce a coordinate system whose origin coincides with the centre of the circle, as shown in Figure 11.8, so that $\vec{r}$ corresponds to the position of the rocket relative to the origin. The force of thrust and the tension are also shown in the diagram. We choose the direction of the x axis such that the rocket was located at the intersection of the x axis and the circle at time, $t = 0$.

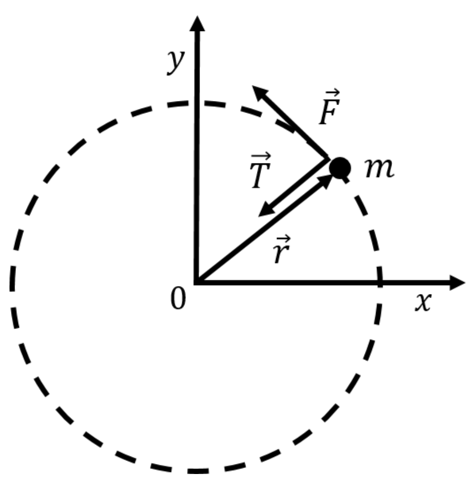


Figure 11.8: Coordinate system to describe the motion of the rocket.

The net torque on the rocket about the point of rotation is given by the cross-product between the thrust force, $\vec F$, and the position vector, $\vec r$:

$$\vec\tau^{net} = \vec r \times \vec F$$

and will point in the positive z direction (as given by the right hand rule). $\vec r$ and $\vec F$ are perpendicular, so the magnitude of the net torque is given by:

$$\tau^{net} = rF\sin(90^\circ) = RF$$

where R is the magnitude of $\vec r$. The net torque vector is thus:

$$\vec\tau^{net} = RF\hat z$$

Applying the rotational version of Newton's Second Law allows us to determine the angular acceleration:

$$\begin{aligned}
\vec\tau^{net} &= mr^2\vec\alpha \\
RF\hat z &= mR^2\vec\alpha \\
\therefore \vec\alpha &= \frac{F}{mR}\hat z
\end{aligned}$$

The angular acceleration vector points in the positive z direction (as does the net torque), and indicates that the rocket is accelerating in the counter-clockwise direction about the z axis.

After a period of time t, the rocket will have covered an angular displacement, $\Delta\theta$, given by:

$$\begin{aligned}
\Delta\theta = \theta(t) - \theta_0 &= \omega_0 t + \frac{1}{2}\alpha t^2 \\
&= \frac{1}{2}\frac{F}{mR}t^2
\end{aligned}$$

The linear displacement, Δs, that corresponds to this angular displacement is:

$$\Delta s = R\Delta\theta = \frac{1}{2}\frac{F}{m}t^2$$

Discussion: The formula that we found for the total linear displacement is the same that we would have found if the particle were moving in a straight line with a net force F applied to it (as the particle would have a constant acceleration given by F/m).

11.3 Torque

The torque associated with a force is a mathematical tool to describe how much a particular force will cause a particle (or solid) object to rotate about a given point or a given axis of rotation. A torque is **only defined relative to an axis or point of rotation**. It never makes sense to say "the torque is ...", and one should always say "the torque about this axis/point of rotation is ... ". Angular quantities (torque, angular velocity, angular displacement, etc) are only ever defined relative to a specific axis or point of rotation.

Mathematically, the torque vector from a force, $\vec{F}$, exerted at a position, $\vec{r}$, relative to the axis or point of rotation is defined as:

$$\vec{\tau} = \vec{r} \times \vec{F}$$

Note that the torque from a given force increases if that force is further from the axis of rotation (if $\vec{r}$ has a bigger magnitude).

Consider the solid disk of radius, r, depicted in Figure 11.9. The disk can rotate about an axis that passes through the centre of the disk and that is perpendicular to the plane of the disk. A force, $\vec{F}$, is exerted on the edge of the disk as shown.

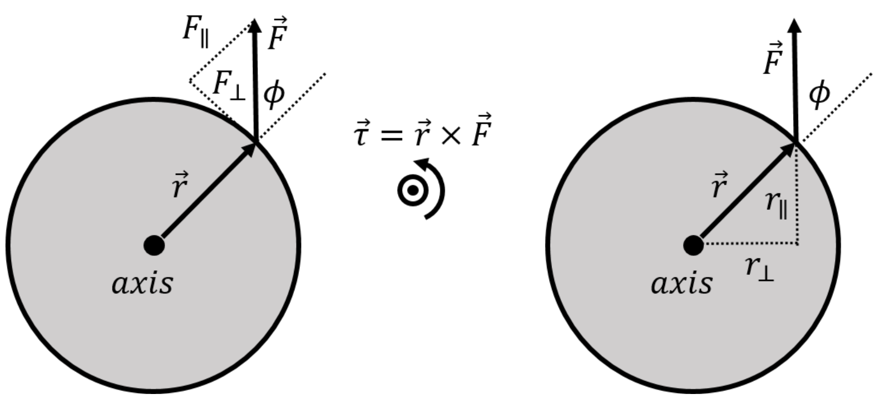


Figure 11.9: A force exerted on the perimeter of a disk that can rotate about an axis that is perpendicular to the disk and that passes through its centre. We can determine the resulting torque by considering either the component of $\vec{F}$ that is perpendicular to $\vec{r}$ (left panel) or the component of $\vec{r}$ that is perpendicular to $\vec{F}$ (right panel). The torque vector, $\vec{\tau}$, is out of the page, as illustrated in the centre.

Intuitively, that force will cause the disk to rotate in the counter-clockwise direction. The torque from the force $\vec{F}$ about the axis as rotation is given by:

$$\vec{\tau} = \vec{r} \times \vec{F}$$

where the vector $\vec{r}$ is perpendicular to the axis of rotation and goes from the axis of rotation to the point where $\vec{F}$ is exerted. The direction of the torque vector is out of the page (right hand rule, see Figure 11.9), and will thus lead to an angular acceleration that is also out of the page, which corresponds to the counter-clockwise direction, as anticipated.

We can break up the force into components that are parallel ($F_{\parallel}$) and perpendicular ($F_{\perp}$) to the vector $\vec{r}$, as shown on the left panel of Figure 11.9. Only the component of the force that is perpendicular to $\vec{r}$ will contribute to rotating the disk. Imagine that the force is from a string that you have attached to the perimeter of the disk; if you pull on the string such that the force is parallel to $\vec{r}$, the disk would not rotate. The magnitude of the torque is given by:

$$\tau = rF\sin\phi \tag{11.7}$$

where ϕ is the angle between $\vec{r}$ and $\vec{F}$, as shown in Figure 11.9. $F\sin\phi$ is precisely the component of $\vec{F}$ that is perpendicular to $\vec{r}$, so we could also write the magnitude of the torque as:

$$\tau = rF_{\perp}$$

which highlights that only the component of the force that is perpendicular to $\vec{r}$ contributes to the torque. Instead of combining the $\sin\phi$ with F to obtain $F_{\perp}$, the component of $\vec{F}$ perpendicular to $\vec{r}$, we can instead combine the $\sin\phi$ with r in Equation 11.7 to obtain $r_{\perp}$, the component of $\vec{r}$ that is perpendicular to $\vec{F}$. This is illustrated in the right panel of Figure 11.9. The magnitude of the torque is thus also given by:

$$\tau = r_{\perp}F$$

The quantity $r_{\perp}$ is called the "lever arm" of the force about a specific axis of rotation.

Emma's Thoughts

Remembering how to maximize the torque about an axis using a pencil

We already know that the greater the force that you apply, the more an object will rotate. Here is an easy way to quickly remind yourself of the two other factors that play a role in whether or not an object will rotate:

Torque about an axis increases if the force is applied further from the axis of rotation.

First, pinch the centre of your pencil. Try to make the pencil rotate by pushing right next to where you are pinching. Try making the pencil rotate again, by pushing near the eraser. You should notice that it is much easier to make the pencil rotate by pushing near the eraser, as it is further from the axis of rotation (the pinch).

Torque about an axis is maximized if the force is applied perpendicular to the object.

Next, you should try pushing on the top of the eraser of your pencil, parallel to the pencil. The pencil will not rotate. Now, try pushing on the eraser, but perpendicular to the pencil. In this case, the pencil will rotate.

If you are ever having trouble remembering the factors involved in maximizing torque about an axis, just grab your pencil case and do this quick exercise.

Checkpoint 11-5

Why is the handle of a door placed on the side of the door that is opposite to the hinges?

A) Because it increases the lever arm of a force used to rotate the door about the handle.
B) Because it increases the perpendicular component of force used to rotate the door about the hinges.
C) Because it increases the lever arm of a force used to rotate the door about the hinges.
D) Because it would be inconvenient if the handle were next to the hinges.

11.4 Rotation about an axis versus rotation about a point

When defining angular quantities (torque, angular acceleration, etc.), it is important to identify whether these are defined relative to an axis or to a point of rotation. This, in turn, determines the vector $\vec{r}$ that is involved in the definition of the angular quantities.

Consider a disk of radius r with a force, $\vec{F}$ exerted on its perimeter, as illustrated in Figure 11.10. The disk can only rotate about an axis that is perpendicular to the disk and that goes through the centre of the disk, like a wheel mounted on an axle. The force has a component, $\vec{F}_{plane}$, that lies in the plane perpendicular to the axis of rotation, and a component, $\vec{F}_{axis}$, that is parallel to axis of rotation.

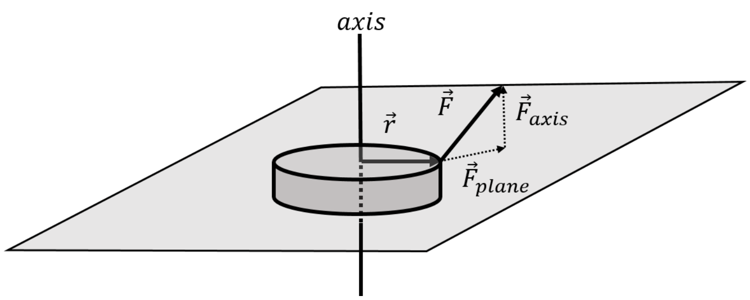


Figure 11.10: A force exerted on disk that can only rotate about an axis through its centre and perpendicular to its plane. Only the component of $\vec{F}$ that is in the plane perpendicular to the axis of rotation, $\vec{F}_{plane}$, will contribute to the torque about the axis of rotation.

The vector $\vec{r}$ is **always defined to be perpendicular to the axis of rotation and to go from the axis of rotation to the point where the force $\vec{F}$ is exerted**, as illustrated. The torque obtained by taking the cross product:

$$\vec{\tau} = \vec{r} \times \vec{F}$$

will be perpendicular to both $\vec{r}$ and $\vec{F}$, and will thus not be parallel to the axis of rotation. **Only the component of the torque that is parallel to the axis of rotation** will contribute to rotating the disk about the axis. Only the component of the force that lies in the plane perpendicular to the axis of rotation, $\vec{F}_{plane}$, will contribute to the component of the torque about that axis of rotation. Thus, when we need to determine the torque about an axis of rotation, we can **consider vectors $\vec{r}$ and $\vec{F}$ that lie in the plane perpendicular to the axis of rotation**. The torque of $\vec{F}$ relative to the axis of rotation is thus:

$$\vec{\tau}_{axis} = \vec{r} \times \vec{F}_{plane}$$

Furthermore, only the component of $\vec{F}_{plane}$ that is perpendicular to $\vec{r}$ will contribute to that torque, as we saw in the previous section.

In general, solid objects such as a disk can only rotate about an axis. In that case, one can consider only the components of forces that lie in the plane perpendicular to the axis of rotation in order to calculate the components of the torques about that axis that are parallel to that axis.

A point particle may be able to rotate about any axis that goes through a point of rotation. The net torque vector on the particle about that point will indicate the direction of the axis about which the particle would rotate. This is illustrated in the left panel of Figure 11.11.

Instead, if the particle were constrained to rotate about the z axis (e.g. if the particle is on a track), then we would use the component of the torque vector that is parallel to the z axis to describe its motion, as illustrated in the right panel. The z component of the torque could be determined by using only the components of the forces that lie in the plane perpendicular to the axis, and defining the vector $\vec{r}$ from the axis to the particle rather than from the point of rotation to the particle.

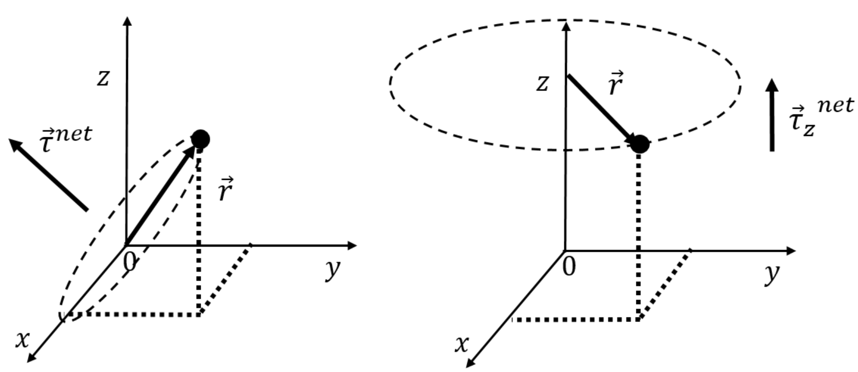


Figure 11.11: Left panel: a particle rotating about a circle centred at the origin with an axis determined from the net torque vector. Right panel: a particle that is constrained to rotate about the z axis.

Example 11-2

A force given by $\vec F = F_x\hat x + F_y\hat y + F_z\hat z$ is exerted at a position $\vec r = r_x\hat x + r_y\hat y + r_z\hat z$. Calculate the torque about the z axis as well as the torque about the origin.

Solution

To calculate the torque about the z axis, we need take the cross-product between the components of the vectors $\vec r$ and $\vec F$ that lie in the $x - y$ plane, since that is the plane perpendicular to the axis of rotation (the z axis). This gives:

$$\vec\tau_z = (r_x\hat x + r_y\hat y) \times (F_x\hat x + F_y\hat y) = (r_xF_y - r_yF_y)\hat z$$

If instead we want to calculate the torque about the origin, we take the cross-product between the two vectors:

$$\begin{aligned}\vec\tau &= (r_x\hat x + r_y\hat y + r_z\hat z) \times (F_x\hat x + F_y\hat y + F_z\hat z)\\ &= (r_yF_z - r_zF_y)\hat x + (r_zF_x - r_xF_z)\hat y + (r_xF_y - r_yF_y)\hat z\end{aligned}$$

If a particle were located at the given position, the force would cause the particle to (instantaneously) rotate about an axis that goes through the origin and is parallel to the torque vector.

Discussion: This example highlights the difference between calculating the torque about an axis of rotation and determining the torque about a point. When calculating the torque about an axis that goes through the origin, we only consider the components of the vectors $\vec r$ and $\vec F$ that are in the plane perpendicular to the axis of rotation. This would correspond to a situation in which the particle is constrained to move in a plane that is perpendicular to the axis of rotation. Instead, if we calculate the torque about the origin, then the torque vector determines the axis of rotation through the origin about which the particle would rotate. In this case, since the axis of rotation is the z axis, and the point of rotation was the origin, the torque about the z axis was simply the z component of the torque calculated about the origin.

11.5 Rotational dynamics for a solid object

We now consider the rotational dynamics for a solid object about a specific axis of rotation. Just as we did in Chapter 10, we model a solid object as a system made of many particles of mass m_i. Because all of the points in a solid must move in unison, they all **rotate about an axis of rotation instead of a point**. We describe the position of each particle i by a vector $\vec r_i$ that is **perpendicular to the axis of rotation and goes from the axis to the corresponding particle**, as shown in Figure 11.12.

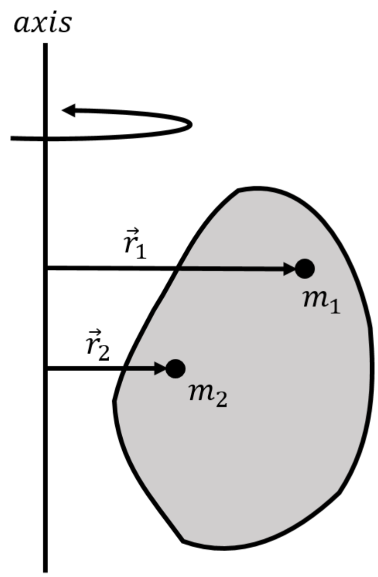


Figure 11.12: Two point particles that are part of a large solid object and their position vectors relative to an axis of rotation.

We wish to model the motion of the object as it rotates about a specific axis. Thus, when considering the net torque on any particle i, we only consider the component of the particle's net torque that is parallel to the axis of rotation (that component of torque that comes from forces that are in the plane perpendicular to the rotation axis).

We can write the rotational version of Newton's Second Law for particle, i, with mass m_i, and position vector $\vec{r}_i$ relative to the rotation axis:

$$\sum_k \vec{\tau}_{ik} = \vec{\tau}_i^{net} = m_i r_i^2 \vec{\alpha}_i$$

where $\vec{\tau}_{ik}$ is the k-th torque on particle i. $\vec{\tau}_i^{net}$ is the net torque on the particle **about the axis of rotation** and $\vec{\alpha}_i$ is the particle's angular acceleration about that axis.

We can divide the torques exerted on a particle into internal and external torques. Internal torques are those exerted by another particle in the system, whereas external torques are exerted by something external to the system. If particle 1 exerts a torque $\vec{\tau}$ on particle 2, particle 2 will exert an equal and opposite torque, $-\vec{\tau}$ on particle 1.

Indeed, consider the two particles that exert an equal and opposite force (Newton's Third Law), $\vec{F}$, on each other, and an arbitrary point/axis of rotation, as illustrated in Figure 11.13. The torque on particle 1 from the force exerted by particle 2 will have the same magnitude as the torque on particle 2 from the force by particle 1. This is because both forces have the same magnitude and they are co-linear, which results in them having the same lever arm. The torque vector from each force will be in opposite directions, because the forces are in opposite direction. Newton's Third Law thus also holds for torques.

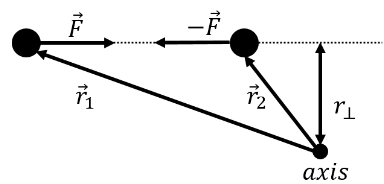


Figure 11.13: Two particles will exert equal and opposite torques on each other due to Newton's Third Law; the forces exerted by each particle on the other are co-linear and will thus have the same lever arm relative to any point/axis of rotation.

We can sum together the equations for each particle i:

$$\begin{aligned}\vec{\tau}_1^{net} + \vec{\tau}_2^{net} + \vec{\tau}_3^{net} + \ldots &= m_1 r_1^2 \vec{\alpha}_1 + m_2 r_2^2 \vec{\alpha}_2 + m_3 r_3^2 \vec{\alpha}_3 + \ldots \\ \sum_i \vec{\tau}_i^{net} &= \sum_i m_i r_i^2 \vec{\alpha}_i\end{aligned}$$

where the sum over all of the torques exerted on each particle will be equal to the net external torque exerted on all of the particles, since the sum of the internal torques, $\vec{\tau}_i^{int}$, will be zero:

$$\sum_i \vec{\tau}_i^{net} = \sum_i \vec{\tau}_i^{int} + \sum_i \vec{\tau}_i^{ext} = \sum_i \vec{\tau}_i^{ext} = \vec{\tau}^{ext}$$

where $\vec{\tau}^{ext}$ is the net external torque on the system.

All of the particles are part of the same rigid body, and cannot move relative to each other. Furthermore, they must all move around circles that are centred about the axis of rotation and in a plane perpendicular to that axis. They must thus all have the same angular acceleration[1], $\vec{\alpha}_i = \vec{\alpha}_1 = \vec{\alpha}_2 = \cdots = \vec{\alpha}$. We can thus factor the angular acceleration, $\vec{\alpha}$, out of the sum.

We can thus write Newton's Second Law for rotational dynamics of a solid object as:

$$\begin{aligned}\sum_i \vec{\tau}_i^{net} &= \sum_i m_i r_i^2 \vec{\alpha}_i \\ \therefore \vec{\tau}^{ext} &= \left(\sum_i m_i r_i^2\right) \vec{\alpha}\end{aligned}$$

The term in parentheses describes how the various masses are distributed relative to the axis of rotation. The term in parenthesis is called the **moment of inertia of the object**, and usually denoted with the letter, I:

$$\boxed{I = \sum_i m_i r_i^2} \tag{11.8}$$

[1]They will have different linear accelerations, but the angular acceleration (and angular velocity) will be the same for all particles if they are moving in unison.

The moment of inertia is a property of the object **relative to a specific axis of rotation**. Re-writing Newton's Second Law for the rotational dynamics of solid objects using the moment of inertia:

$$\boxed{\vec{\tau}^{ext} = I\vec{\alpha}} \tag{11.9}$$

The net torque exerted on an object in the direction of the axis of rotation is thus equal to its moment of inertia about that axis multiplied by its angular acceleration about that axis. In other words, the moment of inertia describes how the object will resist rotational motion given a net torque. An object with a smaller moment of inertia will have a larger angular acceleration for a given torque. Again, this is analogous to the linear case, where the acceleration of an object given a net force is determined by its inertial mass.

Example 11-3

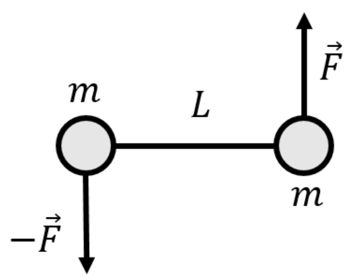


Figure 11.14: A dumbbell made of two small identical masses separated by a distance L.

Two small point masses, m, are connected by a mass-less rod of length L to form a dumbbell, as illustrated in Figure 11.14. A net force of magnitude F is exerted on each mass, in opposite directions, as illustrated in the Figure.

a) What is the linear acceleration of the centre of mass of the dumbbell?
b) What is the angular acceleration of the dumbbell relative to an axis that goes through its centre of mass and is perpendicular to the page?
c) What is the angular acceleration of the dumbbell relative to an axis that goes through one of the masses and is perpendicular to the page?

Solution

We model the dumbbell as a rigid body made of two point masses held at a fixed distance.

a) The linear acceleration of the centre of mass must be zero, because the net force on the dumbbell is zero. However, just because the centre of mass does not move does not mean that all parts of the dumbbell are immobile.

b) First, we calculate the angular acceleration relative to an axis that is perpendicular the page and goes through the centre of mass. The centre of mass is located midway between the two masses, as illustrated in Figure 11.15. We also define a coordinate system as shown, such that the z axis is out of the page.

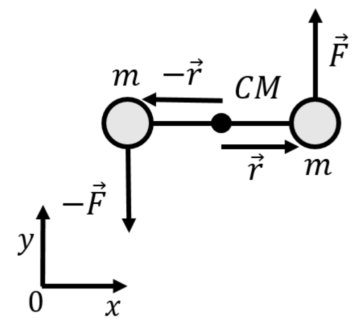


Figure 11.15: The dumbbell rotating about the centre of mass.

The vector from the axis of rotation to each mass will have the same magnitude, r, but different directions. The net external torque on the dumbbell relative to the axis that goes through the centre of mass, $\vec{\tau}^{ext}$, which is equal to the sum of the torques from each force:

$$\begin{aligned}\vec{\tau}^{ext} &= \vec{r} \times \vec{F} + (-\vec{r}) \times (-\vec{F}) \\ &= 2(\vec{r} \times \vec{F}) = 2(r\hat{x} \times F\hat{y}) = 2rF(\hat{x} \times \hat{y}) = 2rF\hat{z} \\ &= LF\hat{z}\end{aligned}$$

where we used the fact that $2r = L$. The net torque is thus non zero and in the positive z direction; the dumbbell will have an angular acceleration that is parallel to the net torque, and thus will accelerate in the counter-clockwise direction.
The moment of inertia of the dumbbell relative to the axis through the centre of mass is given by:

$$I = \sum_i m_i r_i^2 = mr^2 + mr^2 = 2mr^2 = \frac{1}{2}mL^2$$

Using Newton's Second Law for rotational dynamics, we find the angular acceleration to be:

$$\begin{aligned}\vec{\tau}^{ext} &= I\vec{\alpha} \\ LF\hat{z} &= \frac{1}{2}mL^2\vec{\alpha} \\ \therefore \vec{\alpha} &= \frac{2F}{mL}\hat{z}\end{aligned}$$

Because the centre of mass is fixed (the sum of the forces is zero), the two ends of the dumbbell will rotate about an axis that goes through the centre of mass.

This is a feature of all situations in which the net force on an object is zero and the net torque about an axis that goes through the centre of mass is non-zero.

c) Let us now calculate the angular acceleration of the dumbbell about an axis that goes through one of the masses, as illustrated in Figure 11.16.

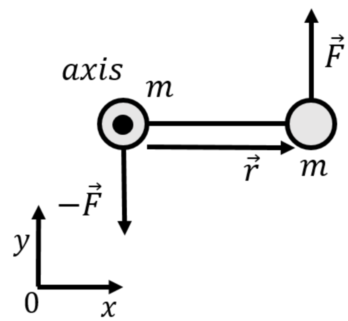


Figure 11.16: The dumbbell rotating about one of its ends.

We first calculate the net torque on the dumbbell. The vector that goes from the axis of rotation to the force exerted on the mass that coincides with the rotation axis is zero. Thus, only the force exerted on the mass that is not at the rotation axis contributes to the net torque:

$$\vec{\tau}^{ext} = \vec{r} \times \vec{F} = LF\hat{z}$$

The moment of inertia of the dumbbell about this axis is:

$$I = \sum_i m_i r_i^2 = m(0)^2 + m(r^2) = mL^2$$

which is larger than it was about the centre of mass. Again, the angular acceleration is found using Newton's Second Law for rotational dynamics:

$$\begin{aligned}
\vec{\tau}^{ext} &= I\vec{\alpha} \\
LF\hat{z} &= mL^2\vec{\alpha} \\
\therefore \vec{\alpha} &= \frac{F}{mL}\hat{z}
\end{aligned}$$

We find that the angular acceleration is smaller about an axis that goes through one of the mass than it is about an axis through the centre of mass. Because the centre of mass of the dumbbell is fixed, we can only think of the dumbbell as instantaneously rotating about one of its ends; that is, the motion of the dumbbell will not be such that one mass rotates about the other; this is only true instantaneously.

Discussion: This simple example illustrates several key features about rotational dynamics:

- If the sum of the forces on an object is zero, it does not mean that the entire object is stationary; it only implies that the centre of mass is stationary (or rather, moving with a constant velocity, but we can always choose to model the system in a frame of reference where the centre of mass is stationary).

- If the sum of the forces on an object is zero, and the sum of the external torques is non-zero, the object will rotate about an axis that goes through the centre of mass. That is, all points on the object will move along circles that are centred on an axis that goes through the centre of mass.
- We can model the rotating object about any axis that we choose. In general, the net external torque and the moment of inertia will depend on the choice of axis, as will the resulting angular acceleration.
- When determining the motion of the centre of mass, we can draw a free-body diagram, and the location of where the forces are exerted do not matter.
- When determining how the object rotates, we cannot use a free-body diagram, because it matters where the forces are applied (as the torque from a given force depends on the location where the force is applied relative to the axis of rotation).

11.6 Moment of inertia

In order to model how an object rotates about an axis, we use Newton's Second Law for rotational dynamics:

$$\vec\tau^{ext} = I\vec\alpha$$

where $\vec\tau^{ext}$ is the net external torque exerted on the object about the axis of rotation, $\vec\alpha$ is the angular acceleration of the object, and I is the moment of inertia of the object (about the axis). If we consider the object as being made of many particles of mass m_i each located at a position $\vec r_i$ relative to the axis of rotation, the moment of inertia is defined as:

$$I = \sum_i m_i r_i^2$$

Consider, for example, the moment of inertia of a uniform rod of mass M and length L that is rotated about an axis perpendicular to the rod that pass through one of the ends of the rod, as depicted in Figure 11.17.

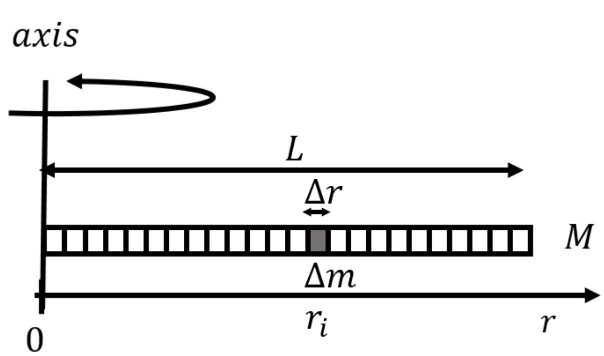


Figure 11.17: A rod of length L and mass M being rotated about an axis perpendicular to the rod that goes through one of its ends.

We introduce the linear mass density of the rod, λ, as the mass per unit length:

$$\lambda = \frac{M}{L}$$

We model the rod as being made of many small mass elements of mass Δm, of length Δr, at a location r_i, as illustrated in Figure 11.17. Using the linear mass density, the mass element, Δm, has a mass of:

$$\Delta m = \lambda \Delta r$$

The rod is made of many such mass elements, and the moment of inertia of the rod is thus given by:

$$I = \sum_i \Delta m r_i^2 = \sum_i \lambda \Delta r r_i^2$$

If we take the limit in which the length of the mass element is infinitesimally small ($\Delta r \to dr$) the sum can be written as an integral over the dimension of the rod:

$$\begin{aligned} I &= \int_0^L \lambda r_i^2 dr = \frac{1}{3}\lambda L^3 = \frac{1}{3}\left(\frac{M}{L}\right) L^3 \\ &= \frac{1}{3} M L^2 \end{aligned}$$

where we re-expressed the linear mass density in terms of the mass and length of the rod. In general, we can write the moment of inertia of a continuous object as:

$$I = \int r^2 dm$$

where dm is a small mass element that makes up the object, r is the distance from that mass element to the axis of rotation, and the integral is over the dimension of the object. As we did above, we would usually set up this integral so that dm is expressed in terms of r so that we can take an integral over r.

Example 11-4

Calculate the moment of inertia of a uniform thin ring of mass M and radius R, rotated about an axis that goes through its centre and is perpendicular to the disk.

Solution

We take a small mass element dm of the ring, as shown in Figure 11.18.

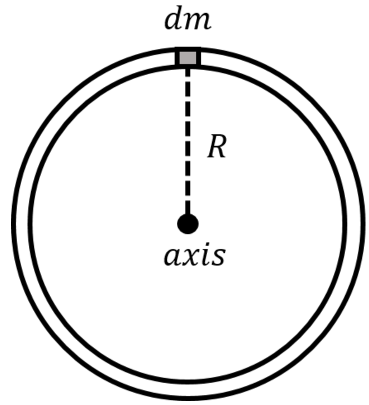


Figure 11.18: A small mass element on a ring.

The moment of inertia is given by:

$$I = \int dmr^2$$

In this case, each mass element around the ring will be the same distance away from the axis of rotation. The value r^2 in the integral is a constant over the whole ring, and so can be taken out of the integral:

$$I = \int dmr^2 = R^2 \int dm$$

where we used the fact that the ring has a radius R, so the distance r of each mass element to the axis of rotation is R. The integral:

$$\int dm$$

just means "sum all of the mass elements, dm", and is thus equal to M, the total mass of the ring. The moment of inertia of the ring is thus:

$$I = R^2 \int dm = MR^2$$

11.6.1 The parallel axis theorem

The moment of inertia of a solid object can be difficult to calculate, especially if the object is not symmetric. The parallel axis theorem allows us to determine the moment of inertia of an object about an axis, if we already know the moment of inertia of the object about an axis that is parallel and goes through the centre of mass of the object.

Consider an object for which we know the moment of inertia, I_{CM}, about an axis that goes through the object's centre of mass. We define a coordinate system such that the origin is located at the centre of mass, and the z axis is parallel to the axis about which we know the moment of inertia, as illustrated in Figure 11.19.

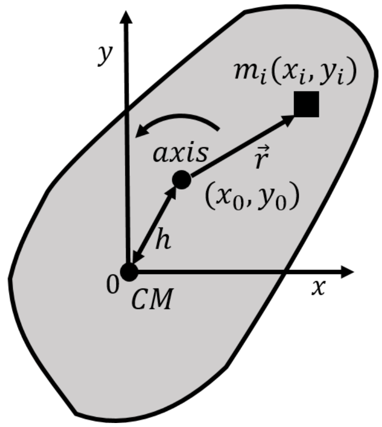


Figure 11.19: An object with a coordinate system whose origin is at the object's centre of mass, and for which we know the moment of inertia about the z axis. We wish to determine the object's moment of inertia through a second axis, parallel to the z axis, but located a distance h away from the centre of mass.

We wish to determine the moment of inertia for the object for an axis that is parallel to the z axis, but goes through a point with coordinates (x_0, y_0) located a distance h away from the centre of mass. The moment of inertia about an axis parallel to the z axis and that goes through that point, I_h is given by:

$$I_h = \sum_i m_i r_i^2$$

where m_i is a mass element of the object located at a distance r_i from the axis of rotation. If the mass element is located at a position (x_i, y_i) relative to the centre of mass, we can write the distance r_i in terms of the position of the mass element, and of the position of the axis of rotation:

$$r_i^2 = (x_i - x_0)^2 + (y_i - y_0)^2 = x_i^2 - 2x_i x_0 + x_0^2 + y_i^2 - 2y_i y_0 + y_0^2$$

Note that:

$$x_0^2 + y_0^2 = h^2$$

The moment of inertia, I_h, can thus be written as:

$$\begin{aligned}
I_h &= \sum_i m_i r_i^2 = \sum_i (m_i(x_i^2 + y_i^2) - 2x_0 m_i x_i - 2y_0 m_i y_i + m_i h^2) \\
&= \sum_i m_i(x_i^2 + y_i^2) + h^2 \sum_i m_i - 2x_0 \sum_i m_i x_i - 2y_0 \sum_i m_i y_i
\end{aligned}$$

where we broke the sum up into several sums, and factored constant terms (h, x_0, y_0) out of the sums, since these constants do not depend on which mass element we are considering. The first term is the moment of inertia about the centre of mass, since $x_i^2 + y_i^2$ is the distance to the centre of mass. The second term is h^2 times the total mass of the object, since the sum of all the m_i is just the mass, M, of the object. Now consider the term:

$$-2x_0 \sum_i m_i x_i$$

The sum, $\sum m_i x_i$ is the numerator in the definition of the x coordinate of the centre of mass! The sum is thus zero, because we choose the origin to be located at the centre of mass. The last two terms in the sum are thus identically zero, because they correspond to the x and y coordinates of the centre of mass!

We can thus write the parallel axis theorem:

$$\boxed{I_h = I_{CM} + Mh^2} \tag{11.10}$$

where I_{CM} is the moment of inertia of an object of mass M about an axis that goes through the centre of mass and, I_h, is the moment of inertia about a second axis that is parallel to the first and a distance h away.

Example 11-5

In the previous section, we calculated the moment of inertia of a rod of length L and mass M through an axis that is perpendicular to the rod and through one of its ends, and found that it was given by:

$$I = \frac{1}{3}ML^2$$

What is the moment of inertia of the rod about an axis that is perpendicular to the rod and goes through its centre of mass?

Solution

In this case, we know the moment of inertia through an axis that does not go through the centre of mass. The centre of mass is located a distance $h = L/2$ away from the point about which we know the moment of inertia, I_h.

Using the parallel axis theorem, we can find the moment of inertia through the centre of mass:

$$\begin{aligned} I_{CM} &= I_h - Mh^2 \\ &= \frac{1}{3}ML^2 - M\left(\frac{L}{2}\right)^2 = \frac{1}{12}ML^2 \end{aligned}$$

Discussion: We find that the moment of inertia about the centre of mass is smaller than the moment of inertia about the end of the rod. This makes sense because when rotating the rod about its end, more of its mass is further away from the axis of rotation, which results in a larger moment of inertia.

11.7 Equilibrium

In this section, we consider the conditions under which an object is in static or dynamic equilibrium. An object is in equilibrium if it does not rotate when viewed in a frame of reference where the object's centre of mass is stationary (or moving at constant velocity).

11.7.1 Static equilibrium

An object is in static equilibrium, if **both the sum of the external forces exerted on the object and the sum of the external torques (about any axis) are zero**. If the object is in static equilibrium the centre of mass will have no acceleration and the object will have no angular acceleration. In the centre of mass frame of reference, the object is immobile.

Example 11-6

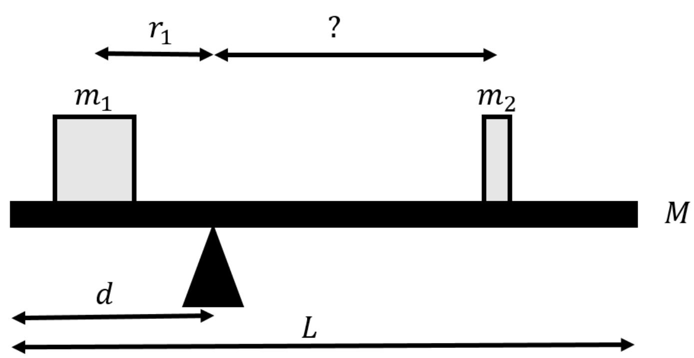


Figure 11.20: Two masses on a balance.

Two masses, m_1 and m_2 are placed on a balance as shown in Figure 11.20. The balance is made of a plank of mass M and length L that is placed on a fulcrum that is a distance d from one of the edges of the plank. If mass m_1 is placed at a distance r_1 from the fulcrum, how far should mass m_2 be placed on the other side of the plank in order for the balance to be in equilibrium?

Solution

We can consider the plank as the object that is in static equilibrium. Thus, the sum of the forces and the sum of the torques on the plank must be zero. We first start by identifying the forces that are exerted on the plank; these are:

1. $\vec{F}_g$, the weight of the plank, exerted at the centre of mass of the plank.
2. $\vec{F}_1$, a force equal to the weight of mass m_1, exerted at the location of m_1.
3. $\vec{F}_2$, a force equal to the weight of mass m_2, exerted at the location of m_2.
4. $\vec{N}$, a normal force exerted by the fulcrum.

The forces are illustrated in Figure 11.21 along with our choice of coordinate system.

The z axis is not illustrated, and is directed out of the page.

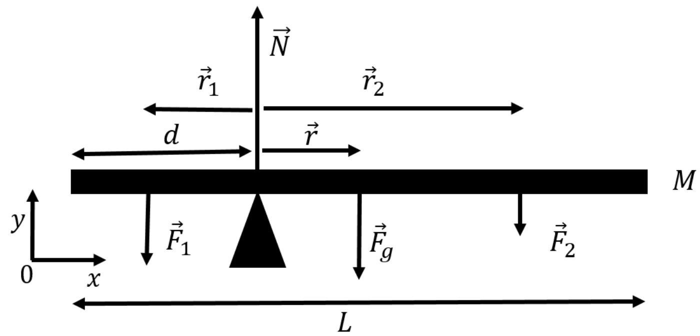


Figure 11.21: Forces exerted on the plank.

All of the forces are in the y direction, so we only write the y component of Newton's Second Law (with zero acceleration), which allows us to determine the magnitude of the normal force:

$$\begin{aligned} \sum F_y = N - Mg - m_1g - m_2g = 0 \\ \therefore N = (M + m_1 + m_2)g \end{aligned}$$

Because the plank is in static equilibrium, the sum of the torques must also be zero. We can choose the axis of rotation about which to calculate the torques. We choose an axis that is parallel to the z axis (out of the page) and goes through the fulcrum. In general, since we can choose the axis of rotation, it is usually convenient to choose an axis that goes through a point where at least one force is being exerted, because the torque from that force will be zero (its lever arm will be zero). Furthermore, since all of the forces are in the xy plane, the net torque on the plank will be in the z direction, so it makes sense to choose an axis in that direction.

The torques from the weight of the plank and from the force exerted by mass m_2 will be in the negative z direction, and the torque from the force exerted by mass m_1 will be in the positive z direction. The normal force will not result in any torque, because it is exerted at the axis of rotation and has a lever arm of zero.

We define $\vec{r}_1$ as the vector from the fulcrum to mass m_1. The torque, $\vec{\tau}_1$, from the force exerted by mass m_1 is given by:

$$\begin{aligned} \vec{\tau}_1 &= \vec{r}_1 \times \vec{F}_1 = (-r_1\hat{x}) \times (-F_1\hat{y}) \\ &= r_1F_1(\hat{x} \times \hat{y}) = r_1F_1\hat{z} = r_1m_1g\hat{z} \end{aligned}$$

where we used the fact that the magnitude of $\vec{F}_1$ is m_1g. Similarly, the torques from the force exerted by m_2, $\vec\tau_2$, and by the weight, $\vec\tau_g$, are given by:

$$\begin{aligned}
\vec\tau_2 &= \vec r_2 \times \vec F_2 = -m_2 g r_2 \hat z\\
\vec\tau_g &= \vec r \times \vec F_g = -rMg\hat z = -\left(\frac{L}{2} - d\right) Mg\hat z
\end{aligned}$$

where $\frac{L}{2} - d$ is the distance between the fulcrum and where the weight of the plank is exerted. We require that the z component of the net torque be equal to zero (since all of the torques are in the z direction), which allows us to determine r_2:

$$\begin{aligned}
\sum \tau_z &= \tau_{1z} + \tau_{2z} + \tau_{gz} = 0\\
m_1 g r_1 - m_2 g r_2 - \left(\frac{L}{2} - d\right) Mg &= 0\\
\therefore r_2 &= \frac{1}{m_2}\left(m_1 r_1 - \left(\frac{L}{2} - d\right) M\right)
\end{aligned}$$

Note that because we chose to calculate the torques about a point that goes through the fulcrum, in this case, we did not need to determine the value of the normal force which we obtained from Newton's Second Law.

Discussion: This example highlights the fact that when an object is in static equilibrium, we can choose a convenient axis about which to calculate the torques. In this case, by calculating the torques about the fulcrum, we did not need to consider the torque from the normal force. If we had chosen a different point, then the torque from the normal force would have been non-zero, and we would have used Newton's Second Law to express the normal force in terms of the other quantities. Physically, if we had placed the fulcrum at the centre of the plank $d = L/2$, then we would have found that $m_1 r_1 = m_2 r_2$, the well known equation for a balance. This equation, of course, comes from requiring that the torques from the forces exerted by m_1 and m_2 are equal in magnitude and opposite in direction.

11.7.2 Dynamic equilibrium

Review Topics

Before proceeding, you may wish to review Section 5.6 on inertial forces.

When an object is in dynamic equilibrium, its centre of mass is accelerating, but the object is not rotating when viewed from its centre of mass frame of reference. Thus, the sum of the external forces exerted on the object is not zero, while the net external torque exerted on the object is zero, in the frame of reference of the centre of mass.

Consider, for example, a speed skater going around a circular track or radius R, and leaning into the centre making an angle θ with the ice, as depicted in Figure 11.22. The skater's centre of mass is accelerating, because she is going around a circle, so the net force on the skater is not zero. However, in the reference frame of the skater, the skater is not rotating;

she is thus in dynamic equilibrium.

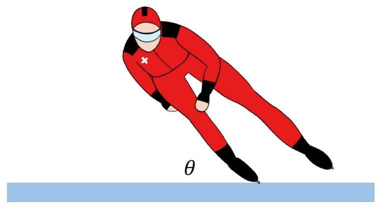


Figure 11.22: A speed skater leaning in as she goes around a circle.

The forces on the skater are:

1. $\vec{F}_g$, her weight, exerted at her centre of mass with magnitude, Mg.
2. $\vec{N}$, a normal force, exerted by the ice upwards on her skates.
3. $\vec{f}_s$, a force of static friction, exerted towards the centre of the circle, by the ice on her skates.

The forces are illustrated in Figure 11.23 along with our choice of coordinate system.

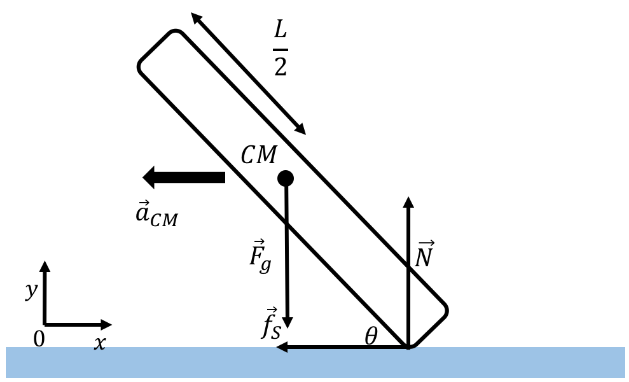


Figure 11.23: Forces on the speed skater from Figure 11.22.

The sum of the forces exerted on the skater must be towards the centre of the circle and equal to the mass of the skater times her centripetal acceleration (which is the acceleration of her centre of mass, $\vec{a}_{CM}$). The x and y components of Newton's Second Law are thus given by:

$$\sum F_x = -f_s = -ma_{CM} - m\frac{v^2}{R}$$
$$\sum F_y = N - mg = 0$$

All of the forces exerted on the skater are in the xy plane, so we consider torques about an axis that is co-linear with the z axis. Consider the torques about an axis through the point of contact between the skates and the ice; there is a net torque in the counter-clockwise direction due to the weight of the skater (the weight is the only force that can result in a torque about the point of contact with the ice). We expect that the skater would topple over, however, this must not be a correct model for the skater, since we know that it is possible for her to lean in without falling.

Consider, instead, the sum of the torques about an axis through her centre of mass. If the skater has a length L and the centre of mass is in the middle of the skater, the sum of the torques about the centre of mass is given by the torques from the normal forces and the force of friction:

$$\sum \tau = \tau_{Nz} + \tau_{f_s z} = \frac{L}{2}\cos\theta N - \frac{L}{2}\sin\theta f_s$$

About the centre of mass, the torques must be zero for the skater not to rotate, and this would give a relation between the force of static friction and the normal force.

Why do we get an incorrect model when we take the torques about the point of contact between the ice and the skater? In order to determine if the skater is rotating, we need to be in the same reference frame as the skater. However, the frame of reference of the skater is not an inertial frame of reference, since the skater is accelerating. We can still model the forces on the skater in the non-accelerating frame of reference, **as long as we include the inertial force,** $-m\vec{a}_{CM}$, in that frame of reference. In the frame of reference of the skater, there is an additional inertial force, $-m\vec{a}_{CM}$, in order for the sum of the forces to be zero (in the frame of reference of the skater, the sum of the forces must be zero since the skater is not accelerating in that frame of reference). The additional inertial force is exerted at the centre of mass, as illustrated in Figure 11.24.

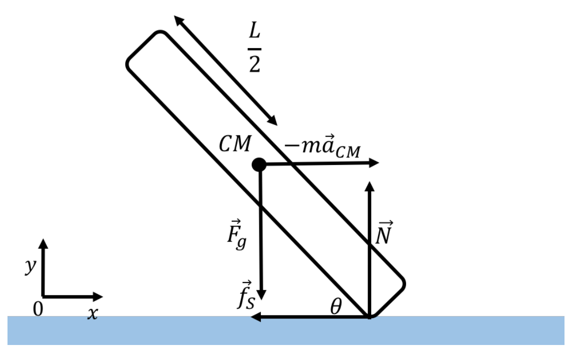


Figure 11.24: Forces on the speed skater from Figure 11.22 as seen in the accelerating frame of reference of the centre of mass.

The reason that our model worked when taking the torques about the centre of mass is that the inertial force, exerted at the centre of mass, does not result in a torque (since it has a lever arm of zero). Our model was technically wrong, but if we take the torques about the centre of mass, then we do not need to worry about the inertial force. If we include the additional inertial force, then we can take the torques about any point, just as in the static equilibrium case.

11.8 Summary

Key Takeaways

We can describe the kinematics of rotational motion using vectors to indicate both an axis of rotation and the direction of rotation about that axis. If a particle with velocity vector, $\vec{v}$, is rotating in an circle about an axis, then its angular velocity vector, $\vec{\omega}$, relative to that axis is defined as:

$$\vec{\omega} = \frac{1}{r^2}\vec{r} \times \vec{v}$$

where $\vec{r}$ is a vector from the axis of rotation to the particle. The particle rotates in a circle that lies in the plane defined by $\vec{r}$ and $\vec{v}$, perpendicular to the axis of rotation. The direction of the angular velocity vector is co-linear with the axis of rotation and the direction of rotation is given by the right-hand rule for axial vectors.

One can define the angular velocity of a particle relative to a point of rotation, even if the particle is not moving in a circle. In that case, the angular velocity corresponds to the angular velocity of the particle as if it were instantaneously moving about a circle.

If a particle moving around a circle has a tangential acceleration, $\vec{a}_s$, then its angular acceleration vector is defined as:

$$\vec{\alpha} = \frac{1}{r^2}\vec{r} \times \vec{a}_s$$

The torque from a force, $\vec{F}$, exerted at a position $\vec{r}$, relative to an axis (or point) of rotation is defined as:

$$\vec{\tau} = \vec{r} \times \vec{F}$$

Torque is analogous to force in that it is used to model the causes of motion. Torques are only ever defined relative to an axis or point of rotation. The torque vector will be co-linear with the axis about which the object on which the force is exerted would rotate as a result of that force.

The magnitude of the torque can be written using either the component of the force, $F_\perp$ perpendicular to the vector $\vec{r}$, or the lever arm, $r_\perp$, of the force relative to the axis of rotation:

$$\begin{aligned}\tau &= rF\sin\phi \\ &= rF_\perp \\ &= r_\perp F\end{aligned}$$

where ϕ is the angle between the vectors $\vec{r}$ and $\vec{F}$ when these are placed "tail to tail".

Using rotational/angular quantities, we can modify Newton's Second Law to describe rotational dynamics about a given axis (or point) of rotation. For a point particle, this gives:

$$\vec{\tau}^{net} = mr^2\vec{\alpha}$$

where $\vec{\tau}^{net}$ is the net torque on the particle (the sum of the torques from each force exerted on the particle) about the axis, and $\vec{\alpha}$ is the resulting angular acceleration about that axis.

For an object (either continuous or made of point particles), the rotational version of Newton's Second Law for rotation about a specific axis is given by:

$$\vec{\tau}^{net} = I\vec{\alpha}$$

where I is the moment of inertia of the object about that axis.

The moment of inertia of an object about an axis of rotation is given by

$$I = \sum_i m_i r_i^2$$

if the object is modelled as a system of point particles of mass m_i each a distance r_i from the axis of rotation. For a continuous object, the moment of inertia is given by:

$$I = \int r^2 dm$$

where dm is a small mass element a distance r from the axis of rotation and the integral is over the dimension of the object. Generally, one can set up the integral by expressing dm in terms of r using the density of the object, and then integrating r over the dimension of the object.

If the moment of inertia of an object of mass M about an axis that goes through the centre of mass is given by I_{CM}, then the moment of inertia, I_h, of the object through an axis that is parallel and a distance h from the centre of mass is given by the parallel axis theorem:

$$I_h = I_{CM} + Mh^2 \quad \text{Parallel axis theorem}$$

Objects are in equilibrium if they are not rotating when viewed in their centre of mass frame of reference. Thus, for an object to be in equilibrium, the sum of the torques on the object, in the centre of mass reference frame, must be zero.

An object is in static equilibrium if the centre of mass is not accelerating, and thus the sum of the external forces on the object is zero. To model the torques on an object in static equilibrium, one can choose the axis about which to calculate the torques. A good choice is to choose an axis that is perpendicular to the plane in which the forces on the object are exerted (if such a plane exists), and to choose the axis to go through a point where at least one force is exerted (so that torques exerted at that point are identically zero).

An object is in dynamic equilibrium if the centre of mass is accelerating, but the object does not rotate when viewed in the frame of reference of its centre of mass. In dynamic equilibrium, if one models the torques exerted on the object about an axis that does not go through the centre of mass, then one must remember to include an inertial force exerted at the centre of mass.

Important Equations

Angular quantities:

$$\vec{\omega} = \frac{1}{r^2}\vec{r} \times \vec{v}$$
$$\vec{\alpha} = \frac{1}{r^2}\vec{r} \times \vec{a}_\perp$$
$$\vec{v}_s = \vec{\omega} \times \vec{r}$$
$$\vec{a}_s = \vec{\alpha} \times \vec{r}$$

Torque from a force:

$$\begin{aligned}\vec{\tau} &= \vec{r} \times \vec{F}\\ \tau &= rF\sin\phi\\ &= rF_\perp\\ &= r_\perp F\end{aligned}$$

Newton's Second Law for a point particle about a given axis of rotation:

$$\vec{\tau}^{net} = mr^2\vec{\alpha}$$

Newton's Second Law for rotation about an axis:

$$\vec{\tau}^{net} = I\vec{\alpha}$$

Moment of Inertia:

$$I = \sum_i m_i r_i^2$$
$$I = \int r^2 dm$$

Parallel Axis Theorem:

$$I_h = I_{CM} + Mh^2$$

11.9 Thinking about the material

Reflect and research

1. Compare the steering wheels of a small car and a large transport truck. What are the differences, and why?
2. List 2 kitchen utensils that use torque to "get the job done".

To try at home

1. Take a large textbook and consider the 3 axes that are parallel to the sides of the textbook and go through the centre of mass. By rotating the book along the three axes successively, determine the axis about which the moment of inertia of the textbook is the largest.
2. Confirm that the moment of inertia of a rod is smaller if the rod is rotated about its centre of mass than if it is rotated by one of its ends.

To try in the lab

1. Propose an experiment to measure the moment of inertia of an object and to compare that to a model prediction.

11.10 Sample problems and solutions

11.10.1 Problems

Problem 11-1: Calculate the moment of inertia of a uniform disk of mass M and radius R, rotated about an axis that goes through its centre and is perpendicular to the disk. (Solution)

Problem 11-2:

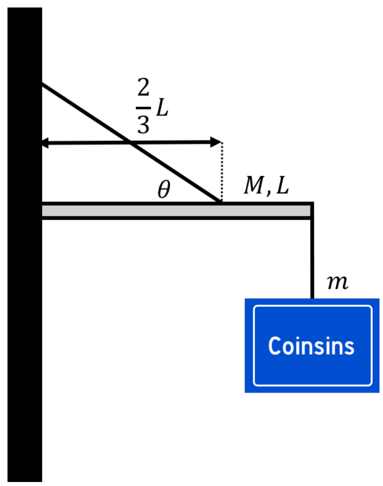


Figure 11.25: A sign is suspended on a horizontal bar of mass M and length L.

A sign holder is built by attaching a bar of mass M and length L to a wall using a hinge that allows the bar to rotate in the vertical plane. The sign of mass m is attached to the end of the bar that is opposite fo the wall. The bar is held up by a rope that is attached to the wall on one end and to the bar on the other end, two thirds of the length of the bar from the wall, as illustrated in Figure 11.25. The rope makes an angle θ with respect to the horizontal bar. Find the tension in the rope and the magnitude of the force exerted by the hinge onto the bar. (Solution)

11.10.2 Solutions

Solution to problem 11-1: We need to split up the disk into mass elements, dm, that we can sum together to obtain the moment of inertia of the disk. We can choose a ring of radius r and radial thickness dr for the shape of our mass element, as depicted in Figure 11.26.

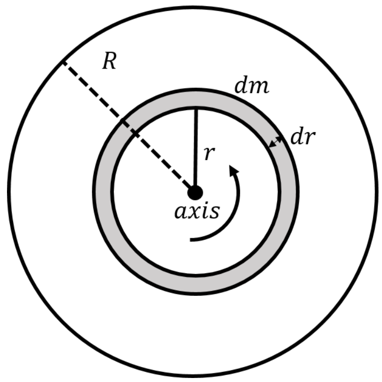


Figure 11.26: A mass element, dm, in the shape of a ring of radius r and radial thickness dr.

We can define a surface mass density, σ, equal to the mass per unit area of the disk:

$$\sigma = \frac{M}{\pi R^2}$$

The mass of the ring shaped element is thus given by:

$$dm = \sigma 2\pi r dr$$

where $2\pi r dr$ is the area of the mass element. You can imagine unfolding the mass element into a rectangle of height dr and of length $2\pi r$ to obtain its area. Now that we have expressed the mass element in terms of r, we can proceed to calculate the moment of inertia of the disk.

We know from Example 11-4, that the infinitesimal moment of inertia, dI, of a ring of radius r and infinitesimal mass, dm, about its axis of symmetry is given by:

$$dI = dmr^2$$

The moment of inertia of the disk, is found by summing the moments of inertia of the infinitesimal rings:

$$\begin{aligned}
I &= \int dI = \int dmr^2 = \int_0^R \sigma 2\pi r dr r^2 = 2\pi\sigma \int_0^R r^3 dr \\
&= 2\pi\sigma \frac{1}{4} R^4 = 2\pi \left(\frac{M}{\pi R^2}\right) \frac{1}{4} R^4 \\
&= \frac{1}{2} M R^2
\end{aligned}$$

where we removed the surface mass density by expressing it in term of the total mass and radius of the disk.

Discussion: The moment of inertia of a disk of mass M and radius R is half of that of a ring of radius R and mass M. It is thus easier to rotate the disk than the ring.

Solution to problem 11-2: The whole system does not move and so it is in static equilibrium. In order to determine the forces exerted on the bar by the rope and the hinge, we model the bar as being in static equilibrium. The forces exerted on the bar are:

- $\vec{F}_g$, the weight of the bar, with magnitude Mg, exerted at the bar's centre of mass.
- $\vec{F}_m$, a downwards forced exerted by the sign at the end of the bar, with magnitude mg.
- $\vec{T}$, a force of tension exerted by the rope at a distance $2/3L$ from the wall.
- $\vec{R}$, a force exerted by the hinge on the bar at the end next to the wall[2]. We expect that the force from the hinge will have both a horizontal component, R_x, and a vertical component, R_y, in order for the net force on the bar to be zero.

The forces are illustrated in Figure 11.27 along with our choice of coordinate system (and the z axis, not shown, points out of the page).

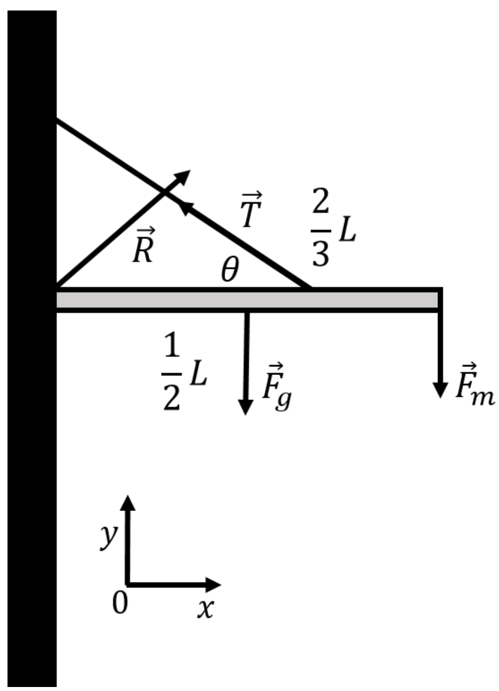


Figure 11.27: Forces on the bar that is holding the sign of mass m.

[2]We chose the letter R for "Reaction", as this is the force of reaction from the hinge as the bar pushes against the hinge.

We start by writing out the x and y components of Newton's Second Law (with zero acceleration):

$$\begin{aligned}
\sum F_x &= R_x - T\cos\theta = 0\\
\sum F_y &= R_y + T\sin\theta - Mg - mg = 0
\end{aligned}$$

We can choose the axis about which to calculate the torques. Since all of the forces are in the xy plane, we choose to calculate the torques about an axis parallel to the z axis that goes through the hinge on the wall. The force from the hinge, $\vec R$, will thus result in a torque of zero (since it has a lever arm of zero). The torque from each force about the hinge is given by:

$$\begin{aligned}
\vec\tau_M &= \vec r_M \times \vec F_g = \left(\frac{L}{2}\hat x\right)\times(-Mg\hat y) = -Mg\frac{L}{2}\hat z\\
\vec\tau_T &= \vec r_T \times \vec T = \left(\frac{L}{3}\hat x\right)\times(-T\cos\theta\hat x + T\sin\theta\hat y) = T\sin\theta\frac{L}{3}\hat z\\
\vec\tau_m &= \vec r_m \times \vec F_m = (L\hat x)\times(-mg\hat y) = -mgL\hat z
\end{aligned}$$

The sum of the torques in the z direction must be zero for static equilibrium, which allows us to determine the magnitude of the force of tension:

$$\begin{aligned}
\sum \tau_z &= \tau_{Mz} + \tau_{Tz} + \tau_{mz} = 0\\
-Mg\frac{L}{2} + T\sin\theta\frac{L}{3} - mgL &= 0\\
-Mg\frac{1}{2} + T\sin\theta\frac{1}{3} - mg &= 0\\
\therefore T &= \frac{3g}{\sin\theta}\left(m + \frac{M}{2}\right)
\end{aligned}$$

Using the x and y components of Newton's Second Law, we can now use the tension to determine the x and y components of the force exerted by the hinge:

$$\begin{aligned}
R_x &= T\cos\theta = \frac{3g}{\tan\theta}\left(m + \frac{M}{2}\right)\\
R_y &= (M+m)g - T\sin\theta = (M+m)g - 3g\left(m + \frac{M}{2}\right) = -\left(2m + \frac{M}{2}\right)g
\end{aligned}$$

We find that the y component of the force from the hinge is in the negative y direction, so **our diagram in Figure 11.27 is wrong**! If you removed the hinge on the wall and instead held that end of the bar with your hand, you would feel that the end of the bar is trying to go into the wall and upwards, as the bar tries to rotate with the opposite end moving downwards due to the weight of the sign. You would have to push in the positive x and negative y direction to keep the bar from moving.

Discussion: In this example, we saw that we needed to use both the sum of the forces and the sum of the torques in order to determine the forces on the bar.

12 Rotational energy and momentum

In this chapter, we extend our description of rotational dynamics to include the rotational equivalents of kinetic energy and momentum. We also develop the framework for describing the motion of rolling objects. We will see that many of the relations that hold for linear quantities also hold for angular quantities.

Learning Objectives

- Understand how to define the rotational kinetic energy of an object as well as its total kinetic energy.
- Understand how to model rolling motion, and what slipping means in the context of rolling motion.
- Understand how to define the angular momentum of an object and when it is conserved.

Think About It

How can you model the motion of a downwards going yo-yo?

A) It is similar to that of an object falling with a force of drag.
B) It is similar to that of an object rolling down an incline.
C) It is similar to that of an object sliding down an incline.
D) It is similar to that of an object rotating about a fixed axis of rotation.

12.1 Rotational kinetic energy of an object

In this section, we show how to define the rotational kinetic energy of an object that is rotating about a stationary axis in an inertial frame of reference. Consider a solid object that is rotating about an axis with angular velocity, $\vec{\omega}$, as depicted in Figure 12.1.

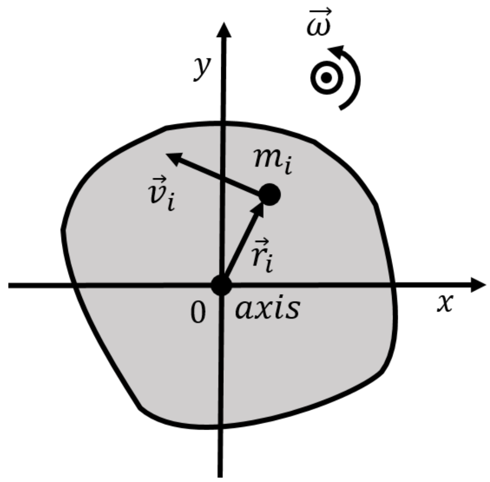


Figure 12.1: An object rotating about an axis that passes through its centre of mass.

We can model the object as being composed of many point particles, each with a mass m_i, located at a position $\vec{r}_i$, with velocity $\vec{v}_i$ relative to the axis of rotation. We choose a coordinate system whose origin is on the axis of rotation and whose z axis is co-linear with the axis of rotation, as depicted in Figure 12.1.

Each particle of mass m_i in the object has a kinetic energy, K_i:

$$K_i = \frac{1}{2}m_i v_i^2$$

We can sum the kinetic energy of each particle together to get the total rotational kinetic energy, K_{rot}, of the object:

$$K_{rot} = \sum_i \frac{1}{2}m_i v_i^2$$

Although each particle will have a different velocity, $\vec{v}_i$, they will all have the same angular velocity, $\vec{\omega}$. For any particle, located a distance r_i from the axis of rotation, their velocity is related to the angular velocity of the object by:

$$\begin{aligned}\vec{v}_i &= \vec{\omega} \times \vec{r}_i \\ v_i &= \omega r_i\end{aligned}$$

where $\vec{\omega}$ and $\vec{r}_i$ are always perpendicular to each other, since $\vec{\omega}$ is out of the plane of the page. Furthermore, the velocity vector, $\vec{v}_i$, will always be perpendicular to $\vec{r}_i$, since all particles are moving in circles centred about the axis of rotation. We can thus write the total rotational kinetic energy of the object using the angular speed:

$$\begin{aligned}K_{rot} &= \sum_i \frac{1}{2}m_i v_i^2 = \sum_i \frac{1}{2}m_i r_i^2\omega^2 = \frac{1}{2}\omega^2 \sum_i m_i r_i^2 \\ &= \frac{1}{2}I\omega^2\end{aligned}$$

where we factored ω and the one half out of the sum, as these are the same for each particle i. We then recognized that the remaining sum is simply the definition of the object's moment of inertia about the axis:

$$I = \sum_i m r_i^2$$

Thus, the rotational kinetic energy of an object rotating with angular speed ω about an axis that is stationary in an inertial frame of reference is given by:

$$\boxed{K_{rot} = \frac{1}{2} I \omega^2} \tag{12.1}$$

where I is the object's moment of inertia about that axis. The rotational kinetic energy is functionally very similar to the linear kinetic energy; instead of mass, we use the moment of inertia, and instead of speed squared, we use angular speed squared.

12.1.1 Work on a rotating object

We can calculate the work done by a force exerted on an object rotating about a stationary axis in an inertial frame of reference. Let $\vec F$ be a force exerted at position, $\vec r$, relative to the axis of rotation at some instant in time, and let the force be exerted in the plane perpendicular to the axis of rotation, as illustrated in Figure 12.2. Because the object is rotating about the given axis, only the component of the force that is tangent to the circle about which the point where the force is exerted can do work (only the component of the force that is parallel to the displacement can do work).

The work done by the force as the object rotates by a certain angle is given by:

$$W = \int \vec F \cdot d\vec l = \int F_\perp dl$$

where $d\vec l$ is a small displacement along the (circular) path followed by the point where the force is exerted, as illustrated in Figure 12.2. $F_\perp$ is the component of $\vec F$ that is perpendicular to the vector, $\vec r$, from the axis of rotation to the location where the force is exerted ($F_\perp$ is the component of $\vec F$ that is tangent to the circle).

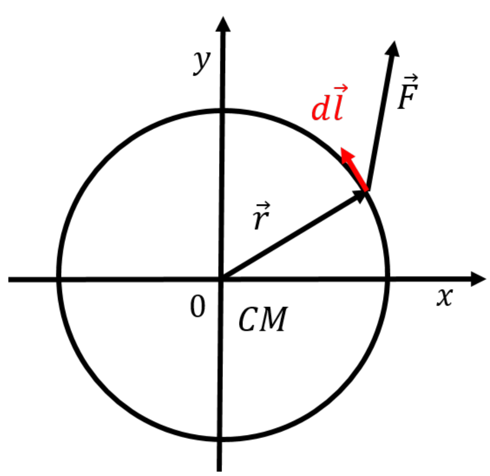


Figure 12.2: Calculating the work done by a force on a rotating object.

At some instant in time, when the force is exerted at position, $\vec r$, consider the scalar product between the torque from the force, $\vec\tau$, and an infinitesimal angular displacement, $d\vec\theta$, about the axis of rotation:

$$\vec\tau \cdot d\vec\theta = (\vec r \times \vec F) \cdot \left(\frac{1}{r^2}\vec r \times d\vec l\right)$$

The vectors $\vec\tau$ and $d\vec\theta$ are parallel to the axis of rotation (because $\vec F$ and $d\vec l$ are in the plane perpendicular to the axis of rotation), so their scalar product will be equal to the product of their magnitudes. The vector $\vec r \times \vec F$ has a magnitude of:

$$\vec r \times \vec F = rF_\perp$$

where $F_\perp$ is the component of the force tangent to the circle. The vector $\vec r \times d\vec l$ has a magnitude:

$$\vec r \times d\vec l = rdl$$

since $\vec r$ and $d\vec l$ are always perpendicular. The scalar product $\vec\tau \cdot d\vec\theta$ is thus equal to:

$$\vec\tau \cdot d\vec\theta = rF_\perp \frac{1}{r^2} rdl = F_\perp dl$$

The work done by a force when an object rotates about an axis can thus be written in terms of its torque about that axis and the corresponding angular displacement from θ_1 to θ_2:

$$W = \int_{\theta_1}^{\theta_2} \vec\tau \cdot d\vec\theta \tag{12.2}$$

The net work done on an object through an angular displacement from θ_1 to θ_2 can thus be written using the net torque $\vec\tau^{net}$ exerted on the object:

$$W^{net} = \int_{\theta_1}^{\theta_2} \vec\tau^{net} \cdot d\vec\theta$$

We can re-arrange this using Newton's Second Law for rotational dynamics:

$$\begin{aligned}\vec\tau^{net} &= I\vec\alpha \\ &= I\frac{d\vec\omega}{dt} = I\frac{d\omega}{d\theta}\frac{d\vec\theta}{dt} = I\frac{d\omega}{d\theta}\vec\omega\end{aligned}$$

which allows us to write the integral over a change in angular velocity instead of angular displacement:

$$\begin{aligned}W^{net} &= \int_{\theta_1}^{\theta_2} \vec\tau^{net} \cdot d\vec\theta = \int_{\theta_1}^{\theta_2} I\frac{d\omega}{d\theta}\vec\omega \cdot d\vec\theta \\ &= \int_{\omega_1}^{\omega_2} I\omega d\omega = \frac{1}{2}I\omega_2^2 - \frac{1}{2}I\omega_1^2\end{aligned}$$

where we used the fact that $\vec{\omega}$ are $d\vec{\theta}$ are parallel. We thus find that the Work-Energy Theorem can also be applied to find the change in rotational kinetic energy resulting from the net work done by a torque:

$$\boxed{W^{net} = \int_{\theta_1}^{\theta_2} \vec{\tau}^{net} \cdot d\vec{\theta} = \Delta K_{rot}} \tag{12.3}$$

If a constant torque, $\vec{\tau}$, is exerted on an object that is rotating at constant angular velocity, $\vec{\omega}$, then the rate at which that work is being done is given by:

$$P = \frac{dW}{dt} = \frac{d}{dt}\vec{\tau} \cdot d\vec{\theta} = \vec{\tau} \cdot \frac{d\vec{\theta}}{dt} = \vec{\tau} \cdot \vec{\omega}$$

This is very similar to the power, $P = \vec{F} \cdot \vec{v}$, with which a force does work on an object moving with constant velocity, except that instead of force we use torque, and instead of velocity, we use angular velocity.

12.1.2 Total kinetic energy of an object

In the frame of reference of the centre of mass, an object rotating about an axis through its centre of mass with angular velocity, $\vec{\omega}$, will have rotational kinetic energy, K_{rot}, given by:

$$K_{rot} = \frac{1}{2} I_{CM} \omega^2$$

where I_{CM} is the moment of inertia of the object about the axis through its centre of mass.

We wish to determine the kinetic energy of the object in an inertial frame of reference where the object's centre of mass is moving with a velocity $\vec{v}_{cm}$; that is, in a frame where the axis of rotation is moving with the velocity of the centre of mass. We model the object as being composed of particles of mass, m_i, each located at position, $\vec{r}_i$, relative to the axis of rotation through the centre of mass. The velocity, $\vec{v}_i$, of a particle i, in this frame of reference, is given by:

$$\vec{v}_i = \vec{\omega} \times \vec{r}_i + \vec{v}_{CM}$$

where $\vec{\omega} \times \vec{r}_i$ is the velocity of the particle as seen in the centre of mass (due to rotation). The kinetic energy of particle i, K_i, is given by:

$$K_i = \frac{1}{2} m_i v_i^2 = \frac{1}{2} m_i (\vec{v}_i \cdot \vec{v}_i)$$

where we expressed the speed of the particle squared using a scalar product of the velocity of the particle with itself. The total kinetic energy of the object is found by summing the kinetic energies of all of the particles:

$$\begin{aligned}
K_{tot} &= \sum \frac{1}{2} m_i (\vec v_i \cdot \vec v_i) \\
&= \frac{1}{2} \sum_i m_i (\vec\omega \times \vec r_i + \vec v_{CM}) \cdot (\vec\omega \times \vec r_i + \vec v_{CM}) \\
&= \frac{1}{2} \sum_i m_i (\vec\omega \times \vec r_i) \cdot (\vec\omega \times \vec r_i) + \frac{1}{2} \sum_i m_i (\vec v_{CM}) \cdot (\vec v_{CM}) + \sum_i m_i (\vec\omega \times \vec r_i) \cdot (\vec v_{CM}) \\
&= \frac{1}{2} \sum_i m_i \omega^2 r_i^2 + \frac{1}{2} \sum_i m_i v_{CM}^2 + \sum_i m_i (\vec\omega \times \vec r_i) \cdot (\vec v_{CM}) \\
&= \frac{1}{2} I_{CM} \omega^2 + \frac{1}{2} M v_{CM}^2 + \sum_i m_i (\vec\omega \times \vec r_i) \cdot (\vec v_{CM})
\end{aligned}$$

where the first term is the rotational kinetic energy that we found earlier. The second term, called the "translational kinetic energy", can be thought of as the kinetic energy of the whole system with mass $M = \sum m_i$, due to the translational motion of the centre of mass. The last term is identically zero; we can re-order the scalar product and factor $\vec v_{CM}$ out of the sum:

$$\begin{aligned}
\sum_i m_i (\vec\omega \times \vec r_i) \cdot (\vec v_{CM}) &= (\vec v_{CM}) \cdot \sum_i m_i (\vec\omega \times \vec r_i) \\
&= (\vec v_{CM}) \cdot \sum_i m_i \vec v\,'_i
\end{aligned}$$

where $v'_i = \vec\omega \times \vec r_i$ is the velocity of particle i in the center of mass frame of reference. But the sum:

$$\sum_i m_i \vec v\,'_i$$

is the numerator for the definition of the velocity of the centre of mass, which, in the centre of mass frame of reference is identically zero!

Thus, the total kinetic energy of an object of mass, M, that is rotating about an axis through its centre of mass with angular velocity, ω, and whose centre of mass is moving with velocity, $\vec v_{CM}$, is given by:

$$\boxed{K_{tot} = K_{rot} + K_{trans} = \frac{1}{2} I_{CM} \omega^2 + \frac{1}{2} M v_{CM}^2} \tag{12.4}$$

The total kinetic energy can be thought of as the sum of the rotational and kinetic energies.

12.2 Rolling motion

In this section, we examine how to model the motion of an object that is rolling along a surface, such as the motion of a bicycle wheel. Consider the motion of a wheel of radius, R, rotating with angular velocity, $\vec\omega$, about an axis perpendicular to the wheel and through its centre of mass, **as observed in the centre of mass frame**. This is illustrated in Figure 12.3.

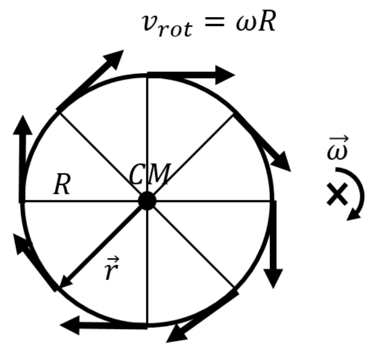


Figure 12.3: A wheel rotating with angular velocity $\vec{\omega}$ about an axis through its centre of mass.

In the frame of reference of the centre of mass, each point on the edge of the wheel has a velocity, $\vec{v}_{rot}$, due to rotation given by:

$$\vec{v}_{rot} = \vec{\omega} \times \vec{r}$$

where $\vec{r}$ is a vector (of magnitude R) from the centre of mass to the corresponding point on the edge of the wheel (shown in Figure 12.3 for a point on the lower left of the wheel). The vector $\vec{r}$ is always perpendicular to $\vec{\omega}$, so that the speed of all points on the edge, as measured in the frame of reference of the centre of mass, is the same:

$$v_{rot} = \omega R \tag{12.5}$$

as illustrated in Figure 12.3.

Now, suppose that the whole wheel is moving, as it rolls on the ground, such that the centre of mass of the wheel moves with a velocity, $\vec{v}_{CM}$, as illustrated in Figure 12.4.

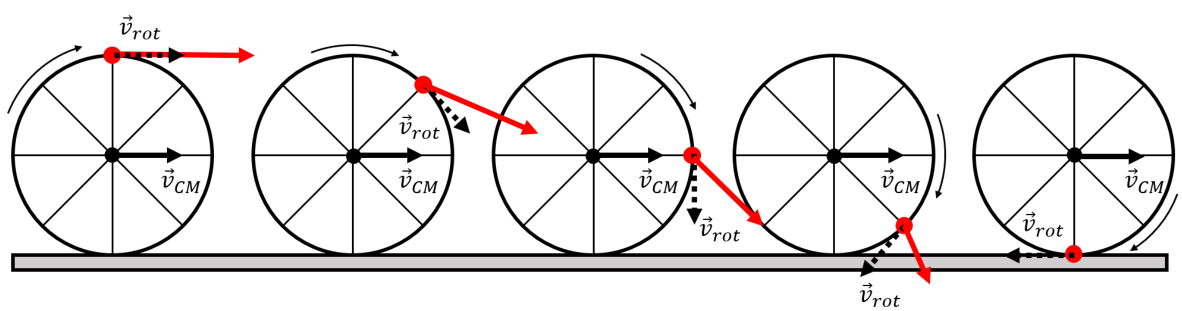


Figure 12.4: A wheel rolling without slipping on the ground, with the centre of mass moving with velocity $\vec{v}_{CM}$. The wheel is shown at different instants in time, as the point shown in red moves around the centre of mass.

In the frame of reference of the ground, each point on the edge of the wheel will have a velocity $\vec{v}$ given by:

$$\vec{v} = \vec{v}_{rot} + \vec{v}_{CM}$$

That is, in the frame reference of the ground, each point will have a velocity obtained by (vectorially) adding its velocity relative to the centre of mass, $\vec{v}_{rot}$, and the velocity of the centre of mass relative to the ground, $\vec{v}_{CM}$. This is illustrated in Figure 12.4 for one specific point, shown in red. The red vector corresponds to the velocity of the red point as the wheel rotates, and is obtained by adding the velocity of the centre of mass, $\vec{v}_{CM}$, and the velocity, $\vec{v}_{rot}$, relative to the centre of mass (shown as the dashed vector, tangent to the edge of the wheel).

Consider, specifically, the instant in time when the red point is at the bottom of the wheel, where the wheel makes contact with the ground. **If the wheel is not slipping with respect to the ground**, then the point is, at that instant, at rest relative to the ground. We call this type of motion "rolling without slipping"; the point on the rotating object that is in contact with the ground is instantaneously at rest relative to the ground. This is the scenario illustrated in Figure 12.4.

For the point in contact with the ground, the vectors $\vec{v}_{rot}$ and $\vec{v}_{CM}$ are anti-parallel, horizontal, and must sum to zero. Writing out the horizontal component of the velocity of that point (choosing the positive direction to be in the direction of the velocity of the centre of mass):

$$\begin{aligned} v &= -v_{rot} + v_{CM} = 0 \\ \therefore v_{rot} &= v_{CM} \end{aligned}$$

and we find that, for rolling without slipping, the speed due to rotation about the centre of mass has to be equal to the speed of the centre of mass. The speed due to rotation about the centre of mass can be expressed using the angular velocity of the wheel about the centre of mass (Equation 12.5). For rolling without slipping, we thus have the following relationship between angular velocity and the speed of the centre of mass:

$$\boxed{\omega R = v_{CM}} \quad \text{(rolling without slipping)} \tag{12.6}$$

It makes sense for the angular velocity to be related to the speed of the centre of mass. The faster the wheel rotates, the faster the centre of mass will move. If the wheel is slipping with respect to the ground, then the point of contact is no longer stationary relative to the ground, and there is no relation between the angular velocity and the speed of the centre of mass. For rolling with slipping, imagine the motion of your bicycle wheel as you try to ride your bike on a slick sheet of ice.

For rolling without slipping, the magnitude of the linear acceleration of the centre of mass, a_{CM}, is similarly related to the magnitude of the angular acceleration of the wheel, α, about the centre of mass:

$$\begin{aligned} a_{CM} &= \frac{dv_{CM}}{dt} = \frac{d}{dt}\omega R = R\frac{d\omega}{dt} \\ \therefore a_{CM} &= R\alpha \end{aligned}$$

Checkpoint 12-1

For rolling without slipping (Figure 12.4), the speed of the point on the wheel that is in contact with the ground is 0. What is the speed of the point at the top of the wheel?

A) 0.
B) v_{CM}.
C) $2v_{CM}$.
D) None of the above.

Example 12-1

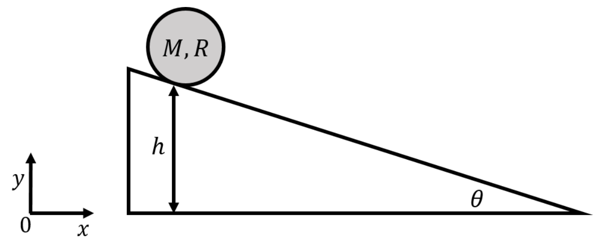


Figure 12.5: A disk rolling without slipping down an incline.

A disk of mass M and radius R is placed on an incline at a height h above the ground. The incline makes an angle θ with respect to the horizontal, as shown in Figure 12.5. If the disk starts at rest and rolls without slipping down the incline, what speed will the centre of mass have when the disk reaches the bottom of the incline?

Solution

We can use the conservation of mechanical energy to determine the speed of the centre of mass at the bottom of the incline, as there are no non-conservative forces doing work on the disk. If we choose to define gravitational potential energy such that it is zero at the bottom of the incline, we can write the total mechanical energy of the disk at the top of the incline as:

$$E = K + U = (0) + Mgh$$

where the kinetic energy is zero, since the disk starts at rest[a]. At the bottom of the incline, the disk will have only kinetic energy, since the potential energy at the bottom is defined to be zero. The kinetic energy of the disk will have a component from the rotation of the disk about the centre of mass, with angular speed ω, and a component from the translation of the centre of mass with speed v_{CM}. The mechanical energy at the bottom of the incline is thus:

$$E' = K' + U = K'_{rot} + K'_{trans} + (0) = \frac{1}{2}I_{CM}\omega^2 + \frac{1}{2}Mv_{cm}^2$$

Since the disk is rolling without slipping, its angular speed is related to the speed of centre of mass:

$$\omega = \frac{v_{CM}}{R}$$

The moment of inertia of the disk about its centre of mass is given by:

$$I_{CM} = \frac{1}{2}MR^2$$

We can thus write the mechanical energy at the bottom of the incline as:

$$\begin{aligned}
E' &= \frac{1}{2}I_{CM}\omega^2 + \frac{1}{2}Mv_{cm}^2 \\
&= \frac{1}{2}\left(\frac{1}{2}MR^2\right)\left(\frac{v_{CM}}{R}\right)^2 + \frac{1}{2}Mv_{cm}^2 \\
&= \frac{3}{4}Mv_{cm}^2
\end{aligned}$$

Applying conservation of energy allows us to determine the speed of the centre of mass at the bottom of the incline:

$$\begin{aligned}
E &= E' \\
Mgh &= \frac{3}{4}Mv_{cm}^2 \\
\therefore v_{CM} &= \sqrt{\frac{4}{3}gh}
\end{aligned}$$

Discussion: This example showed how we can use the conservation of energy to model the motion of an object that is rolling without slipping. The constraint of rolling without slipping allowed for the angular speed of the object to be related to the speed of its centre of mass.

[a] Technically, the potential energy should be taken for the height of the centre of mass, which is a distance $h_{CM} = h + R\cos\theta$ from the ground at the top of the incline, and a height $h'_{CM} = R$ at the bottom of the incline. The net difference in height for the centre of mass is thus $h_{CM} - h'_{CM} = h + R(1 - \cos\theta)$. If we assume that h is much bigger than R, then this is negligible, otherwise, that is what we should use instead of h for the potential energy.

Checkpoint 12-2

A hoop, a disk, and a sphere roll without slipping down an incline. If they are all released at the same time, in what order will they arrive at the bottom?

A) Hoop, disk, sphere.
B) Sphere, disk, hoop.
C) Disk, sphere, hoop.
D) Disk, hoop, sphere.

12.2.1 The instantaneous axis of rotation

When an object is rolling without slipping, we can model its motion as the superposition of rotation about the centre of mass and translational motion of the centre of mass, as in the previous section. However, because the point of contact between the rolling object and the ground is stationary, we can also model the motion as if the object were instantaneously rotating with angular velocity, $\vec\omega$, about a stationary axis through the point of contact. That is, we can model the motion as rotation only, with no translation, if we choose an axis of rotation through the point of contact between the ground and the wheel.

We call the axis through the point of contact the "instantaneous axis of rotation", since, instantaneously, it appears as if the whole wheel is rotating about that point. This is illustrated in Figure 12.6, which shows, in red, the velocity vector for each point on the edge of the wheel, relative to the instantaneous axis of rotation. Because the axis of rotation is fixed to the ground, the velocity of each point about that axis of rotation corresponds to the same velocity relative to the ground that is depicted in Figure 12.4.

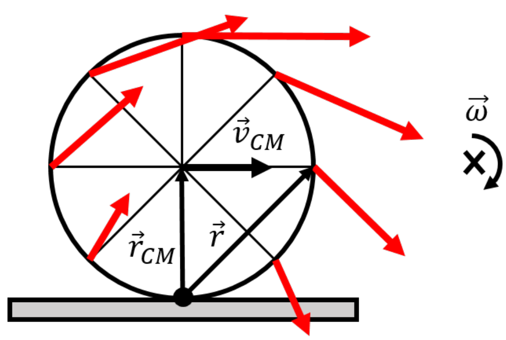


Figure 12.6: A wheel that is rolling without slipping, as viewed if rotating about the instantaneous axis of rotation that passes through the point of contact with the ground.

In particular, the angular velocity, $\vec\omega$, about the instantaneous axis of rotation is the same as when we model the motion as translation plus rotation about the centre of mass ,as in the previous section. Indeed, relative to the instantaneous axis of rotation, the centre of mass must still have a velocity $\vec{v}_{CM}$, which is given by:

$$\vec{v}_{CM} = \vec{\omega} \times \vec{r}_{CM}$$
$$\therefore v_{CM} = \omega R$$

where $\vec{r}_{CM}$ is the vector from the axis of rotation to the centre of mass. This is the same condition for rolling without slipping that we found before. Similarly, the velocity of any point on the wheel, relative to the ground, is given by:

$$\vec{v} = \vec{\omega} \times \vec{r}$$

where $\vec{r}$ is the vector from the axis of rotation to the point of interest (shown in Figure 12.6 for the point on the right side of the wheel). In particular, the velocity vector (in red) for any point is always perpendicular to the vector $\vec{r}$ for that point, which was not necessarily obvious when modelling the motion as rotation plus translation, as in Figure 12.4.

Example 12-2

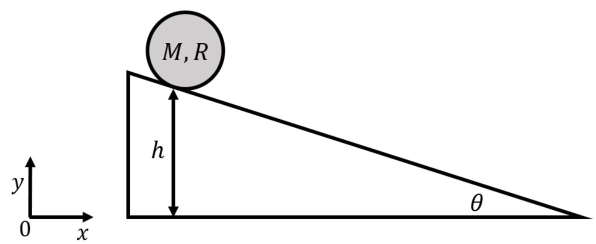


Figure 12.7: A disk rolling without slipping down an incline.

A disk of mass M and radius R is placed on an incline at a height h above the ground. The incline makes an angle θ with respect to the horizontal, as shown in Figure 12.7. What is the angular acceleration of the disk, about an axis through its centre of mass, as it rolls without slipping down the slope?

Solution

In order to determine the angular acceleration of the disk about the centre of mass, we need to model the forces that are exerted on the disk. The forces exerted on the disk are:

1. $\vec{F}_g$, the weight of the disk, exerted downwards at the centre of mass, with magnitude Mg.
2. $\vec{N}$, a normal force perpendicular to the incline, exerted by the incline at the point of contact with the disk.
3. $\vec{f}_s$, a force of static friction parallel to the incline, exerted by the incline at the point of contact with the disk. Without this force, the disk would simply slide down the incline without rotating.

These forces are illustrated in Figure 12.8, along with the acceleration of the centre of mass, and our choice of coordinate system (we choose the x axis parallel to the acceleration of the centre of mass, to facilitate applying Newton's Second Law).

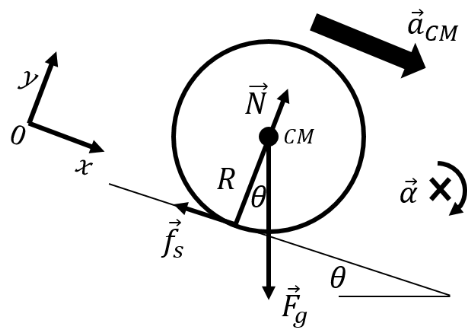


Figure 12.8: The forces on the disk rolling without slipping down an incline.

The angular acceleration of the disk about the centre of mass, $\vec{\alpha}$ is given by Newton's Second Law for rotational dynamics:

$$\vec{\tau}^{ext} = I_{CM}\vec{\alpha}$$

where $\vec{\tau}^{ext}$ is the net external torque on the disk about the centre of mass (which will be in the negative z direction).

The only force that can exert a torque about the centre of mass is the force of static friction. Gravity has a lever arm of zero and the normal force is anti-parallel to the vector that goes from the centre of mass to the point where the force is exerted. The net torque about the centre of mass is thus:

$$\vec{\tau}^{ext} = \vec{\tau}_{f_s} = \vec{r}_{f_s} \times \vec{f}_s = -Rf_s\hat{z}$$

The angular acceleration will thus be in the negative z direction, and the magnitude is given by:

$$\alpha = \frac{\tau^{ext}}{I_{CM}} = \frac{Rf_s}{\frac{1}{2}MR^2} = \frac{2f_s}{MR}$$

However, we do not know the magnitude of the force of static friction. We can use the x and y components of Newton's Second Law to determine it (with acceleration of the centre of mass in the x direction):

$$\sum F_x = F_g \sin\theta - f_s = Ma_{CM}$$
$$\sum F_y = N - F_g \cos\theta = 0$$

Because the disk is rolling without slipping, the acceleration of the centre of mass is related to the angular acceleration of the disk:

$$a_{cm} = \alpha R$$

The x component of Newton's Second Law can thus be used to determine the magnitude of the force of static friction in terms of the angular acceleration:

$$\begin{aligned}
Mg\sin\theta - f_s &= M\alpha R\\
\therefore f_s &= Mg\sin\theta - M\alpha R
\end{aligned}$$

We can then substitute out the force of friction from our previous formula for the angular acceleration:

$$\begin{aligned}
\alpha &= \frac{2f_s}{MR}\\
&= \frac{2Mg\sin\theta - 2M\alpha R}{MR} = \frac{2g\sin\theta}{R} - 2\alpha\\
\therefore \alpha &= \frac{2g\sin\theta}{3R}
\end{aligned}$$

Instead of modelling the motion of the disk as rotation about the centre of mass and translation of the center of mass, we can also model it about the instantaneous axis of rotation.

The angular acceleration about the instantaneous axis of rotation will be the same as the angular acceleration about the centre of mass. About the instantaneous axis of rotation, only the force of gravity can exert a torque, since the normal force and the force of friction both have a lever arm of zero. The torque from the force of gravity, about the instantaneous axis of rotation is:

$$\vec{\tau}_g = -F_g R\sin\theta\hat{z} = -MgR\sin\theta\hat{z}$$

The torque from the force of gravity is equal to the moment of inertia of the disk about the instantaneous axis of rotation, I, multiplied by its angular acceleration:

$$\begin{aligned}
\tau^{ext} &= \tau_g = I\alpha\\
\therefore \alpha &= \frac{\tau_g}{I} = \frac{MgR\sin\theta}{I}
\end{aligned}$$

The moment of inertia about the instantaneous axis of rotation is easily found using the parallel axis theorem:

$$I = I_{CM} + MR^2 = \frac{1}{2}MR^2 + MR^2 = \frac{3}{2}MR^2$$

This allows us to find the angular acceleration of the disk:

$$\begin{aligned}\alpha &= \frac{MgR\sin\theta}{I} = \frac{MgR\sin\theta}{\frac{3}{2}MR^2}\\ &= \frac{2g\sin\theta}{3R}\end{aligned}$$

as we found previously, but in this case, we did not need to use Newton's Second Law to determine the force of friction.

Discussion: We saw that we can model the dynamics of the rolling body using either an axis through the centre of mass, or an axis through the instantaneous axis of rotation. The latter was easier in this case, because it did not require using Newton's Second Law.

By using an axis through the centre of mass to model the motion of the disk, it was clear that the force of static friction is required in order for the disk to rotate. Without the force of static friction, the disk would slide along the surface of the incline. The disk could still rotate if there is a force of kinetic friction that causes a torque that rotates the disk. If the surface were completely frictionless, the disk would simply slide down the incline, and we could model it as a sliding block. If the incline is too steep the force of static friction is no longer sufficient to provide the necessary torque required for the angular acceleration to be that which corresponds to rolling without slipping, and the disk would slip.

12.3 Angular momentum

In this section, we show that we can define a quantity called "angular momentum" as the rotational equivalent of the linear momentum.

12.3.1 Angular momentum of a particle

The angular momentum relative to a point of rotation, $\vec{L}$, of a particle with linear momentum, $\vec{p}$, is defined as:

$$\boxed{\vec{L} = \vec{r}\times\vec{p}} \tag{12.7}$$

where $\vec{r}$ is the vector from the point of rotation to the particle, and the linear momentum, $\vec{p}$, is defined relative to an inertial frame of reference in which the point of rotation is at rest.

Consider the time-derivative of angular momentum (where we have to use the product rule for derivatives):

$$\begin{aligned}\frac{d\vec{L}}{dt} &= \frac{d}{dt}(\vec{r}\times\vec{p})\\ &= \frac{d\vec{r}}{dt}\times\vec{p}+\vec{r}\times\frac{d\vec{p}}{dt}\\ &= \vec{v}\times\vec{p}+\vec{r}\times\frac{d\vec{p}}{dt}\end{aligned}$$

The first term is zero, since $\vec{v}$ is parallel to $\vec{p}$ by definition. Recall Newton's Second Law written using linear momentum:

$$\frac{d\vec{p}}{dt} = \vec{F}^{net}$$

where $\vec{F}^{net}$ is the net force on the particle relative to the point of rotation. The rate of change of angular momentum is thus given by:

$$\begin{aligned}\frac{d\vec{L}}{dt} &= \vec{r}\times\frac{d\vec{p}}{dt}\\ &= \vec{r}\times\vec{F}^{net}\end{aligned}$$

where the term on the right is the net torque on the particle. Thus, the rate of change of angular momentum is given by:

$$\boxed{\frac{d\vec{L}}{dt} = \vec{\tau}^{net}} \tag{12.8}$$

which is analogous to the linear case, but we used angular momentum instead of linear momentum and net torque instead of net force. The net torque on a particle is thus equal to the rate of change of its angular momentum. In particular, the angular momentum of a particle will remain constant (not change with time) if the net torque on the particle is zero.

We can also define the angular momentum of a particle using only angular quantities:

$$\vec{L} = \vec{r}\times\vec{p} = m\vec{r}\times\vec{v} = mr^2\vec{\omega}$$

where we factored the mass m out of the momentum and used the definition $\vec{\omega} = 1/r^2(\vec{r}\times\vec{v})$. We can think of mr^2 as the moment of inertia, I, of the particle and write:

$$\boxed{\vec{L} = mr^2\vec{\omega} = I\vec{\omega}} \tag{12.9}$$

which is a close analogue to the definition of linear momentum, but we use moment of inertia instead of mass and angular velocity instead of velocity.

The angular momentum is thus parallel to the angular velocity of the particle about the point of rotation. If no net torque is exerted on the particle about that point, then the particle's angular momentum about that point will remain constant. We can also consider the torque and angular momentum about an axis instead of a point; in that case, we would simply take the components of torque and angular momentum that are parallel to that axis.

Example 12-3

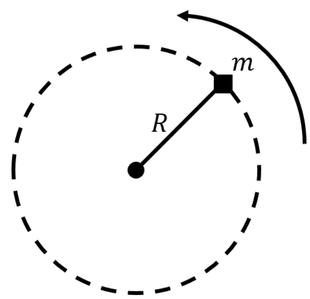


Figure 12.9: A small block attached to a mass-less string moving in a horizontal circle on a table.

A small block of mass m attached to a mass-less string is moving along a circle of radius R on a horizontal table, as depicted from above in Figure 12.9. If the table is frictionless: are the block's linear and/or angular momentum with respect to the axis of rotation conserved? If there is friction between the table and the block, are the block's linear and/or angular momentum with respect to the axis of rotation conserved? What can you say about the kinetic energy of the block in the two cases?

Solution

If there is no friction between the block and the table, then the forces exerted on the block are:

1. $\vec{F}_g$, the block's weight, exerted downwards, with magnitude mg.
2. $\vec{N}$, a normal force, exerted upwards, with magnitude mg.
3. $\vec{T}$, a force of tension, exerted towards the centre of the circle.

All of these forces are perpendicular to the (tangential) displacement of the block along the circle. Thus, there can be no work done on the block and its speed, v, must remain constant. The kinetic energy of the block must thus remain constant.

The sum of the forces on the block must be towards the centre of the circle, since the block is in uniform circular motion. The linear momentum of the block cannot be conserved if there is a net force on the block (and clearly, the block's velocity vector changes direction as it goes around the circle).

The forces of weight and the normal force are both outside of the plane of motion, and thus cannot exert a torque along the axis of rotation. They are also equal and opposite in magnitude so the net torque from those two forces is always zero (since the net force from those forces is zero). The force of tension is always anti-parallel to the vector $\vec{r}$, from the axis of rotation to the particle, and cannot result in a torque about the rotation axis. Thus, the net torque on the block is zero and its angular momentum must be conserved.

If there is kinetic friction exerted by the table on the block, then there is an additional force, $\vec{f}_s$, exerted on the block in the direction opposite of motion (tangent to the circle, in the opposite direction from the block's velocity).

The force of friction will do negative work on the block, slowing it down and reducing its kinetic energy, which is no longer conserved. The net force on the block is non-zero, so its linear momentum is still not conserved. Finally, the force of friction, which is always perpendicular to $\vec{r}$, will result in a torque that reduces the angular velocity of the block. The block's angular momentum is thus no longer conserved when there is friction between the table and the block.

Discussion: In this example, we saw that kinetic energy, linear momentum, and angular momentum are all conserved under different conditions. Kinetic energy is conserved if no net work is done on the block. Linear momentum is conserved if the net force on the block is zero. Angular momentum is conserved if the net torque on the block is zero. By introducing angular momentum, we are able to use a new conserved quantity to help us model rotational dynamics.

Example 12-4

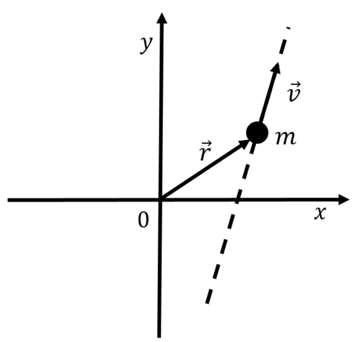


Figure 12.10: A particle moving in a straight line.

A particle is moving with constant velocity $\vec{v}$ (in a straight line) relative to a coordinate system in an inertial frame of reference, as shown in Figure 12.10. Show that its angular momentum about the origin is conserved.

Solution

In this case, the particle is moving in a straight line, but we can still define its angular momentum relative to the origin. If $\vec r$ is the position vector of the particle relative to the origin, its angular momentum is:

$$\vec L = \vec r \times \vec p$$

We can take the time derivative of the angular momentum to see if it changes with time:

$$\begin{aligned}\frac{d\vec L}{dt} &= = \frac{d}{dt}(\vec r \times \vec p)\\ &= \frac{d\vec r}{dt} \times \vec p + \vec r \times \frac{d\vec p}{dt}\\ &= \vec v \times \vec p + \vec r \times \frac{d\vec p}{dt}\end{aligned}$$

The first term is zero because $\vec v$ and $\vec p$ are parallel (so their cross-product must be zero). The second term is zero because the particle's momentum is constant in time (since its velocity is constant). Thus, the particle's angular momentum does not change with time, and it is conserved.

Discussion: Of course, we expected this result since no net torque is exerted on the particle. It is however worth highlighting that a particle does not need to be rotating for its angular momentum about a given axis to be defined or conserved; all that matters is that there is no net torque on the particle relative to that axis.

12.3.2 Angular momentum of an object or system

Consider a system made of many particles of mass, m_i, each with a position, $\vec r_i$, and velocity, $\vec v_i$, relative to a point of rotation that is fixed in an inertial frame of reference.

We can write Newton's Second Law using the angular momentum, $\vec L_i$, for particle i:

$$\frac{d\vec L_i}{dt} = \vec \tau_i^{net}$$

where $\vec \tau_i^{net}$ is the net torque exerted on particle i. We can sum each side of this equation for all of the particles in the system:

$$\begin{aligned}\frac{d\vec L_1}{dt} + \frac{d\vec L_2}{dt} + \frac{d\vec L_3}{dt} + \dots &= \vec \tau_1^{net} + \vec \tau_2^{net} + \vec \tau_3^{net} + \dots\\ \therefore \frac{d}{dt}\sum_i \vec L_i &= \sum_i \vec \tau_i^{net}\end{aligned}$$

The sum of all of the torques on all of the particles will include a sum over torques that are internal to the system and torques that are external to the system. The sum over internal torques is zero:

$$\sum_i \vec{\tau}_i^{net} = \sum_i \vec{\tau}_i^{int} + \sum_i \vec{\tau}_i^{ext} = \sum_i \vec{\tau}_i^{ext} = \vec{\tau}^{ext}$$

where we defined, $\vec{\tau}^{ext}$, to be the net external torque exerted on the system. We also introduce the total angular momentum of the system, $\vec{L}$, as the sum of the angular momenta of the individual particles:

$$\vec{L} = \sum_i \vec{L}_i$$

The rate of change of the total angular momentum of the system is then given by:

$$\boxed{\frac{d\vec{L}}{dt} = \vec{\tau}^{ext}} \tag{12.10}$$

Up to this point, we did not require that the system be a solid object, so the particles in the system can move relative to each other. For example, the particles could be the Sun, planets, and everything else that is in our Solar System. The total angular momentum of all of the bodies in the Solar System (say, relative to the Sun) is conserved if there is no net torque on the solar system relative to the Sun (i.e. if there is no torque about the Sun exerted on any of the bodies in the system that is not exerted by one of the other bodies in the system).

Now, consider a solid object that is modelled as a system of many particles of mass, m_i, at position, $\vec{r}_i$, with velocity, $\vec{v}_i$, relative to a fixed axis of rotation. We can define the angular momentum of a single particle as (Equation 12.9):

$$\vec{L}_i = m_i r_i^2 \vec{\omega}_i^2$$

The total momentum of the system is the sum of the angular momenta of the individual particles:

$$\vec{L} = \sum_i \vec{L}_i = \sum_i m_i r_i^2 \vec{\omega}_i^2$$

Because all of the particles are part of the same object, they must all move in unison and have the same angular velocity, $\vec{\omega}$, relative to the axis of rotation. We can thus define the angular momentum about the rotation axis for a solid object with angular velocity, $\vec{\omega}$, as:

$$\boxed{\vec{L} = \left(\sum_i m_i r_i^2\right)\vec{\omega} = I\vec{\omega}} \tag{12.11}$$

where we recognized that the sum in parentheses is simply the moment of inertia of the object relative to the axis of rotation. Again, it should be emphasized that this is the total angular momentum of the object about an axis of rotation, and not about a point.

Visualizing the torque and angular momentum of a system can be challenging because it almost always requires visualizing something in three dimensions. Consider a wheel (e.g. a bicycle wheel) that is spinning about horizontal axle which you hold with your hands, as illustrated in the left panel of Figure 12.11 (without the hands). Imagine that you are holding onto the axle so that the wheel is front of you, your right hand is to the right of the wheel and your left hand is to the left of the wheel.

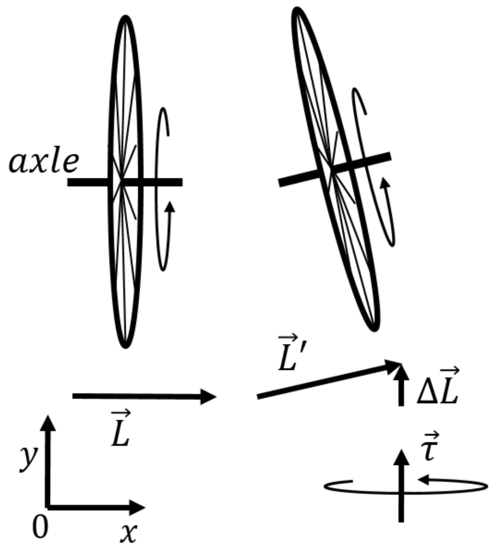


Figure 12.11: A wheel rotating on an axle, with a horizontal angular velocity (left). If you try to tilt the axle as shown in the right panel, changing the angular momentum of the wheel, you will also need to exert a torque in the vertical direction (shown at the bottom right).

We define a coordinate system as shown so that the wheel is spinning as shown in the left panel, with angular velocity (and angular momentum) in the positive x direction (the top of the wheel is coming towards you).

You then try to lift your right hand while lowering your left hand in order to tilt the rotation axis, as shown in the right panel. In doing so, you change the direction of the angular momentum (and angular velocity) of the wheel such that the angular momentum, $\vec{L}'$, now has a vertical component, $\Delta\vec{L}$, as shown. The torque that is required in order to change the angular momentum is given by:

$$\vec{\tau} = \frac{d\vec{L}}{dt} \sim \frac{\Delta\vec{L}}{\Delta t}$$

where Δt is the time that it takes to change the axis of rotation. The torque required in order to change the axis of rotation is directed in the same direction as $\Delta\vec{L}$ (the positive y direction). That is, you will not be able to simply tilt the axle as shown; if you want to tilt the axle, you will also need to push forward with you right hand and pull backwards with your left hand to exert the required torque (shown in the bottom right of the figure)! If you simply try to tilt the rotation axis, your right hand will be pushed towards you and your left hand away from you, as a reaction to the torque that would otherwise be required to tilt the axis!

12.3.3 Conservation of angular momentum

In the previous section, we saw that the net external torque that is exerted on an object (or system) is equal to the rate of change of its angular momentum:

$$\frac{d\vec{L}}{dt} = \vec{\tau}^{ext}$$

where the angular momentum and torque are measured about the same axis or point of rotation, fixed in an inertial frame of reference.

The total angular momentum of a system about a point of rotation is conserved (i.e. does not change with time) if there is no net external torque exerted on the system about that point. If one makes the system large enough, then all of the torques can be taken to be internal, and the angular momentum of the system is conserved. The angular momentum of the Universe about a fixed point is thus conserved.

Conservation of angular momentum is another conservation law that we derived from Newton's Second Law. In the modern formulation of physics, we understand that the conservation of angular momentum is associated with rotational symmetry of Newton's Second Law; it does not matter from which "angle" we model a system, we can always use Newton's Second Law. Similarly, conservation of linear momentum is associated with translational symmetry and conservation of energy is associated with the fact that Newton's Second Law does not change with time. Angular momentum is fundamentally different than linear momentum and energy, and is conserved under different conditions. The angular momentum of a system about a given axis/point is conserved if there is no net torque on the system about that axis/point.

Example 12-5

During a spin, a figure skater brings his arms close to his body and increases his angular velocity from ω_1 to ω_2. By what fraction did his moment of inertia decrease in doing so?

Solution

We can consider the rotation axis to be vertical through the centre of the skater. When the figure skater is spinning, there is no net external torque on him. Thus, his angular momentum is conserved as he bring his arms in. As he bring his arms in, his moment of inertia decreases, since he is bringing the mass of his arms closer to the axis of rotation. If I_1 and I_2 are the moments of inertia of the skater before and after brining his arms in, respectively, we can write the angular momentum about his axis of rotation as:

$$L_1 = I_1\omega_1$$
$$L_2 = I_2\omega_2$$

Since there is no external torque on the skater, the angular momentum is the same before and after he changes his moment of inertia:

$$
\begin{aligned}
L_1 &= L_2 \\
I_1\omega_1 &= I_2\omega_2 \\
\therefore \frac{I_1}{I_2} &= \frac{\omega_2}{\omega_1}
\end{aligned}
$$

Discussion: A spinning figure skater is a good example of the conservation of angular momentum. By changing their shape, they can change their moment of inertia and thus their angular velocity.

Example 12-6

Show that Kepler's Second Law is equivalent to a statement about conservation of the angular momentum of a planet orbiting the Sun.

Solution

Kepler's Second Law states that in a period of time Δt, the area, ΔA, that is swept out by a planet is constant, regardless of where it is along its orbit. In other words:

$$\frac{\Delta A}{\Delta t} = \text{constant}$$

Figure 12.12 shows a planet in an elliptical orbit around the sun.

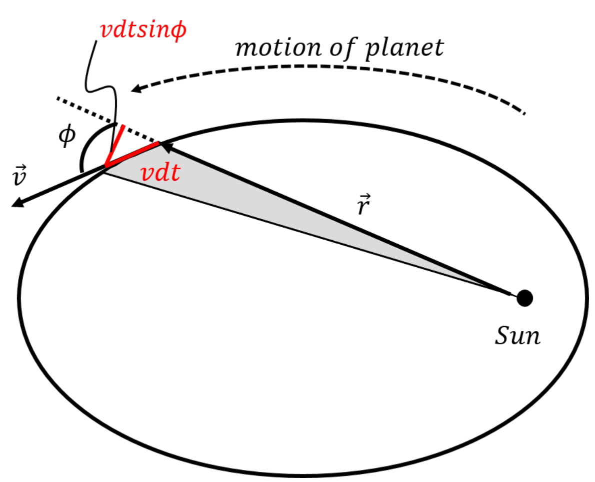


Figure 12.12: The area swept out by a planet in a period of time dt.

At some point in time, the planet has a velocity vector $\vec{v}$ and position vector $\vec{r}$ relative to the Sun. In a small period of time dt, the planet will move along a short distance vdt, which we can take as a straight line if dt is small enough. Let ϕ be the angle between the velocity and position vectors when these are tail to tail, as illustrated.

The small amount of area, dA, swept out by the planet in a period of time dt, is given by the area of the right angle triangle with height r and base $vdt \sin\phi$[a]:

$$dA = \frac{1}{2} rvdt \sin\phi$$

The rate at which the area is swept out is thus:

$$\frac{dA}{dt} = \frac{1}{2} rv \sin\phi$$

Consider now the magnitude of the planet's angular momentum about the Sun:

$$L = rp \sin\phi = rmv \sin\phi$$

where the mass of the planet is m. The rate at which the planet sweeps out the area can be written in terms of the angular momentum of the planet:

$$\frac{dA}{dt} = \frac{1}{2} rv \sin\phi = \frac{L}{2m}$$

The only force exerted on the planet is the gravitational force from the Sun. That force is always anti-parallel to the vector $\vec{r}$ from the Sun to the planet, and cannot result in a torque on the planet about the Sun. Thus, the angular momentum of the planet about the Sun must be conserved, and L is constant. In turn, this means that the rate at which area is swept out by the planet, which is proportional to L, is also constant. Thus, Kepler's Second Law is equivalent to saying that the angular momentum of a planet relative to the Sun is constant.

[a] This is only exact in the limit of $dt \to 0$, when the small area from the extra piece outside of the ellipse vanishes.

12.4 Summary

Key Takeaways

If an object is rotating with angular speed, ω, about an axis that is fixed in an inertial frame of reference, the rotational kinetic energy of that object is given by:

$$K_{rot} = \frac{1}{2}I\omega^2$$

where I is the moment of inertia of that object about the axis of rotation.

The net work done by the net torque exerted on an object about a fixed axis or rotation in an inertial frame of reference is equal to object's change in rotational kinetic energy:

$$W = \int_{\theta_1}^{\theta_2} \vec{\tau}^{net} \cdot d\vec{\theta} = \frac{1}{2}I\omega_2^2 - \frac{1}{2}I\omega_1^2$$

If a torque, $\vec{\tau}$, about a stationary axis is exerted on an object that is rotating with a constant angular velocity, $\vec{\omega}$, about that axis, then the torque does work at a rate:

$$P = \vec{\tau} \cdot \vec{\omega}$$

If an object of mass, M, is rotating about an axis through its centre of mass, and the centre of mass of is moving with speed, v_{CM}, relative to an inertial frame of reference, then the total kinetic energy of the object is given by:

$$K_{tot} = K_{rot} + K_{trans} = \frac{1}{2}I_{CM}\omega^2 + \frac{1}{2}Mv_{CM}^2$$

where, ω, is the angular speed of the object about the centre of mass, and, I_{CM}, is the moment of inertia of the object about the centre of mass. The two terms in the kinetic energy come from the rotation about the centre of mass (K_{rot}), and the translational motion of the centre of mass (K_{trans}).

An object is said to be rolling without slipping on a surface if the point on the object that is in contact with the surface is instantaneously at rest relative to the surface. We can model an object that is rolling without slipping by superimposing rotational motion about the centre of mass with translational motion of the centre of mass. The angular speed, ω, and the angular acceleration, α, of the object about an axis through its centre of mass are related to the speed, v_{CM}, and linear acceleration, a_{CM}, of the centre of mass, respectively:

$$v_{CM} = \omega R$$
$$a_{CM} = \alpha R$$

These conditions are equivalent to stating that the object is rolling without slipping.

When an object is rolling without slipping, we can also model its motion as if it were instantaneously rotating about an axis that goes through the point of contact between the object and the ground (the instantaneous axis of rotation). The angular speed (and acceleration) about the instantaneous axis of rotation are the same as they are when the object is modelled as rotating about its (moving) centre of mass.

An object can only be rolling without slipping if there is a force of static friction exerted by the surface on the object. Without this force, the object would slip along the surface.

We can define the angular momentum of a particle, $\vec{L}$, about a point in an inertial frame of reference as:

$$\vec{L} = \vec{r} \times \vec{p}$$

where, $\vec{r}$, is the vector from the point to the particle, and, $\vec{p}$, is the linear momentum of the particle. If the particle has an angular velocity, $\vec{\omega}$, relative to an axis of rotation its angular momentum about that axis can be written as:

$$\vec{L} = mr^2\vec{\omega} = I\vec{\omega}$$

where, r, is the distance between the particle and the axis of rotation, and $I = mr^2$, can be thought of as the moment of inertia of the particle about that axis.

We can write the equivalent of Newton's Second Law for the rotational dynamics of a particle using angular momentum:

$$\frac{d\vec{L}}{dt} = \vec{\tau}^{net}$$

where, $\vec{\tau}^{net}$, is the net torque on the particle about the same point used to define angular momentum. That point must be in an inertial frame of reference.

The rate of change of the total angular momentum for a system of particles, $\vec{L} = \vec{L}_1 + \vec{L}_2 + \dots$, about a given point is given by:

$$\frac{d\vec{L}}{dt} = \vec{\tau}^{ext}$$

where, $\vec{\tau}^{ext}$, is the net external torque on the system about the point of rotation. If the net external torque of the system is zero, then the total angular momentum of the system is constant (conserved). Again, the point of rotation must be in an inertial frame of reference[a].

For a solid object, in which all of the particles must move in unison, we can define the angular momentum of the object about a stationary axis to be:

$$\vec{L} = I\vec{\omega}$$

where, $\vec{\omega}$, is the angular velocity of the object about that axis, and, I, is the object's corresponding moment of inertia about that axis.

Many of the relations that exist between linear quantities have an analogue relation between the corresponding angular quantities, as summarized in the table below:

Name	**Linear**	**Angular**	**Correspondence**
Displacement	s	$\vec{\theta}$	$d\vec{\theta} = \frac{1}{r^2}\vec{r} \times d\vec{s}$
Velocity	$\vec{v}$	$\vec{\omega}$	$\vec{\omega} = \frac{1}{r^2}\vec{r} \times \vec{v}$, $v_s = \vec{\omega} \times \vec{r}$[b]
Acceleration	$\vec{a}$	$\vec{\alpha}$	$\vec{\alpha} = \frac{1}{r^2}\vec{r} \times \vec{a}$, $a_s = \vec{\alpha} \times \vec{r}$[c]
Inertia	m	I	$I = \sum_i m_i r_i^2$
Momentum	$\vec{p} = m\vec{v}$	$\vec{L} = I\vec{\omega}$	$\vec{L} = \vec{r} \times \vec{p}$
Newton's Second Law	$\vec{F}^{ext} = m\vec{a}_{CM}$	$\vec{\tau}^{ext} = I\vec{\alpha}$	$\vec{F} \to \vec{\tau}$, $m \to I$, $\vec{a} \to \vec{\alpha}$
Newton's Second Law	$\frac{d\vec{p}}{dt} = \vec{F}^{ext}$	$\frac{d\vec{L}}{dt} = \vec{\tau}^{ext}$	$\vec{F} \to \vec{\tau}$, $\vec{p} \to \vec{L}$
Kinetic energy	$\frac{1}{2}mv^2$	$\frac{1}{2}I\omega^2$	$m \to I$, $v \to \omega$
Power	$\vec{F} \cdot \vec{v}$	$\vec{\tau} \cdot \vec{\omega}$	$\vec{F} \to \vec{\tau}$, $\vec{v} \to \vec{\omega}$

[a] Technically, if the point is the centre of mass, then this is valid even in an accelerating frame of reference.

[b] This corresponds to the component of velocity perpendicular to $\vec{r}$.

[c] This corresponds to the component of acceleration perpendicular to $\vec{r}$.

Important Equations

Rotational kinetic energy of a rotating object:

$$K_{rot} = \frac{1}{2}I\omega^2$$

Total kinetic energy:

$$K_{tot} = K_{rot} + K_{trans} = \frac{1}{2}I_{CM}\omega^2 + \frac{1}{2}Mv_{CM}^2$$

Work:

$$W = \int_{\theta_1}^{\theta_2} \vec{\tau}^{net} \cdot d\vec{\theta} = \frac{1}{2}I\omega_2^2 - \frac{1}{2}I\omega_1^2$$

Power:

$$P = \vec{\tau} \cdot \vec{\omega}$$

Angular momentum:

$$\vec{L} = \vec{r} \times \vec{p}$$

$$\vec{L} = mr^2\vec{\omega} = I\vec{\omega}$$

$$\frac{d\vec{L}}{dt} = \vec{\tau}^{net}$$

$$\frac{d\vec{L}}{dt} = \vec{\tau}^{ext}$$

$$\vec{L} = I\vec{\omega}$$

12.5 Thinking about the material

12.5.1 Reflect and research

Reflect and research

1. How can a bicycle move forward? Draw the external forces on the bicycle that are required for the wheels to turn.
2. Does conservation of angular momentum play a role in being able to remain upright on a bicyle? If yes, how?
3. How does an anti-lock braking system (ABS) provide better breaking for your car? What is the physics behind this?

To try at home

1. Describe how you can qualitatively confirm conservation of angular momentum.

Reflect and research

1. Propose an experiment to measure the critical angle of an incline, above which a given object cannot roll without slipping, and compare this to a model prediction.
2. Propose an experiment to test the conservation of angular momentum of a rotating object.
3. Propose an experiment to test whether an object with constant velocity can impart angular momentum to another object.

12.6 Sample problems and solutions

12.6.1 Problems

Problem 12-1: A yo-yo can be modelled as two uniform disks, of radius R_2, attached to either side of a smaller uniform disk of radius R_1, as in Figure 12.13. We can assume that all three disks have a mass m. A mass-less string is wrapped around the smaller disk and then the yo-yo is released. What is the acceleration of the centre of mass of the yo-yo as it falls and the string unwinds?

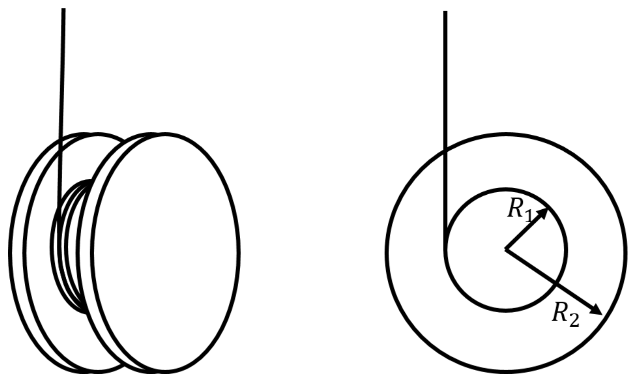


Figure 12.13: Left: Side view of the yo-yo. Right: Front view of the yo-yo, modelled as two disks of radius of R_2 attached to either side of a disk of radius R_1.

(Solution)

Problem 12-2:

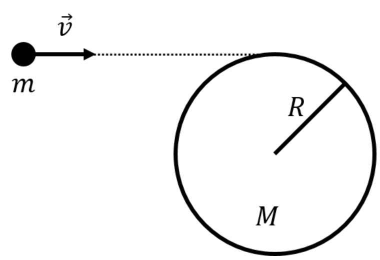


Figure 12.14: A projectile of mass m is about to collide with a disk that can spin about its axis of symmetry. View from above.

A projectile of mass m is fired towards a stationary disk of radius R and mass M that lies on a horizontal table, as depicted from above in Figure 12.14. The disk is in the horizontal plane and can rotate about a vertical axis through its centre. The axle about which the disk rotates is attached to the table and cannot move. The projectile's velocity, $\vec{v}$, is horizontal and such that the projectile embeds itself at the edge of the disk. What is the angular velocity of the disk, about its centre, after the projectile has embedded itself into the disk? Was the collision elastic? Was linear momentum conserved during the collision? (Solution)

12.6.2 Solutions

Solution to problem 12-1:

The forces acting on the yo-yo are:

- $\vec{F}_g$, its weight, with magnitude $3mg$.
- $\vec{T}$, a force of tension from the string.

The forces, where they are exerted, and our choice of coordinate system are shown in Figure 12.15.

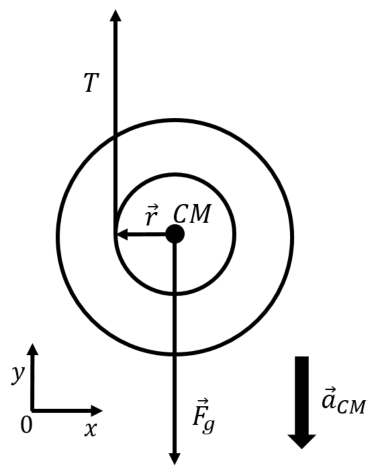


Figure 12.15: Free body diagram for the yo-yo.

The yo-yo can be modelled as rolling without slipping, as if it were rolling along the string that unwinds. The torque about the centre of mass is provided by the tension in the string. The angular acceleration of the yo-yo, α, will be related to the linear acceleration of the centre of mass, $\vec{a}_{CM}$, since this is rolling without slipping:

$$a_{CM} = \alpha R_1$$

where R_1 is the radius that is analogous to rolling motion. Since the torque from the force of gravity is zero, we can write Newton's Second Law for rotational quantities as:

$$\vec{\tau}^{ext} = I\vec{\alpha}$$
$$TR_1 = I\alpha$$

where TR_1 is the magnitude of the torque from the force of tension, since the tension is perpendicular to the vector $\vec{r}$ between the centre of mass and the point where the tension is exerted. The moment of inertia of the yo-yo about its centre of mass is the sum of the moments of inertia of the three disks about their axis of symmetry:

$$I = \frac{1}{2}MR_2^2 + \frac{1}{2}MR_2^2 + \frac{1}{2}MR_1^2 = \frac{1}{2}M(2R_2^2 + R_1^2)$$

We can also write Newton's Second Law in the vertical direction for the yo-yo (of mass $3M$):

$$\begin{aligned}\sum F_y = -F_g + T &= -3Ma_{CM}\\ -3Mg + T &= -3Ma_{CM}\end{aligned}$$

where we a_{CM} is the magnitude of the acceleration of the centre of mass (since we included the sign in the first equation).

We can eliminate the unknown force of tension from the equations by substitution. Using the equation from Newton's Second Law:

$$T = 3M(g - a_{CM})$$

and substituting this into the rotational equation:

$$\begin{aligned}TR_1 &= I\alpha\\ 3M(g - a_{CM})R_1 &= I\alpha\end{aligned}$$

We can solve for a_{CM} by using the condition for rolling without slipping ($\alpha R_1 = a_{CM}$):

$$\begin{aligned}3M(g - a_{CM})R_1 &= I\frac{a_{CM}}{R_1}\\ \frac{I}{R_1}a_{CM} + 3MR_1a_{CM} &= 3MgR_1\\ a_{CM}\left(\frac{I}{R_1} + 3MR_1\right) &= 3MgR_1\\ a_{CM} &= \frac{3MgR_1}{\frac{I}{R_1} + 3MR_1}\\ &= \frac{3MgR_1}{\frac{\frac{1}{2}M(2R_2^2+R_1^2)}{R_1} + 3MR_1}\\ &= \left(\frac{3R_1^2}{\frac{1}{2}(2R_2^2 + R_1^2) + 3R_1^2}\right)g\\ \therefore a_{CM} &= \left(\frac{3R_1^2}{R_2^2 + \frac{7}{2}R_1^2}\right)g\end{aligned}$$

Solution to problem 12-2: We consider the projectile and disk as a system, and a rotation axis that passes through the centre of disk. There are no external torques exerted on the system about the rotation axis, so the angular momentum of the system must be conserved through the collision. Before the collision, only the projectile has angular momentum about the axis of rotation, so the magnitude of the angular momentum before the collision is:

$$L = rp\sin\phi$$

where ϕ is the angle between the particle's momentum, $\vec{p} = m\vec{v}$, and a vector, $\vec{r}$, from the axis of rotation to the particle. We can calculate the particle's angular momentum just before the collision, so that $\vec{r}$ is the vector from the centre of the circle to the point where the particle collides (with magnitude R, and perpendicular to $\vec{v}$). The initial angular momentum of the system is thus:

$$L = rp = Rmv$$

After the collision, the projectile is embedded in the disk. The resulting object has a moment of inertia given by:

$$I = I_{disk} + I_{particle} = \frac{1}{2}MR^2 + mR^2$$

After the collision, the angular momentum of the disk with the embedded projectile is given by:

$$L' = I\omega = \left(\frac{1}{2}M + m\right) R^2\omega$$

Using conservation of angular momentum, the angular velocity of the disk after the collision is:

$$\begin{aligned}
L &= L' \\
Rmv &= \left(\frac{1}{2}M + m\right) R^2\omega \\
\therefore \omega &= \frac{mv}{\left(\frac{1}{2}M + m\right) R}
\end{aligned}$$

We do not expect that mechanical energy is conserved during the collision, since the projectile embeds itself, which must cost energy. The mechanical energy before the collision is given by the kinetic energy of the projectile:

$$E = \frac{1}{2}mv^2$$

After the collision, the kinetic energy is the rotational kinetic energy of the disk with embedded projectile about the axis of rotation:

$$\begin{aligned}
E' &= \frac{1}{2}I\omega^2 = \frac{1}{2}\left(\frac{1}{2}M + m\right) R^2 \left(\frac{mv}{\left(\frac{1}{2}M + m\right) R}\right)^2 \\
&= \frac{1}{2}\frac{m^2}{\left(\frac{1}{2}M + m\right)}v^2
\end{aligned}$$

We can see that E' is less than E, by taking their ratio:

$$\begin{aligned}
\frac{E'}{E} &= \frac{\frac{1}{2}\frac{m^2}{\left(\frac{1}{2}M+m\right)}v^2}{\frac{1}{2}mv^2} \\
&= \frac{m}{\left(\frac{1}{2}M + m\right)} < 1
\end{aligned}$$

and we confirm that mechanical energy is not conserved in the collision (and that energy was lost since one had to deform the projectile and disk).

Linear momentum is clearly not conserved since the final linear momentum is zero, whereas before the collision, it is $\vec{p} = m\vec{v}$. The centre of mass of the disk+projectile system moves before the collision and not after. There must thus be a net external force that is exerted on the system. That force is exerted by the table onto the axle of disk, as the disk would otherwise recoil when hit with the projectile.

Discussion: In this example, we used conservation of angular momentum to model a collision. The collision is inelastic, because the projectile embeds itself into the disk. The linear momentum is not conserved through the collision because the axle about which the disk rotates must exert a force on the disk to prevent it from recoiling.

13 Simple harmonic motion

In this chapter, we look at oscillating systems that undergo "simple harmonic motion", such as the motion of a mass attached to a spring. Many systems in the physical world, such as an oscillating pendulum, can be described by the same mathematical formalism that describes the motion of a mass attached to a spring.

Learning Objectives

- Understand how to model the position, velocity, and acceleration of a mass attached to a spring.
- Understand the conditions under which a system undergoes simple harmonic motion.
- Understand how to model the motion of a pendulum when it undergoes simple harmonic motion.

Think About It

What do the motion of a mass attached to a spring, a cork bobbing in the water, and a pendulum have in common?

13.1 The motion of a spring-mass system

As an example of simple harmonic motion, we first consider the motion of a block of mass m that can slide without friction along a horizontal surface. The mass is attached to a spring with spring constant k which is attached to a wall on the other end. We introduce a one-dimensional coordinate system to describe the position of the mass, such that the x axis is co-linear with the motion, the origin is located where the spring is at rest, and the positive direction corresponds to the spring being extended. This "spring-mass system" is illustrated in Figure 13.1.

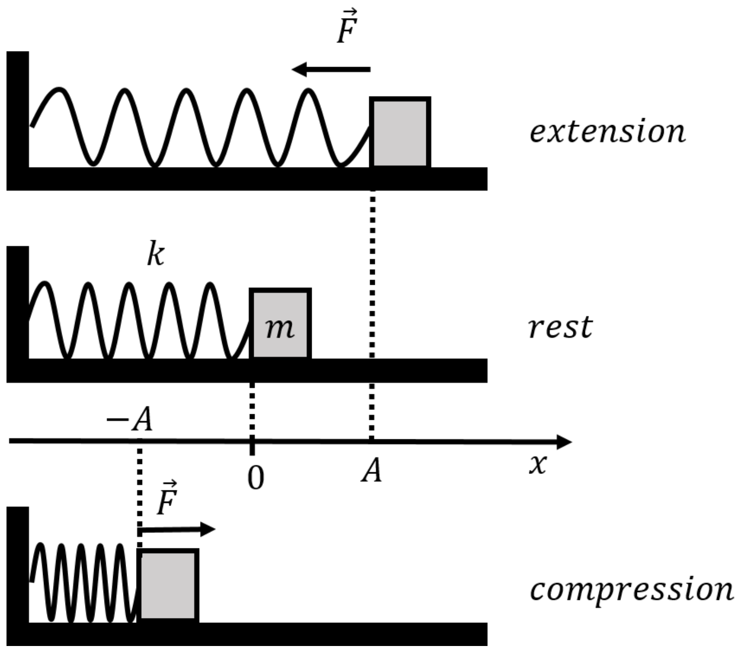


Figure 13.1: A horizontal spring-mass system oscillating about the origin with an amplitude A.

We assume that the force exerted by the spring on the mass is given by Hooke's Law:

$$\vec{F} = -kx\hat{x}$$

where x is the position of the mass. The only other forces exerted on the mass are its weight and the normal force from the horizontal surface, which are equal in magnitude and opposite in direction. Therefore, the net force on the mass is the force from the spring.

As we saw in Section 8.4, if the spring is compressed (or extended) by a distance A relative to the rest position, and the mass is then released, the mass will oscillate back and forth between $x = \pm A$[1], which is illustrated in Figure 13.1. We call A the "amplitude of the motion". When the mass is at $x = \pm A$, its speed is zero, as these points correspond to the location where the mass "turns around".

13.1.1 Description using energy

We can describe the motion of the mass using energy, since the mechanical energy of the mass is conserved. At any position, x, the mechanical energy, E, of the mass will have a term from the potential energy, U, associated with the spring force, and kinetic energy, K:

$$E = U + K = \frac{1}{2}kx^2 + \frac{1}{2}mv^2$$

We can find the mechanical energy, E, by evaluating the energy at one of the turning points. At these points, the kinetic energy of the mass is zero, so $E = U(x = A) = 1/2kA^2$. We

[1] As long as there is no friction to reduce the mechanical energy of the mass.

can then write the expression for mechanical energy as:

$$\boxed{\frac{1}{2}kx^2 + \frac{1}{2}mv^2 = \frac{1}{2}kA^2} \tag{13.1}$$

We can thus always know the speed, v, of the mass at any position, x, if we know the amplitude A:

$$v(x) = \sqrt{\frac{k(A^2 - x^2)}{m}}$$

Checkpoint 13-1

If you double the amplitude of the motion of a mass attached to a spring, its maximum speed will be:

A) double.
B) $\sqrt{2}$ times greater.
C) the same.
D) halved.

13.1.2 Kinematics of simple harmonic motion

We can use Newton's Second Law to obtain the position, $x(t)$, velocity, $v(t)$, and acceleration, $a(t)$, of the mass as a function of time. The x component of Newton's Second Law for the mass attached to the spring can be written:

$$\sum F_x = -kx = ma$$

We can write the acceleration in Newton's Second Law more explicitly as the second derivative of the position, $x(t)$, with respect to time. If we do this, we can see that Newton's Second Law for the mass attached to the spring is a differential equation for the function $x(t)$ (we call it an "equation of motion"):

$$\begin{aligned} ma &= -kx \\ m\frac{d^2x}{dt^2} &= -kx \\ \therefore \frac{d^2x}{dt^2} &= -\frac{k}{m}x \end{aligned} \tag{13.2}$$

We want to find the position function, $x(t)$. Equation 13.2 tells us that the second derivative of $x(t)$ with respect to time must equal the negative of the $x(t)$ function multiplied by a constant, k/m. Without having taken a course on differential equations, it might not be obvious what the function $x(t)$ could be. Several, equivalent functions can satisfy this equation. One possible choice, which we present here as a guess, is[2]:

$$\boxed{x(t) = A\cos(\omega t + \phi)} \tag{13.3}$$

[2]Other possible guesses that work are $A\sin(\omega t + \phi)$, and $x(t) = A\cos(\omega t) + B\sin(\omega t)$.

where A, ω, and ϕ are constants that we need to determine. We can take the second order derivative with respect to time of the function above to verify that it indeed "solves" the differential equation:

$$\begin{aligned}
x(t) &= A\cos(\omega t + \phi) \\
\frac{d}{dt}x(t) &= -A\omega\sin(\omega t + \phi) \\
\frac{d^2}{dt^2}x(t) &= \frac{d}{dt}\left(-A\omega\sin(\omega t + \phi)\right) = -A\omega^2\cos(\omega t + \phi) \\
\therefore \frac{d^2}{dt^2}x(t) &= -\omega^2 x(t)
\end{aligned}$$

The last equation has exactly the same form as Equation 13.2, which we obtained from Newton's Second Law, if we define ω as:

$$\boxed{\omega = \sqrt{\frac{k}{m}}} \tag{13.4}$$

We call ω the "angular frequency" of the spring-mass system. We have found that our guess for $x(t)$ satisfies the differential equation.

Checkpoint 13-2

What is the SI unit for angular frequency?

A) Hz
B) rad/s
C) $\mathrm{N}^{1/2}\mathrm{m}^{-1/2}\mathrm{kg}^{-1/2}$
D) All of the above

Olivia's Thoughts

In Chapter 3, we found, $x(t)$, from a function, $a(t)$, by using simple integration. You may be wondering why we can't do the same thing in order to find $x(t)$ for the mass-spring system. The difference is that, before, the acceleration was a function of time. Here, the acceleration is a function of x. This means that we have to use a different method to solve for $x(t)$, which is why we are making these "guesses" to solve a differential equation.

We still need to identify what the constants A and ϕ have to do with the motion of the mass. The constant A is the maximal value that $x(t)$ can take (when the cosine is equal to 1). This corresponds to the amplitude of the motion of the mass, which we already had labelled, A. The constant, ϕ, is called the "phase" and depends on when we choose $t = 0$

to be. Suppose that we define time $t = 0$ to be when the mass is at $x = A$; in that case:

$$\begin{aligned}
x(t=0) &= A \\
A\cos(\omega t + \phi) &= A \\
A\cos(\omega(0) + \phi) &= A \\
\cos(\phi) &= 1 \\
\therefore \phi &= 0
\end{aligned}$$

If we define $t = 0$ to be when the mass is at $x = A$, then the phase, ϕ, is zero. In general, the value of ϕ can take any value between $-\pi$ and $+\pi$[3] and, physically, corresponds to our choice of when $t = 0$ (i.e. the position of the mass when we choose $t = 0$).

Since we have determined the position as a function of time for the mass, its velocity and acceleration as a function of time are easily found by taking the corresponding time derivatives:

$$\begin{aligned}
x(t) &= A\cos(\omega t + \phi) \\
v(t) &= \frac{d}{dt}x(t) = -A\omega\sin(\omega t + \phi) \\
a(t) &= \frac{d}{dt}v(t) = -A\omega^2\cos(\omega t + \phi)
\end{aligned}$$

Checkpoint 13-3

What is the value of ϕ if we choose $t = 0$ to be when the mass is at $x = 0$ and moving in the positive x direction?

A) π
B) $-\pi$
C) $\pi/2$
D) $-\pi/2$

The position of the mass is described by a sinusoidal function of time; we call this type of motion "simple harmonic motion". The position and velocity as a function of time for a spring-mass system with $m = 1\,\text{kg}$, $k = 4\,\text{N/m}$, $A = 10\,\text{m}$ are shown in Figure 13.2 for two different choices of the phase, $\phi = 0$ and $\phi = \pi/2$.

[3]The argument to the cosine function is in radians, since the angular frequency is usually defined in radians per second. The value of ϕ is constrained to be within that range, since the cosine function is periodic with a period 2π.

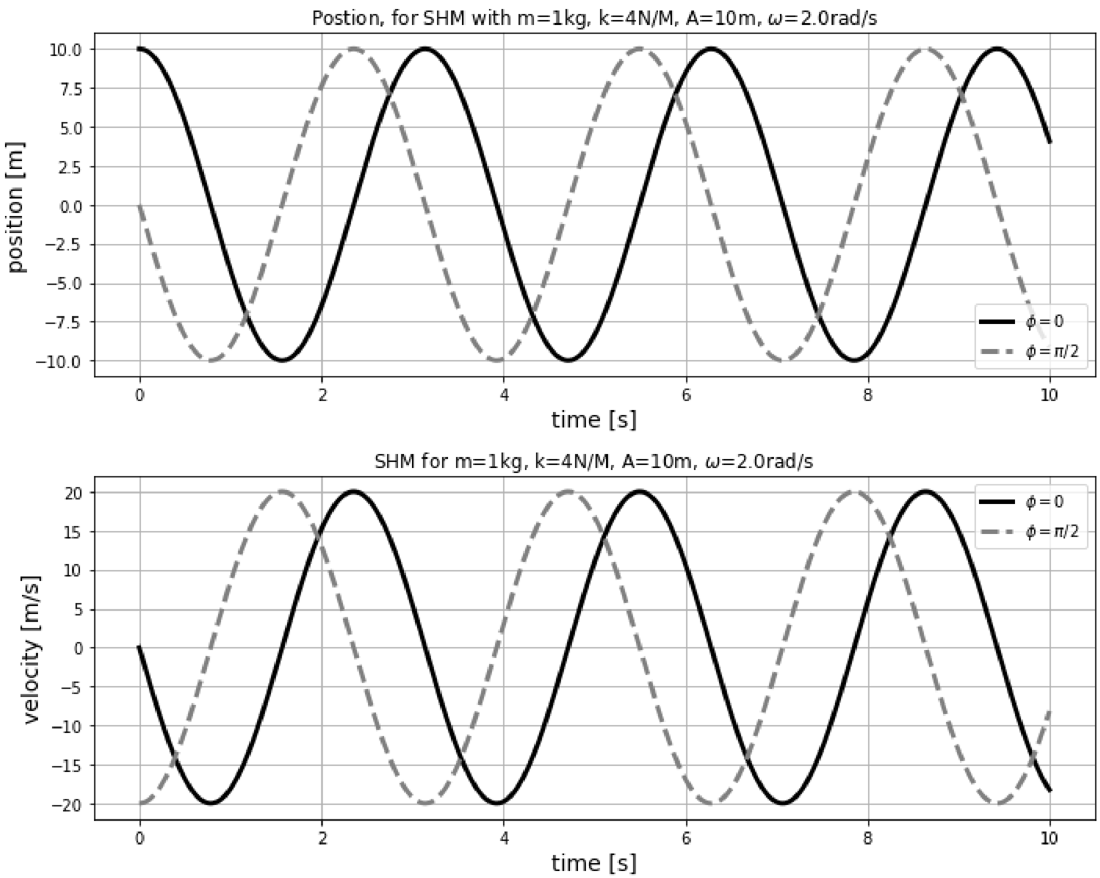


Figure 13.2: Position and velocity as a function of time for a mass-spring system for two different values of the phase, ϕ.

We can make a few observations about the position and velocity illustrated in Figure 13.2:

- Changing the phase, ϕ, results in an horizontal shift of the functions. A positive phase results in a shift of the functions to the left.
- The highest speed corresponds to a position of $x = 0$ and the largest position, $x = \pm A$, corresponds to a speed of zero.
- $\phi = 0$ corresponds to the "initial condition" at $t = 0$, where the position of the mass is $x = A$ and its speed is $v = 0$.
- $\phi = \pi/2$ corresponds to the "initial condition" at $t = 0$, where the position of the mass is $x = 0$ and its velocity is in the negative direction, and with maximal amplitude.
- The position is always between $x = \pm A$, and the velocity is always between $v = \pm A\omega$.

The motion of the spring is clearly periodic. If the period of the motion is T, then the position of the mass at time t will be the same as its position at $t + T$. The period of the motion, T, is easily found:

$$\boxed{T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}}} \tag{13.5}$$

And the corresponding frequency is given by:

$$\boxed{f = \frac{1}{T} = \frac{\omega}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{k}{m}}} \tag{13.6}$$

It should now be clear why ω is called the angular frequency, since it is related to the frequency of the motion.

Checkpoint 13-4

In order to double the oscillation period of a spring-mass system, you can

A) double the ratio of the mass over the spring constant.
B) quadruple the mass.
C) halve the spring constant.
D) All of the above.

13.1.3 Analogy with uniform circular motion

We can make an analogy between the mathematical description of the motion of a spring-mass system and that of uniform circular motion. Consider a particle that is moving along a circle of radius A, with constant angular speed ω, as illustrated in Figure 13.3.

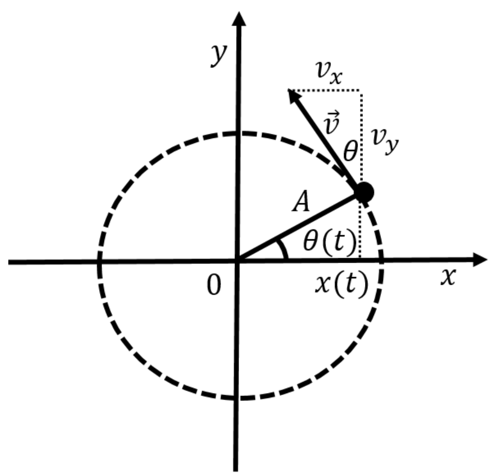


Figure 13.3: Uniform circular motion of a particle along a circle of radius A with constant angular speed ω.

The angular position, $\theta(t)$, of the particle is given by:

$$\theta(t) = \theta_0 + \omega t$$

if the particle was located at an angular position θ_0 at $t = 0$ ($\theta_0 = 0$ in Figure 13.3). The x coordinate of the particle is given by:

$$x(t) = A\cos(\theta(t)) = A\cos(\theta_0 + \omega t)$$

We can see that the x coordinate of the particle has the same functional form as the position for simple harmonic motion. The same is true for the particle's velocity. The magnitude of the particle's velocity is given by:

$$v = \omega r = \omega A$$

where $r = A$ is the radius of the circle. The x component of the particle's velocity is easily found from the figure and is given by:

$$v_x(t) = -v\sin(\theta(t)) = -\omega A\sin(\theta_0 + \omega t)$$

We can visualize simple harmonic motion as if it were the projection onto the x axis of uniform circular motion with angular speed ω about a circle with radius A. The phase ϕ corresponds to the angular position of the particle around the circle, θ_0, at time $t = 0$. When the particle crosses the y axis ($x = 0$), its velocity is in the x direction, so the x component of the velocity is maximal. When the particle crosses the x axis ($x = \pm A$), the x component of the velocity is zero.

Olivia's Thoughts

Here's a visualization of uniform circular motion projected onto the x axis:

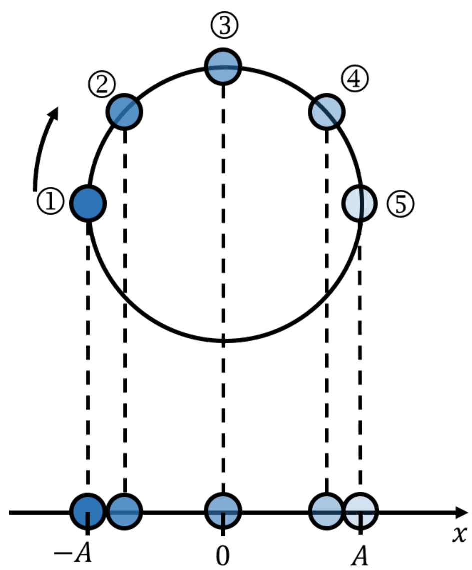


Figure 13.4: Projecting the motion of a ball around a circle onto the x axis.

Figure 13.4 shows a ball moving at a constant speed around a circle of radius A. In this diagram, I have taken snapshots of the ball's motion at regular time intervals as the ball moves from Position 1 to Position 5. Since the speed is constant, the balls are evenly spaced out around the circle. At the bottom of the figure, you can see what it would look like if we only considered the motion in the x direction (this is the projection of the motion onto the x axis). You could also think of this as what the motion would look like if you looked up at the circle from below. As you can see, this projection looks a lot like the motion of a mass on a spring. The motion of the ball is constrained between $-A$ and $+A$ (the turning points), and the velocity of the ball, in the x direction, will be highest when $x = 0$. There are tons of videos online that show animations of this concept, just look up "SHM as a projection of circular motion" and you will get lots of different ways to visualize this.

13.2 Vertical spring-mass system

Consider the vertical spring-mass system illustrated in Figure 13.5.

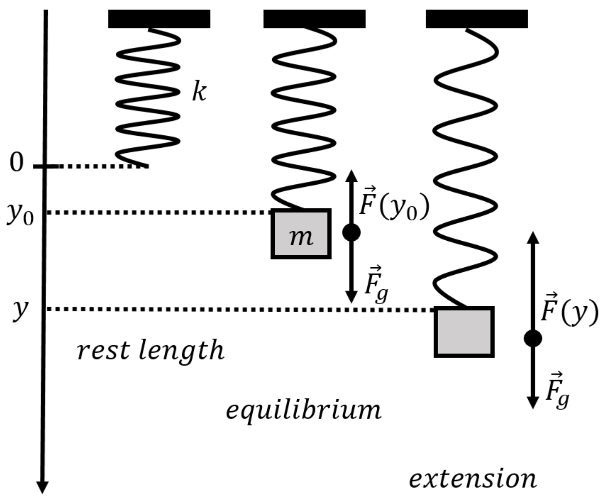


Figure 13.5: A vertical spring-mass system.

When no mass is attached to the spring, the spring is at rest (we assume that the spring has no mass). We choose the origin of a one-dimensional vertical coordinate system (y axis) to be located at the rest length of the spring (left panel of Figure 13.5). When a mass m is attached to the spring, the spring will extend and the end of the spring will move to a new equilibrium position, y_0, given by the condition that the net force on the mass m is zero. The only forces exerted on the mass are the force from the spring and its weight. The condition for the equilibrium is thus:

$$\begin{aligned}\sum F_y = F_g - F(y_0) &= 0\\ mg - ky_0 &= 0\\ \therefore mg &= ky_0\end{aligned}$$

Now, consider the forces on the mass at some position y when the spring is extended downwards relative to the equilibrium position (right panel of Figure 13.5). Newton's Second Law at that position can be written as:

$$\begin{aligned}\sum F_y = mg - ky &= ma\\ \therefore m\frac{d^2y}{dt^2} &= mg - ky\end{aligned}$$

Note that the net force on the mass will always be in the direction so as to "restore" the position of the mass back to the equilibrium position, y_0. If the mass had been moved upwards relative to y_0, the net force would be downwards.

We can substitute the equilibrium condition, $mg = ky_0$, into the equation that we obtained from Newton's Second Law:

$$\begin{aligned}
m\frac{d^2y}{dt^2} &= mg - ky \\
m\frac{d^2y}{dt^2} &= ky_0 - ky \\
m\frac{d^2y}{dt^2} &= -k(y - y_0) \\
\therefore \frac{d^2y}{dt^2} &= -\frac{k}{m}(y - y_0)
\end{aligned}$$

Consider a new variable, $y' = y - y_0$. This is the same as defining a new y' axis that is shifted downwards by y_0; in other words, this the same as defining a new y' axis whose origin is at y_0 (the equilibrium position) rather than at the position where the spring is at rest. Noting that the second time derivative of $y'(t)$ is the same as that for $y(t)$:

$$\frac{d^2y}{dt^2} = \frac{d^2}{dt^2}(y' + y_0) = \frac{d^2y'}{dt^2}$$

we can write the equation of motion for the mass, but using $y'(t)$ to describe its position:

$$\frac{d^2y'}{dt^2} = \frac{k}{m}y'$$

This is the same equation as that for the simple harmonic motion of a horizontal spring-mass system (Equation 13.2), but with the **origin located at the equilibrium position** instead of at the rest length of the spring. In other words, a vertical spring-mass system will undergo simple harmonic motion in the vertical direction about the equilibrium position. In general, a spring-mass system will undergo simple harmonic motion if a constant force that is co-linear with the spring force is exerted on the mass (in this case, gravity). That motion will be centred about a point of equilibrium where the net force on the mass is zero rather than where the spring is at its rest position.

Checkpoint 13-5

How does the period of motion of a vertical spring-mass system compare to the period of a horizontal system (assuming the mass and spring constant are the same)?

A) The period of the vertical system will be larger.
B) The period of the vertical system will be smaller.
C) The period will be the same.

13.2.1 Two-spring-mass system

Consider a horizontal spring-mass system composed of a single mass, m, attached to two different springs with spring constants k_1 and k_2, as shown in Figure 13.6.

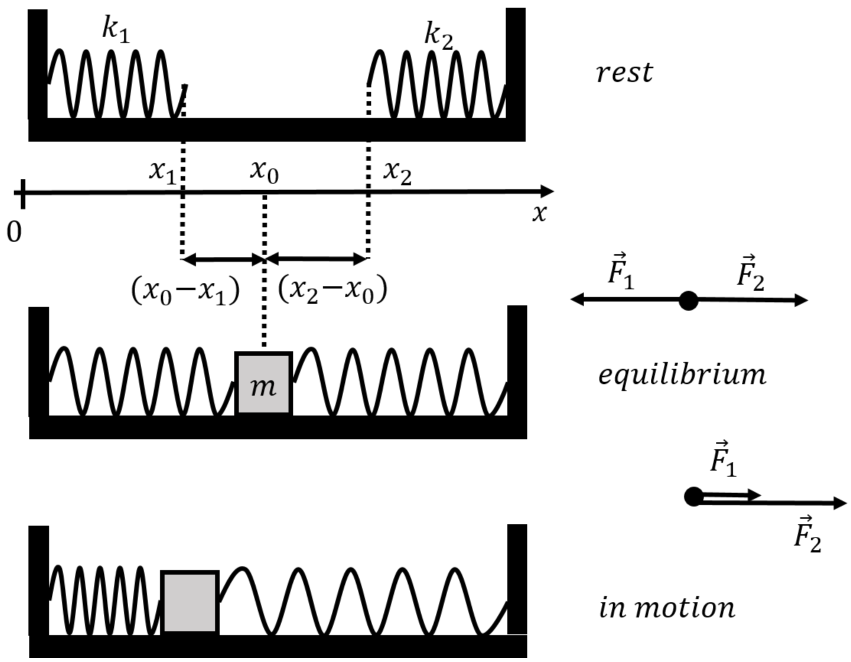


Figure 13.6: A mass attached to two different springs.

We introduce a horizontal coordinate system, such that the end of the spring with spring constant k_1 is at position x_1 when it is at rest, and the end of the k_2 spring is at x_2 when it is as rest, as shown in the top panel. A mass m is then attached to the two springs, and x_0 corresponds to the equilibrium position of the mass when the net force from the two springs is zero. We will assume that the length of the mass is negligible, so that the ends of both springs are also at position x_0 at equilibrium. You can see in the middle panel of Figure 13.6 that both springs are in extension when in the equilibrium position. It is possible to have an equilibrium where both springs are in compression, if both springs are long enough to extend past x_0 when they are at rest.

If we assume that both springs are in extension at equilibrium, as shown in the figure, then the condition for equilibrium is given by requiring that the sum of the forces on the mass is zero when the mass is located at x_0. The extension of the spring on the left is $x_0 - x_1$, and the extension of the spring on the right is $x_2 - x_0$:

$$\begin{aligned}
\sum F_x = -k_1(x_0 - x_1) + k_2(x_2 - x_0) &= 0 \\
-k_1x_0 + k_1x_1 + k_2x_2 - k_2x_0 &= 0 \\
-(k_1 + k_2)x_0 + k_1x_1 + k_2x_2 &= 0 \\
\therefore k_1x_1 + k_2x_2 &= (k_1 + k_2)x_0
\end{aligned}$$

Note that if the mass is displaced from x_0 in any direction, the net force on the mass will be in the direction of the equilibrium position, and will act to "restore" the position of the mass back to x_0.

When the mass is at some position x, as shown in the bottom panel (for the k_1 spring in compression and the k_2 spring in extension), Newton's Second Law for the mass is:

$$\begin{aligned}
-k_1(x-x_1)+k_2(x_2-x) &= ma \\
-k_1x+k_1x_1+k_2x_2-k_2x &= m\frac{d^2x}{dt^2} \\
-(k_1+k_2)x+k_1x_1+k_2x_2 &= m\frac{d^2x}{dt^2}
\end{aligned}$$

Note that, mathematically, this equation is of the form $-kx+C=ma$, which is the same form of the equation that we had for the vertical spring-mass system (with $C=mg$), so we expect that this will also lead to simple harmonic motion. We can use the equilibrium condition ($k_1x_1+k_2x_2=(k_1+k_2)x_0$) to re-write this equation:

$$\begin{aligned}
-(k_1+k_2)x+k_1x_1+k_2x_2 &= m\frac{d^2x}{dt^2} \\
-(k_1+k_2)x+(k_1+k_2)x_0 &= m\frac{d^2x}{dt^2} \\
\therefore -(k_1+k_2)(x-x_0) &= m\frac{d^2x}{dt^2}
\end{aligned}$$

Let us define $k=k_1+k_2$ as the "effective" spring constant from the two springs combined. We can also define a new coordinate, $x'=x-x_0$, which simply corresponds to a new x axis whose origin is located at the equilibrium position (in a way that is exactly analogous to what we did in the vertical spring-mass system). We can thus write Newton's Second Law as:

$$\begin{aligned}
-(k_1+k_2)(x-x_0) &= m\frac{d^2x}{dt^2} \\
-kx' &= m\frac{d^2x'}{dt^2} \\
\therefore \frac{d^2x'}{dt^2} &= -\frac{k}{m}x'
\end{aligned}$$

and we find that the motion of the mass attached to two springs is described by the same equation of motion for simple harmonic motion as that of a mass attached to a single spring. In this case, the mass will oscillate about the equilibrium position, x_0, with a an effective spring constant $k=k_1+k_2$. Combining the two springs in this way is thus equivalent to having a single spring, but with spring constant $k=k_1+k_2$. The angular frequency of the oscillations is given by:

$$\omega=\sqrt{\frac{k}{m}}=\sqrt{\frac{k_1+k_2}{m}}$$

13.3 Simple harmonic motion

In the previous sections, we modelled the motion of a mass attached to a spring and found that its position, $x(t)$, was described by the following differential equation:

$$\boxed{\frac{d^2x}{dt} = -\omega^2 x} \tag{13.7}$$

A possible solution to that equation was given by:

$$\boxed{x(t) = A\cos(\omega t + \phi)} \tag{13.8}$$

We then saw that the motion of a vertical spring-mass system, as well as that of a mass attached to two springs, could also be described by Equation 13.7. Any physical system that can described by Equation 13.7 is said to undergo "simple harmonic motion", or to be a "simple harmonic oscillator". If we find that the physical model of a system leads to Equation 13.7, then we immediately know that the position of system can be described by Equation 13.8.

The key physical characteristic of a simple harmonic oscillator is that there is a "restoring force" whose magnitude is proportional to the displacement from the equilibrium position. A restoring force is a force that acts to place the system back in equilibrium, and is thus always in the direction that is opposite of the displacement relative to an equilibrium position. In the three systems that we considered so far, the net force on the mass was always such that it would restore the mass back to the equilibrium position, where the net force on the mass is zero.

Many systems in nature are well modelled as simple harmonic oscillators. Some examples are: the motion of a pendulum as it oscillates, the motion of a buoy bobbing up and down in the sea, the motion of electrons in a shorted capacitor, and the vibrations of atoms in a molecule.

13.4 The motion of a pendulum

In this section, we show how and when the motion of a pendulum can be described as simple harmonic motion. Consider the simple pendulum that is constructed from a mass-less string of length, L, attached to a fixed point on one end and to a point mass m on the other, as illustrated in Figure 13.7.

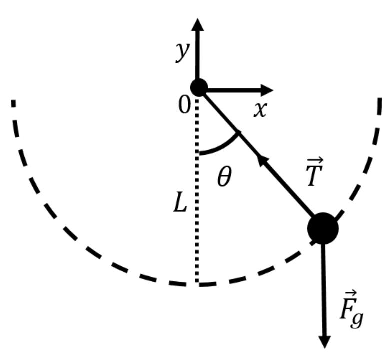


Figure 13.7: A simple pendulum which oscillates in a vertical plane.

The pendulum can swing in the vertical plane, and we have shown our choice of coordinate system (the z axis, not shown, is out of the page). The only two forces on the mass are the tension from the string and its weight. We can describe the position of the mass by the angle, $\theta(t)$, that the string makes with the vertical. We can model the dynamics of the simple pendulum by considering the net torque and angular acceleration about the axis of rotation that is perpendicular to the plane of the page and that goes through the point on the string that is fixed.

The force of tension cannot create a torque on the mass about the axis of rotation, as it is anti-parallel to the vector from the point of rotation to the mass. The net torque is thus the torque from the force of gravity:

$$\begin{aligned}\vec{\tau}^{net} &= \vec{\tau}_g \\ &= \vec{r} \times \vec{F}_g = (L\sin\theta\hat{x} - L\cos\theta\hat{y}) \times (-mg\hat{y}) \\ &= -mgL\sin\theta\hat{z}\end{aligned}$$

where L is the magnitude of the vector, $\vec{r}$, from the axis of rotation to where the force of gravity is exerted. The net torque is equal to the angular acceleration, α, multiplied by the moment of inertia, I, of the mass:

$$\begin{aligned}\vec{\tau}^{net} &= I\vec{\alpha} \\ -mgL\sin\theta\hat{z} &= mL^2\vec{\alpha} \\ -g\sin\theta\hat{z} &= L\vec{\alpha}\end{aligned}$$

where $I = ML^2$ is the moment of inertia for a point mass a distance L away from the axis of rotation. For the position illustrated in Figure 13.7, the angular acceleration of the pendulum is in the negative z direction (into the page) and corresponds to a clockwise motion for the pendulum, as we would expect. The angular acceleration is the second time derivative of the angle, θ:

$$\alpha = \frac{d^2\theta}{dt^2}$$

We can thus re-write the equation that we obtained from the rotational dynamics version of Newton's Second Law as:

$$\begin{aligned}-g\sin\theta\hat{z} &= L\vec{\alpha} \\ \frac{d^2\theta}{dt^2} &= -\frac{g}{L}\sin\theta\end{aligned}$$

where we only used the magnitudes in the second equation, since all of the angular quantities are in the z direction. This equation of motion for $\theta(t)$ almost looks like the equation for simple harmonic oscillation for the angle θ (except that we have $\sin\theta$ instead of θ). However, consider the "the small angle approximation"[4] for the sine function:

$$\sin\theta \approx \theta$$

If the oscillations of the pendulum are "small", such that the small angle approximation is valid, then the equation of motion for the pendulum is:

$$\begin{aligned}\frac{d^2\theta}{dt^2} &= -\frac{g}{L}\sin\theta \approx -\frac{g}{L}\theta \\ \therefore \frac{d^2\theta}{dt^2} &= -\frac{g}{L}\theta \quad \text{(for small } \theta\text{)}\end{aligned}$$

and the angle that the pendulum makes with the vertical is described by the equation for simple harmonic oscillation with angular frequency:

$$\omega = \sqrt{\frac{g}{L}}$$

The angle, θ, as a function of time is thus described by the function:

$$\theta(t) = \theta_{max}\cos(\omega t + \phi)$$

where θ_{max} is the maximal amplitude of the oscillations and ϕ is a phase that depends on when we choose to define $t = 0$.

Checkpoint 13-6

Kaiden built a grandfather clock using a simple pendulum, but he found that the period was twice as large as as he wanted it to be. In order to halve the period of the pendulum, he can

A) change the mass.
B) halve the length of the string.
C) quarter the length of the string.
D) double the length of the string.
E) quadruple the length of the string.

13.4.1 The physical pendulum

A physical pendulum is defined as any object that is allowed to rotate in the vertical plane about some axis that goes through the object, as illustrated in Figure 13.8.

[4]Look up the Maclaurin/Taylor series for the sine function!

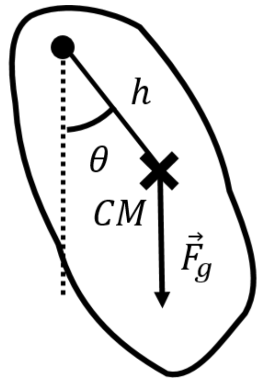


Figure 13.8: A physical pendulum which oscillates in a vertical plane about an axis through the object.

The only forces exerted on the pendulum are its weight (exerted at its centre of mass) and a contact force exerted at the axis of rotation. The physical pendulum can be modelled in exactly the same way as the simple pendulum, except that we use the moment of inertia of the object about the axis of rotation. Only the weight results in a torque about the rotation axis, since the contact force is exerted at the rotation axis:

$$\tau^{net} = \tau_g = I\alpha$$
$$-mgh\sin\theta = I\alpha = I\frac{d^2\theta}{dt^2}$$

where h is the distance from the axis of rotation to the centre of mass. In the small angle approximation, this becomes:

$$\frac{d^2\theta}{dt^2} = -\frac{mgh}{I}\theta \quad \text{(for small } \theta)$$

and we find that the physical pendulum oscillates with an angular frequency:

$$\omega = \sqrt{\frac{mgh}{I}}$$

13.5 Summary

Key Takeaways

The equation of motion for the position, $x(t)$, of the mass in a one-dimensional spring-mass system with no friction can be written:

$$\frac{d^2x}{dt^2} = -\sqrt{\frac{k}{m}}x = -\omega^2 x$$

and has a solution:

$$x(t) = A\cos(\omega t + \phi)$$

where A is the amplitude of the motion, ϕ is the phase, which depends on our choice of initial conditions (when we choose time $t = 0$), and ω:

$$\omega = \sqrt{\frac{k}{m}}$$

is the angular frequency of the motion. The mass will oscillate about an equilibrium position with a period, T, and frequency, f, given by:

$$\begin{aligned} T &= \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}} \\ f &= \frac{1}{T} = \frac{\omega}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \end{aligned}$$

The velocity and acceleration of the mass are found by taking the time derivatives of the position $x(t)$:

$$\begin{aligned} x(t) &= A\cos(\omega t + \phi) \\ v(t) &= \frac{d}{dt}x(t) = -A\omega\sin(\omega t + \phi) \\ a(t) &= \frac{d^2}{dt^2}x(t) = \frac{d}{dt}\left(-A\omega\sin(\omega t + \phi)\right) = -A\omega^2\cos(\omega t + \phi) \end{aligned}$$

The total mechanical energy of the mass, at some position x, is given by:

$$E = U + K = \frac{1}{2}kx^2 + \frac{1}{2}mv^2 = \frac{1}{2}kA^2$$

and is conserved.

Any system that can be described by the equation of motion:

$$\frac{d^2x}{dt^2} = -\omega^2 x$$

is said to be a simple harmonic oscillator, and its position will be described by:

$$x(t) = A\cos(\omega t + \phi)$$

A simple harmonic oscillator will always oscillate about an equilibrium position, where the net force on the oscillator is zero. The net force on a simple harmonic oscillator is always directed towards the equilibrium position, and has a magnitude proportional to the distance of the oscillator from its equilibrium position. The force is called a restoring force. A vertical spring-mass system, and a mass attached to two springs will both undergo simple harmonic motion about their respective equilibrium position.

A simple pendulum will undergo simple harmonic oscillations, if the amplitude of the oscillations is small. The angular frequency for the oscillations of a simple pendulum only depends on the length of the pendulum:

$$\omega = \sqrt{\frac{g}{L}}$$

This is valid in the small angle approximation, where:

$$\sin\theta \approx \theta$$

A physical pendulum of mass m which oscillates about an axis through the object will also undergo simple harmonic oscillation in the small angle approximation. The angular frequency of the oscillations for a physical pendulum is given by:

$$\omega = \sqrt{\frac{mgh}{I}}$$

where h is the distance between the centre of mass and the axis of rotation, and I is the moment of inertia of the object about the rotation axis.

Important Equations

Position, velocity, and acceleration for SHM:

$$x(t) = A\cos(\omega t + \phi)$$
$$v(t) = \frac{d}{dt}x(t) = -A\omega\sin(\omega t + \phi)$$
$$a(t) = \frac{d^2}{dt^2}x(t) = -A\omega^2\cos(\omega t + \phi)$$

Period and frequency:

$$\omega = \sqrt{\frac{k}{m}}$$
$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}}$$
$$f = \frac{1}{T} = \frac{\omega}{2\pi} = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$$

Mechanical energy:

$$E = U + K = \frac{1}{2}kx^2 + \frac{1}{2}mv^2 = \frac{1}{2}kA^2$$

Simple pendulum (small angles):

$$\omega = \sqrt{\frac{g}{L}}$$

Physical pendulum (small angles):

$$\omega = \sqrt{\frac{mgh}{I}}$$

13.6 Thinking about the material

Reflect and research

1. What is an example of a system that is a simple harmonic oscillator (not covered in this this chapter)? What is the restoring force for that system?
2. What happens to the motion of a mass-spring system in the presence of friction? Sketch out the position as a function of time.
3. What is a "damped" harmonic oscillator?
4. What is a coupled oscillator? Find a video of a coupled oscillator online and describe the motion.
5. How do the shock absorbers on a car relate to simple harmonic motion?

To try at home

1. Compare values of θ and $\sin\theta$ to see when the small angle approximation holds. Does it matter if θ is expressed in radians?
2. Build a simple pendulum and describe the motion. Is it simple harmonic motion? Is it damped simple harmonic motion? Does the frequency depend on the length of the pendulum as expected?

To try in the lab

1. Theory lab: what is the function $x(t)$ if there is a frictional force, proportional to velocity, $-bv$, exerted on the spring mass system?
2. Propose an experiment to test whether the period of the motion of pendulum depends on the amplitude of the motion.
3. Propose an experiment to test whether a physical pendulum is well-described by simple harmonic motion.

13.7 Sample problems and solutions

13.7.1 Problems

Problem 13-1: Ty ($m = 30\,\text{kg}$) is trying out a new piece of equipment at his local playground. The equipment consists of a platform that is connected to two springs. The top spring ($k_1 = 2400\,\text{N/m}$) connects the platform to the playground structure and the bottom spring ($k_2 = 3480\,\text{N/m}$) (Figure 13.9) connects it to the ground. When no one is standing on the platform the platform is 50 cm off the ground. When Ty is standing on the platform, he oscillates up and down, and the lowest point that the platform reaches is 35 cm off the ground. Show that this is simple harmonic motion and determine what Ty's maximum speed will be. (Solution)

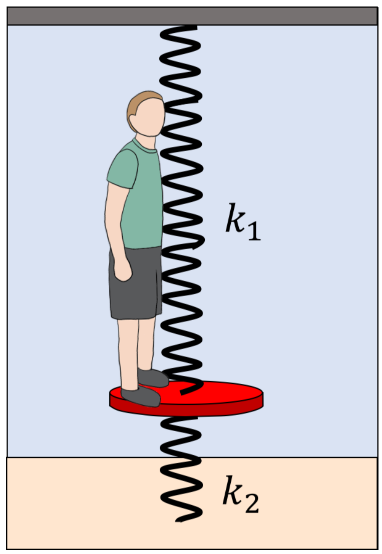


Figure 13.9: Playground equipment made of a platform connected to two vertical springs.

Problem 13-2: A torsional pendulum consists of a horizontal rod suspended from a vertical wire. When the rod is rotated so that it is displaced an angle θ from equilibrium, the wire (which is now twisted) provides a restoring torque about the axis of the wire given by:

$$\tau = -\kappa\theta$$

where κ is the torsion coefficient, which depends on the stiffness of the wire. You may notice that this formula closely resembles Hooke's law.

a) You construct a torsional pendulum by attaching two small spherical masses (you can assume they are point masses, each of mass m) to the ends of a thin (mass-less) rod of length L and attaching a wire to the centre of the rod (Figure 13.10). When you displace one of the masses by an angle θ and release it, you find that it oscillates with a period T. Find an expression for the torsion coefficient, κ, in term of T, m, and L.

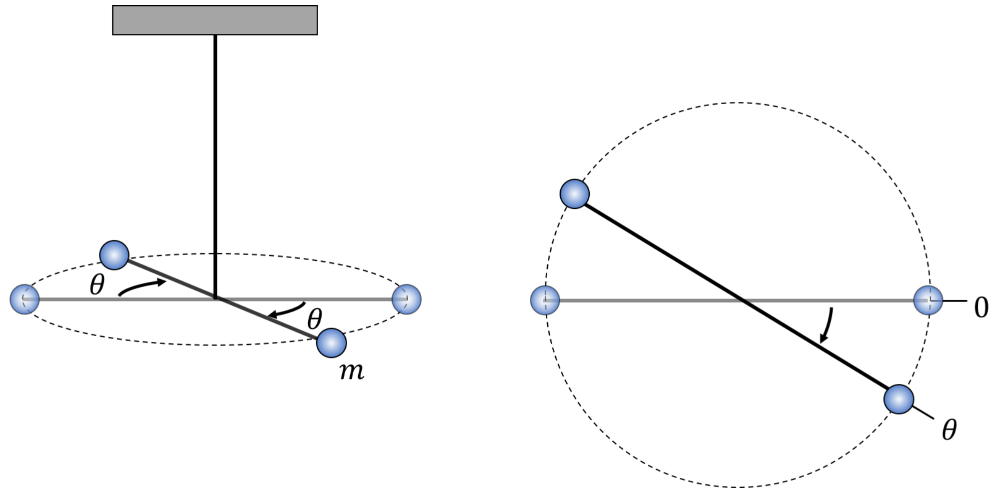


Figure 13.10: A torsional pendulum. The right side shows a top view.

b) You place two very large spheres, each of mass M, near each of the small spheres (as shown in Figure 13.11). Each of the small spheres will be acted on by a force of gravity from the **nearest** large sphere. The pendulum is at equilibrium when it is deflected an angle β from its original equilibrium position. At the new equilibrium, the displacement vectors connecting the centres of large and small spheres have a magnitude d and are essentially perpendicular to the rod. Find an expression for the universal gravitational constant G, in terms of the masses, the length of the rod, and the period measured in part a).
Fun fact! This set-up resembles an experiment performed by Henry Cavendish that was first used to determine the value for G and to test Newton's Universal Theory of Gravity.

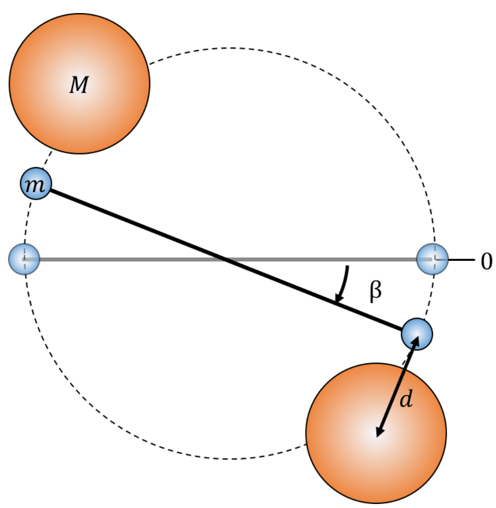


Figure 13.11: Two very large spheres are placed near each of the small masses on the torsional pendulum (top view). At the new equilibrium, each small mass is a distance d from the nearest large mass.

(Solution)

13.7.2 Solutions

Solution to problem 13-1: First, we need to solve for the new equilibrium position of the platform, x_0, when Ty is standing on the platform. We define the x axis so that the origin is 50 cm above the ground (the equilibrium position when no one is standing on the platform) and choose the positive direction to be downwards (Figure 13.12).

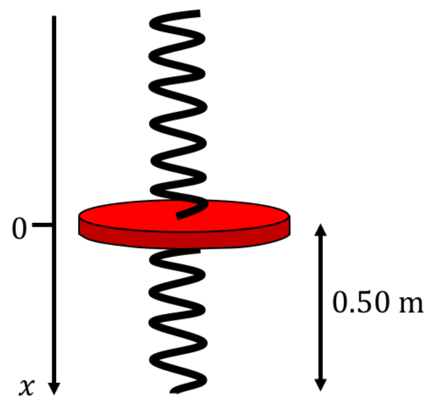


Figure 13.12: The platform when no one is standing on it.

Even though we do not know the mass of the platform, or the actual resting lengths of the spring, we do not need to know these, since we can model the platform with nobody on it as a single spring with spring constant $k = k_1 + k_2$ and rest position $x = 0$.

When Ty is standing on the platform, the sum of the forces is given by his weight and the force from the "effective spring":

$$\sum F = mg - (k_1 + k_2)x$$

where we noted that, when the platform moves down, both the top and bottom spring will exert a force upwards (Figure 13.13).

At equilibrium, the sum of the forces is equal to zero. We can use this to solve for the displacement at x_0:

$$0 = mg - (k_1 + k_2)x_0$$
$$\therefore x_0 = \frac{mg}{k_1 + k_2} = \frac{(30\,\text{kg})(9.8\,\text{m/s}^2)}{(2400\,\text{Nm}) + (3480\,\text{Nm})} = 0.05\,\text{m}$$

We will confirm that this is a simple harmonic oscillator by showing that the system's motion can be described by the equation:

$$\frac{d^2x}{dt^2} = -\omega^2 x$$

For some position x below equilibrium, we can rewrite Newton's second law as:

$$ma = mg - (k_1 + k_2)x$$
$$m\frac{d^2x}{dt^2} = mg - (k_1 + k_2)x$$

In order to show that this is simple harmonic motion, we need to combine the right hand side of the equation into one term. We found earlier that $mg = (k_1 + k_2)x_0$, which we can use here:

$$\begin{aligned}
m\frac{d^2x}{dt^2} &= (k_1 + k_2)x_0 - (k_1 + k_2)x \\
\frac{d^2x}{dt^2} &= \frac{(k_1 + k_2)}{m}(x_0 - x) \\
\frac{d^2x}{dt^2} &= -\frac{(k_1 + k_2)}{m}(x - x_0)
\end{aligned}$$

We now define an x' axis such that $x' = x - x_0$. This means that the origin of the x' axis is at the new equilibrium position:

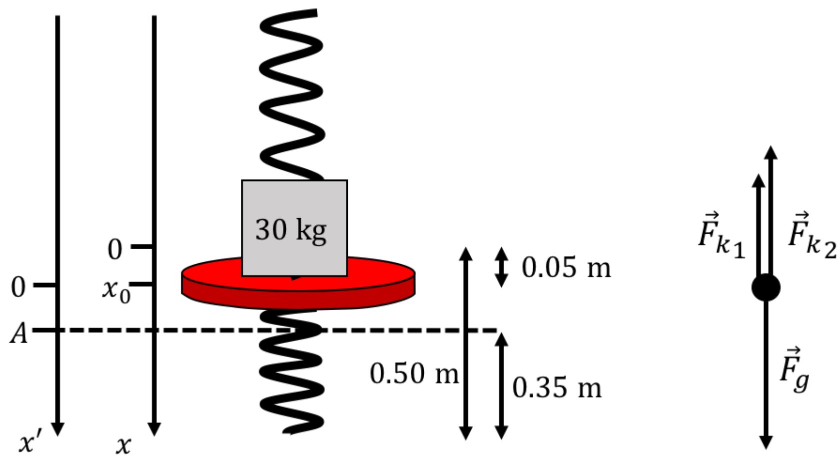


Figure 13.13: The forces acting on the platform and our new coordinate system.

We can now rewrite our expression using the x' axis:

$$\frac{d^2x}{dt^2} = -\frac{(k_1 + k_2)}{m}x'$$

This equation tells us that this is simple harmonic motion about the new equilibrium position, where $\omega = \sqrt{(k_1 + k_2)/m}$. We know that the lowest point that the platform reaches is 35 cm above the ground, which, on our x' axis, corresponds to $x' = 10\,\text{cm}$ (Figure 13.13). Thus, the amplitude of the oscillation is $A = 0.1\,\text{m}$. Because this is simple harmonic motion, we know that the position of the platform can be described by the following function:

$$x'(t) = A\cos(\omega t + \phi)$$

We set $t = 0$ to be when the platform is at its lowest point ($x' = A$). The value of ϕ is thus:

$$\begin{aligned}
x'(0) &= A\cos(\omega(0) + \phi) \\
A &= A\cos(\phi) \\
1 &= \cos(\phi) \\
\therefore \phi &= 0
\end{aligned}$$

The velocity is given by:

$$\begin{aligned} v(t) = \frac{d}{dt}x(t) &= -A\omega\sin(\omega t + \phi) \\ &= -A\omega\sin(\omega t) \end{aligned}$$

The speed will be maximized when $\sin(\omega t) = 1 \quad \text{or} -1$. So, the maximum speed will be:

$$\begin{aligned} |v| &= A\omega \\ |v| &= A\sqrt{\frac{(k_1 + k_2)}{m}} \\ |v| &= (0.1\,\text{m})\sqrt{\frac{(2400\,\text{Nm} + 3480\,\text{Nm})}{30\,\text{kg}}} \\ |v| &= 1.4\,\text{m/s} \end{aligned}$$

Solution to problem 13-2:

(a) The only force that creates a torque on the masses is the restoring force from the twisting of the wire. The rotational dynamics version of Newton's Second Law relates this torque to the angular acceleration, α of the rod:

$$I\alpha = -\kappa\theta$$

where I is the moment of inertia of the rod. Rewriting α more explicitly as the second time derivative of the angle, we get:

$$\begin{aligned} I\frac{d^2\theta}{dt^2} &= -\kappa\theta \\ \frac{d^2\theta}{dt^2} &= -\frac{\kappa}{I}\theta \end{aligned}$$

By inspection, we can see that the torsional pendulum is a simple harmonic oscillator, where $\omega = \sqrt{\kappa/I}$. The period of the motion is therefore:

$$\begin{aligned} T &= \frac{2\pi}{\omega} \\ T &= 2\pi\sqrt{\frac{I}{\kappa}} \end{aligned}$$

We can rearrange this expression to get κ:

$$\begin{aligned} T^2 &= \frac{4\pi^2 I}{\kappa} \\ \kappa &= \frac{4\pi^2 I}{T^2} \end{aligned}$$

The moment of inertia for one of the masses is $m(L/2)^2$, where $L/2$ is the distance from the mass to the axis of rotation. The moment of inertia for the two masses attached to the mass-less rod is:

$$I = 2m\left(\frac{L}{2}\right)^2 = \frac{mL^2}{2}$$

Putting this into our expression for κ:

$$\kappa = \frac{2\pi^2 mL^2}{T^2}$$

(b) The two forces that provide torques for the small spheres are gravity and the force exerted by the twisting wire. Each of the small spheres will experience a force due to gravity from the nearest large sphere. At equilibrium, the force due to gravity on one of the small spheres is therefore:

$$F_g = \frac{GMm}{d^2}$$

Assuming that, at equilibrium, the force vector is perpendicular to the rod, the torque from one of the large spheres is just the force multiplied by the distance to the axis of rotation. Since there are two large spheres, each of which creates a torque on the pendulum, the total torque due to gravity is:

$$\begin{aligned}\tau_g &= 2F_g\frac{L}{2}\\ &= F_gL\\ &= \frac{GMm}{d^2}L\end{aligned}$$

(Note that τg is the torque due to gravity **at equilibrium only**). We can use Newton's second law for the pendulum to find an expression for G. At equilibrium, the net torque is equal to zero, and the angle of deflection is β:

$$\begin{aligned}\tau_{net} &= \tau_{wire} - \tau_g\\ 0 &= \tau_{wire} - \tau_g\\ \tau_g &= \tau_{wire}\\ \frac{GMm}{d^2}L &= \kappa\beta\\ \therefore G &= \frac{\kappa\beta d^2}{LMm}\end{aligned}$$

Using our expression for κ found in part a), this becomes:

$$G = \frac{2\pi^2 L\beta d^2}{MT^2}$$

14 Waves

In this chapter we introduce the tools to describe waves. Waves arise in many different physical systems (the ocean, a string, electromagnetism, etc.), and can be described by a common mathematical framework.

Learning Objectives

- Understand the definition of different types of waves.
- Understand how to mathematically describe travelling and standing waves.
- Understand how to model the propagation of a pulse on a rope.
- Understand how to model the energy transported by a wave.
- Understand how to model the interference of waves.
- Understand how standing waves form and how to model them.

Think About It

Two waves travel down two identical strings (Figure 14.1). The frequency of the first wave is twice that of the second wave. Which wave will be faster?

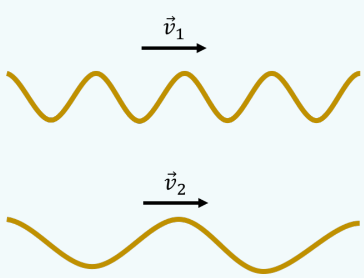


Figure 14.1: Two waves travelling down two identical strings.

A) The first wave.
B) The second wave.
C) The speeds will be the same.

14.1 Characteristics of a wave

14.1.1 Definition and types of waves

A travelling wave is a **disturbance that travels through** a medium. Consider the waves made by fans at a soccer game, as in Figure 14.2. The fans can be thought of as the medium through which the wave propagates. The elements of the medium may oscillate about an equilibrium position (the fans move a short distance up and down), but they do not travel with the wave (the fans do not move horizontally with the wave).

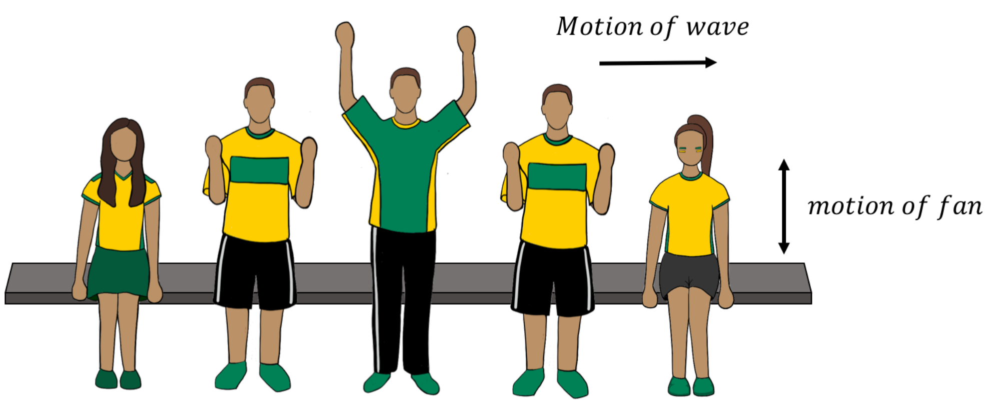


Figure 14.2: A transverse wave made by soccer fans moving up and down.

Consider the ripples (waves) made by a rock dropped in a pond (Figure 14.3). The ripples travel outwards from where the rock was dropped, but the water itself does not move outwards. The individual water molecules will move in small circles about an equilibrium position, but they do not move along with the waves.

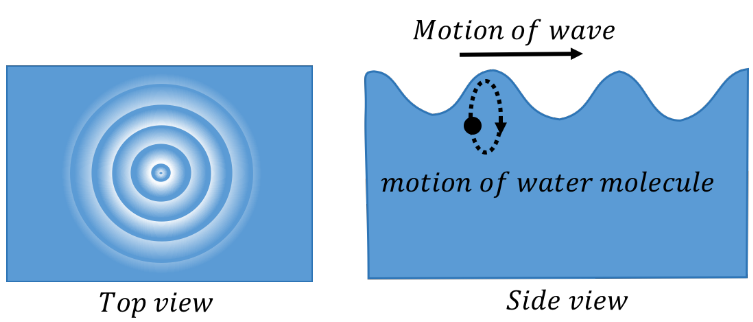


Figure 14.3: A transverse wave travelling through water. The left panel shows the view from above as ripples move outwards. The right panel shows the motion of an individual water molecule as the wave is viewed from the side.

We can distinguish between two classes of waves, based on the motion of the medium through which it propagates. With **transverse waves**, the elements of the medium oscillate back and forth in a direction perpendicular to the motion of the wave. For example, if you attach a horizontal rope to a wall and move the other end up and down (Figure 14.4), you can create a disturbance (a wave) that travels horizontally along the rope. The parts of the rope do not move horizontally; they only move up and down, about some equilibrium position.

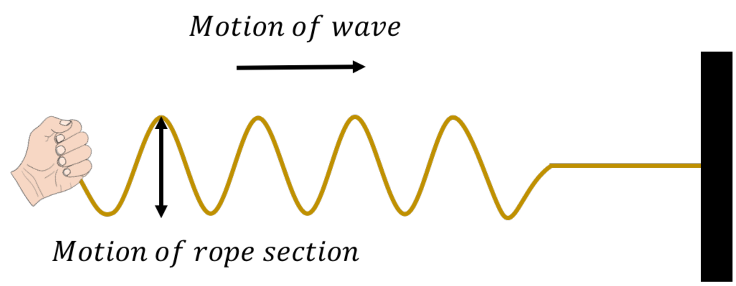


Figure 14.4: A transverse wave travelling through a rope. The wave is created by moving one end of the rope up and down.

With **longitudinal waves**, the elements of the medium oscillate back and forth in the same direction as the motion of the wave. If you clap your hands, you will create a pressure disturbance in the air that will propagate; this is what we call sound (a sound wave). The air molecules oscillate about an equilibrium position in the same direction as the wave propagates, but they do not move with the wave.

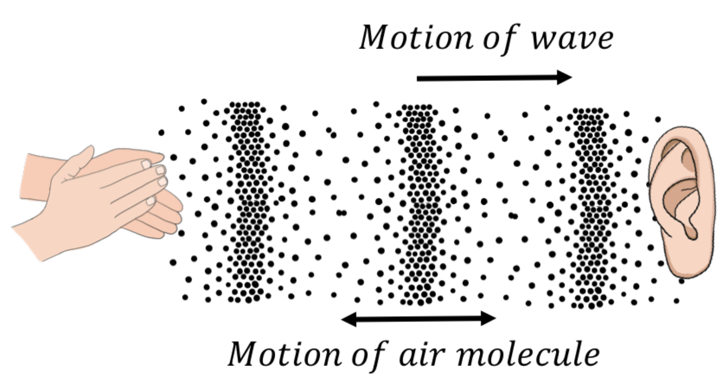


Figure 14.5: A longitudinal sound wave travelling through the air. The air molecules move back and forth in the same direction as the wave, but they oscillate about an equilibrium position instead of moving with the wave.

Furthermore, we can distinguish between "travelling waves", in which a disturbance propagates through a medium, and "standing waves", which do not transport energy through the medium (for example, a vibrating string on a violin).

Checkpoint 14-1

Are the waves propagating through a slinky when you compress and elongate it (Figure 14.6) transverse or longitudinal?

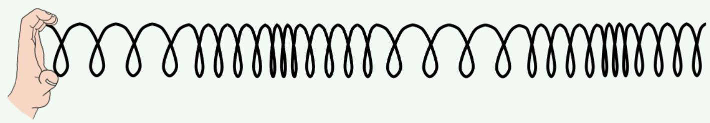


Figure 14.6: A wave travelling through a slinky. The wave is created when you compress or elongate the slinky

A) Transverse
B) Longitudinal

Physically, a wave can only propagate through a medium if the medium can be deformed. When a particle in the medium is disturbed from its equilibrium position, it will experience a restoring force that acts to bring it back to its equilibrium position. Often, if the displacement of the particle from the equilibrium is small, the magnitude of that force is proportional to the displacement. Thus, as we will see, we can model the propagation of waves by treating the particles in the medium as simple harmonic oscillators.

A source of energy is required in order to deform the medium and generate a wave. For example, that source of energy could be a speaker creating sound waves by pushing a membrane back and forth; speakers require energy, and are often rated by the electrical power that they convert into sound waves (e.g. a 50 W speaker consumes 50 W of electrical power to produce sound).

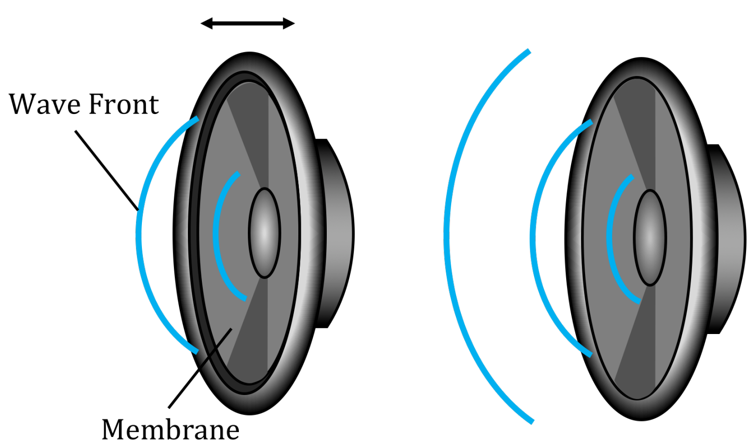


Figure 14.7: A speaker creating sound sound waves. The membrane vibrates back and forth which deforms the air to create sound waves that propagate through the air.

14.1.2 Description of a wave

In this chapter, we will mostly discuss how to describe sinusoidal waves; those for which the displacement of particles in the medium can be described by a sinusoidally-varying function of position. As we will see, more complicated waves can always be described as if they are the combination of multiple sine waves. We can use several quantities to describe a travelling wave, which are illustrated in Figure 14.8:

- The **wavelength**, λ, is the distance between two successive maxima ("peaks") or minima ("troughs") in the wave.
- The **amplitude**, A, is the maximal distance that a particle in the medium is displaced from its equilibrium position.
- The **velocity**, $\vec{v}$, is the velocity with which the disturbance propagates through the medium.
- The **period**, T, is the time it takes for two successive maxima (or minima) to pass through the same point in the medium.
- The **frequency**, f, is the inverse of the period ($f = 1/T$).

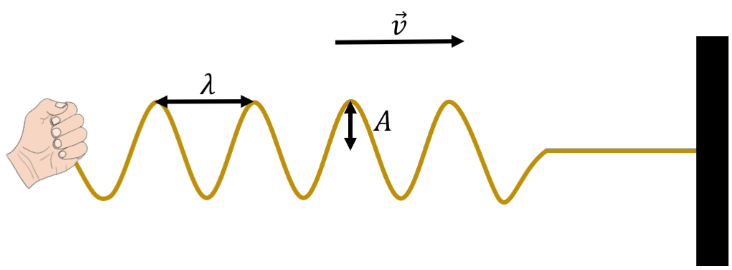


Figure 14.8: Wavelength, velocity, and amplitude for a transverse wave on a rope.

The wavelength, speed, and period of the wave are related, since the amount of time that it takes for two successive maxima of the wave to pass through a given point will depend on the speed of the wave and the distance between maxima, λ. Since it takes a time, T, for two maxima a distance λ apart to pass through a given point in the medium, the speed of the wave is given by:

$$\boxed{v = \frac{\lambda}{T} = \lambda f} \tag{14.1}$$

Thus, of the three quantities (speed, period/frequency, and wavelength), only two are independent, as the third quantity must depend on the value of the other two. **The speed of a wave depends on the properties of the medium through which the wave propagates and not on the mechanism that is generating the wave**. For example, the speed of sound waves depends on the pressure, density, and temperature of the air through which they propagate, and not on what is making the sound. When a mechanism generates a wave, that mechanism usually determines the frequency of the wave (e.g. frequency with which the hand in Figure 14.8 moves up and down), the speed is determined by the medium, and the wavelength can be determined from Equation 14.1.

Checkpoint 14-2

What can you say about the sound emitted by a cello versus that emitted by a violin?

A) The sound from the violin has a higher frequency.
B) The sound from the cello has a longer wavelength.
C) The sound from both instruments propagates at the same speed.
D) All of the above.

14.2 Mathematical description of a wave

In order to describe the motion of a wave through a medium, we can describe the motion of the individual particles of the medium as the wave passes through. Specifically, we describe the position of each particle using its displacement, D, from its equilibrium position. Consider our rope example in which a sine wave is propagating through a medium (the rope) in the positive x direction, as shown in Figure 14.9

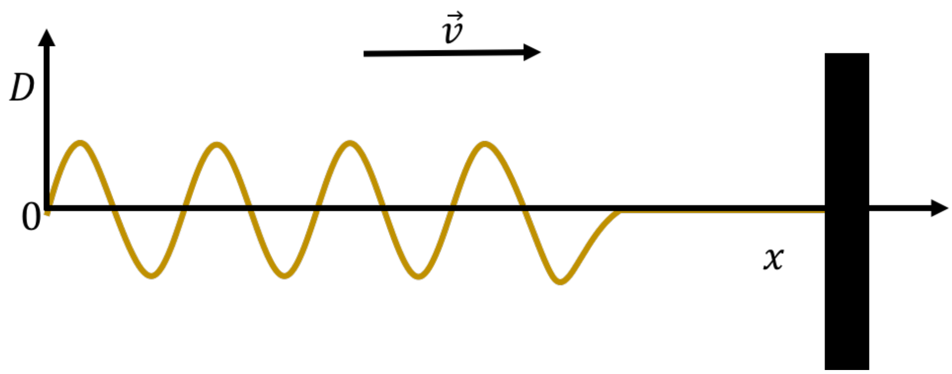


Figure 14.9: The displacement (D) of points at different positions (x) on a rope as a sine wave passes through.

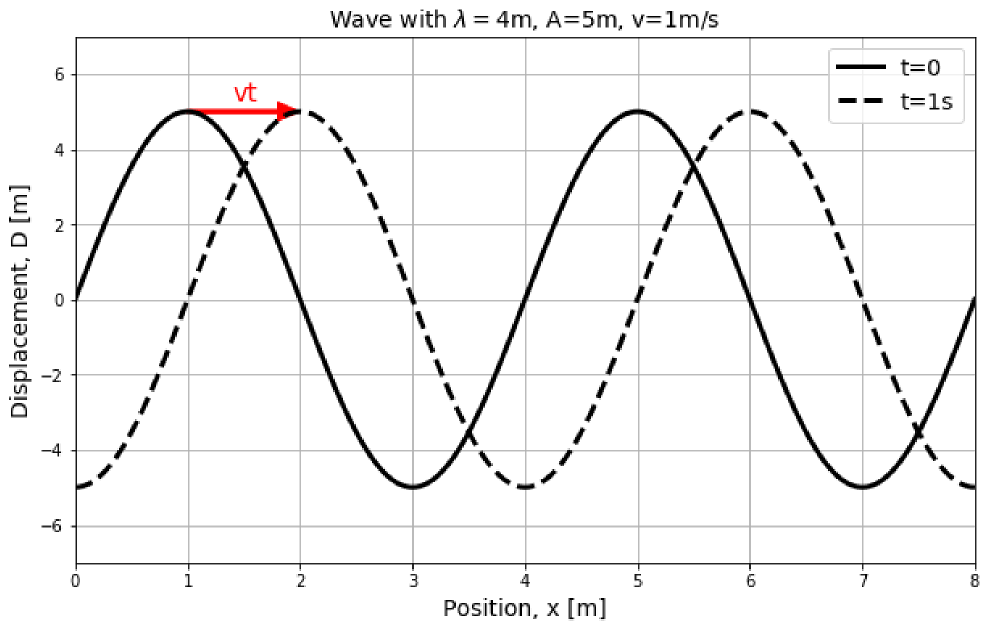


Figure 14.10: Displacement as a function of position for particles in a medium as a wave passes through. The dotted line shows the dipsplacement as a function of time 1 s *after the solid line, and corresponds to a wave travelling towards the right.*

The displacement, D, of each point at position, x, in the medium is shown on the vertical axis of Figure 14.10. The solid black line corresponds to a snapshot of the wave at time $t = 0$. The wave has an amplitude, $A = 5\,\text{m}$, a velocity, $v = 1\,\text{m/s}$, and a wavelength, $\lambda = 4\,\text{m}$. The dotted line corresponds to a snapshot of the wave one second later, at $t = 1\,\text{s}$, when the wave has moved to the right by a distance $vt = 1\,\text{m}$.

It is important to note that Figure 14.10 is not restricted to describing transverse waves, even if the illustration suggests that the particles' displacements (vertical axis) are perpendicular to the direction of propagation of the wave (horizontal). The quantity, D, that is plotted on the vertical axis corresponds to the displacement of a particle from its equilibrium position. That displacement could correspond to the longitudinal displacement of a particle in a longitudinal wave.

At time $t = 0$ (solid line), the displacement of each point in the medium, $D(x, t = 0)$, as a function of their distance from the origin, x, can be described by a sine function:

$$D(x, t = 0) = A \sin\left(\frac{2\pi}{\lambda}x\right) \tag{14.2}$$

This corresponds to the displacement being 0 at the origin and at any position, x, that is a multiple of the wavelength, λ.

If the wave moves with velocity v in the positive x direction, then at time t, the sine function in Figure 14.10 will have shifted to the right by an amount vt (dotted line). The displacement of a point located at position x at time t will be the same as the displacement of the point at position $x - vt$ at time $t = 0$. For example, in Figure 14.10 the displacement of the point $x = 2\,\text{m}$ at time $t = 1\,\text{s}$ is the same as the displacement of the point at position $x - vt = 1\,\text{m}$ at $t = 0$.

We can state this condition as:

$$D(x,t) = D(x - vt, t = 0)$$

That is, at some time t, the displacement of a point at position x is found by finding the position of the point at $x-vt$ at $t = 0$. We already have an equation to find the displacement of a point at $t = 0$. Using the above condition, we can modify Equation 14.2 to write a function for the displacement of a point at position x at time t:

$$D(x,t) = A\sin\left(\frac{2\pi}{\lambda}(x - vt)\right)$$

Noting that $v/\lambda = 1/T$, we can write this as:

$$D(x,t) = A\sin\left(\frac{2\pi x}{\lambda} - \frac{2\pi t}{T}\right)$$

In the above derivation, we assumed that at time $t = 0$, the displacement at $x = 0$ was $D(x = 0, t = 0) = 0$. In general, the displacement could have any value at $x = 0$ and $t = 0$, so we can allow the wave to shift left or right by including a phase, ϕ, which can be determined from the displacement at $x = 0$ and $t = 0$:

$$\boxed{D(x,t) = A\sin\left(\frac{2\pi x}{\lambda} - \frac{2\pi t}{T} + \phi\right)} \tag{14.3}$$

where $\phi = 0$ corresponds to the displacement being zero at $x = 0$ and $t = 0$.

Checkpoint 14-3

What is the value of the phase ϕ if the displacement of the point at $x = 0$ is $D = A/2$ at time $t = 0$?

A) $\pi/6$.
B) $\pi/4$.
C) $\pi/3$.
D) $\pi/2$.

The equation above is written in terms of the wavelength, λ, and period, T, of the wave. Often, one uses the "wave number", k, and the "angular frequency", ω, to describe the wave. These are defined as:

$$k = \frac{2\pi}{\lambda} \tag{14.4}$$

$$\omega = \frac{2\pi}{T} \tag{14.5}$$

Using the wave number and the angular frequency removes the factors of 2π in the expression for $D(x,t)$, which can now be written as:

$$\boxed{D(x,t) = A\sin(kx - \omega t + \phi)} \tag{14.6}$$

It is important to note that the wave number, k, has no relation to the spring constant that we used for springs.

Using Equation 14.1, we can also relate the wave number and angular frequency to the speed of the wave:

$$v = \frac{\lambda}{T} = \frac{\frac{2\pi}{k}}{\frac{2\pi}{\omega}} = \frac{\omega}{k}$$

14.2.1 The wave equation

In Chapter 13, we saw that any physical system whose position, x, satisfies the following equation:

$$\frac{d^2x}{dt^2} = -\omega^2 x$$

will undergo simple harmonic motion with angular frequency ω, and that $x(t)$ can be modelled as:

$$x(t) = A\cos(\omega t + \phi)$$

Similarly, any system, where the displacement of a particle as a function of position and time, $D(x,t)$, satisfies the following equation:

$$\frac{\partial^2 D}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 D}{\partial t^2} \tag{14.7}$$

is described by a wave that propagates with a speed v. The equation above is called the "one-dimensional wave equation" and would be obtained from modelling the dynamics of the system, just as the equation of motion for a simple harmonic oscillator can be obtained from Newton's Second Law. For the harmonic oscillator, the properties of the system (e.g. mass and spring constant) determine the angular frequency, ω. For a wave, the properties of the medium determine the speed of the wave, v.

We use partial derivatives in the wave equation instead of total derivatives because $D(x,t)$ is multi-variate. A possible solution to the one-dimensional wave equation is:

$$D(x,t) = A\sin(kx - \omega t + \phi)$$

which is the function that we used in the previous section to describe a sine wave.

Furthermore, if multiple solutions to the wave equation, $D_1(x,t)$, $D_2(x,t)$, etc, exist, then any linear combination, $D(x,t)$, of the solutions will also be a solution to the wave equation:

$$D(x,t) = a_1 D_1(x,t) + a_2 D_2(x,t) + a_3 D_3(x,t) + \dots$$

This last property is called "the superposition principle", and is the result of the wave equation being linear in D (it does not depend on D^2, for example). It is easy to check, for example, that if $D_1(x,t)$ and $D_2(x,t)$ satisfy the wave equation, so does their sum.

In three dimensions, the displacement of a particle in the medium depends on its three spatial coordinates, $D(x,y,z,t)$, and the wave equation in Cartesian coordinates is given by:

$$\frac{\partial^2 D}{\partial x^2} + \frac{\partial^2 D}{\partial y^2} + \frac{\partial^2 D}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 D}{\partial t^2}$$

There are many functions that can satisfy this equation, and the best choice will depend on the physical system being modelled and the properties of the wave that one wishes to describe.

14.3 Waves on a rope

In this section, we model the motion of transverse waves on a rope, as this provides insight into many properties of waves that extend to waves propagating in other media.

14.3.1 A pulse on a rope

We start by modelling how a single pulse propagates down a horizontal rope that is under a tension, F_T[1]. A wave is generally considered to be a regular series of alternating upwards and downwards pulses propagating down the rope. Modelling the propagation of a pulse is thus equivalent to modelling the propagation of a wave. Figure 14.11 shows how one can generate a pulse in a taught horizontal rope by raising (and then lowering) one end of the rope.

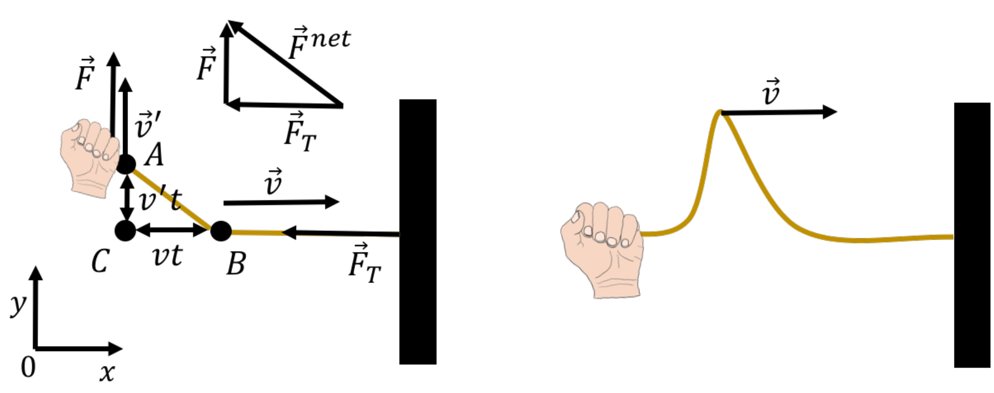


Figure 14.11: (Left:) Pulling upwards and then downwards on a horizontal rope causes a pulse to form and propagate. After a short period of time, a pulse is seen propagating down the rope (right).

[1] We do not use T for tension, so as to not confuse with the period of a wave.

We can model the propagation speed of the pulse by considering the speed, v, of point B that is shown in the left panel of Figure 14.11. Note that point B is not a particle of the rope, and is, instead, the location of the "front" of the disturbance that the pulse causes on the rope. We model the rope as being under a horizontal force of tension, $\vec{F}_T$, and the pulse is started by exerting a vertical force, $\vec{F}$, to move the end (point A) of the rope upwards with a speed, v'. Thus, by pulling upwards on the rope with a force, $\vec{F}$, at a speed v', we can start a disturbance in the rope that will propagate with speed v.

In a short amount of time, t, the point A on the rope will have moved up by a distance $v't$, whereas point B will have moved to the right by a distance vt. If t is small enough, we can consider the points A, B, and C to form the corners of a triangle. That triangle is similar to the triangle that is made by vectorially summing the applied force $\vec{F}$ and the tension $\vec{F}_T$, as shown in the top left of Figure 14.11. In this case, we mean the geometry term "similar", which describes two triangles which have the same angles. We can thus write:

$$\begin{aligned}\frac{F}{F_T} &= \frac{v't}{vt} = \frac{v'}{v}\\ \therefore F &= F_T\frac{v'}{v}\end{aligned}$$

Consider the section of rope with length vt that we have raised by applying that force (we assume that the distance AB is approximately equal to the distance BC). If the rope has a mass per unit length μ, then the mass of the rope element that was raised (between points A and B) has a mass, m, given by:

$$m = \mu vt$$

The vertical component of the momentum of that section of rope, with vertical speed given by v', is thus:

$$p = mv' = \mu vtv'$$

If the vertical force, $\vec{F}$, was exerted for a length of time, t, on the mass element, it will give it a vertical impulse, Ft, equal to the change in the vertical momentum of the mass element:

$$\begin{aligned}Ft &= \Delta p\\ Ft &= \mu vtv'\\ \therefore F &= \mu vv'\end{aligned}$$

We can equate this expression for F with that obtained from the similar triangles to obtain an expression for the speed, v, of the pulse:

$$\begin{aligned}\mu vv' &= F_T\frac{v'}{v}\\ \therefore v &= \sqrt{\frac{F_T}{\mu}}\end{aligned}$$

The speed of a pulse (and wave) propagating through a rope with linear mass density, μ, under a tension, F_T, is given by:

$$\boxed{v = \sqrt{\frac{F_T}{\mu}}} \tag{14.8}$$

If the tension in the rope is higher, the pulse will propagate faster. If the linear mass density of the rope is higher, then the pulse will propagate slower.

14.3.2 Reflection and transmission

In this section, we examine what happens when a pulse travelling down a rope arrives at the end of the rope. First, consider the case illustrated in Figure 14.12 where the end of the rope is fixed to a wall.

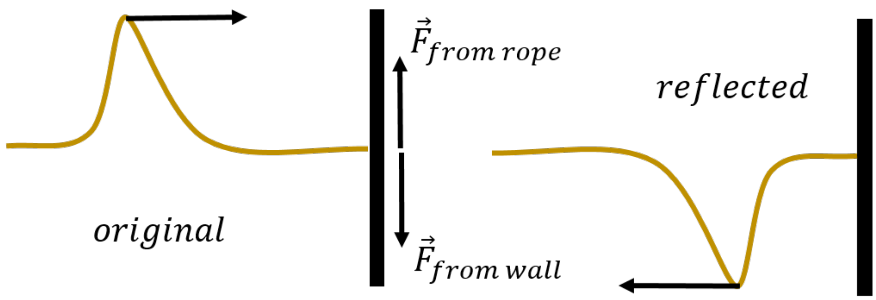


Figure 14.12: When the end of the rope is held fixed, the reflected pulse will be inverted.

When the pulse arrives at the wall, the rope will exert an upwards force on the wall, $\vec{F}_{from\,rope}$. By Newton's Third Law, the wall will then exert a downwards force on the rope, $\vec{F}_{from\,wall}$. The downwards force exerted on the rope will cause a downwards pulse to form, and the reflected pulse will be inverted compared to the initial pulse that arrived at the wall.

Now, consider the case when the end of the rope has a ring attached to it, so that it can slide freely up and down a post, as illustrated in Figure 14.13.

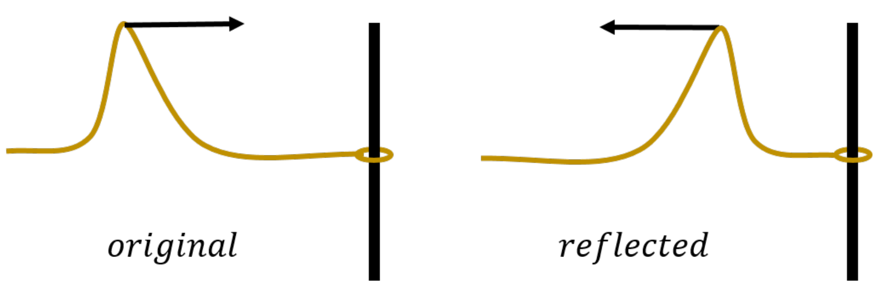


Figure 14.13: When the end of the rope is free, the reflected pulse will be upright.

In this case, the end of the rope will move up as the pulse arrives, which will then create a reflected pulse that is in the same orientation as the incoming pulse.

Finally, consider a pulse that propagates down a rope of mass per unit length μ_1 that is tied to a second rope with mass per unit length μ_2, which have the same tension. When the pulse arrives at the interface between the two media (the two ropes), part of the pulse will be reflected back, and part will be transmitted into the second medium (Figure 14.14).

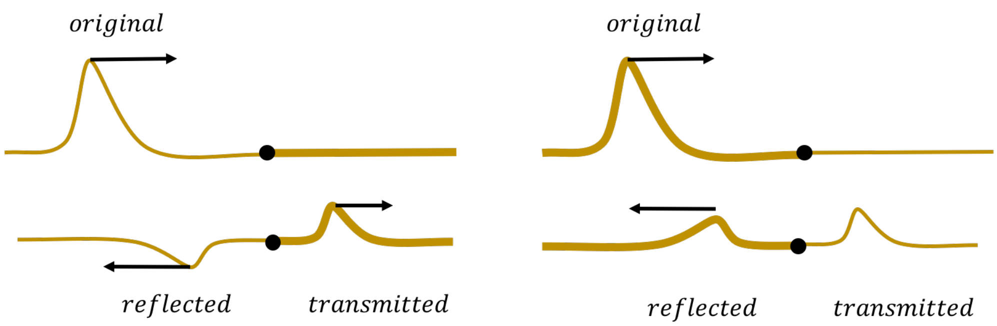


Figure 14.14: A pulse can be both reflected and transmitted as it changes medium. Left panel: The pulse is transmitted from a lighter rope to a heavier rope. Right panel: The pulse is transmitted from a heavier rope to a lighter rope

By considering the boundary conditions, one can derive the coefficient of reflection, R (see Problem 14-2 for the derivation). This coefficient is the ratio of the amplitude of the reflected pulse to the amplitude of initial pulse. The ratio is found to be:

$$R = \frac{\sqrt{\mu_1} - \sqrt{\mu_2}}{\sqrt{\mu_1} + \sqrt{\mu_2}}$$

When the pulse moves from a lighter rope to a heavier rope ($\mu_1 < \mu_2$), the reflected pulse will be inverted ($R < 0$). When the pulse moves from a heavier rope to a lighter rope ($\mu_1 > \mu_2$), the reflected pulse will stay upright ($R > 0$).

When the end of the rope is fixed to a wall (as in Figure 14.12), this represents a limiting case in which the linear mass density of the second material approaches infinity ($\mu_2 \to \infty$):

$$R = \lim_{\mu_2 \to \infty} \frac{\sqrt{\mu_1} - \sqrt{\mu_2}}{\sqrt{\mu_1} + \sqrt{\mu_2}} = \frac{-\sqrt{\mu_2}}{\sqrt{\mu_2}} = -1$$

which means that the amplitude of the reflected pulse will have the same magnitude as the initial pulse but will be in the opposite direction. When the end of the rope is free (Figure 14.13), this represents another limiting case, where $\mu_2 \to 0$:

$$R = \lim_{\mu_2 \to 0} \frac{\sqrt{\mu_1} - \sqrt{\mu_2}}{\sqrt{\mu_1} + \sqrt{\mu_2}} = \frac{\sqrt{\mu_1}}{\sqrt{\mu_1}} = 1$$

which means that the amplitude of the reflected pulse will be in the same direction and have the same amplitude as the initial pulse.

Checkpoint 14-4

A wave propagates from a light rope to a heavier rope that is attached to the light rope (as the pulse illustrated in Figure 14.14). What can you say about the wavelength of the wave on either side of the interface?

A) It is the same in both sections of rope.
B) The wavelength in the heavy section of rope is longer.
C) The wavelength in the light section of rope is longer.

14.3.3 The wave equation for a rope

In this section, we show how to use Newton's Second Law to derive the wave equation for transverse waves travelling down a rope with linear mass density, μ, under a tension, F_T. Consider a small section of the rope, with mass dm, and length dx, as a wave passes through that section of the rope, as illustrated in Figure 14.15.

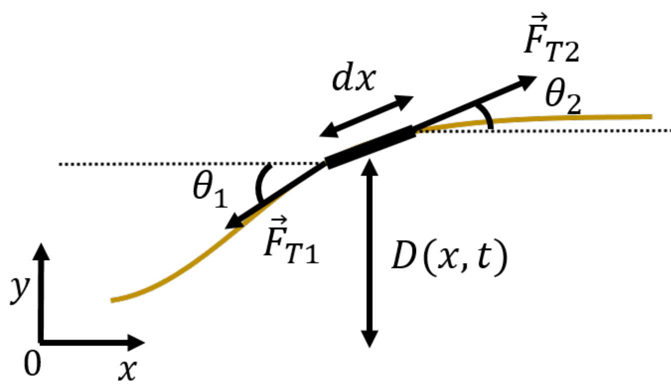


Figure 14.15: A small section of rope under tension as a wave passes through.

We assume that the weight of the mass element is negligible compared to the force of tension that is in the rope. Thus, the only forces exerted on the mass element are those from the tension in the rope, pulling on the mass element from each side, with forces, $\vec{F}_{T1}$ and $\vec{F}_{T2}$. In general, the forces from tension on either side of the mass element will have different directions and make different angles, θ, with the horizontal, although their magnitude is the same. Let $D(x,t)$ be the vertical displacement of the mass element located at position x. We can write the y (vertical) component of Newton's Second Law for the mass element, dm, as:

$$\begin{aligned}\sum F_y = F_{T2y} - F_{T1y} &= (dm)a_y \\ F_T\sin\theta_2 - F_T\sin\theta_1 &= dm\frac{\partial^2 D}{\partial t^2} \\ F_T(\sin\theta_2 - \sin\theta_1) &= dm\frac{\partial^2 D}{\partial t^2}\end{aligned}$$

where we used the fact that the force of tension has a magnitude, F_T, on either side of the mass element, and that the acceleration of the mass in the vertical direction is the second time-derivative of $D(x,t)$, since for a transverse wave, this corresponds to the y position of a particle. We now make the small angle approximation:

$$\sin\theta \approx \tan\theta = \frac{\partial D}{\partial x}$$

in which the sine of the angle is approximately equal to the tangent of the angle, which is equal to the slope of the rope. Applying this approximation to Newton's Second Law:

$$F_T\left(\frac{\partial D}{\partial x}\bigg|_{right} - \frac{\partial D}{\partial x}\bigg|_{left}\right) = dm\frac{\partial^2 D}{\partial t^2}$$

where we indicated that the term in parentheses is the difference in the slope of the rope between the right side and the left side of the mass element. If the rope has linear mass density, μ, then the mass of the rope element can be expressed in terms of its length, dx:

$$dm = \mu dx$$

Replacing dm in the equation gives:

$$\begin{aligned}
F_T\left(\frac{\partial D}{\partial x}\bigg|_{right} - \frac{\partial D}{\partial x}\bigg|_{left}\right) &= \mu dx\frac{\partial^2 D}{\partial t^2}\\
F_T\left(\frac{\frac{\partial D}{\partial x}\Big|_{right} - \frac{\partial D}{\partial x}\Big|_{left}}{dx}\right) &= \mu\frac{\partial^2 D}{\partial t^2}
\end{aligned}$$

The term in parentheses is the difference in the first derivatives of $D(x,t)$ with respect to x, divided by the distance, dx, between which those derivatives are evaluated. This is precisely the definition of the second derivative with respect to x, so we can write:

$$\begin{aligned}
F_T\frac{\partial^2 D}{\partial x^2} &= \mu\frac{\partial^2 D}{\partial t^2}\\
\therefore \frac{\partial^2 D}{\partial x^2} &= \frac{\mu}{F_T}\frac{\partial^2 D}{\partial t^2}
\end{aligned}$$

which is precisely the wave equation:

$$\frac{\partial^2 D}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 D}{\partial t^2}$$

with speed:

$$v = \sqrt{\frac{F_T}{\mu}}$$

as we found earlier. Thus, we find that the speed of the propagation of the wave is related to the dynamics of modelling the system, and is not related to the wave itself.

14.4 The speed of a wave

In the previous section we found that the speed of a transverse wave in a rope is related to the ratio of the tension in the rope to the linear mass density of the rope:

$$v = \sqrt{\frac{F_T}{\mu}}$$

The speed of a wave in any medium is usually given by a ratio, where the numerator is a measure of how easy it is to deform the medium, and the denominator is measure of the inertia of the medium. For a rope, the tension is a measure of how stiff the rope is. A higher tension makes it more difficult to disturb the rope from equilibrium and it will "snap back" faster when disturbed, so the pulse will propagate faster. The heavier the rope, the harder it will be for the disturbance to propagate as the rope has more inertia, which will slow down the pulse.

The only way that a wave can propagate through a medium is if that medium can be deformed and the particles in the medium can be displaced from their equilibrium position, much like simple harmonic oscillators. The wave will propagate faster if those oscillators have a stiff spring constant and there is a strong force trying to restore them to equilibrium. However, if those oscillators have a large inertia, even with a large restoring force, they will accelerate back to their equilibrium with a smaller acceleration.

In general, the speed of a wave is given by:

$$v = \sqrt{\frac{\text{Stiffness of medium}}{\text{Inertia of medium}}}$$

For example, the speed of longitudinal pressure waves in a solid is given by:

$$v = \sqrt{\frac{E}{\rho}}$$

where E is the "elastic (or Young's) modulus" for the material, and ρ is the density of the material. The elastic modulus of a solid is a measure of the material's resistance to being deformed when a force (or pressure) is exerted on it. The more easily it is deformed, the lower its elastic modulus will be.

For the propagation of longitudinal pressure waves through a fluid, the speed is given by:

$$v = \sqrt{\frac{B}{\rho}}$$

where B is the bulk modulus of the liquid, and ρ its density.

Checkpoint 14-5

A wave will propagate faster through...

A) ice.
B) water.

14.5 Energy transported by a wave

In this section, we examine how to model the energy that is transported by waves. Although no material moves along with a wave, mechanical energy can be transported by a wave, as evidenced by the damage caused by the waves from an earthquake.

14.5.1 A wave as being made of simple harmonic oscillators

Consider a wave that is propagating through a medium. We can model the motion of one of the particles in the medium as if it were the motion of a simple harmonic oscillator[2]. This is illustrated in Figure 14.16, which shows the displacement **as a function of time** for a point in the medium located at the origin when a wave passes through that point. The displacement of that point, at $x = 0$, if we choose $\phi = 0$, is given by:

$$D(x = 0, t) = A\sin(-\omega t)$$

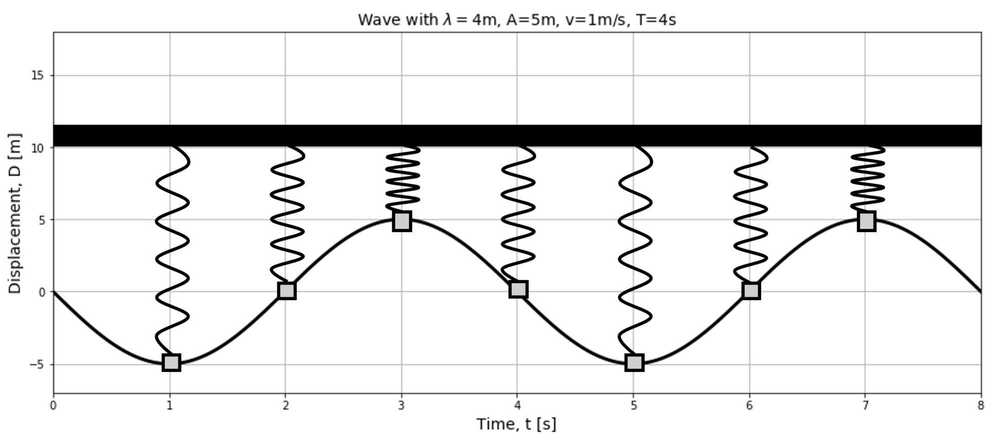


Figure 14.16: The displacement as a function of time for one particle in the medium (at $x = 0$) is identical to the motion of a simple harmonic oscillator.

The displacement of the particle in the medium is described by the same equation as the position of a simple harmonic oscillator, with the same angular frequency ω, as that of the wave.

We can also view a snapshot of the wave in time, and model the **different** points in the medium as different oscillators that all have different displacements. This is shown in Figure 14.17.

[2]If the medium has a linear restoring force or if the amplitude of the oscillations is small.

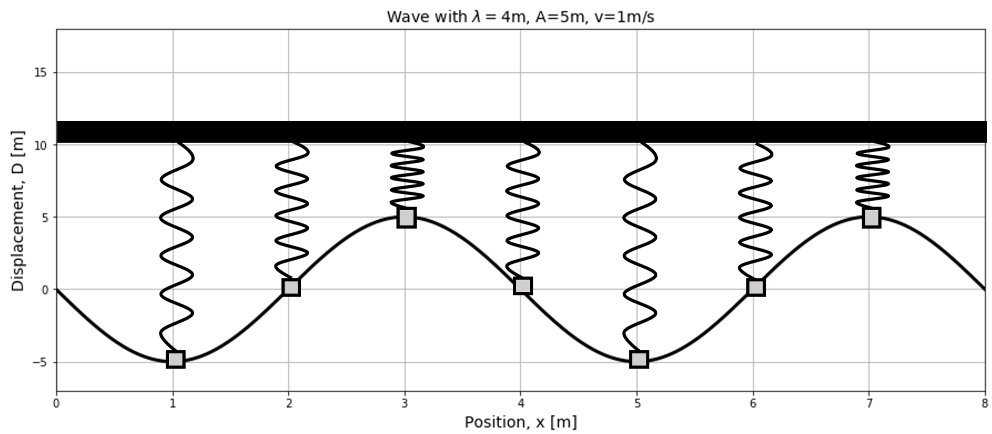


Figure 14.17: The displacement as a function of position for different points in a medium. Each point in the medium can be modelled as a simple harmonic oscillator.

14.5.2 Energy transported in a one dimensional wave

In this section, we show how to describe the energy transported by a one-dimensional wave along a rope. We model each particle in the rope through which the wave propagates as a small simple harmonic oscillator with mass m, attached to a spring with an effective spring constant, k_s[3].

Of course, there is no actual spring, but we can still determine an effective spring constant, k_s, from the angular frequency:

$$\omega = \sqrt{\frac{k_s}{m}}$$
$$\therefore k_s = \omega^2 m$$

which corresponds to the spring constant that would give the correct angular frequency for the particle of mass m.

The total mechanical energy of one oscillator, E_m, can be evaluated when the oscillator is at its maximal displacement, A, from its equilibrium, where its kinetic energy is zero:

$$E_m = \frac{1}{2}k_s A^2 = \frac{1}{2}\omega^2 m A^2$$

If the rope is infinitely long, and carries a continuous wave, it will have an infinite amount of energy, as it will correspond to an infinite number of oscillators. Instead, let us calculate how much energy, E_λ, is stored in the wave over one wavelength, λ. To do so, we need to evaluate how many effective oscillators are contained in the rope, over a distance λ, so that we can sum all of their energies together to obtain the energy stored in one wavelength:

$$E_\lambda = \sum \frac{1}{2}\omega^2 m A^2$$

[3]We use k_s for the spring constant, to distinguish it from k, the wave number.

where the sum is over the number of oscillators in one wavelength. Of course, the rope is not actually made of oscillators, but we can model each section of rope of length dx has being an oscillator of mass $dm = \mu dx$, where μ is the linear mass density of the rope. The sum (integral) of the energy of the oscillators over one wavelength can thus be written as:

$$E_\lambda = \int_0^\lambda \frac{1}{2}\omega^2\mu A^2 dx = \frac{1}{2}\omega^2\mu A^2\lambda$$

The energy stored in one wavelength is not a very useful property of a wave, since the total energy in the wave depends on the length of the wave. We can describe the rate at which energy is transmitted by the wave (its power), since we know how long, T, it will take the wave to travel one wavelength, and we just determined how much energy is stored in one wavelength. The average power with which energy is transported by a wave is given by:

$$P = \frac{E_\lambda}{T} = \frac{\frac{1}{2}\omega^2\mu A^2\lambda}{T} = \frac{1}{2}\omega^2\mu A^2 v$$

where T is the period of the wave, and $v = \lambda/T$ is the speed of the wave. The power transmitted by a wave on a rope is thus given by:

$$\boxed{P = \frac{1}{2}\omega^2\mu A^2 v} \tag{14.9}$$

We can see that the power transmitted by a wave goes as the amplitude, A, of the wave squared. It thus takes four times more energy to double the amplitude of waves that are sent down a rope.

14.5.3 Energy transported in a spherical, three-dimensional, wave

In this section, we show how to model the rate at which energy is transported in spherical three-dimensional waves, such as the sound waves that are generated when you clap your hands. A spherical sound wave is a pressure disturbance in the air that propagates spherically outwards from a point of emission. We can think of thin spherical shells containing air that expand and contract about their equilibrium position as the wave moves through the shells. The motion of each shell is similar to that of a simple harmonic oscillator of mass dm, where dm is the mass of air in the oscillating shell.

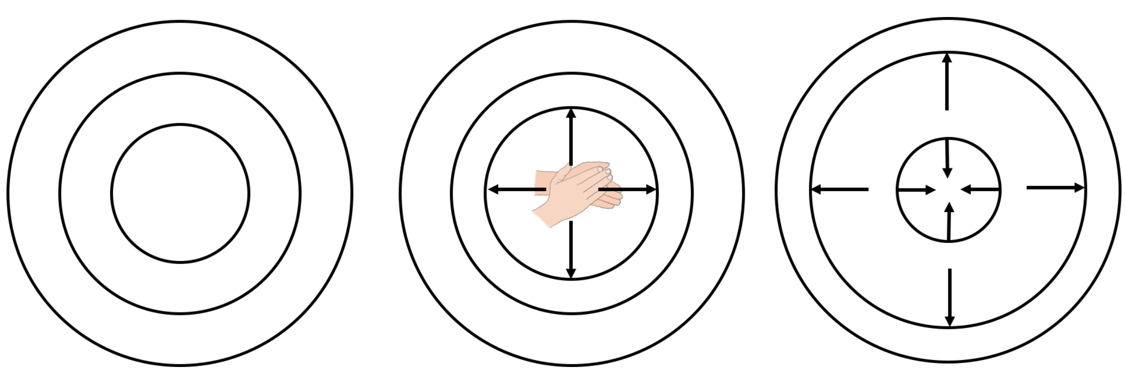


Figure 14.18: Left: We divide the air into thin spherical shells. Here we represent three shells (the black circles). Center: When you clap, the innermost shell is given energy and expands. Right: Energy is transferred to the next shell, which expands as the first shell contracts. This is how the wave propagates outwards. When the shells are closer together, the air molecules are closer together and exert a pressure that tries to expand the shell.

Consider a shell at a radial position, r, from the source, with thickness dr, and mass dm:

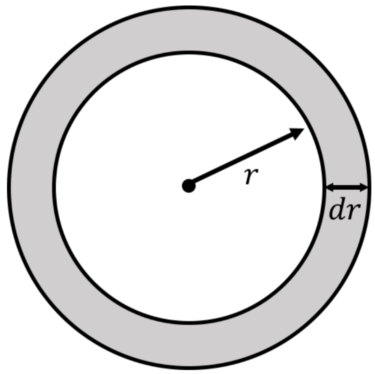


Figure 14.19: A spherical shell at radial position r with thickness dr

If the medium has a density, ρ, then the mass of the shell is given by:

$$dm = \rho dV = \rho 4\pi r^2 dr$$

where $dV = 4\pi r^2 dr$ is the volume of the shell. Again, if we model each shell as a simple harmonic oscillator with mass dm, then the energy, dE, stored in that oscillating shell is given by:

$$dE = \frac{1}{2}k_s A^2 = \frac{1}{2}\omega^2 dm A^2 = \frac{1}{2}\omega^2 A^2 \rho 4\pi r^2 dr = 2\pi\rho\omega^2 A^2 r^2 dr$$

where ω is the angular frequency of the wave, and A is the amplitude of the wave. We expressed the effective spring constant, k_s, in terms of the angular frequency of the simple harmonic oscillator and its mass, as we did in the previous section. It now makes less sense to determine the energy that is stored in one wavelength of the wave because the energy, dE, stored in one shell depends on the location, r, of that shell. This was not the case for a one-dimensional wave, where the energy stored in one oscillator did not depend on the position of that oscillator.

The rate at which energy is transported by the wave is given by:

$$P = \frac{dE}{dt}$$

We can use the Chain Rule to change this into a derivative over r:

$$P = \frac{dE}{dr}\frac{dr}{dt} = \frac{dE}{dr}v$$

where $\frac{dr}{dt} = v$ is the speed of the wave (the rate of change of the radius of a shell). The power transmitted by the spherical wave is thus given by:

$$P = \frac{dE}{dr}v = 2\pi\rho\omega^2 A^2 r^2 v$$

where the power appears to depends on how far you are from the source (r).

Suppose that you have a 50 W speaker emitting sound; each radial shell emanating from the speaker must transport energy at a rate of 50 W. This is simply a statement that the energy radiated by the speaker has to move from one shell to the next and be conserved. Since the power transported by a shell appears to depend on the radius of the shell, if the power transmitted by each shell is the same, then the amplitude of the wave in each shell must decrease, so that the power does not actually depend on the radius of the shell. In particular, for a spherical wave, the amplitude will decrease as a function of distance from the source:

$$\begin{aligned} P &= \text{constant} \\ \therefore A &= \frac{1}{r}\sqrt{\frac{P}{2\pi\rho\omega^2 v}} \end{aligned}$$

This is very different from the propagation of a one-dimensional wave, in which the amplitude does not change with distance. In practice, if there are energy losses due to, say, friction, then the amplitude of a one-dimensional wave would also decrease with distance from the source, but this is a different effect.

Olivia's Thoughts

Here's a slightly different way to think about why the amplitude of the wave decreases as you get further from the source. When a spherical wave travels outwards, energy is passed from one shell to the next. The outer shells are bigger than the inner shells, and so they will contain more particles. Because of conservation of energy, when the energy is transferred from one shell to the next, the total energy stays the same. In the outer shells, the energy must be shared between a greater number of particles, so each particle gets less energy, and therefore oscillates with a smaller amplitude than the particles in the previous shell did.

To remember this, imagine the shells in Figure 14.18 are circles of kids standing side by side. The innermost circle has 10 kids and the outermost circle has 100 kids. If you have 100 candies, and you give them to the kids in the innermost circle, each will get 10 so they will get really hyper and start jumping around a lot. If you instead give the 100 candies to the kids in the outermost circle, each will only get one. The kids will only get a little bit hyper and jump around less.

The "intensity of a wave", I, is defined as the power per unit area that is transmitted by the wave. For a spherical wave front at radial position r, with area $4\pi r^2$, the intensity of the wave is defined as:

$$I = \frac{P}{4\pi r^2} = \frac{1}{2}\rho\omega^2 A^2 v$$

Usually, the intensity of a wave is something that you can measure, as it corresponds to the power delivered into some measuring device with a known surface area. For example, we cannot directly measure the total power that is transported by the waves from an earthquake, as we would need an instrument that could encompass the entire resulting wave. Instead, we can measure the intensity of waves from the earthquake by measuring how much power is delivered into some instrument with a known surface area. By knowing our distance from the earthquake, we could then determine the total power output of the earthquake.

The intensity is a measure of how much energy is delivered per unit area by a wave and goes down as the square of the distance from the source (since $A \propto 1/r$). If the source of the wave is an earthquake, then your house will have four times less damage than your friend's, if your house is located only twice as far from the epicentre as your friend's. You will cause four times less damage to your ears if you move only twice as far away from the stage at a rock concert.

14.6 Superposition of waves and interference

In this section, we consider what happens when two (or more) different waves propagate in a medium and interfere with each other. The superposition principle states that if $D_1(x,t)$, $D_2(x,t)$, etc, are functions that satisfy the wave equation, then any linear combination of these functions, $D(x,t)$:

$$D(x,t) = a_1 D_1(x,t) + a_2 D_2(x,t) + a_3 D_3(x,t) + \dots$$

will also satisfy the wave equation.

Suppose that you hold one end of a rope and shake it with a specific frequency, creating waves that are described by:

$$D_1(x,t) = A_1 \sin(k_1 x - \omega_1 t + \phi_1)$$

Your friend, at the other end of the rope shakes the rope with a different frequency, creating waves that propagate in the opposite direction and that are described by:

$$D_2(x,t) = A_2 \sin(k_2 x + \omega_2 t + \phi_2)$$

The superposition principle states that the net displacement at any position x at some time t can be found by summing the displacement from the two waves together:

$$D(x,t) = A_1 \sin(k_1 x - \omega_1 t + \phi_1) + A_2 \sin(k_2 x + \omega_2 t + \phi_2)$$

The superposition of waves is illustrated in Figure 14.20, which shows three waves, and their resulting sum in the bottom most panel.

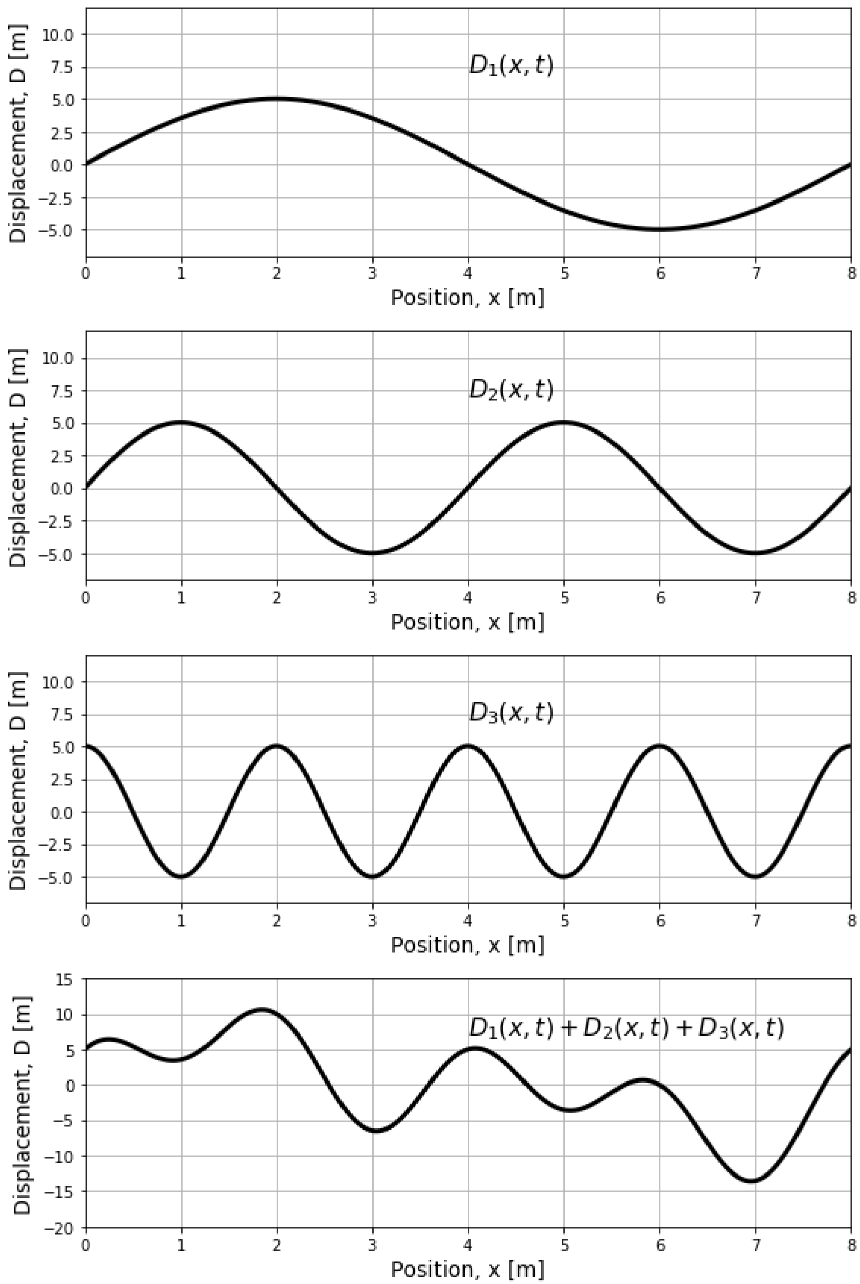


Figure 14.20: The superposition of three waves to create a resulting wave shown in the bottom panel. The waves are shown as the displacement as a function of position at a fixed instant in time.

The resulting wave is created by the "interference" of the three waves, and mathematically is simply a sum of the three individual waves at each position (and instant in time). The resulting wave in this example has a rather complicated shape, that is no longer described by a sine function. However, by the superposition principle, it is a valid solution to the wave equation[4].

[4]Fourier's Theorem states that any periodic function can be described as the linear combination of sine

The individual waves in the top three panels of Figure 14.20 all have an amplitude of $5\,\text{m}$. The resulting wave, at some points (e.g. at $x = 2\,\text{m}$), has an amplitude that is larger than any of the individual waves; we say that, at those positions, the individual waves have "constructively interfered". In other locations (e.g. at $x = 6\,\text{m}$), the resulting wave has a smaller amplitude than the individual waves, and we say that the individual waves have "destructively interfered". The interference between waves can be observed easily on a water surface, for example by observing the constructive and destructive interference pattern of waves that originate from two pebbles being dropped at the same time a certain distance apart. Constructive interference between waves is also thought to be behind some reports of gigantic waves observed out at sea.

If two waves have the same wavelength and amplitude, it is possible for them to completely destructively interfere, resulting in no net wave. Similarly, they can also completely constructively interfere, resulting in a wave with a larger amplitude. Complete destructive and constructive interference are illustrated in the left and right panel of Figure 14.21, respectively.

(or cosine) functions. This is the reason why we focused on using a sine function to describe a wave.

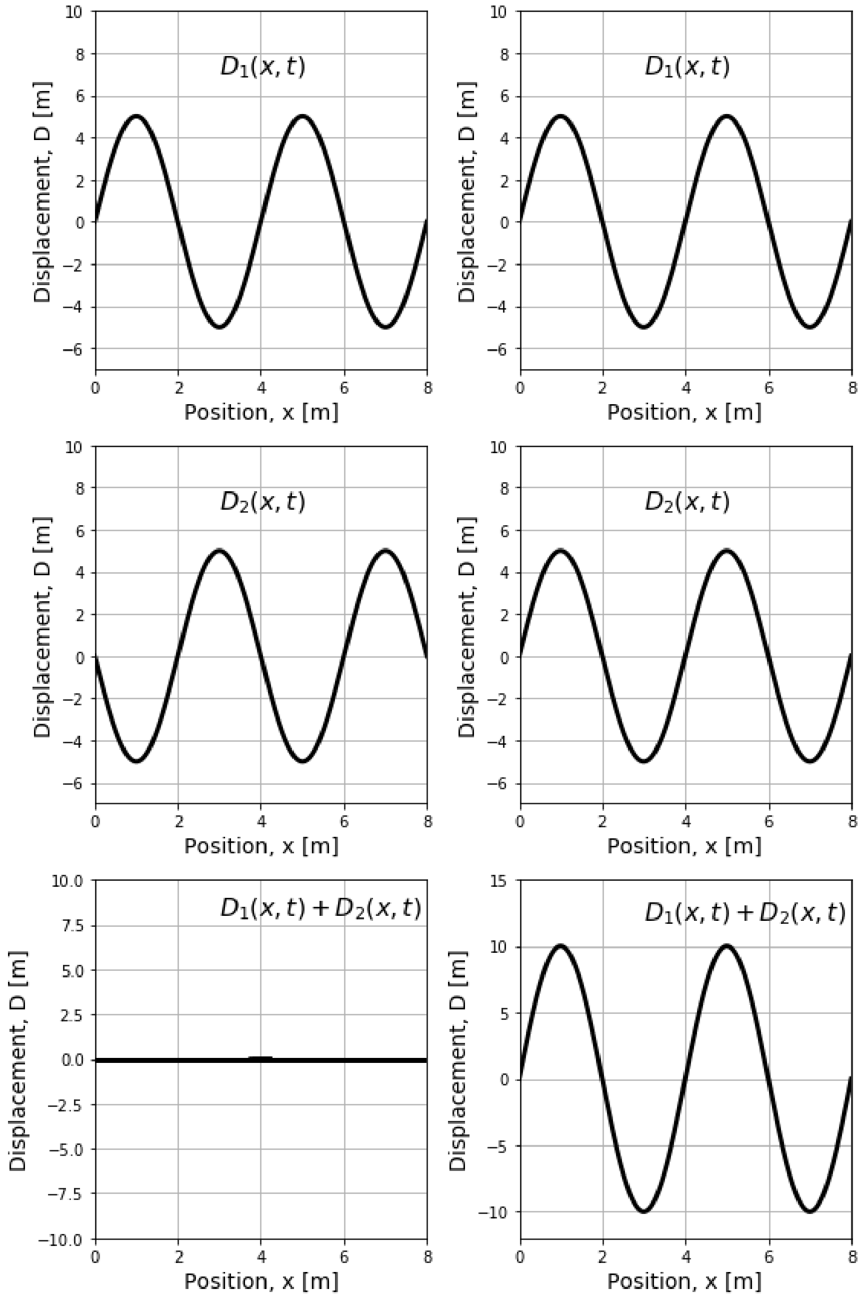


Figure 14.21: Destructive (left) and constructive (right) interference of waves.

14.7 Standing waves

As we saw in the last section, when waves have the same frequency, it is possible for them to interfere completely, either destructively or constructively. Waves of the same frequency that interfere can be generated by propagating waves along a string, as the reflected waves from the end of the string will have the same frequency as, and interfere with, the original waves. In general, the resulting wave will be quite complicated, but if you "choose" the frequency (or wavelength) of the generated waves precisely, then the waves will interfere and create a "standing wave". The standing wave is named this way because it does not appear to propagate along the string. Instead, each point on the string will oscillate with an amplitude that depends on where the point is located along on the string. In contrast, for a travelling wave, all of the points oscillate with the same amplitude.

Three standing waves of different frequencies (wavelengths) are illustrated in Figure 14.22.

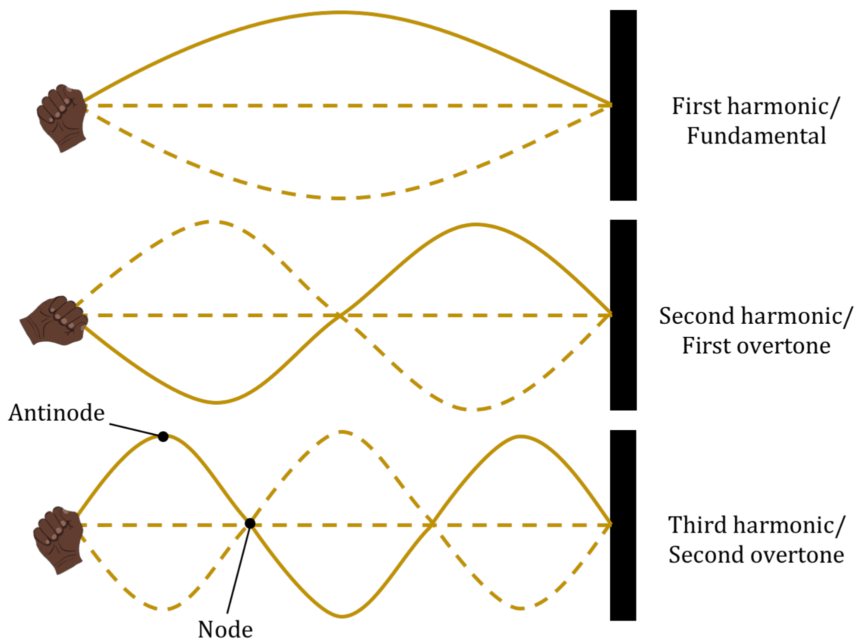


Figure 14.22: The first three standing waves on a string.

The solid line in each of the three panels corresponds to one particular snapshot of the standing wave at a particular instant in time. The dashed lines correspond to snapshots at different times. In particular, there is a time where the displacement of all points on the string is zero. Each point on the string vibrates with a different amplitude, which corresponds to the solid line (and the opposite dashed line). Certain points do not oscillate at all; these are called "nodes". The points at the end of the string are always nodes. Certain points vibrate with a maximal amplitude; these are called "anti-nodes".

In general, if you pluck a taught string (such as a guitar string), you will create a complicated wave, equivalent to many sine waves with different frequencies, that propagate outwards from the point where the string was plucked. Those sine waves will be reflected by the ends of the string and interfere with each other. Most of the waves will interfere in a complicated way and decay away. Those waves that have the correct frequency to create standing waves will persist on the string for a longer period of time. The string will eventually vibrate as a superposition of the fundamental frequency (the standing wave with one anti-node, also called the first harmonic), and the higher "harmonics" (those standing waves with more anti-nodes).

The wavelength of the fundamental standing wave for a string of length, L, is given by the condition:

$$\lambda = 2L$$

In general, the nth harmonic will have a wavelength of:

$$\boxed{\lambda_n = \frac{2L}{n} \qquad n = 1, 2, 3, \dots} \tag{14.10}$$

The corresponding frequency is give by:

$$\boxed{f_n = \frac{nv}{2L}} \tag{14.11}$$

where $v = f\lambda$ is the speed of the waves on the string.

A standing wave is the result of two waves of the same frequency and amplitude travelling in opposite directions. Thus, there is no energy that is transmitted by a standing wave (e.g. through the nodes at the end of the string). Although we described standing waves for a string, these are not restricted to one dimensional waves. For example, the membrane of a drum can also support standing waves.

Checkpoint 14-6

A standing wave (composed of two travelling waves) has a maximum amplitude A. What must the amplitude A_0 of each travelling wave be?

A) $A_0 = 1/4A$
B) $A_0 = 1/2A$
C) $A_0 = A$
D) $A_0 = 2A$

In general, most objects can be characterized by a harmonic (or "resonant") frequency that corresponds to the standing waves that can exist in the object. If that object is, say, shaken, many waves will propagate through the object and cancel out, except those that have the resonant frequency. Relatively small vibrations, if at the correct frequency, can lead to large standing waves that can result in damage to the object.

14.7.1 Mathematical description of a standing wave

A standing wave is the result of two identical waves, travelling in opposite directions, interfering. Consider the waves described by $D_1(x, t)$ and $D_2(x, t)$ that are modelled as follows:

$$D_1(x, t) = A \sin(kx - \omega t)$$
$$D_2(x, t) = A \sin(kx + \omega t)$$

These two waves are identical, but travel in opposite directions (due to the sign in front of the ωt). The superposition of these waves is given by:

$$\begin{aligned}
D(x,t) &= D_1(x,t) + D_2(x,t) \\
&= A\Big(\sin(kx - \omega t) + \sin(kx + \omega t) \Big)
\end{aligned}$$

We can use the following trigonometric identity to combine these into a single term:

$$\sin\theta_1 + \sin\theta_2 = 2\sin\left(\frac{\theta_1 + \theta_2}{2}\right)\cos\left(\frac{\theta_1 - \theta_2}{2}\right)$$

The resulting wave is thus given by:

$$\begin{aligned}
D(x,t) &= 2A\sin\left(\frac{kx - \omega t + kx + \omega t}{2}\right)\cos\left(\frac{kx - \omega t - kx - \omega t}{2}\right) \\
&= 2A\sin(kx)\cos(\omega t)
\end{aligned}$$

If this wave describes the wave on a string of length L with both ends held fixed, and we set the origin of our coordinate system at one end of the string, then we require that the displacement at $x = 0$ and $x = L$ is always zero. The first condition is always true, and the second requires that:

$$\begin{aligned}
D(x = L, t) &= 0 \\
\sin(kL) &= 0 \\
\therefore kL &= n\pi \qquad n = 1, 2, 3, \dots
\end{aligned}$$

and kL must be a multiple of 2π. In terms of the wavelength, λ, this gives:

$$\begin{aligned}
\frac{2\pi}{\lambda}L &= n\pi \\
\therefore \lambda &= \frac{2L}{n}
\end{aligned}$$

as we argued before, for the wavelength of the n-th harmonic. The standing wave for the n-th harmonic is thus described by

$$\boxed{D(x,t) = 2A\sin\left(\frac{n\pi}{L}x\right)\cos(\omega t)} \tag{14.12}$$

A point at position x will behave like a simple harmonic oscillator and oscillate with an amplitude given by:

$$A(x) = 2A\sin\left(\frac{n\pi}{L}x\right)$$

Each point on the string will vibrate with the same angular frequency, ω, but with a different amplitude, depending on their position. For the n-th harmonic, the nodes of the standing wave are located at:

$$\begin{aligned}\sin\left(\frac{n\pi}{L}x\right) &= 0\\ \frac{n\pi}{L}x &= m\pi \qquad m = 0, 1, 2, \dots\\ \therefore x &= m\frac{L}{n}\end{aligned}$$

Thus, for example, the second node ($m = 2$) of the third harmonic ($n = 3$), is located at $x = 2L/3$, as can be seen in the bottom panel of Figure 14.22. The anti-nodes are located at:

$$\begin{aligned}\frac{n\pi}{L}x &= m\frac{\pi}{2} \qquad m = 1, 3, 5, 7, \dots\\ \therefore x &= m\frac{L}{2n}\end{aligned}$$

where, for example, the first anti-node of the first harmonic is located at $x = L/2$, as can be seen in the top panel of Figure 14.22.

Checkpoint 14-7

A standing wave on a string (fixed at both ends) has a fundamental frequency f. If you quadruple the tension in the string, how can you change the length of the string so that the fundamental frequency remains the same?

A) half the length.
B) double the length.
C) triple the length.
D) quadruple the length.

Olivia's Thoughts

Let's take another look at the equation for a standing wave. In this section, we saw that the equation for a standing wave is given by:

$$D(x, t) = 2A\sin(kx)\cos(\omega t)$$

We can rearrange this equation to get:

$$D(x, t) = \underbrace{2A\cos(\omega t)}_{\text{amplitude}}\sin(kx)$$

This looks like the equation for a stationary wave (the displacement is a function of x) with an amplitude $2A\cos(\omega t)$. We know that $\cos(\omega t)$ will give a value that ranges between -1 and 1, so we can just think of $\cos(\omega t)$ as a scaling term that modifies the amplitude of the wave.

When we look at a standing wave, this is exactly what we see - a wave whose amplitude is always changing but that does not travel one way or the other. Figure 14.23 shows a few snapshots of what the wave looks like at different times.

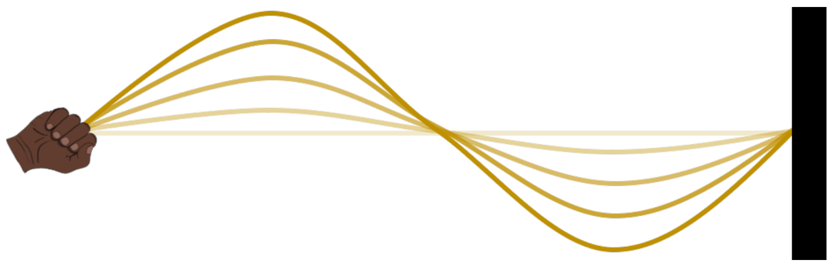


Figure 14.23: A standing wave as a stationary wave whose amplitude changes over time

We can see from the equation that the maximum amplitude will be $2A$. This makes sense when we remember that the standing wave is made of two travelling waves of amplitude A. As these waves move, there will be moments when they completely constructively interfere, which is when the amplitude of the standing wave is maximized. When they completely destructively interfere, the amplitude is zero.

14.8 Summary

Key Takeaways

A travelling wave is the propagation of a disturbance with a speed, v, through a medium. Particles in the medium oscillate back and forth, about an equilibrium position, as a wave passes through the medium, but they are not carried with the wave. Only energy is transmitted by a wave.

In a transverse wave, the particles in the medium oscillate in a direction that is perpendicular to the velocity of the wave. In a longitudinal wave, the particles of the medium oscillate in a direction that is co-linear with the velocity of the wave.

A sine wave is described by it frequency, f, its wavelength, λ, its amplitude, A, and its speed, v. We can also use the period of the wave, T, in lieu of the frequency. The frequency and wavelength of a wave are related to each other by the speed of the wave:

$$v = \lambda f$$

Mathematically, a one-dimensional travelling sine wave moving in the positive x direction can be described by:

$$D(x,t) = A\sin(kx - \omega t + \phi)$$

where $D(x,t)$ is the displacement of the particle in the medium at position x at time t. ϕ is the phase of the wave and depends on our choice of when $t = 0$. k is the wave number of the wave, and ω its angular frequency. These are related to the wavelength and frequency, respectively:

$$k = \frac{2\pi}{\lambda}$$
$$\omega = 2\pi f = \frac{2\pi}{T}$$

If a dynamical model (e.g. Newton's Second Law) of a system/medium leads to an equation with the following form:

$$\frac{\partial^2 D}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 D}{\partial t^2}$$

then waves with a speed of v can propagate through the system/medium.

The speed of a wave on a rope of linear mass density, μ, under a tension, F_T, is given by:

$$v = \sqrt{\frac{F_T}{\mu}}$$

Generally, the speed of a wave in a medium depends on the elasticity of the medium when it is deformed and the inertia of the particles in the medium. In order for a wave to propagate through a medium, the particles in the medium must be able to be displaced from their equilibrium position.

A pulse travelling through a rope will get reflected at the end of the rope and travel back in the opposite direction. If the end of the rope is fixed, the reflected pulse will be inverted. If the end of the rope can move, the reflected pulse will be in the same orientation as the incoming pulse.

A one-dimensional wave in a rope of linear mass density, μ, will transfer energy at an average rate:

$$P = \frac{1}{2}\omega^2\mu A^2 v$$

A three dimensional spherical wave through a medium with density ρ will transfer energy at an average rate:

$$P = 2\pi\rho\omega^2 r^2 v$$

at a distance r from the source of the wave. The amplitude of a spherical wave will decrease as the distance away from the source increases:

$$A = \frac{1}{r}\sqrt{\frac{P}{2\pi\rho\omega^2 v}}$$

The intensity of a spherical wave is defined as the power per unit area transferred by the wave, and is given by:

$$I = \frac{P}{4\pi r^2} = \frac{1}{2}\rho\omega^2 A^2 v$$

The superposition principle states that if $D_1(x,t)$, $D_2(x,t)$, ..., are functions that satisfy the wave equation, then any linear combination of these functions, $D(x,t)$:

$$D(x,t) = a_1 D_1(x,t) + a_2 D_2(x,t) + a_3 D_3(x,t) + \dots$$

will also satisfy the wave equation.

Different waves can interfere constructively or destructively in a medium, and the resulting wave is given by the sum of the functions describing the interfering waves.

Standing waves are formed when waves of the same frequency and amplitude travelling in opposite directions interfere. For standing waves on a string, the displacement of a particle on the string is given by:

$$D(x,t) = 2A\sin\left(\frac{n\pi}{L}x\right)\cos(\omega t)$$

where n is the number of the harmonic and L is the length of the string. In particular, a particle at position x will move up and down as a simple harmonic oscillator with amplitude:

$$A(x) = 2A\sin\left(\frac{n\pi}{L}x\right)$$

The condition for a standing wave to exist on a string is that the length of the string must be equal to a multiple of half of the wavelength of the standing wave:

$$\begin{aligned} L &= n\frac{\lambda}{2} \qquad n = 1, 2, 3, \ldots \\ \lambda &= \frac{2L}{n} \\ f &= \frac{nv}{2L} \end{aligned}$$

Important Equations

Travelling 1d waves:

$$
\begin{aligned}
D(x,t) &= A\sin(kx - \omega t + \phi) \\
k &= \frac{2\pi}{\lambda} \\
\omega &= 2\pi f = \frac{2\pi}{T} \\
v &= \lambda f
\end{aligned}
$$

Wave equation:

$$\frac{\partial^2 D}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 D}{\partial t^2}$$

Wave velocity:

$$v = \sqrt{\frac{F_T}{\mu}} \quad v = \sqrt{\frac{E}{\rho}} \quad v = \sqrt{\frac{B}{\rho}}$$

Power (1d wave in a rope):

$$P = \frac{1}{2}\omega^2 \mu A^2 v$$

Spherical waves:

$$
\begin{aligned}
P &= 2\pi\rho\omega^2 r^2 v \\
A &= \frac{1}{r}\sqrt{\frac{P}{2\pi\rho\omega^2 v}} \\
I &= \frac{P}{4\pi r^2} = \frac{1}{2}\rho\omega^2 A^2 v
\end{aligned}
$$

Standing waves:

$$
\begin{aligned}
D(x,t) &= 2A\sin\left(\frac{n\pi}{L}x\right)\cos(\omega t) \\
A(x) &= 2A\sin\left(\frac{n\pi}{L}x\right)
\end{aligned}
$$

Standing waves on a string (both ends fixed):

$$
\begin{aligned}
L &= n\frac{\lambda}{2} \qquad n = 1, 2, 3, \ldots \\
\lambda &= \frac{2L}{n} \\
f &= \frac{nv}{2L}
\end{aligned}
$$

14.9 Thinking about the material

Reflect and research

1. Look up a video of the Tacoma Narrows bridge failing, and explain what happened.
2. Why do airlines ask you to turn off your electronic devices during take-off?
3. Is it true that there is no sound in space?
4. What type of wave was first observed in 2015?

To try at home

1. Confirm that the reflected pulse from a rope on a string is inverted when the end of the rope is fixed.
2. Think of different ways you could create a standing wave at home and try one of them out. How many harmonics can you create? How can you modify your set-up to create more harmonics?

To try in the lab

1. Propose an experiment to verify the relation $v = \lambda f$.

14.10 Sample problems and solutions

14.10.1 Problems

Problem 14-1: A clarinet can be modelled as an air column that is open at one end and closed at the other end, as in Figure 14.24.

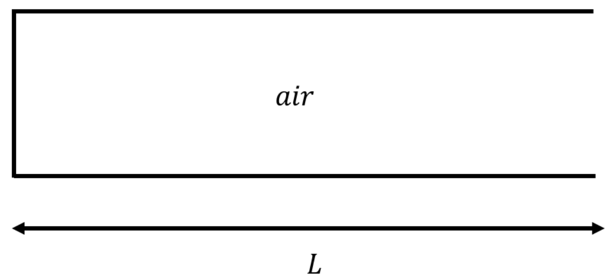


Figure 14.24: A clarinet (of length L) modelled as an air column that is closed at one end.

a) Draw the first three harmonics for a clarinet (draw the maximum displacement of the air molecules as a function of distance in the clarinet).
b) Find an expression for the wavelength of the n^{th} harmonic for a clarinet of length L.
c) If a clarinet is $60\,\text{cm}long$, what is the lowest frequency note it can produce?

(Solution)

Problem 14-2: A pulse propagates down a rope of mass per unit length μ_1 that is tied to a second rope with a mass per unit length μ_2 (Figure 14.25). The tensions in the ropes are equal in magnitude.

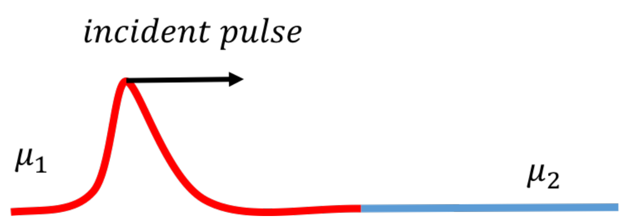


Figure 14.25: An incident pulse propagates through a rope connected to a another rope with a different linear mass density. When it reaches the boundary, part of the pulse will be reflected and part will be transmitted.

a) Write the displacements of the incident pulse, the reflected pulse, and the transmitted pulse in the form $D(x,t) = D(a(t \pm x/v))$, where a is some constant that you need to determine, and the choice of $+$ or $-$ depends on the direction that the pulse is travelling in.
b) The reflection coefficient, R, is the ratio of the amplitude of the reflected pulse to the amplitude of the incident pulse. Using the boundary conditions, show that the

reflection coefficient is given by:

$$R = \frac{\sqrt{\mu_1} - \sqrt{\mu_2}}{\sqrt{\mu_1} + \sqrt{\mu_2}}$$

Note: The boundary is the interface between the two ropes. By "using the boundary conditions", we mean that you should think about what must be true at the boundary for this problem to make sense. Boundary conditions are often more obvious than you think!

(Solution)

14.10.2 Solutions

Solution to problem 14-1:

(a) The first three harmonics are are shown in Figure 14.26.

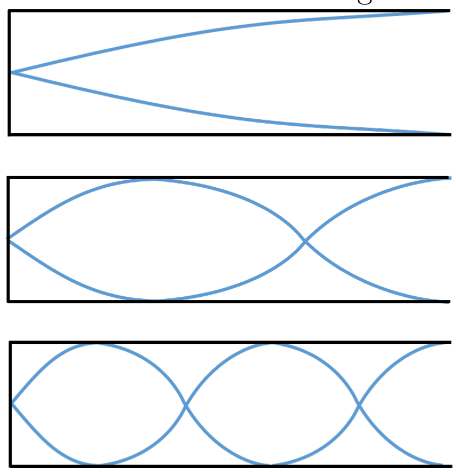


Figure 14.26: The first three harmonics for a clarinet. There is a node at the fixed end and an anti-node at the free end.

(b) The equation for a standing wave is:

$$D(x,t) = 2A\sin(kx)cos(\omega t)$$

We let the fixed end be at $x = 0$. At the fixed end, the displacement is equal to zero. At the free end ($x = L$) the displacement is maximized. The first condition is always true. The second condition will be met when:

$$\begin{aligned} \sin(kL) &= 1 \\ \therefore kL &= \pi/2, 3\pi/2, ... \end{aligned}$$

This condition can be expressed as:

$$\begin{aligned} kL &= \frac{(2n-1)\pi}{2} \\ \frac{2\pi L}{\lambda} &= \frac{(2n-1)\pi}{2} \\ \therefore \lambda &= \frac{4L}{2n-1} \end{aligned}$$

where, in the second line, we used $k = 2\pi/\lambda$. We can check that this formula works for the first three harmonics:

$$\begin{aligned}
n = 1: \quad \lambda &= \frac{4L}{2(1) - 1} \\
L &= \frac{1}{4}\lambda \\
n = 2: \quad \lambda &= \frac{4L}{2(2) - 1} \\
L &= \frac{3}{4}\lambda \\
n = 3: \quad \lambda &= \frac{4L}{2(3) - 1} \\
L &= \frac{5}{4}\lambda
\end{aligned}$$

Referring back to our diagram (Figure 14.26), we can see that our formula holds true for the first three harmonics (i.e. for the first harmonic, the length of the clarinet is equal to 1/4 of a wavelength, etc.)

(c) We found that the wavelength for the n^{th} wavelength is given by:

$$\lambda = \frac{4L}{2n - 1}$$

Writing λ in terms of the velocity, v, and frequency, f, gives:

$$\begin{aligned}
\frac{v}{f} &= \frac{4L}{2n - 1} \\
\therefore f &= \frac{v(2n - 1)}{4L}
\end{aligned}$$

From this formula, we can see that, if we want to find the lowest frequency, we want $n = 1$. The length of the clarinet is $0.6\,\mathrm{m}$, and v is the speed of sound in air which is $343\,\mathrm{m/s}$ at room temperature. Using these values, the lowest frequency is:

$$\begin{aligned}
f &= \frac{(343\,\mathrm{m/s})(2(1) - 1)}{4(0.6\,\mathrm{m})} \\
f &= 143\,\mathrm{Hz}
\end{aligned}$$

Discussion: This frequency is close to the D_3 note, which has a frequency of $144\,\mathrm{Hz}$, so this answer makes sense. However, the value we found differs from the true value. Why might this be?

Solution to problem 14-2:

(a) We let the incident pulse move in the positive x direction (Figure 14.27), and set $x = 0$ to be where the ropes connect.

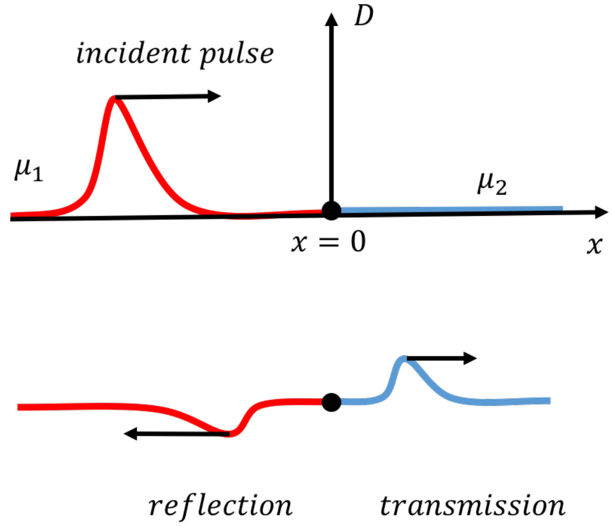


Figure 14.27: An incident pulse propagates through a rope connected to a another rope with a different linear mass density. When it reaches the boundary, part of the pulse is reflected and part is transmitted. Whether the reflected pulse is inverted or upright will depend on the reflection coefficient.

The incident pulse (denoted by i) is a travelling wave, moving in one dimension in the positive x direction. The incident pulse can thus be described by the function:

$$D_I(x,t) = A_I \cos(k_1 x - \omega t)$$

We will use the formulas $k = 2\pi/\lambda$ and $\omega = 2\pi f$ to rewrite this equation in the form $D = (a(t \pm x/v))$. The frequency, f, of the wave will be the same in both ropes. The velocity of the wave, and therefore its wavelength, depends on the mass density of the rope. Since the incident wave travels through the first rope (μ_1), its velocity will be v_1 and its wavelength will be λ_1. The incident wave can thus be described by:

$$\begin{aligned}
D_I &= A_I \cos\left(\frac{2\pi}{\lambda_1}x - 2\pi f t\right)\\
&= A_I \cos\left(2\pi\left(\frac{1}{\lambda_1}x - ft\right)\right)\\
&= A_I \cos\left(2\pi f\left(\frac{x}{v_1} - t\right)\right)\\
&= A_I \cos\left(-2\pi f\left(t - \frac{x}{v_1}\right)\right)\\
D_I &= A_I \cos\left(2\pi f\left(t - \frac{x}{v_1}\right)\right)
\end{aligned}$$

where we used $v = f\lambda$, and noted that $\cos(-x) = \cos(x)$.

The transmitted wave (denoted by the subscript T) will also travel in the positive x

direction, but its speed will be v_2, since it travels through the second rope:

$$D_T = A_T \cos\left(2\pi f\left(t - \frac{x}{v_2}\right)\right)$$

The reflected wave (denoted by R) will travel in the $-x$ direction and at the same speed as the incident pulse.

$$D_R = A_R \cos\left(2\pi f\left(t + \frac{x}{v_1}\right)\right)$$

(b) We will consider the boundary conditions at the interface between the two ropes. One boundary condition is that the rope must be continuous. As a result, the vertical displacement on the $-x$ side of the boundary must be the same as the vertical displacement on the $+x$ side of the boundary at every instant:

$$D_{-x} = D_{+x} \quad \text{at } x = 0$$

The amplitude on the $+x$ side is equal to the amplitude of the transmitted pulse. For the $-x$ side of the boundary, we have to take into account that the incident and reflected pulses will superimpose (when the front of the incident pulse reaches the boundary, it will be reflected and interfere with the end of the incident pulse). This boundary condition can thus be expressed as:

$$A_I + A_R = A_T$$

The slope of the rope must also be continuous at the boundary. Since the incident and reflected pulses superimpose, and the principle of superposition states that the net displacement is the sum of the displacement of these two waves, we can write:

$$\begin{aligned}
\frac{\partial}{\partial x}(D_I + D_R)\bigg|_{x=0} &= \frac{\partial}{\partial x}D_T\bigg|_{x=0} \\
\frac{\partial}{\partial x}D_I\bigg|_{x=0} + \frac{\partial}{\partial x}D_R\bigg|_{x=0} &= \frac{\partial}{\partial x}D_T\bigg|_{x=0}
\end{aligned}$$

Using our equations for the incident, transmitted, and reflected pulses found in part a), and taking the appropriate partial derivatives, this equation becomes:

$$\begin{aligned}
(A_I/v_1)\sin\left(2\pi f\left(t - \frac{x}{v_1}\right)\right)\bigg|_{x=0} + (-A_R/v_1)\sin\left(2\pi f\left(t + \frac{x}{v_1}\right)\right)\bigg|_{x=0} &= \\
(A_T/v_2)\sin\left(2\pi f\left(t - \frac{x}{v_2}\right)\right)\bigg|_{x=0}
\end{aligned}$$

Evaluating at $x = 0$ gives:

$$\begin{aligned}
(A_I/v_1)\sin(2\pi ft) + (-A_R/v_1)\sin(2\pi ft) &= (A_T/v_2)\sin(2\pi ft) \\
\frac{A_I}{v_1} - \frac{A_R}{v_1} &= \frac{A_T}{v_2}
\end{aligned}$$

Using our first condition, $A_I + A_R = A_T$, we get:

$$\frac{A_I}{v_1} - \frac{A_R}{v_1} = \frac{A_I}{v_2} + \frac{A_R}{v_2}$$

Now, we can rearrange to find the reflection coefficient, $R = A_R/A_I$:

$$A_I\left(\frac{v_2 - v_1}{v_1 v_2}\right) = A_R\left(\frac{v_2 + v_1}{v_1 v_2}\right)$$
$$R = \frac{v_2 - v_1}{v_2 + v_1}$$

Since the velocities in the first and second rope are $v_1 = \sqrt{F_T/\mu_1}$ and $v_2 = \sqrt{F_T/\mu_2}$, respectively, the reflection coefficient can be written as:

$$R = \frac{\sqrt{\frac{F_T}{\mu_2}} - \sqrt{\frac{F_T}{\mu_1}}}{\sqrt{\frac{F_T}{\mu_2}} + \sqrt{\frac{F_T}{\mu_1}}}$$
$$= \frac{\sqrt{F_T}}{\sqrt{F_T}} \cdot \frac{\frac{1}{\sqrt{\mu_2}} - \frac{1}{\sqrt{\mu_1}}}{\frac{1}{\sqrt{\mu_2}} + \frac{1}{\sqrt{\mu_1}}}$$
$$\therefore R = \frac{\sqrt{\mu_1} - \sqrt{\mu_2}}{\sqrt{\mu_1} + \sqrt{\mu_2}}$$

as desired.

15 Fluid mechanics

In this chapter, we introduce the tools required to model the dynamics of fluids. This will allow us to model how objects can float, how water flows through a pipe, and how airplane wings create lift. We will start by introducing the concept of pressure and modelling static fluids (hydrostatics) before developing models for fluids that flow (hydrodynamics). Fluids are generally defined as the phase of matter in which atoms (or molecules) are only loosely bound to each other, such as in gases or liquids. Most of the formalism that we develop will apply to any fluid (gas, liquid, plasma), although we will often restrict ourselves to modelling the most simple situations (e.g. laminar flow of an incompressible liquid).

Learning Objectives

- Understand the concept of pressure, and how pressure is modelled in a fluid.
- Understand how to model the pressure gradient due to gravity.
- Understand Pascal's Principle and how to model hydraulic lifts and pressure sensing devices.
- Understand how a pressure gradient leads to a force of buoyancy.
- Understand the difference between laminar and turbulent flow.
- Understand the equation of continuity, and the concepts of mass and volumetric flow.
- Understand how to apply Bernoulli's Principle to model the speed and pressure within a flowing fluid.
- Understand how to model the resistance to flow in a pipe using the viscosity of a fluid.

Think About It

You are sailing, and the wind is blowing from the north. You want to travel upwind (north). In what direction should you point your boat/sail?

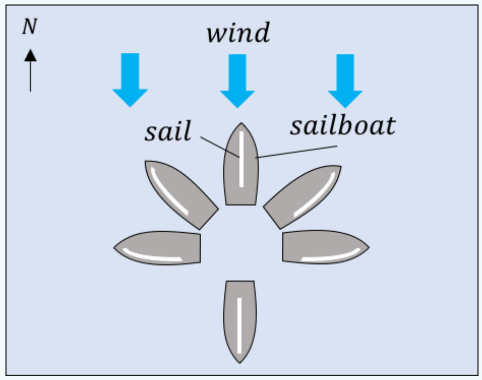


Figure 15.1: Possible directions you can point your sailboat.

A) North
B) South
C) Point either East or West
D) Alternate between North-east and North-west
E) You cannot go upwind.

15.1 Pressure

The pressure exerted by a force, $\vec{F}$, over a surface with area, A, is a scalar quantity, P, defined as:

$$P = \frac{F_{\perp}}{A}$$

where $F_{\perp}$ is the component of the force perpendicular to the surface. The SI unit for pressure is the Pascal (Pa). Pressure is related to the area, A, over which a force is exerted, and can be thought of as a measure of how concentrated that force is. For example, a force of $10\,\text{N}$ exerted through a needle (a small area) will result in a much larger pressure than if that force was exerted by a flat hand (a larger area).

When a force is exerted on a fluid, it creates pressure that we model as being **everywhere in the fluid**. For each element in the fluid, the pressure from the surrounding fluid exerts an inwards force on the element from **all directions** (see Figure 15.2). In reaction, the element exerts an outwards force in all directions, and these forces act on neighbouring elements.

This is somewhat analogous to the tension that exists everywhere in a rope, where each element of the rope experiences forces from the neighbouring elements in rope that try to "pull it apart". Pressure can be thought of as a "negative" tension, in that the material under pressure is experiencing forces trying to collapse the element onto itself, rather than trying to pull it apart. To create a tension in a rope, one would exert an outwards force on the rope (in order to stretch it), so that the rope exerts an inwards force in reaction. In order to create pressure in a fluid, one must exert an inwards force on the fluid, which then exerts an outwards force in reaction.

If we consider a small cubic volume of fluid, as depicted in the centre of Figure 15.2, that element of fluid will experience inwards forces in all directions from the pressure in the surrounding fluid, as illustrated by the arrows. If the forces from the pressure result in no net force on the fluid element, then we say that the fluid is in hydrostatic equilibrium, and the element of fluid will be at rest in an inertial frame of reference.

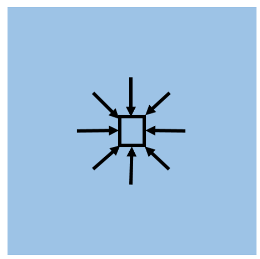


Figure 15.2: A small element inside of a fluid with pressure will experience no net force from the pressure in the fluid, since the force associated with the pressure in the fluid is exerted in all directions.

Consider, instead, an element of fluid that at the edge of a container for the fluid (e.g. a cup of water), as depicted in Figure 15.3.

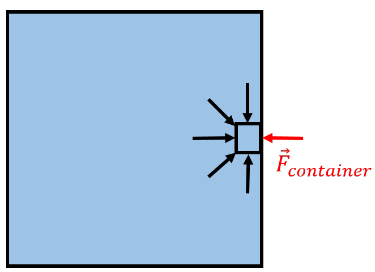


Figure 15.3: At the edge of a container, a small element of fluid will exert an outwards force on the container, and the container will exert an inwards force on the element of fluid.

In this case, there is no fluid on the right-hand side of the fluid element to exert a force towards the left. If the fluid element is in equilibrium, it must then be the container that exerts that force, $\vec{F}_{container}$, on the fluid. By Newton's Third Law, the element of fluid exerts an outwards force on the container. This is true at all points on the surface of container, which will all experience an outwards force from the pressure of the fluid. If the pressure is constant over a surface, the magnitude of the outwards force on the surface will be equal to the pressure of the fluid multiplied by the area of that surface.

If you place an empty sealed tin can under water, the water will exert a pressure on all of the surfaces of the tin can that leads to a net inwards force on all surfaces of the tin can.

If the water pressure is high enough, the tin can will get crushed. If, on the other hand, the tin can is allowed to fill with water, it will not get crushed, as the water inside the tin can will have the same pressure as the water outside the tin can and will exert an equal net outwards force on all surfaces of the tin can. The net force on each surface of the can will be zero, and the tin can will not get crushed, no matter how high the water pressure is.

In general, if there is an interface with fluid on either side of it a different pressures, it is the **difference in pressure** on either side of the interface that determines the net force exerted on the interface, rather than the absolute pressure.

Checkpoint 15-1

You place a tin can on a table, and use a pump to create a vacuum inside of the can. You observe that the tin can gets crushed. Which explanation is correct?

A) By sucking the air out of the can, you also suck in on the walls of the can.
B) You lower the pressure inside the can so that the air outside the can exerts a larger inwards force on the can than the outwards force from the air inside the can.
C) You lower the pressure inside the can so that the air inside the can exerts a pulling force on the walls of the can.
D) All of the above are all valid ways to model this.

15.1.1 The effect of gravity

When discussing Figure 15.2, we argued that the fluid exerts an equal force, from all directions, on the fluid element, so that the net force on the fluid element is zero. This is not quite correct in the presence of gravity, where the fluid element will have a weight. Thus, if the fluid element is to be in equilibrium, the upwards force (and pressure) from the fluid below must be higher than that from the fluid above the fluid element.

Figure 15.4 shows an element of fluid that has a height h and a surface area A in the horizontal plane. The pressure, P_2, in the fluid below the fluid element must be higher than the pressure, P_1, above the fluid element, if the fluid element is in equilibrium.

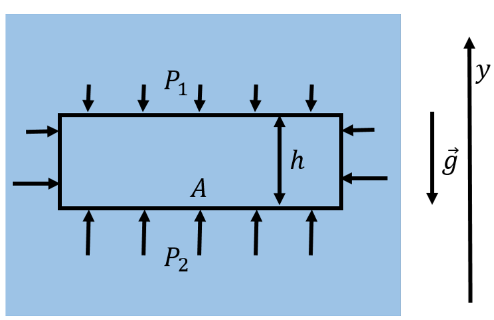


Figure 15.4: In the presence of gravity, the pressure below an element of fluid must be larger if the fluid element is to remain in equilibrium.

The element of fluid has a total mass, m, given by:

$$m = \rho V = \rho A h$$

where, $V = Ah$, is the volume of the fluid, and, ρ, its density.

The net (horizontal) force exerted by the external fluid on the fluid element is zero along the vertical surfaces. Let P_1 be the pressure in the fluid above the fluid element, and P_2 be the pressure below the fluid element. If we choose a y axis that is positive upwards and the fluid element does not accelerate in the vertical direction, then the y component of Newton's Second Law, written for the fluid element, is:

$$\begin{aligned}
\sum F_y = F_2 - F_1 - mg &= 0 \\
P_2 A - P_1 A - mg &= 0 \\
P_2 A - P_1 A - \rho A h g &= 0 \\
\therefore P_2 - P_1 &= \rho g h
\end{aligned}$$

where we used the fact that the force resulting from a pressure is given by the pressure multiplied by the area over which it is exerted. We thus find that the difference in pressure due to gravity in a fluid between two positions, y_2 and y_1, is given by:

$$\boxed{P(y_2) - P(y_1) = -\rho g (y_2 - y_1)} \tag{15.1}$$

where the y axis is defined to increase in the upwards direction. Since the pressure in the fluid depends on the location in the fluid, we say that there is a "pressure gradient" in the fluid.

Checkpoint 15-2

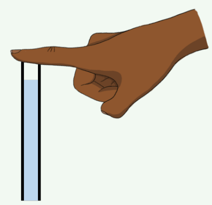


Figure 15.5: Holding water in a vertical straw.

You use your finger to block off the top end of a straw and then remove the straw from a glass of water. What is the most correct description of why the water stays in the straw (Figure 15.5) before you release your finger?

A) The straw cannot have vacuum inside of it; unless the finger is removed to let air in to replace the water, the water will remain in the straw.
B) There is a small amount of vacuum above the water that sucks the water upwards and prevents it from dropping.
C) The pressure of the air in the straw below the water is higher than the pressure of the air in the straw above the water.
D) The pressure of the air in the straw below the water is lower than the pressure of the air in the straw above the water.

We have assumed that the density of the fluid, ρ, is constant, and that the fluid cannot be compressed. This is a very good approximation for a liquid such as water, but not for a gas, whose density will depend on its pressure. If the fluid were a gas (e.g. a column of air in our atmosphere), both the density and the pressure will change as a function of height. We can easily take this into account in our model, if we consider the fluid element to have a very small height, dy, instead of the finite height, h, as in the derivation above. A fluid element with an infinitesimal height, dy, is illustrated in Figure 15.6.

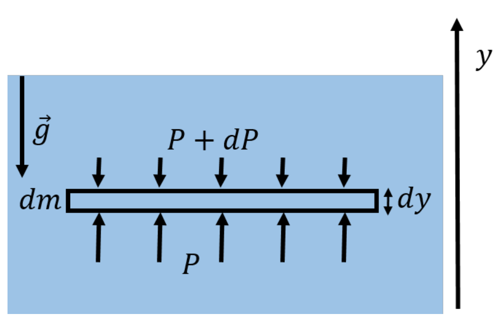


Figure 15.6: Pressure gradient from gravity on an infinitesimal fluid element.

In the very small height, dy, the density of the fluid, ρ, can be taken to be constant, and the infinitesimal element of fluid will have a mass dm:

$$dm = \rho A dy$$

We can model the pressure exerted by the fluid above the fluid element as $P + dP$, and the pressure exerted by the fluid below as P, where dP is a small (negative) change in pressure[1]. The y component of Newton's Second Law written for the infinitesimal fluid element is thus:

$$\begin{aligned}\sum F_y &= PA - (P + dP)A - dmg = 0\\ PA - PA - dPA - \rho A dy g &= 0\\ \therefore -dP - \rho g dy &= 0\end{aligned}$$

We can thus determine how pressure changes with height, y:

$$\boxed{\frac{dP}{dy} = -\rho g} \tag{15.2}$$

This tells us that the rate of change of pressure with increasing y is negative; in other words, the pressure decreases as the elevation increases, as we had already concluded. We can integrate the equation to obtain the change in pressure in going from y_1 to y_2:

$$\begin{aligned}dP &= -\rho g dy\\ \int_{P_1}^{P_2} dp &= -\int_{y_1}^{y_2} \rho g dy\\ \therefore P_2 - P_1 &= -\int_{y_1}^{y_2} \rho g dy\end{aligned}$$

If the density, ρ, is constant, then this leads to Equation 15.1. Note that, thus far, we have only modelled how pressure in a fluid changes with height, but we have not determined the absolute pressure in a fluid.

Example 15-1

If we assume that the density of air is proportional to its pressure, how does the density of air change with altitude?

Solution

We know that the rate of change of pressure with altitude (position y, where positive y is defined to be upwards) is given by:

$$\frac{dP}{dy} = -\rho g$$

Since we can assume that the density is proportional to the pressure, we can introduce

[1] We placed the dP on the top part of the fluid, even though the pressure is higher on the bottom part of the fluid, because the y axis increases upwards. We are really interested in the change in pressure, dP, that corresponds to a change in height, dy, along the positive y direction.

an arbitrary constant, a, and state that:

$$\begin{aligned}\rho &= aP\\ \therefore \frac{dP}{dy} &= \frac{d}{dy}\frac{1}{a}\rho = \frac{1}{a}\frac{d\rho}{dy}\end{aligned}$$

where the constant a can be evaluated if we know the pressure and density at some point. We can thus write that the rate of change of the density with position y is given by:

$$\begin{aligned}\frac{1}{a}\frac{d\rho}{dy} &= -\rho g\\ \therefore \frac{d\rho}{dy} &= -ag\rho\end{aligned}$$

This is a separable differential equation for ρ, allowing us to separate the variables and integrate from, say, an altitude of $y = 0$, where the density is ρ_0, to an altitude y, where the density is ρ:

$$\begin{aligned}\frac{d\rho}{\rho} &= -agdy\\ \int_{\rho_0}^{\rho}\frac{d\rho}{\rho} &= -\int_0^y agdy\\ \ln(\rho) - \ln(\rho_0) &= -agy\\ \ln\left(\frac{\rho}{\rho_0}\right) &= -agy\end{aligned}$$

We can take the exponential on each side of the equation to get rid of the logarithm:

$$\begin{aligned}\frac{\rho}{\rho_0} &= e^{-agy}\\ \therefore \rho(y) &= \rho_0 e^{-agy}\end{aligned}$$

We thus find that the density of the air decreases exponentially with altitude. This is why it is more difficult to breathe at high altitude. Since we assumed that the density of the air is proportional to its pressure, the air pressure will also decrease exponentially with increasing altitude:

$$P(y) = P_0 e^{-agy}$$

where P_0 is the pressure at an altitude of $y = 0$. If we know P_0 and ρ_0, then the constant a is given by:

$$a = \frac{\rho_0}{P_0}$$

Discussion: If we applied this model to the Earth's atmosphere, our model would only provide qualitative agreement, as the density of the air also depends on its temperature and other factors. Nonetheless, it is interesting that, based on the simple requirement that an element of air be in hydrostatic equilibrium, we are able to obtain a reasonable description of how pressure and density change with altitude in the Earth's atmosphere.

15.1.2 Pascal's Principle

Pascal's Principle states that **if an external pressure is exerted on a fluid, the pressure everywhere in the fluid increases by that amount**. For example, if a fluid is contained in a piston with a cross-section area, A, and a force, F, is exerted on the piston (Figure 15.7), then the pressure everywhere in the fluid increase by F/A.

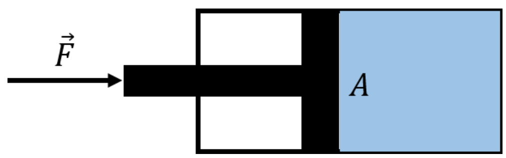


Figure 15.7: A force exerted on the piston will increase the pressure everywhere in the fluid.

If we wish to determine the absolute pressure in the water at some depth, h, in the ocean, we need to include the fact that the Earth's atmosphere exerts a net downwards force on the surface of the ocean in addition to the fact that the pressure changes with depth due to gravity. The pressure from the air in the Earth's atmosphere is called "atmospheric pressure", and depends on a variety of conditions, such as the weather. The average pressure from the atmosphere is $P_0 = 1.013 \times 10^5$ Pa. If the atmospheric pressure is P_0 at the surface of the ocean, then the pressure at some depth, h, is given by:

$$P(h) = P_0 + \rho g h$$

where ρ is the density of water. As a consequence, the pressure at any depth, h, in a fluid is the same everywhere at that depth in the fluid.

Checkpoint 15-3

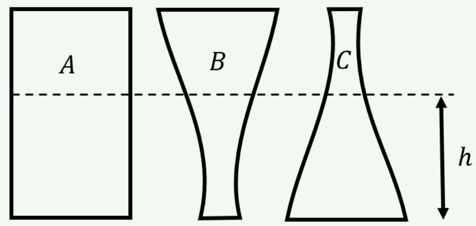


Figure 15.8: Three glasses with different shapes.

You fill the three glasses in Figure 15.8 such that the liquid reaches a height h above the bottom of the glass. What can you say about the pressure of the liquid at the bottom of each glass?

A) It is highest for glass A.
B) It is highest for glass B.
C) It is highest for glass C.
D) It is the same for all glasses.
E) It is only the same for all glasses if we can neglect atmospheric pressure.

Example 15-2

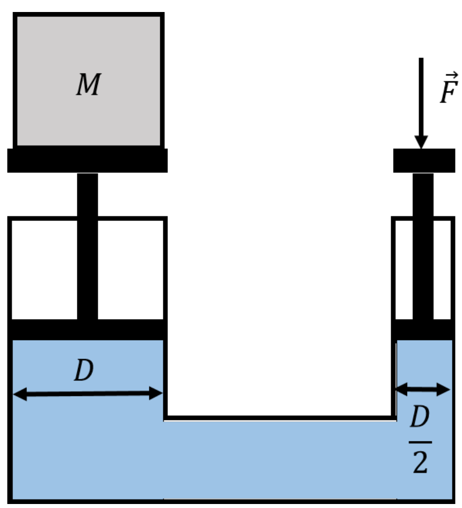


Figure 15.9: A force exerted on the piston of a hydraulic lift in order to lift a mass M.

A hydraulic lift exploits Pascal's principle in order to use a small force to exert a large force. The hydraulic lift in Figure 15.9 shows a lift that is constructed by having a fluid between two vertical movable pistons. The pistons are cylindrical and the diameter of their cross-sections are D and $D/2$. A mass, M, is placed on the piston with the larger diameter. What is the magnitude of the force, $\vec{F}$, that must be applied on the smaller piston in order to lift the mass, M?

Solution

If a force $\vec{F}$ is applied to the small piston, then the pressure in the fluid will increase by:

$$\Delta P = \frac{F}{A} = \frac{F}{\pi\frac{D^2}{4}} = \frac{4F}{\pi D^2}$$

This will result in a net upwards force, $\vec{F}'$, on the large piston, with a magnitude:

$$F' = \Delta P A' = \Delta P \pi D^2 = \frac{4F}{\pi D^2}\pi D^2 = 4F$$

Thus the force on the large piston will be four times that exerted on the small piston. One only needs to exert a force with a magnitude of $Mg/4$ in order to lift the mass, M.

15.1.3 Measuring pressure

In this section, we describe how one can design instruments to measure pressure. The most straightforward device is a manometer, which is constructed using a U-shaped tube filled with a fluid of known density, ρ, as shown in Figure 15.10.

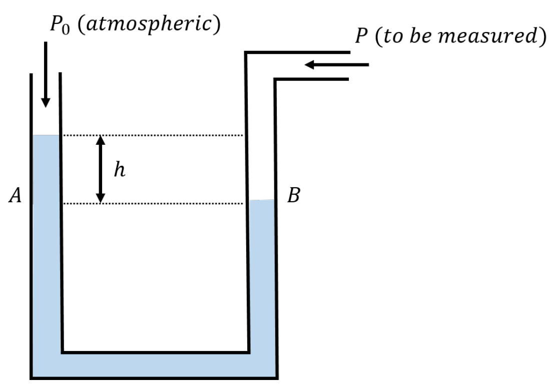


Figure 15.10: A manometer can measure the difference between a pressure P and atmospheric pressure, P_0. That difference is called "gauge pressure".

A manometer can be used to measure a pressure P relative to atmospheric pressure, P_0. One end of the tube is open to atmospheric pressure, and the other is connected to the fluid (e.g. a gas) for which we want to measure the pressure. If the pressure being measured is larger than atmospheric pressure, the fluid in the manometer will experience a greater downwards force on the side of the pressure to be measured than on the side open to atmospheric pressure, as shown in Figure 15.10. There will be a difference, h, in the level of the fluid on each side of the tube, which is directly proportional to the difference in pressure between the two sides of the tube.

Consider the point in the fluid at location B in Figure 15.10, where the pressure is $P_B = P$, the pressure to be measured. The point in the fluid at location A, which is at the same height in the fluid, must have the same pressure as point B. We can write the pressure at point A, P_A, as the sum of the atmospheric pressure and the pressure from the column of water of height, h:

$$P_A = P_0 + \rho g h$$

Since this must also be equal to the pressure at point B, we can find the difference between the pressure we want to measure and atmospheric pressure:

$$\begin{aligned} P_A &= P_B \\ P_0 + \rho g h &= P \\ \therefore P - P_0 &= \rho g h \end{aligned}$$

The difference between a pressure and the atmospheric pressure is called “gauge pressure”, and is all that we can measure if we do not know the absolute value of the atmospheric pressure. Using a manometer, the gauge pressure is given by $\rho g h$, whereas the “absolute pressure”, P, is given by adding the atmospheric pressure to the gauge pressure, $P = P_0 + \rho g h$. Most pressure measuring devices (“pressure gauges”), measure pressure relative to atmospheric pressure, using a similar mechanism.

The atmospheric pressure at a location on Earth varies based on the weather. A barometer is an instrument designed to measure the atmospheric pressure. A simple barometer can be built from a manometer, with one end closed, as illustrated in Figure 15.11.

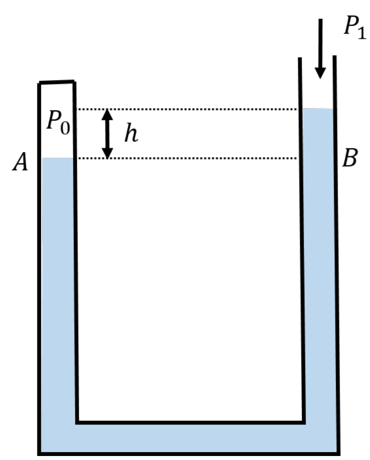


Figure 15.11: A barometer constructed from a manometer to measure relative changes in atmospheric pressure.

One end of the manometer is sealed on a day where the atmospheric pressure is, say, P_0, while the other end of the tube is left open. The height difference, h, between the fluid in either side of the tube is a measure of how different the current atmospheric pressure, P_1, is relative to the pressure, P_0, when the manometer was sealed. In Figure 15.11, the barometer is shown on a day where the atmospheric pressure is lower than on the day the manometer was sealed. The difference in pressure is given by:

$$P_1 = P_0 + \rho g h$$

if we define h to be positive when the side with the pressure P_0 is higher (so h is negative in Figure 15.11 and P_1 is less than P_0).

We can also measure the absolute atmospheric pressure if we evacuate the air out of the sealed end of the tube, so that $P_0 = 0$. When doing so, the difference in height between the fluid on either side of the manometer is a measure of the absolute atmospheric pressure.

Example 15-3

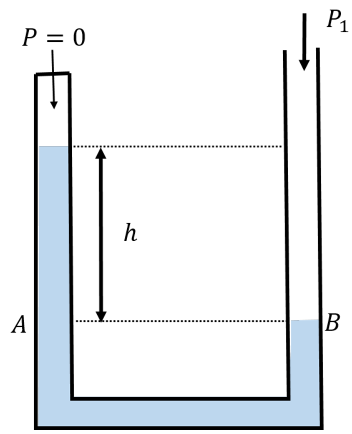


Figure 15.12: A barometer constructed from a manometer to measure absolute atmospheric pressure.

Using a manometer filled with water ($\rho = 1 \times 10^3\,\text{kg/m}^3$), you construct a barometer to measure the absolute atmospheric pressure by evacuating the air from one side of the manometer, as shown in Figure 15.12. What is the difference in height, h, when the atmospheric pressure is "nominal", $P_1 = 1.013 \times 10^5\,\text{Pa}$?

Solution

The pressure, P_1, on the open side of the manometer is given by:

$$P_B = P_A$$
$$P_1 = P_0 + \rho g h = \rho g h$$

if the sealed side of the manometer has a pressure, $P_0 = 0$, above the fluid. If $P_1 =$

1.013×10^5 Pa, we can find the height, h:

$$h = \frac{P_1}{\rho g} = \frac{(1.013 \times 10^5\,\text{Pa})}{(1000\,\text{kg/m}^3)(9.8\,\text{m/s}^2)} = 10.3\,\text{m}$$

Discussion: The difference in height is about 10 m when the atmospheric pressure is nominal. This means that the manometer needs to be at least this tall to measure absolute atmospheric pressure, which is not practical to construct! If, instead, one uses a liquid with a higher density than that of water, then this height can be reduced substantially. Traditionally, barometers have been built using mercury, which has a density of ($\rho_{Hg} = 13.6 \times 10^3\,\text{kg/m}^3$), so that the height difference at nominal atmospheric pressure is 760 mm. This is a much easier instrument to build (apart from the safety concerns of using mercury). For this reason, an often-used unit of pressure is "mm of mercury", which corresponds to the height difference in a manometer that is built using mercury.

Checkpoint 15-4

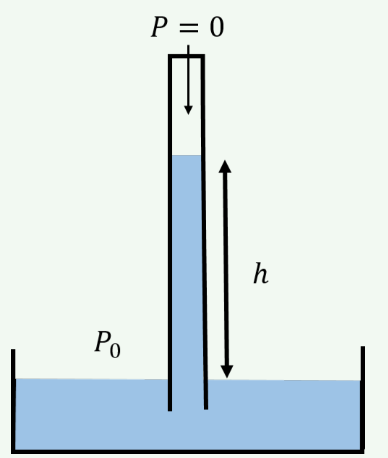


Figure 15.13: A Torricelli barometer.

You build a Torricelli barometer, as illustrated in Figure 15.13, to measure the absolute atmospheric pressure. The sealed vertical tube has a space at the top that is evacuated (a pressure of zero), so that atmospheric pressure on the container of liquid forces the liquid up the tube to a height, h, which is proportional to atmospheric pressure. If you use olive oil as the liquid, what can you say about the height, h, for nominal atmospheric pressure?

A) It is greater than 10.3 m.
B) It is equal to 10.3 m.
C) It is less than 10.3 m.
D) Not enough information to tell.

15.2 Buoyancy

In this section, we examine how the pressure gradient in a fluid leads to a force of buoyancy on an object that is immersed in the fluid.

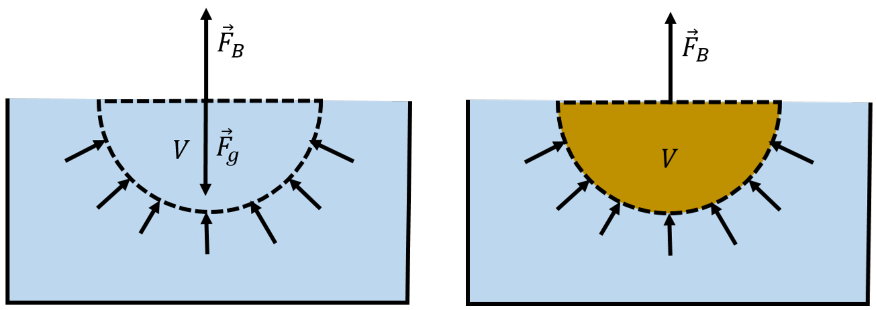


Figure 15.14: (Left:) The weight of a fluid element, $\vec{F}_g$, is supported by the net upwards force from the pressure, $\vec{F}_B$, of the fluid below it. (Right:) If the fluid element is removed and replaced with an object, there will still be the same net upwards force, $\vec{F}_B$, from the pressure of the fluid, which is now exerted on the object.

In the left panel of Figure 15.14, we show a hemi-spherical element of fluid with a volume V. The weight of the element of fluid, $\vec{F}_g$, is supported by the net upwards force, $\vec{F}_B$, exerted by the pressure of the fluid surrounding the fluid element. The mass, M, of the element of fluid is given by:

$$M = \rho V$$

where ρ is the density of the fluid. The net force from the pressure, F_B, must thus have the same magnitude as the weight:

$$F_B = Mg = \rho V g$$

Now, suppose that the fluid element is "displaced" and replaced by the hull of a boat, as shown in the right panel of Figure 15.14. The net upwards force from the pressure of the fluid must remain the same, F_B, but that force is now exerted on the hull of the boat. We call that force the force of "buoyancy", which is the reason that a boat can float and the reason that you feel lighter when walking in a swimming pool than on land.

Thus, if an object displaces a volume, V, of a fluid with density ρ, when immersed in the fluid, that object will experience an upwards force of buoyancy, $\vec{F}_B$, with magnitude:

$$\boxed{F_B = \rho V g} \tag{15.3}$$

This "principle" was originally discovered by Archimedes, who stated that the force of buoyancy is equal to the weight of the displaced fluid. Note that we drew the fluid element at the surface of the fluid, but this is not required, and a force of buoyancy will be present if the object is completely immersed in the liquid. If you refer back to Figure 15.4, you will recall that the net upwards force on an element of fluid must be equal to its weight, even if the fluid element is completely immersed.

Checkpoint 15-5

Does the force of buoyancy on a fully submerged object increase with the depth at which the object is submerged (ignoring any change from the varying value of $\vec{g}$)?

A) Yes, because the force of buoyancy comes from the pressure in the fluid, which increases with depth.
B) No, because the force of buoyancy comes from the difference in pressure above and below the object, which does not increase with depth.

Checkpoint 15-6

You observe that if you pour olive oil slowly into your glass of water, the oil floats above the water. What can you conclude?

A) The mass of a given volume of oil is less than the mass of the same volume of water.
B) The mass of a given volume of oil is more than the mass of the same volume of water.

Example 15-4

You measure the weight of an object by suspending it with a spring scale. When you measure the weight of the object in air, you find that it has a weight W_a. When you measure the weight of the object when it is completely submerged in water, you find that it has a weight W_w. What is the density of the object?

Solution

Given the weight of the object in air, we can easily determine its mass:

$$M = \frac{W_a}{g}$$

However, since we do not know its volume, V, we cannot directly determine its density. When the object is submerged in water, the measured weight will be the actual weight of the object (as measured in air) minus the magnitude of the force of buoyancy exerted by the water:

$$\begin{aligned} W_w &= W_a - \rho_w g V \\ \therefore V &= \frac{W_w - W_a}{\rho_w g} \end{aligned}$$

where ρ_w is the density of water. Given the volume, we can now determine the object's density, ρ:

$$\rho = \frac{M}{V} = \frac{W_a \rho_w g}{g(W_w - W_a)} = \rho_w \frac{W_a}{W_w - W_a}$$

Discussion: By using Archimedes' Principle, we were able to determine the volume, and thus the density of the object, by comparing measurements of its weight in air and in water. This is similar to the method that Archimedes came up with to determine if a crown owned by a general was made of real gold or if some of the gold had been replaced with an equal weight of silver. Archimedes supposedly went to the baths to ponder how to determine if the crown was made of gold and had his Eureka moment when we he noticed the water level in the bath went up as he went into the bath. He realized that denser gold would displace less water than silver for an equal weight.

Olivia's Thoughts

Whether or not an object will float depends on its density. Let's consider an object that is placed in water. The only forces acting on the object are its weight and the force of buoyancy. We want to know when the net force will be zero. I'm going to write out Newton's Second Law for the object, but writing the mass of the object in terms of its density and volume.

$$\begin{aligned} F_g &= F_B \\ m_O g &= F_B \\ \rho_O V_O g &= \rho_W V_W g \end{aligned}$$

where O refers to the object and W refers to the water. Cancelling out the g's, we can write this as:

$$\frac{\rho_O}{\rho_W} = \frac{V_W}{V_0}$$

Consider a solid cube that has the same density as water. In this case, $\rho_0/\rho_W = 1$, and so $V_W/V_0 = 1$. This means that, in order for the cube to float, a volume of water that is equal to the volume of the cube must be displaced. So, the entire cube must be submerged. If you placed this cube 5 m deep in the water, it would stay at this depth.

Now consider a cube whose density is half that of water. We find that $\rho_0/\rho_W = 0.5$, so we must have $V_W/V_0 = 0.5$. In order for the cube to float, only half of it needs to be submerged. If you placed the cube 5 m deep in the water, it would rise to the surface and stop when half of it was above water (after bobbing for a bit).

Finally, what about a cube whose density is 1.5 times the density of water? In this case, one and a half cubes worth of water would have to be displaced in order for the cube to float. Even when the entire cube is submerged, not enough volume has been displaced in order for it to float, so the cube will sink.

Objects like pool noodles or life jackets allow us to float because they have low densities. They have very little mass (they don't add much to the weight) in a relatively large volume (they can displace water to add to the buoyant force). An object with a density less than water will float with some fraction of the object being submerged.

15.3 Hydrodynamics

In the previous sections we developed "hydrostatic" models for fluids when those fluids are at rest (in some inertial reference frame). In this section, we develop "hydrodynamic" models to discuss what happens when fluids flow. We will restrict our models to fluids that flow in a "laminar" fashion, rather than a "turbulent" fashion.

Laminar flow is the flow of a fluid when each particle in the fluid follows a path that can be represented by a line (a "streamline"). Turbulent flow is the flow of a fluid where particles can follow rather complex paths, usually involving "Eddy currents" (little whirlpools). The two types of flow are illustrated in Figure 15.15.

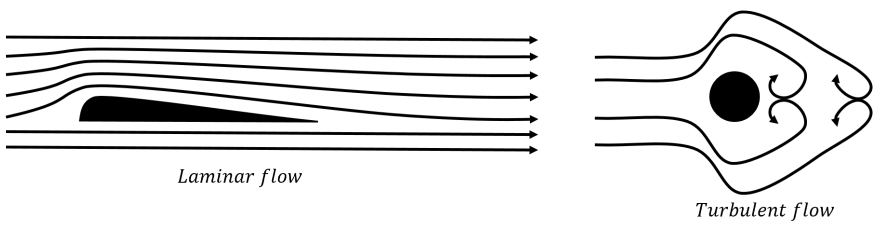


Figure 15.15: Laminar (left) and turbulent (right) flow of a fluid around an object.

15.3.1 Continuity of flow

Consider the laminar flow of a fluid through a pipe whose cross-sectional area narrows from A_1 to A_2 in the direction of flow, as illustrated in Figure 15.16.

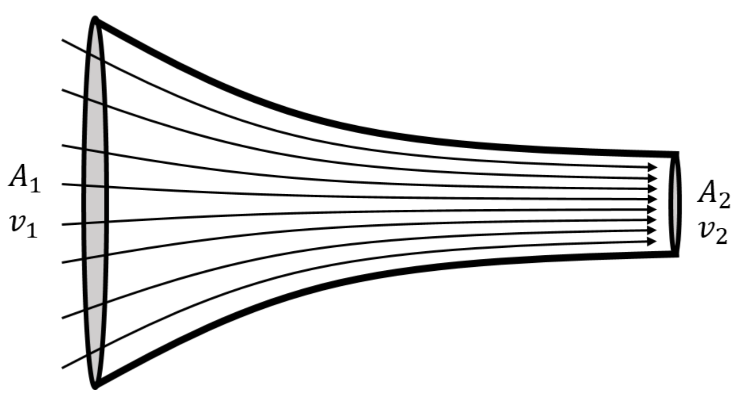


Figure 15.16: Laminar flow of a fluid in a narrowing pipe.

The particles that make up the fluid have a speed v_1 at the wide end of the pipe and speed

v_2 at the narrow end. The **equation of continuity** is based on the premise that the fluid that enters the pipe must exit the pipe, as there is nowhere else for the fluid to go. That is, if during a period of time, Δt, a mass, Δm, of fluid enters the wide end of the pipe, then during that same period of time, the same mass of fluid must exit the narrow end of the pipe.

During a period of time, Δt, the fluid at the wide end of the pipe will travel a distance $l_1 = v_1 \Delta t$. Thus, a volume of fluid, ΔV_1, will enter the wide end of the pipe:

$$\Delta V_1 = A_1 l_1 = A_1 v_1 \Delta t$$

Similarly, during that period of time, a volume ΔV_2 will exit the narrow end of the pipe:

$$\Delta V_2 = A_2 l_2 = A_2 v_2 \Delta t$$

If the fluid is compressible, its density can change. Let ρ_1 be the density of the fluid at the wide end of the pipe and ρ_2 be the density of the fluid at the narrow end. The mass of fluid, Δm, entering the wide end of the pipe is given by:

$$\Delta m = \rho_1 \Delta V_1 = \rho_1 A_1 v_1 \Delta t$$

The mass of fluid exiting the narrow end of the pipe is given by:

$$\Delta m = \rho_2 \Delta V_2 = \rho_2 A_2 v_2 \Delta t$$

The mass of fluid entering the wide end of the pipe must equal the mass exiting the narrow end of the pipe:

$$\rho_1 A_1 v_1 \Delta t = \rho_2 A_2 v_2 \Delta t$$

Leading to the equation of continuity:

$$\boxed{\rho_1 A_1 v_1 = \rho_2 A_2 v_2} \tag{15.4}$$

The quantity $\rho A v$ has dimensions of mass per time, and corresponds to the mass of fluid passing through a cross section A per unit time.

If the fluid is incompressible, as are most liquids, then the density is the same on both sides of the pipe, and the equation simplifies to:

$$\boxed{A_1 v_1 = A_2 v_2} \quad \text{(Incompressible fluid)} \tag{15.5}$$

For a liquid, we can define the "volumetric flow", Q, as:

$$Q = Av$$

where A is the cross-sectional area of the surface through which a fluid with speed, v, flows[2]. Q has the dimension of volume per time, and corresponds to the volume of fluid passing through the cross section A per unit time. For an incompressible fluid, the equation of continuity is thus equivalent to stating that the volumetric flow, Q, of the fluid is a constant.

[2] If the velocity of the fluid is not perpendicular to the surface, then v is the component of the velocity perpendicular to the surface.

Checkpoint 15-7

Figure 15.17: Water flowing out of a faucet.

When water flows out of your faucet, you observe that the stream of water gets narrower as the water moves down, as shown in Figure 15.17. Why is this?

A) The atmospheric pressure increases as the waters moves downwards, so the stream of water is more and more compressed.
B) As the water accelerates due to gravity, the cross-sectional area of the flowing water must reduce in order to preserve a constant flow rate.

Example 15-5

Your garden hose has a diameter of $D = 2\,\text{cm}$. How fast must water flow out of the hose if you are to fill a $5\,\text{l}$ bucket in one minute?

Solution

We need the volume flow rate from the hose to be $Q = 5\,\text{l/min}$. We can convert this to SI units:

$$Q = (5\,\text{l/min})\left(\frac{1}{1000}\text{m}^3/\text{l}\right)\left(\frac{1}{60}\text{min/s}\right) = \frac{5}{6\times 10^4}\text{m}^3/\text{s} = 8.3\times 10^{-5}\,\text{m}^3/\text{s}$$

Since we know the area of the hose, we can determine the speed of the water to achieve the given flow rate:

$$\begin{aligned}Q &= Av = \pi\left(\frac{D}{2}\right)^2 v\\ \therefore v &= \frac{Q}{\pi\left(\frac{D}{2}\right)^2} = \frac{(8.3\times 10^{-5}\,\text{m}^3/\text{s})}{\pi(0.01\,\text{m})^2} = 0.265\,\text{m/s}\end{aligned}$$

15.3.2 Bernoulli's Principle

In this section, we examine how the pressure and speed of a fluid change as a the fluid flows. We will restrict ourselves to discussing the **laminar** flow of an **incompressible** fluid with no friction. Bernoulli was the first to quantitatively describe the flow of incompressible fluids, and we will show in this section how to derive "Bernoulli's Principle".

Consider the laminar flow of an incompressible fluid through a pipe that changes height, from y_1 to y_2, as well as cross-sectional area, from A_1 to A_2, as shown in Figure 15.18. The figure shows an element of fluid, in blue, as it moves through the pipe. The top panel corresponds to the location of the fluid element at time $t = 0$, whereas the bottom panel shows the location of the element of fluid at time $t = \Delta t$.

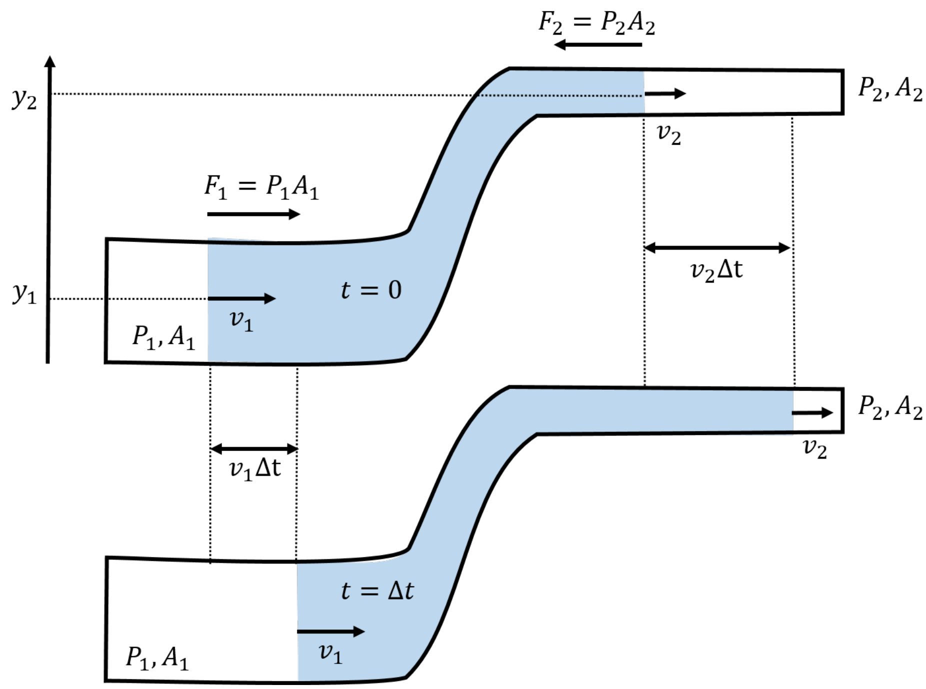


Figure 15.18: Laminar flow of an incompressible fluid through a pipe that changes cross-sectional area and height in the direction of flow. An element of fluid, in blue, is shown at time $t = 0$ (top panel), and then, at a later time, $t = \Delta t$ (bottom panel).

To model how the fluid moves through this pipe, we can use energy and the Work-Energy Theorem. We start by considering the amount of work done on the element of fluid as it moves from the position in the top panel to the position in the bottom panel.

The fluid that is to the left of the element of fluid exerts a pressure, P_1, on the fluid element that leads to a net force, $\vec F_1$, towards the right. Similarly, the fluid to the right of the element of fluid exerts a net force $\vec F_2$ in the opposite direction, due to the pressure P_2 on that side of the fluid element.

In a period of time, Δt, the left part of the fluid element will move a distance $l_1 = v_1\Delta t$, while the right part of the fluid element will move a distance $l_2 = v_2\Delta t$. We can calculate the work done by each force, defining positive work to be in the direction of motion:

$$\begin{aligned}
W_1 &= F_1l_1 = (P_1A_1)(v_1\Delta t)\\
W_2 &= -F_2l_2 = -(P_2A_2)(v_2\Delta t)
\end{aligned}$$

Gravity will also do (negative) work on the fluid as it changes height. In a period of time, Δt, a mass of fluid, Δm, will move from position $y = y_1$ to position $y = y_2$. The mass of fluid that changes height is given by the part of the fluid that moves a distance, l_1, on the right side of the pipe:

$$\Delta m = V_1\rho = A_1 l_1\rho = A_1 v_1 \Delta t\rho$$

Because of the equation of continuity, this is also equal to the mass of fluid that moves a distance, l_2, on the left side of the pipe:

$$\Delta m = V_2\rho = A_2 l_2\rho = A_2 v_2 \Delta t\rho$$

since $A_1 v_1 = A_2 v_2$.

The force of gravity will thus do negative work on that mass element:

$$W_g = -\Delta m g(y_2 - y_1) = -(A_1 v_1 \Delta t\rho)g(y_2 - y_1)$$

The net work done on the element of fluid over the time Δt is thus:

$$W^{net} = W_1 + W_2 + W_g = P_1 A_1 v_1 \Delta t - P_2 A_2 v_2 \Delta t - A_1 v_1 \Delta t\rho g(y_2 - y_1)$$

Note that, because of the equation of continuity, $A_1 v_1 = A_2 v_2$, we can factor out a $A_1 v_1$ from each term:

$$W^{net} = A_1 v_1 \Delta t\Big(P_1 - P_2 - \rho g(y_2 - y_1)\Big)$$

The net work done on the fluid must equal the change in kinetic energy, ΔK, of the mass element, Δm, from one end of the pipe to the other:

$$\begin{aligned}\Delta K &= \frac{1}{2}\Delta m v_2^2 - \frac{1}{2}\Delta m v_1^2\\ &= \frac{1}{2}(A_1 v_1 \Delta t\rho)(v_2^2 - v_1^2)\end{aligned}$$

Using the Work-Energy Theorem, we have:

$$\begin{aligned}W^{net} &= \Delta K\\ A_1 v_1 \Delta t\Big(P_1 - P_2 - \rho g(y_2 - y_1)\Big) &= \frac{1}{2}(A_1 v_1 \Delta t\rho)(v_2^2 - v_1^2)\\ P_1 - P_2 - \rho g(y_2 - y_1) &= \frac{1}{2}\rho v_2^2 - \frac{1}{2}\rho v_1^2\end{aligned}$$

We can re-arrange this so that all the quantities for each side of the pipe are on the same side of the equation:

$$P_1 + \frac{1}{2}\rho v_1^2 + \rho g y_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho g y_2$$

Since the locations 1 and 2 that we chose are arbitrary, we can state that, for laminar incompressible flow, the following quantity evaluated at any position is a constant:

$$\boxed{P + \frac{1}{2}\rho v^2 + \rho g y = \text{constant}} \tag{15.6}$$

This statement is what we call Bernoulli's Equation, and is equivalent to conservation of energy for the fluid. If the fluid is not flowing ($v_1 = v_2 = 0$), then this is equivalent to the statement of hydrostatic equilibrium that we derived in Equation 15.1:

$$\begin{aligned} P_1 + \rho g y_1 &= P_2 + \rho g y_2 \\ \therefore P_2 - P_1 &= -\rho g(y_2 - y_1) \end{aligned}$$

If the flow of the fluid is at constant height ($y_2 = y_1$), then Bernoulli's equation can be written as:

$$P_1 + \frac{1}{2}\rho v_1^2 = P_2 + \frac{1}{2}\rho v_2^2$$

If a fluid is flowing at constant height such that $v_2 > v_1$ (as in Figure 15.16), then $P_2 < P_1$; that is, the **pressure in the fluid is lower if the fluid is flowing faster**. Note that P is the pressure inside the fluid and is not related to the force that would be exerted by the fluid if it were to collide with an object. It makes sense that the fluid has a lower pressure where it is moving faster, because the net force exerted on the fluid is related to the difference in pressure on either side of the fluid. The fluid will accelerate in the direction where pressure decreases, thus it will be moving faster when it is in a region of low pressure.

Bernoulli's principle can be used to describe many phenomena. For example, an airplane wing (technically, an "airfoil") creates lift because the pressure of the air above the wing is lower than the pressure above the wing. This is illustrated in Figure 15.19, which shows that the laminar flow of the air creates a low pressure area above the wing. As the stream lines of air encounter the wing, those that are above the wing get compressed together, which leads to a faster speed of the air above the wing (equation of continuity). The resulting difference in air pressure above and below the wing results in a net upwards force on the wing.

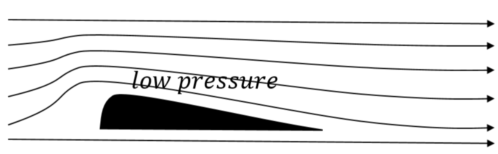


Figure 15.19: Laminar flow of air around a airfoil. The curvature of the asymmetric airfoil forces the streamlines above the airfoil together, increasing the speed of the air due to the continuity equation, and resulting in a low pressure area.

Bernoulli's principle also describes why the roof can be lifted off of a house in high winds (Figure 15.20, left panel). It is not the force of the wind against the roof that blows the roof off of a house; it is the difference in air pressure in the house (normal) and the pressure above the roof (low, due to the flowing wind), that results in a net upwards force on the roof. Bernoulli's principle is also used to construct atomizers which allow liquid in a bottle to be sprayed (Figure 15.20, right panel). For example, perfume bottles often have a bulb connected to a tube/spout. When you squeeze the bulb, it causes the air in the tube to flow quickly, creating a low pressure in the vertical segment of the spout. The liquid is forced up by the pressure in the bottle; once the liquid arrives in the fast flowing air, it is sprayed out along with the air.

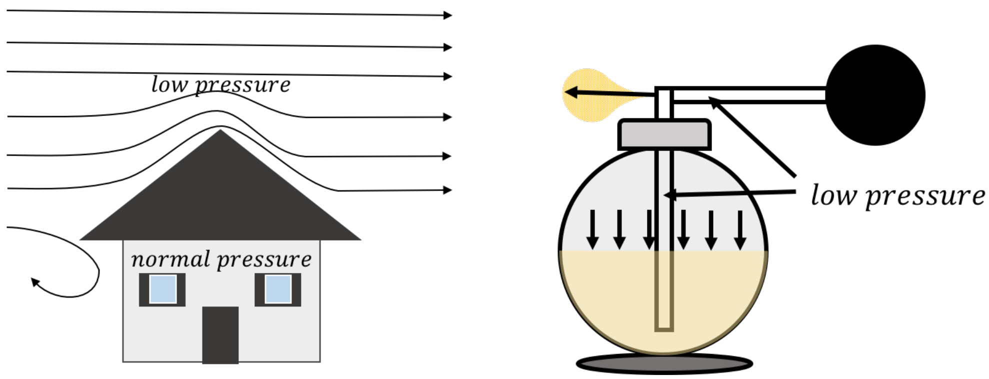


Figure 15.20: (Left:) the wind flowing above a roof creates a low pressure zone above the roof. (Right:) air flowing above a vertical spout in the atomizer creates a low pressure zone; the air pressure in the bottle forces the liquid up the spout.

Checkpoint 15-8

When a high speed train is travelling at constant speed, is there a net force on the windows from air pressure?

A) No, since the windows are stationary relative to the train, there is no net force on them from air pressure.
B) Yes, there is a net outwards force on the windows from air pressure.
C) Yes, there is a net inwards force on the windows from air pressure.

Example 15-6

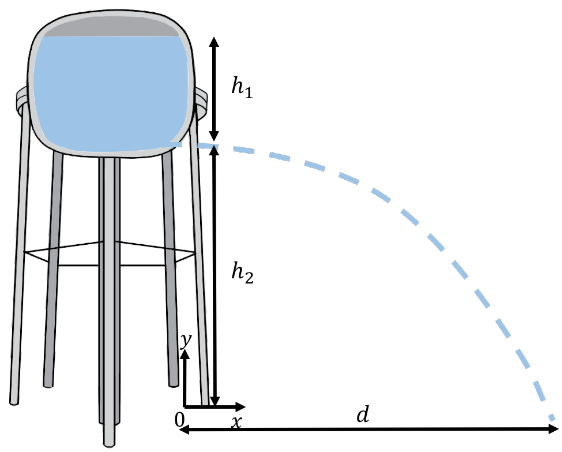


Figure 15.21: Water leaking out of a horizontal hole in a water tank.

A water tower is constructed so that the bottom of the water tank is a height h_2 above the ground, as illustrated in Figure 15.21. The water in the tank is at a height h_1 from the bottom of the tank. A leak from a hole is found at the base of the tank (the water flows horizontally out of the hole). What is the horizontal distance, d, from the bottom of the tower to where the water from the leak hits the ground? Assume that the water level in the tank is constant and that atmospheric pressure does not change appreciably over the height of the tower.

Solution

The pressure in the water tank leads to the water exiting the bottom of the tank with a horizontal velocity of magnitude, v. That water then undergoes projectile motion on its way to the ground. We can first determine the speed of the water exiting the tank and then use the kinematics for projectile motion to model the distance, d.

We model the flow of the water using a two-dimensional coordinate system with a horizontal x axis (positive to the right), and a vertical y axis (positive upwards). We place the origin at the bottom of the water tower, on the ground, below the hole, as shown in Figure 15.21.

At the top of tank, at a height $y = h_1 + h_2$, the water has a speed of zero and is at atmospheric pressure, P_0. At the exit of the hole at the bottom of the tank, at a height $y = h_2$, the water has a speed v_2 and is also at atmospheric pressure. Using Bernoulli's equation at the top (1) and bottom (2) of the tank, we have:

$$P_1 + \frac{1}{2}\rho v_1^2 + \rho g y_1 = P_2 + \frac{1}{2}\rho v_2^2 + \rho g y_2$$
$$P_0 + (0) + \rho g(h_1 + h_2) = P_0 + \frac{1}{2}\rho v^2 + \rho g h_2$$
$$\therefore v_2 = \sqrt{2gh_1}$$

which is exactly the speed that any object falling a distance h_1 would have.

Using kinematics, we can find the time that it would take the water to fall a distance h_2 (where the water's velocity is horizontal as it exits the tank):

$$h_2 = \frac{1}{2}gt^2$$
$$\therefore t = \sqrt{\frac{2h_2}{g}}$$

The distance d covered by the water is thus given by:

$$d = v_2 t = \sqrt{2gh_1}\sqrt{\frac{2h_2}{g}} = \sqrt{4h_1h_2}$$

Discussion: We find that the water coming out of the bottom of the tank, when there is a height, h_1, of water above it providing pressure, will have the same speed as that of a particle which has fallen a distance, h_1. This is because there is no net pressure difference between the top of the water tank and where the water has exited the hole, so gravity is the only force doing work on the water. Gravity will do work at the same rate on particles of water as on any other particle, so the speed of the water particles at the bottom of the tank is the same as if they had fallen a distance, h_1. Again, once the water particles are falling through the air, gravity is the only net force exerted on those particles, so they undergo projectile motion, just as any other particle would.

Example 15-7

You measure that water comes out of your kitchen faucet at a rate of $6\,\mathrm{l/min}$. The faucet has a diameter of $2\,\mathrm{cm}$. At what rate will water flow out of your basement faucet, which has a diameter of $1\,\mathrm{cm}$ and is located a height, $h = 3\,\mathrm{m}$, below your kitchen faucet? Assume that atmospheric pressure, P_0, does not change appreciably between your kitchen and basement.

Solution

The water flows out of the kitchen faucet at a speed, v_1, where the pressure is atmospheric. If the area of the kitchen faucet is A_1 we can determine the speed, v_1, from the given flow rate, $Q_1 = 6\,\mathrm{l/min} = 1\times 10^{-4}\,\mathrm{m^3/s}$:

$$\begin{aligned}
Q_1 &= A_1 v_1\\
v_1 &= \frac{Q_1}{A_1} = \frac{(1\times 10^{-4}\,\mathrm{m^3/s})}{\pi(0.01\,\mathrm{cm})^2} = 0.318\,\mathrm{m/s}
\end{aligned}$$

The water will flow out of the basement faucet with a speed, v_2, where the pressure is also atmospheric, P_0. We can use Bernoulli's equation to relate the flow out of the basement faucet (2) to that at the kitchen faucet (1). We choose the y axis of a vertical coordinate system such that the basement is located at $y_2 = 0$ and the kitchen faucet is located at $y_1 = 3\,\mathrm{m}$:

$$\begin{aligned}
P_1 + \frac{1}{2}\rho v_1^2 + \rho g y_1 &= P_2 + \frac{1}{2}\rho v_2^2 + \rho g y_2\\
P_0 + \frac{1}{2}\rho v_1^2 + \rho g y_1 &= P_0 + \frac{1}{2}\rho v_2^2\\
\frac{1}{2}v_1^2 + g y_1 &= \frac{1}{2}v_2^2\\
\therefore v_2 &= \sqrt{v_1^2 + 2gy_1}\\
&= \sqrt{(0.318\,\mathrm{m/s})^2 + 2(9.8\,\mathrm{m/s^2})(3\,\mathrm{m})} = 7.67\,\mathrm{m/s}
\end{aligned}$$

The corresponding flow rate at the basement faucet will be:

$$Q_2 = A_2 v_2 = \pi(0.005\,\mathrm{m})^2(7.67\,\mathrm{m/s}) = 6.03\times 10^{-4}\,\mathrm{m^3/s} = 36.17\,\mathrm{l/min}$$

Discussion: We find that the flow rate out of the basement faucet is six times that at the kitchen faucet. The speed of the water coming out of the basement faucet is more than 20 times the speed of the water at the kitchen faucet. Although it is true that one gets better water pressure out of a faucet that is lower in the building, this change in flow is unrealistically high, and this is a poor model for flow of water in the pipes of your house.

You can easily verify that the speed of the water in different levels of your house does not vary by a factor near 20 for a 3 m change in height (you could compare the flow rate for two faucets with the same diameter). This is because our model neglects the effect of friction as water flows in the pipes; in reality, there is much greater pressure in the pipes than that due to gravity, as well as a gradient in the pressure in your pipes, that will lead to the flow rates being similar in your kitchen and basement.

15.3.3 Viscosity

So far, we have assumed that fluids flow with no friction. In reality, the particles moving in a fluid exert internal friction on each other called "viscosity". This can be modelled as the friction between different layers of fluid in a laminar flow. For example, you may notice that the water that flows in a wide river flows much faster in the middle of the river than near the river banks, where the water is almost stationary, as shown in Figure 15.22.

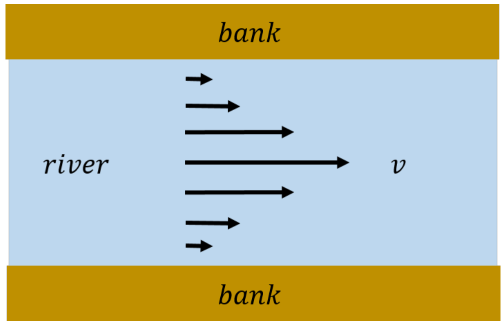


Figure 15.22: Water flowing in a river; the water near the banks is almost immobile due to the viscosity of the water.

One can model the banks of the river as exerting a frictional force on the layer of water that is in contact with the banks. That layer then exerts a frictional force on the next layer closer to the centre of the river, and so on.

One can define a viscosity coefficient, η, based on measuring the force required to pull a plate past another plate when there is a fluid between the plates. Consider two plates that have an area, A, that are a distance l apart, and contain the fluid of interest between them, as illustrated in Figure 15.23.

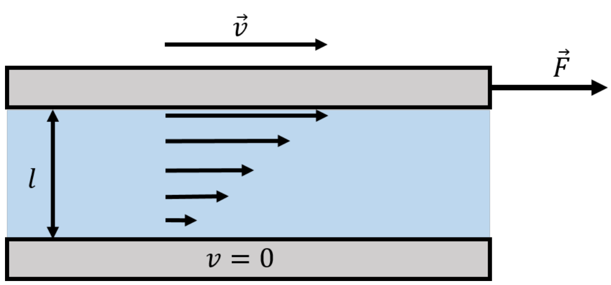


Figure 15.23: A fluid placed between a moving plate (top) and a fixed plate (bottom) in order to measure the viscosity of the fluid.

The viscosity of the fluid is defined based on the force that is required to pull the top plate while the bottom plate remains immobile. The layer of fluid directly below the moving plate will move with the plate at a speed, v, while the layer of fluid immediately in contact will the stationary plate will also be stationary. Moving one plate will thus lead to a gradient (a change) in the speed of the fluid as a function of the position between the two plates. The magnitude of the force, $\vec{F}$, required to move one plate with speed, v, was empirically determined to be proportional to the area of the plates, A, and the speed, v, while being inversely proportional to the distance, l, between the two plates:

$$F \propto A\frac{v}{l}$$

The constant of proportionality is defined as the viscosity, η, of the fluid:

$$\boxed{F = \eta A \frac{v}{l}} \tag{15.7}$$

If the viscosity of the fluid is zero, then no force is required to pull the plate. The more viscous the fluid, the more difficult it is to pull the top plate. You can experiment with this by comparing the force required to move a small piece of paper across the top of a puddle of water and across the top of honey.

The presence of viscosity means that any fluid that flows will lose mechanical energy due to internal friction (which will heat up the fluid). Thus, Bernoulli's equation is not correct if the fluid has viscosity, as a fluid cannot flow through a horizontal pipe without a change in pressure to overcome the losses due to friction.

15.3.4 Poiseuille flow

For the flow of an incompressible viscous fluid through a pipe, one can postulate that the flow rate, Q, is proportional to the change in pressure, ΔP, across the pipe:

$$Q \propto \Delta P$$

where ΔP is taken as the positive difference between the pressure at either end of the pipe. The fluid flows from high pressure to low pressure. We can introduce a constant of proportionality, R, to be the "resistance of the pipe", so that we can write:

$$Q = \frac{\Delta P}{R}$$

where we wrote the constant of proportionality as $1/R$, so that a larger value of R corresponds to a pipe with a higher resistance to flow. That is, for a given pressure difference, as one increases the resistance of the pipe, one decreases the flow rate through that pipe. The relationship above can be used to empirically determine the resistance of a pipe.

The flow through a pipe with a given resistance will be zero if there is no pressure gradient in the fluid along the pipe. Conversely, if there is no flow of fluid in the pipe, the pressure is the same everywhere in the pipe. We can thus also view a drop in pressure in a pipe to be the result of flow of liquid through the pipe. The pressure cannot drop across a horizontal pipe if there is no flow.

When you close the tap on your kitchen faucet, the pressure inside the faucet is close to the pressure in the main water line that supplies your house. As soon as you open the tap and allow water to flow, the pressure in your faucet drops to atmospheric pressure, and the resulting pressure gradient from the main supply forces water to flow out of the faucet. If you try to plug your kitchen faucet with your thumb and stop the flow of water, you will need to exert a force large enough to overcome the pressure that exists in the main water supply. You will find that it is practically impossible to stop the flow of water with your thumb, as the pressure in the main supply needs to be high enough to overcome the resistance of the pipes and still result in a usable flow of water.

Poiseuille first developed a model for the **laminar flow of a liquid through a uniform horizontal cylindrical pipe** of length, L, with a circular cross-section with radius r. He found that the resistance of such a pipe to a fluid of viscosity, η, is given by:

$$R = \frac{8\eta L}{\pi r^4}$$

This makes some intuitive sense, as we expect more resistance (more impedance to flow), if the pipe is longer and if the fluid is more viscous (the resistance is zero if there is no viscosity). We further expect less resistance if the pipe has a larger radius. The resistance found by Poiseuille goes down as the fourth power of the radius. Thus, a pipe that is twice as wide will have a volume flow that is $2^4 = 16$ times larger because of the reduced resistance.

The laminar flow rate, Q, of a viscous fluid through a pipe of length L and radius R, when there is a pressure difference ΔP, is given by:

$$\boxed{Q = \frac{\pi r^4}{8\eta L}\Delta P} \tag{15.8}$$

This is usually referred to as "Poiseuille's Equation".

Checkpoint 15-9

Does the flow rate of water out of a garden hose depend on the length of the hose?

A) No, since the volume of water entering the hose must also exit the hose, it does not matter how long the hose is.
B) Yes, the resistance of the hose depends on its length, so the pressure drop across the hose will change, and so will the flow rate.

Example 15-8

You are modelling the flow of water for a city. Two houses are connected in parallel to the main water supply, so that water from the main supply flows into either house 1 or house 2, and the flows out of each house then join up again at the main supply. The difference in pressure, ΔP, between the entry and exit point of water is the same for each house, and each house can be modelled as having a net resistance, R_1 or R_2, to the flow of water, as illustrated in Figure 15.24. If you model the two houses as being the equivalent of a single "effective" house with an effective resistance R, what is the value of R in terms of R_1 and R_2?

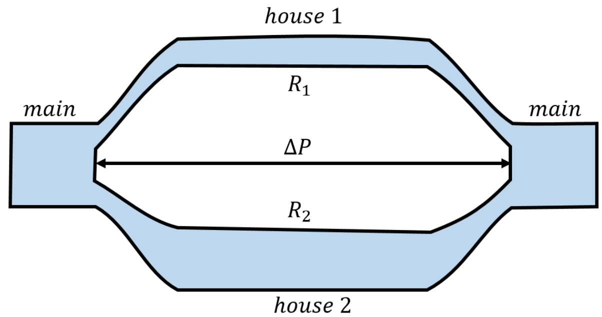


Figure 15.24: Flow of water being separated into two parallel paths that join up again.

Solution

The water from the main will have to flow through either house 1 or house 2. If the flow rate through the main is Q, we require that this be equal to the sum of the flow rates through each house:

$$Q = Q_1 + Q_2$$

The flow through each house is related to the pressure difference, ΔP, across each house (which is the same), as well as the resistance of that house:

$$Q_1 = \frac{\Delta P}{R_1}$$
$$Q_2 = \frac{\Delta P}{R_2}$$

The total flow rate is thus:

$$\begin{aligned} Q = Q_1 + Q_2 &= \frac{\Delta P}{R_1} + \frac{\Delta P}{R_2} \\ &= \Delta P \left(\frac{1}{R_1} + \frac{1}{R2} \right) \end{aligned}$$

We can write this as the flow through an effective resistance, R, with a pressure difference ΔP:

$$Q = \frac{\Delta P}{R}$$
$$\therefore R = \frac{1}{\frac{1}{R_1} + \frac{1}{R2}}$$

Discussion: By requiring that the sum of the flows of water through the houses be the same as the flow rate through the main pipe, we were able to model the two houses as a single effective house with resistance R. You may notice that this is the same relation as the equivalent resistance for two electrical resistors combined in parallel. This is because the flow of electrical current in a resistor can be modelled using similar tools to those required for modelling the flow of a viscous fluid in a pipe.

15.4 Summary

Key Takeaways

The pressure from a force, $\vec{F}$, exerted over a surface with area, A, is a scalar quantity defined as:

$$P = \frac{F_\perp}{A}$$

where $F_\perp$ is the component of the force perpendicular to the surface.

If a force is exerted on the particles in a fluid (e.g. gravity), a pressure will exist everywhere in the fluid. If the fluid is placed in a container, that pressure leads to an external force on all surfaces of the container.

If two fluids at different pressures exist on either side of an interface/object, the net force on that interface/object from the pressures of the fluids will be proportional to the difference in pressure of the fluids on either side.

A fluid is in hydrostatic equilibrium if the sum of the forces on any fluid element is zero. In the presence of gravity, this always leads to a vertical pressure gradient

$$\frac{dP}{dy} = -\rho g$$

where ρ is the density of the fluid, g is the magnitude of the Earth's gravitational field, and the y axis is positive upwards.

If the fluid is incompressible, then the difference in pressure between two points at heights y_1 and y_2 is given by:

$$P(y_2) - P(y_1) = -\rho g(y_2 - y_1)$$

Pascal's Principle states that if an external pressure, P, is applied to one location in a fluid, then the pressure everywhere in the fluid increases by P.

If an object is immersed in a fluid, it will experience a force of buoyancy that is in the opposite direction to the gravitational field in that fluid. The magnitude of the buoyancy force is given by Archimedes' Principle:

$$F_B = \rho V g$$

where, ρ, is the density of the fluid and, V, is the volume of the fluid displaced by the object (i.e. the volume of the part of the object that is immersed in the fluid).

We can distinguish between laminar and turbulent flow of fluids. In laminar flow, individual particles in the fluid follow well-defined streamlines. In turbulent flow, individual particle follow complicated paths that usually involve Eddy currents. In general, it is much easier to model the laminar flow of fluids.

The equation of continuity states that the mass flow rate of a fluid through a closed system must be the same everywhere in the system (no fluid can appear or disappear). For laminar flow of a fluid with density, ρ, flowing at speed, v, through a pipe with cross section, A, the mass flow rate is a constant:

$$\rho A v = \text{constant}$$

A fluid is said to be incompressible if it has constant density. For a fluid of constant density, the volume flow rate, Q, must be constant everywhere in a closed system:

$$Q = Av = \text{constant}$$

Bernoulli's Principle, which is based on the conservation of mechanical energy, states that the following quantity is a constant:

$$P + \frac{1}{2}\rho v^2 + \rho g y = \text{constant}$$

for the laminar flow of a fluid with no viscosity. P is the internal pressure of the fluid, v its speed, and y the height of the fluid relative to a fixed coordinate system. In particular, Bernoulli's Principle implies that, for a constant height, the internal pressure of a fluid must decrease if its speed increases.

Viscosity, η, for the laminar flow of a fluid can be modelled as the result of the internal friction force between layers of the fluid. Because of viscosity, a fluid cannot flow in a horizontal pipe unless there is a difference in pressure across the pipe. Similarly, there will be no horizontal pressure gradient through a fluid unless the fluid is flowing. In general, the volume flow rate, Q, of an incompressible fluid through a pipe with resistance, R, is given by:

$$Q = \frac{\Delta P}{R}$$

For the laminar flow of a fluid with viscosity, η, through a horizontal cylindrical pipe of length, L, and radius, r, the flow rate is given by Poiseuille's equation:

$$Q = \frac{\pi r^4}{8\eta L}\Delta P$$

Important Equations

In the presence of gravity:

$$\frac{dP}{dy} = -\rho g$$
$$P(y_2) - P(y_1) = -\rho g(y_2 - y_1)$$
$$F_B = \rho V g$$

Equation of continuity:

$$\rho A v = \text{constant}$$
$$Q = Av = \text{constant} \quad \text{(if incompressible)}$$

Bernoulli:

$$P + \frac{1}{2}\rho v^2 + \rho g y = \text{constant}$$

Viscosity:

$$Q = \frac{\Delta P}{R}$$
$$Q = \frac{\pi r^4}{8\eta L}\Delta P \quad \text{(Poiseuille)}$$

15.5 Thinking about the material

Reflect and research

1. Does atmospheric pressure increase or decrease when the weather is nice? How come?
2. How does water move from the roots of a tree to the top, for a very tall tree?
3. When did Bernoulli describe the motion of fluids?
4. Where did Bernoulli come from?

To try at home

1. Place your hand in a plastic bag, and immerse your hand with the bag in water. The deeper the column of water, the better. Describe what you feel on your hand in terms of the direction of the force exerted by the water pressure.
2. If you assume that the water that comes out of your bathroom faucet is gravity-fed from a water tank, determine the height of the corresponding water tower relative to your bathroom faucet. Measure the flow rate of water from the faucet to determine the height and discuss whether it makes sense.
3. Try plugging the faucet in your bathroom tap with your thumb. Are you able to completely prevent water from coming out when the tap is open? Estimate the pressure of the water in the pipes leading to your bathroom faucet.
4. In your house/building, measure the flow rate between similar faucets at different heights, and compare with what one would expect from the model from Example 15-7.

To try in the lab

1. Propose an experiment to build a barometer and track the changes in atmospheric pressure as a function of time, and to compare your measurements to those from a weather station.
2. Propose an experiment to characterize how liquid flows in a sponge. Is there a maximum height to which a sponge can draw liquid? How is energy conserved if water is drawn upwards in a sponge?
3. Propose an experiment to measure the resistance of a pipe to the flow of water and compare with the result expected from Poiseuille's equation.
4. Propose an experiment to determiner the viscosity of maple syrup.

15.6 Sample problems and solutions

15.6.1 Problems

Problem 15-1: A man and a woman, Rebecca (57 kg) and Ryan (63 kg), are on a cruise when their ship tragically sinks. They are thrust into the freezing cold ocean. They see a large wooden door floating on the surface of the water, and wonder if they could both survive if they both lay on top of the door. They estimate that the door measures about $2\,\mathrm{m} \times 1\,\mathrm{m} \times 0.12\,\mathrm{m}$. The density of salt water is $\rho_w = 1027\,\mathrm{kg/m^3}$.

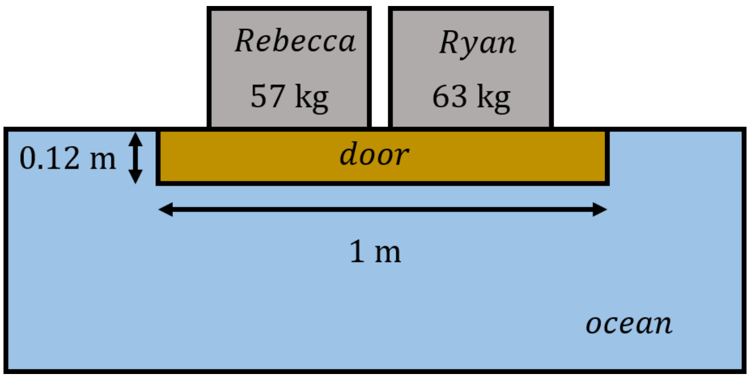


Figure 15.25: Rebecca and Ryan wonder if they can stay above water if they get on top of a floating door.

a) What does the density of the wood have to be in order for Rebecca and Ryan to stay above the surface of the water? (see Figure 15.25)
b) If the door is made of oak ($\rho_d = 750\,\mathrm{kg/m^3}$), will they survive? Can one of them survive?

(Solution)

Problem 15-2: A doctor prescribes an IV drip to a dehydrated patient. She asks a nurse, Rob, to administer 2 l of saline solution ($\eta = 1.0 \times 10^{-3}\,\mathrm{Pa\,s}$, $\rho = 997\,\mathrm{kg/m^3}$) to the patient over 2 hours. An IV drip works by inserting a needle into a vein in a patient's arm. The needle is connected to an IV bag by a tube (Figure 15.26). Lily uses a needle that has a diameter of 0.60 mm and a length of 32 mm. The blood pressure in the patient's veins is 80 mmHg above atmospheric pressure. Note: $1\,\mathrm{mmHg} \approx 133\,\mathrm{Pa}$

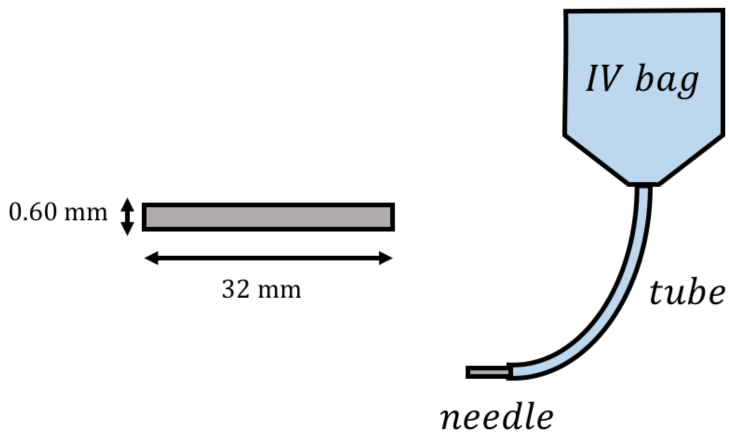


Figure 15.26: Left: A cylindrical IV needle. Right: The IV needle connected to an IV bag by a tube. The free end of the needle goes into the patient's vein.

a) What must the pressure be at the entrance of the needle (the side connected to the saline, not the patient)? Assume that the needle is essentially horizontal and that the diameter of the tube from the IV bag is large enough so that resistance in the vertical tube is negligible. Write your answer in Pascals above atmospheric pressure.
b) How high above the patient's arm should Lily put the IV bag?

(Solution)

15.6.2 Solutions

Solution to problem 15-1:

(a) The forces acting on the door are the force of buoyancy, the door's weight, and the weights of Rebecca and Ryan, as shown in Figure 15.27.

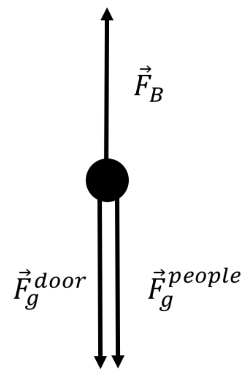


Figure 15.27: The forces acting on the door when Rebecca and Ryan are on top of it.

We can combine the weight of the door and the weight of the people into the total weight, F_g. We choose the y axis to be positive upwards. The sum of the forces on the door in the y direction is given by:

$$\sum F_y = F_B - F_g$$

For the door to float, the net force on the door must be greater than or equal to zero. We want to find the minimum buoyant force for them to float, so we set the net force equal to zero:

$$\begin{aligned} F_g &= F_B \\ (m_R + m_r + m_d)g &= \rho_w V_w g \\ m_R + m_r + m_d &= \rho_w V_w \end{aligned}$$

where the weight includes the mass of Rebecca (m_R), Ryan (m_r) and the door (m_d). We added the subscript W to the right side of the equation to remind ourselves that the buoyant force depends on the density and volume **of the displaced water**. We want to find the maximum density of the wood in order for Rebcca and Ryan to stay above the water's surface. This means that the maximum volume of water that can be displaced is the volume of the door, $V_w = V_d$ (so that the surface of the door is level with the surface of the water, as in Figure 15.25). We can rewrite the mass of the door in terms of its volume and density, and apply our condition that $V_w = V_d$:

$$\begin{aligned} m_R + m_r + \rho_d V_d &= \rho_w V_d \\ \rho_d &= \frac{\rho_w V_d - m_R - m_r}{V_d} \end{aligned}$$

A quick calculation tells us that the volume of the door is $(2\,\mathrm{m})(1\,\mathrm{m})(0.12\,\mathrm{m})$ =

$0.24\,\text{m}^3$. We can now calculate the desired density of the wood:

$$\rho_d = \frac{\rho_w V_d - m_R - m_r}{V_d}$$
$$\rho_d = \frac{(1027\,\text{kg/m}^3)(0.24\,\text{m}^3) - 57\,\text{kg} - 63\,\text{kg}}{0.24\,\text{m}^3}$$
$$\rho_d = 527\,\text{kg/m}^3$$

The maximum density of the wood that would allow them to both float is $527\,\text{kg/m}^3$. Balsa wood has a density that is about $150\,\text{kg/m}^3$, so would allow them to survive. However, it is unlikely that a random floating door is made of balsa wood (although one would choose lighter materials when constructing a ship).

(b) No, they could not both stay on the door because the density of oak is greater than the maximum density of $527\,\text{kg/m}^3$. We can find the amount of mass that can be added to the door (m_A) in order for the person on it to stay above water:

$$\begin{aligned}
F_g &= F_B \\
(m_A + m_d)g &= \rho_w V_w g \\
m_A + \rho_d V_d &= \rho_w V_w \\
m_A + \rho_d V_d &= \rho_w V_d \\
m_A &= V_d(\rho_w - \rho_d)
\end{aligned}$$

where we again used the condition that $V_w = V_d$. We can plug in the appropriate values and solve:

$$m_A = V_d(\rho_w - \rho_d)$$
$$m_A = (0.24\,\text{m}^3)(1027\,\text{kg/m}^3 - 750\,\text{kg/m}^3)$$
$$m_A = 66\,\text{kg}$$

The door can support an additional mass of $66\,\text{kg}$, so either Rebecca or Ryan can survive if the other does not get on the door.

Solution to problem 15-2:

(a) Given that the pressure in the patient's veins is $80\,\text{mmHg}$ above atmospheric pressure, we want to find the pressure required at the other end of the needle so that we get the desired flow rate through the needle. We model the needle as a horizontal cylindrical pipe and assume that the saline solution exhibits laminar flow. We can therefore use Poiseuille's equation:

$$Q = \frac{\pi r^4}{8\eta L}(P_1 - P_2)$$

We let P_1 be the pressure where the needle connects to the tube. Solving for P_1 gives:

$$P_1 = Q\frac{8\eta L}{\pi r^4} + P_2$$

The pressure at the exit of the needle, P_2, is just the blood pressure (80 mmHg+1 atm). The radius of the needle is $0.60\,\text{mm}/2 = 0.30\,\text{mm}$. The flow rate has to be in units of m^3/s. The flow rate in the appropriate units is thus:

$$Q = \frac{2\,\text{l}}{2\,\text{hr}} \cdot \frac{1\,\text{hr}}{3600\,\text{s}} \cdot \frac{0.001\,\text{m}^3}{1\,\text{l}} = 2.8 \times 10^{-7}\,\text{m}^3/\text{s}$$

Using our values, we can calculate P_1:

$$P_1 = (2.8 \times 10^{-7}\,\text{m}^3/\text{s})\frac{8(1.0 \times 10^{-3}\,\text{Pa s})(0.032\,\text{m})}{\pi(3 \times 10^{-4}\,\text{m})^4} + 80\,\text{mmHg} \cdot \frac{133\,\text{Pa}}{1\,\text{mmHg}} + 1\,\text{atm}$$
$$P_1 = 2817\,\text{Pa} + 10\,640\,\text{Pa} + 1\,\text{atm}$$
$$\therefore P_2 = 13\,457\,\text{Pa} \quad \text{above atmospheric pressure}$$

Note that, in the first line, we converted 80 mmHg into Pascals.

(b) We can easily determine the height of the IV bag that is required to give the desired pressure. We choose a coordinate system with a y axis that is vertical (positive upwards) with the origin at the location of the needle (Figure 15.28).

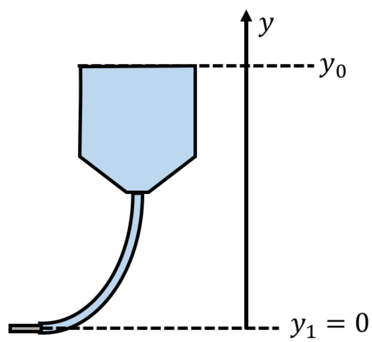


Figure 15.28: The needle is at height 0 and the top of the fluid in the IV bag is at y_0.

At the top of the solution in the IV bag, y_0, the solution has a speed of zero and is at atmospheric pressure, $P_0 = 1\,\text{atm}$. The velocity at the needle is 0, and the pressure is $13\,457\,\text{Pa} + 1\,\text{atm}$. Bernoulli's principle states:

$$P_0 + \frac{1}{2}\rho v_0^2 + \rho g y_0 = P_1 + \frac{1}{2}\rho v_1^2 + \rho g y_1$$

Using our values to solve for y_0, we get:

$$P_0 + \rho g y_0 = P_1$$
$$y_0 = \frac{P_1 - P_0}{\rho g}$$
$$y_0 = \frac{13\,457\,\text{Pa} + 1\,\text{atm} - 1\,\text{atm}}{(997\,\text{kg/m}^3)(9.8\,\text{m/s}^2)}$$
$$y_0 = \frac{13\,457\,\text{Pa}}{(997\,\text{kg/m}^3)(9.8\,\text{m/s}^2)}$$
$$y_0 = 1.4\,\text{m}$$

Therefore, the IV bag should be placed 1.4 m above the patient's arm.

16 Electric charges and fields

In this and subsequent chapters, we start to look at the theories that describe electric and magnetic phenomena. Within the framework for dynamics that was developed by Newton, we will introduce the theories of electromagnetism which describe the electric force, the magnetic force, and how these two interact. This first chapter introduces the description of the electric force, analogously to how we introduced Newton's Universal Theory of Gravity to describe the gravitational force.

Learning Objectives

- Understand the definition of an electric charge.
- Understand the difference between an insulator and a conductor.
- Understand different mechanisms for charging objects.
- Understand Coulomb's model for the electric force.
- Understand the definition of an electric field.
- Understand how to calculate the electric field from a continuous distribution of charge.
- Understand how to model an electric dipole.

Think About It

If you rub a balloon against a carpet and bring it near your head, your hair will stand up and try to touch the balloon.

A) The electric charge of the balloon is opposite of that on your hair.
B) Your hair has no net electric charge, this is an example of charge separation and induction.

16.1 Electric charge

You have likely experienced or heard about electric charge in your life. For example, on a dry Winter day, you might find that after rubbing your bare feet on a polyester carpet you feel a small electric shock upon touching a metallic surface such as a doorknob. This is a manifestation of the electric charge that has built up on you being released onto the doorknob. You probably also have a notion of the existence of positive and negative charges, and that equal charges repel each other whereas opposite charges attract. In this chapter, we develop the description of how these charges can accumulate and how they exert attractive or repulsive forces on each other.

Ordinary matter is made of atoms, which are themselves made of a small positive nucleus (containing positive protons and neutral neutrons) surrounded by a "cloud" of negatively charged electrons. Within a solid object, the atoms in the object can be modelled as being effectively stationary due to inter-atomic forces that hold the atoms together. As a result, the protons (the positively charged part of atoms) can be considered to be fixed in space. The negative electrons, depending on the material, can often move from one atom to another. If an atom looses an electron to another atom, it becomes positive, whereas the atom that acquired the extra electron becomes negative. We define the net charge on an atom (or an object) based on whether there are more protons (positive), more electrons (negative) or an equal amount (neutral). By default, atoms are neutral and have an equal number of protons and electrons. The reason that anything acquires a net electric charge is because it acquired an excess (or deficit) of electrons from another object. We say that "charge is conserved", since the number of electrons does not change and if one object became positively charged, a different object must have become negatively charged by the same amount, so that the total charge (in the Universe) is zero.

When you rub your feet on the carpet, electrons are being removed from one surface (your feet) and deposited on the other (the carpet). Your feet thus acquire a net positive charge (having lost electrons). When you touch a doorknob, the little spark comes from electrons jumping from the doorknob and onto your body. The reason that the electrons leave your feet in the first place is that different materials have different "affinities" for electrons. When you rub two materials together (placing their atoms in close proximity), electrons will transfer to the material with the highest affinity for electrons. This way of creating a net charge is called "charging by friction".

The "triboelectic series" is a list of materials that tend to give up or acquire electrons when they are placed in close contact with each other; some common materials from the series are shown in Figure 16.1. The greatest charge is generated by rubbing together materials that are the furthest away in the diagram. Rubbing silk on a piece of glass result in more charge than rubbing wool on the same piece of glass.

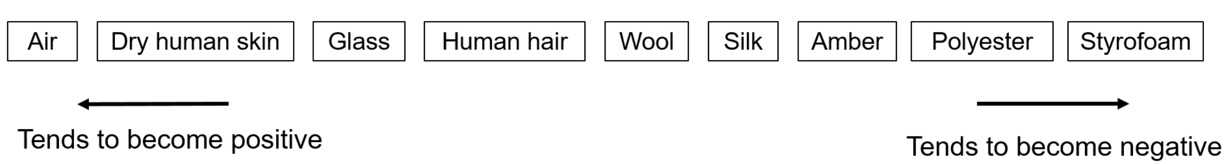


Figure 16.1: A sample of a triboelectric series of materials. The materials on the right-hand side have the greatest affinity to acquire electrons.

Checkpoint 16-1

If you rub a glass rod with silk, which object ends up with an excess of electrons?

A) glass rod.
B) silk.
C) neither, they remain neutral.
D) both will acquire an excess of electrons.

16.1.1 Conductors and insulators

We can broadly classify materials into conductors (such as metals), and insulators (such as wood), depending on how easily the electrons can move around in the material. In a conductor, electrons (rather, the outer electron(s) of an atom) are only loosely bound to their nucleus, and they can thus move around the material freely. In an insulator, the electrons are tightly bound to the nuclei of their atoms and cannot easily move around. There is a third class of materials, semi-conductors, that falls somewhere between a conductor and an insulator. In a semi-conductor, electrons are typically bound to their atoms, but any additional electrons present in the material can move around as if they are in a conductor.

Within a conductor, such as a solid metallic sphere, charges can move around freely. If that sphere has a net charge, for example an excess of electrons, those excess electrons will migrate to the outer surface of the sphere. Electrons in the sphere repel each other and will quickly settle into a position where they are, on average, the furthest from all of the other electrons, which occurs if all of the electrons migrate to the surface. This is illustrated by showing the charges on the surface of the charged sphere in the left panel of Figure 16.2. If an initially neutral conducting sphere is connected to the charged sphere by a conducting wire (right panel of Figure 16.2), some of the electrons will "conduct" (transfer) onto the surface of the neutral sphere, so that, on average, they are further from all other electrons. This way of adding charge to the neutral sphere is called "charging by conduction", and the second sphere will remain charged if the connection between spheres is broken.

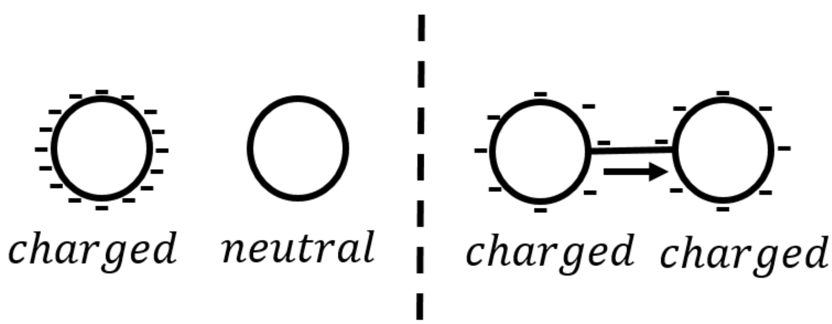


Figure 16.2: Charging by conduction: a neutral conducting sphere is connected to a negatively charged conducting sphere. The charges can "spread out more" if some of the charges move ("conduct") from the charge sphere onto the neutral sphere.

16.1.2 Electrostatic induction

Electrostatic induction allows one to "induce" a charge by using the fact that charges can move freely in a conductor. The left panel of Figure 16.3 shows a (neutral) rod made of a conducting material, with electrons distributed uniformly throughout its volume. In the right panel, a negatively charged sphere is brought next to the rod. Since the rod is conducting, electrons in the rod can easily move and they will thus accumulate on the end of the rod that is furthest from the negative sphere (which repels the electrons). Those electrons will leave positive empty spaces, which can be modelled as positive charges, on the side closest to the sphere. The electrons in the rod will only accumulate for as long as the force from the negative sphere is less than the repulsive force from the electrons that have already accumulated. In practice, such an equilibrium is reached almost instantly. In equilibrium, we say that the rod is "polarized", or that the "charges in the rod have separated", although the rod is overall still neutral.

Note that we can model this as if it where positive charges that move inside of the rod instead of negative charges. The positive charges are attracted to the negative sphere, and thus accumulate on the end of the rod closest to the sphere, leaving a negative charge on the other end. The choice to call electrons "negative" is completely arbitrary. For convenience, we usually build models assuming that positive charges can easily move around, even if, in reality, it is almost always actually (negative) electrons that move.

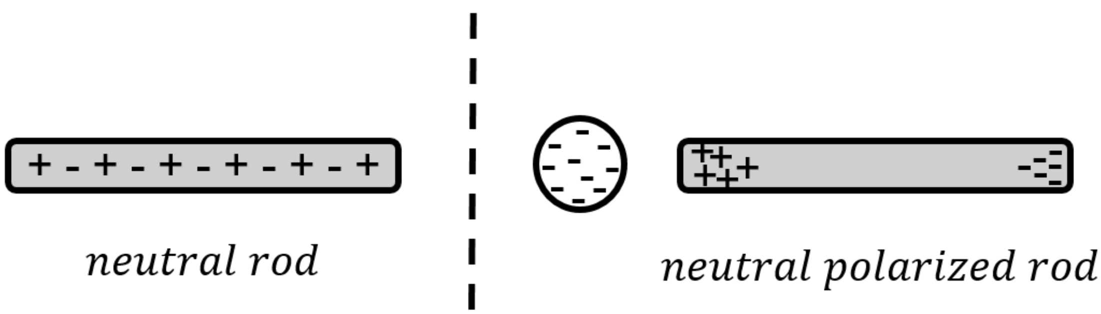


Figure 16.3: Electrostatic induction: when a negatively charged sphere is brought close to a neutral conducting rod, the electrons in the rod, which can move freely, accumulate on the side of the rod furthest from the sphere, leaving an excess of positive charge near the sphere.

We can create a net charge on the polarized rod if we provide a conducting path for charges to leave (or enter) the rod. The Earth can be modelled as a very large reservoir of both positive and negative charges. By connecting the rod to the Earth (we say that we connect the rod to "ground"), we provide a path for the electrons in the rod to be even further from the negatively charged sphere, and they can thus leave the rod entirely in order to go into the ground. This is illustrated in the right-hand panel of Figure 16.4.

If we then disconnect the rod from the ground, it has now acquired an overall positive charge, as in the right hand panel. We call this "charging by induction". We can also think of this in terms of positive charges moving into the rod from the Earth; when we connect the rod to the ground, the positive charges in the Earth can move into the rod and get close to the negatively charged sphere. If we disconnect the rod from the ground, the rod stays positive, just as we conclude when using a model where it is the negative charges that move[1].

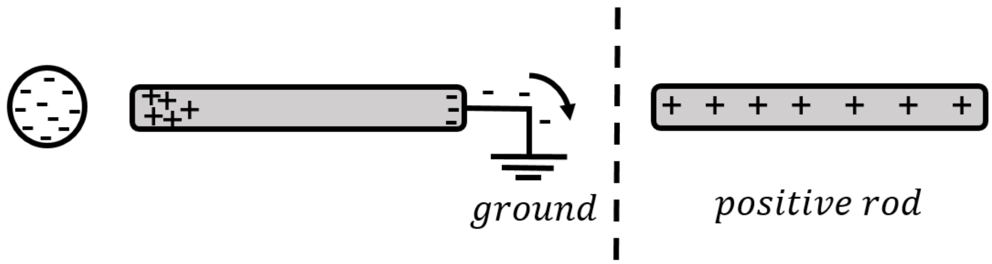


Figure 16.4: Charging by induction: when we connect the polarized rod to the ground, electrons can leave the rod. If we now disconnect the rod from ground, the rod is left with an overall positive charge.

16.2 The Coulomb force

Coulomb was the first to provide a detailed quantitative description of the force between charged objects. Nowadays, we use the (derived) SI unit of "Coulomb" (C) to represent charge. The "charge" of an object corresponds to the net excess (or lack) of electrons on the object. An electron has a charge of $-e = -1.6 \times 10^{-19}\,\mathrm{C}$, which is a very large charge. Thus, an object with a charge of $-1\,\mathrm{C}$ has an excess of about 1×10^{19} electrons on it. If an object has an excess of electrons, it is negatively charged and we indicate this with a negative sign on the charge of the object. An object with a (positive) charge of $1\,\mathrm{C}$ thus has a deficit of 1×10^{19} electrons.

Through careful studies of the force between two charged spheres, Coulomb observed[2] that:

- The force is attractive if the objects have opposite charges and repulsive if the objects have the same charge.
- The force is inversely proportional to the squared distance between spheres.
- The force is larger if the charges involved are larger.

This leads to Coulomb's Law for the electric force (or simply "Coulomb's Law"), $\vec{F}_{12}$, exerted on a point charge Q_1 by another point charge Q_2:

$$\boxed{\vec{F}_{12} = k\frac{Q_1 Q_2}{r^2}\hat{r}_{21}}$$

where $\hat{r}_{21}$ is the unit vector from Q_2 to Q_1 and r is the distance between the two charges, as illustrated in Figure 16.5. $k = 8.99 \times 10^9\,\mathrm{N \cdot m^2/C^2}$ is simply a proportionality constant ("Coulomb's constant") to ensure that the quantity on the right will have units of Newtons

[1] Unless magnetism is involved, it is not possible to distinguish between a flow of positive charges moving in one direction or negative charges moving in the opposite direction.

[2] Others had initially observed the inverse square law for the electric force, but Coulomb was the first to formalize the theory.

when all other quantities are in S.I. units. In some instances, it is more convenient to use the "permittivity of free space", ϵ_0, rather than Coulomb's constant, in which case Coulomb's Law has the form:

$$\vec{F}_{12} = \frac{1}{4\pi\epsilon_0}\frac{Q_1 Q_2}{r^2}\hat{r}_{21}$$

where $\epsilon_0 = \frac{1}{4\pi k} = 8.85 \times 10^{-12}\,\mathrm{C}^2 \cdot \mathrm{N}^{-1} \cdot \mathrm{m}^{-2}$ is a more fundamental constant, as we will see in later chapters.

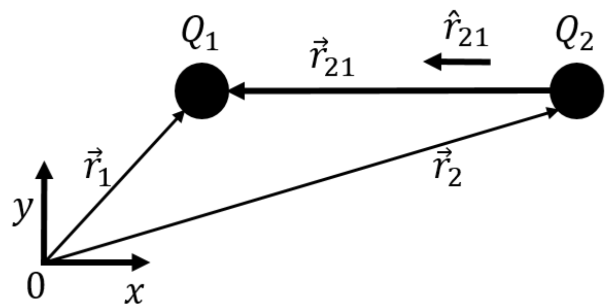


Figure 16.5: Vectors involved in applying Coulomb's Law.

If the two charges have positions $\vec{r}_1$ and $\vec{r}_2$, respectively, then the vector $\hat{r}_{21}$ is given by:

$$\hat{r}_{21} = \frac{\vec{r}_2 - \vec{r}_1}{||\vec{r}_2 - \vec{r}_1||}$$

Coulomb's Law is mathematically identical to the gravitational force in Newton's Universal Theory of Gravity. Rather than quantity of mass determining the strength of the gravitational force, it is the quantity of charge that determines the strength of the electric force. The only major difference is that gravity is always attractive, whereas the Coulomb force can be repulsive.

Checkpoint 16-2

The Coulomb force is conservative.

A) True.
B) False.

The product Q_1Q_2 in the numerator of Coulomb's force is positive if the two charges have the same sign (both positive or both negative) and negative if the charges have opposite signs. Again, referring to Figure 16.5, if the two charges are positive, the force on Q_1 will point in the same direction as $\hat{r}_{21}$ (since all of the scalars are positive in Coulomb's Law) and thus be repulsive. If, instead, the two charges have opposite signs, the product Q_1Q_2 will be negative and the force vector on Q_1 will point in the opposite direction from $\hat{r}_{21}$ and the force is attractive.

Example 16-1

Calculate the magnitude of the electric force between the electron and the proton in a hydrogen atom and compare this to the gravitational force between them.

Solution

We model this by assuming that the electron and proton are point charges a distance of $1\,\text{Å} = 1\times 10^{-10}\,\text{m}$ apart (1 Ångstrom is about the size of the hydrogen atom). The proton and electron have the same charge with magnitude $e = 1.6\times 10^{-19}\,\text{C}$, so the (attractive) electric force between them has a magnitude of:

$$F^e = k\frac{Q_1 Q_2}{r^2} = (9\times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)\frac{(1.6\times 10^{-19}\,\text{C})(1.6\times 10^{-19}\,\text{C})}{(1\times 10^{-10}\,\text{m})^2} = 2.3\times 10^{-8}\,\text{N}$$

which is a small number, but acting on a very small mass. In comparison, the force of gravity between an electron ($m_e = 9.1\times 10^{-31}\,\text{kg}$) and a proton ($m_p = 1.7\times 10^{-27}\,\text{kg}$) is given by:

$$F^g = G\frac{m_e m_p}{r^2} = (6.7\times 10^{-11}\,\text{Nm}^2/\text{kg}^2)\frac{(9.1\times 10^{-31}\,\text{kg})(1.7\times 10^{-27}\,\text{kg})}{(1\times 10^{-10}\,\text{m})^2} = 1.04\times 10^{-47}\,\text{N}$$

Discussion: As we can see, the electric force between an electron and a proton is 39 orders of magnitude larger than the gravitational force! This shows that the gravitational force is extremely weak on the scale of particles and has essentially no effect in particle physics. Indeed, the best current theory of particle physics, and the most precisely tested theory in physics, the "Standard Model", does not need to include gravity in order to provide a spectacularly precise description of particles. One of the big challenges in theoretical physics is nonetheless to develop a theory that integrates the gravitational force with the other forces.

Example 16-2

Three charges, $Q_1 = 1\,\text{nC}$, $Q_2 = -2\,\text{nC}$, and $q = -1\,\text{nC}$, are held fixed at the three corners of an equilateral triangle with sides of length $a = 1\,\text{cm}$, with a coordinate system as shown in Figure 16.6. What is the electric force vector on charge q? (Note that $1\,\text{nC} = 1\times 10^{-9}\,\text{C}$).

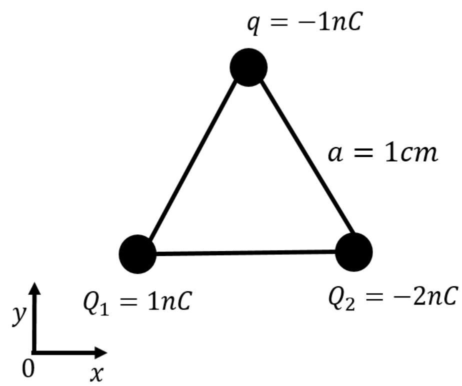


Figure 16.6: Three charges arranged in an equilateral triangle of side a.

Solution

The net electric force on charge q will be the vector sum of the forces from charges Q_1 and Q_2. We thus need to determine the force vectors on q from each charge using Coulomb's Law, and then add those two vectors to obtain the net force on q. The force vectors exerted on q from each charge are illustrated in Figure 16.7.

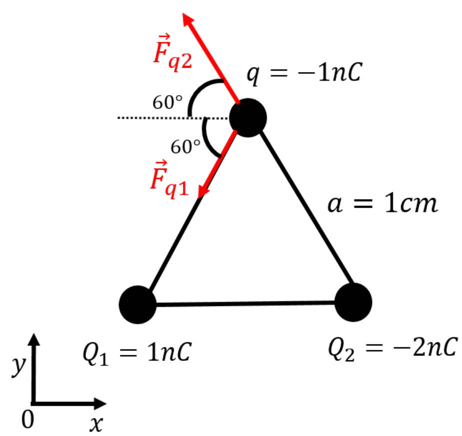


Figure 16.7: Force vectors on charge q.

The force from charge Q_1 has magnitude:

$$F_{q1} = \left| k\frac{Q_1 q}{a^2} \right| = (9 \times 10^9\,\mathrm{N \cdot m^2/C^2})\frac{(1 \times 10^{-9}\,\mathrm{C})(1 \times 10^{-9}\,\mathrm{C})}{(0.01\,\mathrm{m})^2} = 9 \times 10^{-5}\,\mathrm{N}$$

and components:

$$\begin{aligned}\vec{F}_{q1} &= -F_{q1}\cos(60^\circ)\hat{x} - F_{q1}\sin(60^\circ)\hat{y}\\ &= -(4.5\times 10^{-5}\,\text{N})\hat{x} - (7.8\times 10^{-5}\,\text{N})\hat{y}\end{aligned}$$

Similarly, the force on q from Q_2 has magnitude:

$$F_{q2} = \left|k\frac{Q_2 q}{a^2}\right| = (9\times 10^9\,\text{N}\cdot\text{m}^2/\text{C}^2)\frac{(2\times 10^{-9}\,\text{C})(1\times 10^{-9}\,\text{C})}{(0.01\,\text{m})^2} = 1.8\times 10^{-4}\,\text{N}$$

and components:

$$\begin{aligned}\vec{F}_{q2} &= -F_{q2}\cos(60^\circ)\hat{x} + F_{q2}\sin(60^\circ)\hat{y}\\ &= -(9.0\times 10^{-5}\,\text{N})\hat{x} + (1.6\times 10^{-4}\,\text{N})\hat{y}\end{aligned}$$

Finally, we can add the two force vectors together to obtain the net force on q:

$$\begin{aligned}\vec{F}^{net} &= \vec{F}_{q1} + \vec{F}_{q2}\\ &= -(4.5\times 10^{-5}\,\text{N})\hat{x} - (7.8\times 10^{-5}\,\text{N})\hat{y} - (9.0\times 10^{-5}\,\text{N})\hat{x} + (1.6\times 10^{-4}\,\text{N})\hat{y}\\ &= -(13.5\times 10^{-5}\,\text{N})\hat{x} + (8.2\times 10^{-5}\,\text{N})\hat{y}\end{aligned}$$

which has a magnitude of 15.8×10^{-5} N.

Discussion: In this example, we determined the net force on a charge by making use of the superposition principle; namely, that we can treat the forces exerted on q by Q_1 and Q_2 independently, without needing to consider the fact that Q_1 and Q_2 exert forces on each other.

16.3 The electric field

We define the electric field vector, $\vec{E}$, in an analogous way as we defined the gravitational field vector, $\vec{g}$. By defining the gravitational field vector, say, at the surface of the Earth, we can easily calculate the gravitational force exerted by the Earth on any mass, m, without having to use Newton's Universal Theory of Gravity. As you recall, we can define the gravitational field at some position $\vec{r}$, $\vec{g}(\vec{r})$, from a point mass M as the gravitational force per unit mass:

$$\vec{g}(\vec{r}) = -G\frac{M}{r^2}\hat{r}$$

where $\vec{r}$ is a vector from the position of M to where we want to know the gravitational field. As a result, the force exerted on a "test mass", m, located at position $\vec{r}$ relative to mass M is given by:

$$\vec{F}^g = m\vec{g} = -G\frac{Mm}{r^2}\hat{r}$$

which, of course, is the result from Newton's Theory of Gravity. As you recall, we can define the gravitational field for any object that is not a point mass (e.g. the Earth), and use that field to find the force exerted by the Earth on any mass m, without having to re-calculate the gravitational field each time (which requires an integral or Gauss' Law).

We proceed in an analogous was to define the "Electric field", $\vec{E}(\vec{r})$, as the *electric force per unit charge*. If we have a point charge, Q, located at the origin of a coordinate system, then the electric field from that point charge at some position $\vec{r}$ relative to the origin is given by:

$$\boxed{\vec{E}(\vec{r}) = k\frac{Q}{r^2}\hat{r}}$$

If we place a "test charge", q, at position $\vec{r}$ in space, it will experience a force given by:

$$\vec{F}^e = q\vec{E} = k\frac{Qq}{r^2}\hat{r}$$

just as we find from Coulomb's Law.

Checkpoint 16-3

A negative charge is placed at the origin of a coordinate system. At some point in space, the electric field from that charge

A) points towards the origin.
B) points away from the origin.

In Example 16-2, we determined the electric force on charge q, exerted by two other charges Q_1 and Q_2. If we now changed the value of q and wanted to determine the force, we can use the electric field to simplify the process considerably. That is, we can determine the value of the electric field, $\vec{E}$, from Q_1 and Q_2 at the position of q, and then simply multiply that field vector by a charge q to obtain the force on that charge, without having to add force vectors.

Example 16-3

Two charges, $Q_1 = 1\,\text{nC}$, and $Q_2 = -2\,\text{nC}$ are held fixed at two corners of an equilateral triangle with sides of length $a = 1\,\text{cm}$, with a coordinate system as shown in Figure 16.6. What is the electric field vector at the third corner of the triangle?

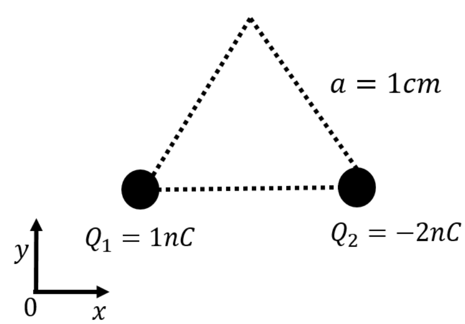


Figure 16.8: Two charges at the corners of an equilateral triangle of side a.

Solution

The net electric field at the third corner of the triangle will be the vector sum of the electric fields from charges Q_1 and Q_2. We thus need to determine the electric field vectors from each charge, and then add those two vectors to obtain the net electric field. The vectors are illustrated in Figure 16.9.

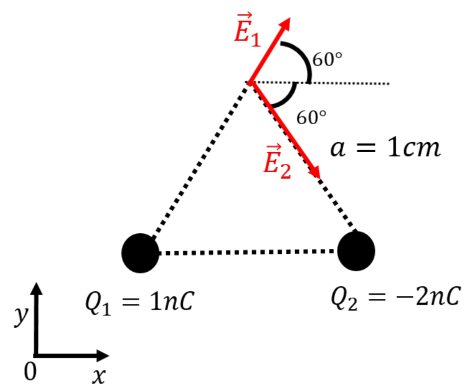


Figure 16.9: Electric field vectors from two charges at the corners of an equilateral triangle of side a.

The electric field from charge Q_1 has magnitude:

$$E_1 = \left|k\frac{Q_1}{a^2}\right| = (9\times 10^9\,\mathrm{N\cdot m^2/C^2})\frac{(1\times 10^{-9}\,\mathrm{C})}{(0.01\,\mathrm{m})^2} = 9\times 10^4\,\mathrm{N/C}$$

and components:

$$\begin{aligned}\vec{E}_1 &= E_1\cos(60^\circ)\hat{x} + E_1\sin(60^\circ)\hat{y}\\ &= (4.5\times 10^4\,\mathrm{N/C})\hat{x} + (7.8\times 10^4\,\mathrm{N/C})\hat{y}\end{aligned}$$

Similarly, the electric field from Q_2 has magnitude:

$$E_2 = \left|k\frac{Q_2}{a^2}\right| = (9\times 10^9\,\mathrm{N\cdot m^2/C^2})\frac{(2\times 10^{-9}\,\mathrm{C})}{(0.01\,\mathrm{m})^2} = 1.8\times 10^5\,\mathrm{N/C}$$

and components:

$$\begin{aligned}\vec{E}_2 &= E_2\cos(60^\circ)\hat{x} - E_2\sin(60^\circ)\hat{y}\\ &= (9.0\times 10^4\,\mathrm{N/C})\hat{x} - (1.6\times 10^5\,\mathrm{N/C})\hat{y}\end{aligned}$$

Finally, we can add the two force vectors together to obtain the net force on q:

$$\begin{aligned}\vec{E}^{net} &= \vec{E}_1 + \vec{E}_2 \\ &= (4.5\times 10^4\,\text{N/C})\hat{x} + (7.8\times 10^4\,\text{N/C})\hat{y} + (9.0\times 10^4\,\text{N/C})\hat{x} - (1.6\times 10^5\,\text{N/C})\hat{y} \\ &= (13.5\times 10^4\,\text{N/C})\hat{x} - (8.2\times 10^4\,\text{N/C})\hat{y}\end{aligned}$$

which has a magnitude of $15.8\times 10^4\,\text{N/C}$. By knowing the electric field at the empty corner of the triangle, we can now calculate the net electric force that would act on any charge placed in that location. For example, if we place a charge $q = -1\,\text{nC}$ (as in Example 16-2), we can easily find the corresponding electric force:

$$\begin{aligned}\vec{F}_q &= q\vec{E} = (-1\,\text{nC})\left[(13.5\times 10^4\,\text{N/C})\hat{x} - (8.2\times 10^4\,\text{N/C})\hat{y}\right] \\ &= -(13.5\times 10^{-5}\,\text{N})\hat{x} + (8.2\times 10^{-5}\,\text{N})\hat{y}\end{aligned}$$

as we found previously. Note that the force on q is in the opposite direction of the electric field vector. This is because q is negative. The **electric field at some point in space thus points in the same direction as the force that a positive test charge would experience**.

Discussion: In this example, we determined the net electric field by making use of the superposition principle; namely, that we can treat the electric fields from Q_1 and Q_2 independently, without needing to consider the fact that Q_1 and Q_2 exert forces on each other. By knowing the electric field at some position in space, we can easily calculate the force vector on any test charge, q, placed at that position. Furthermore, the sign of the charge q will determine in which direction the force will point (parallel to $\vec{E}$ for a positive charge and anti-parallel to $\vec{E}$ for a negative charge).

16.3.1 Visualizing the electric field

Generally, a "field" is something that has a different value at different positions in space. The pressure in a fluid under the presences of gravity is a field: the pressure is different at different heights in the fluid. Since pressure is a scalar quantity (a number), we call it a "scalar field". The electric field is called a "vector field", because it is a vector that is different at each position in space. One way to visualize the electric field is to draw arrows at different positions in space; the length of the arrow is then proportional to the strength of the electric field at that position, and the direction of the arrow then represents the direction of the electric field. The electric field for a point charge is shown using this method in Figure TODO.

One disadvantage of visualizing a vector field with arrows is that the arrows take up space, and it can be challenging to visualize how the field changes magnitude and direction continuously through space. For this reason, one usually uses "field lines" to visualize a vector field. Field lines are continuous lines with the following properties:

- The direction of the vector field at some point in space is tangent to the field line at

that point.
- Field lines have a direction to indicate the direction of the field vector along the tangent.
- The magnitude of the field is proportional to the density of field lines at that point. The more field lines near a location in space, the larger the magnitude of the field vector at that point.

An example of using field lines to represent a vector field in space is shown in Figure 16.10. The corresponding field vector is shown at two different positions in space. At both positions, the vector is tangent to the field line at that position in space and points in the direction of the little arrow drawn at the end of the field lines. The field vector at point A has a larger magnitude than the one at point B, since the field lines are more concentrated at point A than at point B (there are more field lines per unit area at that position in space, the field lines are closer together).

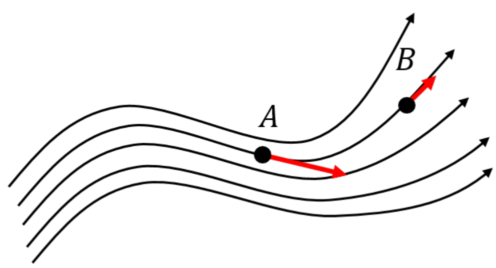


Figure 16.10: An example of determining a field vector from the continuous field lines.

Checkpoint 16-4

It is possible for field lines to cross?

A) Yes.
B) No.

Because the electric field vector always points in the direction of the force that would be exerted on a positive charge, electric field lines will point out from a positive charge and into a negative charge. The electric field lines for a combination of positive and negative charges is illustrated in Figure TODO.

16.3.2 Electric field from a charge distribution

So far, we have only considered Coulomb's Law for point charges (charges that are infinitely small and can be considered to exist at a single point in space). We can use the principle of superposition to determine the electric field from a charged extended/continuous object by modelling that object as being made of many point charges. The electric field from that object is then the sum of the electric field from the point charges that make up that object.

Consider a charged wire that is bent into a semi-circle of radius R, as in Figure 16.11. The wire carries a net electric charge, Q, that is uniformly distributed along the length of the wire. We wish to determine the electric field vector at the centre of the circle.

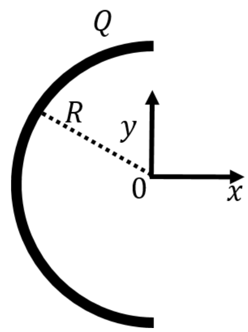


Figure 16.11: A charged wire bent into a semi-circle of radius R.

We start by choosing a very small section of wire and model that section of wire as a point charge with infinitesimal charge dq (as in Figure 16.12). A distance R from that point charge, the electric field from that point charge will have magnitude, dE, given by:

$$dE = k\frac{dq}{R^2}$$

The electric field vector, $d\vec{E}$, from the point charge dq is illustrated in Figure 16.12.

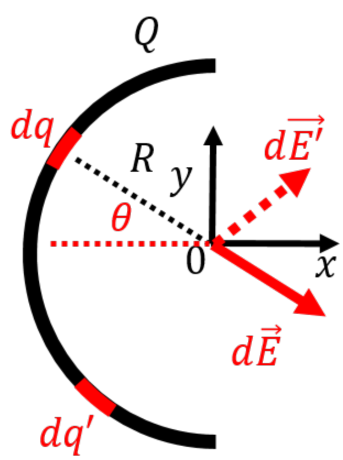


Figure 16.12: Infinitesimal electric fields from point charges along the bent wire.

Using the coordinate system that is show, we define θ as the angle made by the vector from the origin to the point charge dq and the x-axis. The electric field vector from dq is then given by:

$$d\vec{E} = dE\cos\theta\hat{x} - dE\sin\theta\hat{y}$$

The total electric field at the origin will be obtained by summing the electric fields from the different dq over the entire semi-circle:

$$\begin{aligned}\vec{E} &= \int d\vec{E} = \int (dE\cos\theta\hat{x} - dE\sin\theta\hat{y}) \\ &= \left(\int dE\cos\theta\right)\hat{x} - \left(\int dE\sin\theta\right)\hat{y} \\ \therefore E_x &= \int dE\cos\theta \\ \therefore E_y &= -\int dE\sin\theta\end{aligned}$$

We are thus left with two integrals to solve for the x and y components of the electric field, respectively. Before jumping into solving the integrals, it is useful to think about the symmetry of the problem. Specifically, consider a second point charge, dq', located symmetrically about the x-axis from charge dq, as illustrated in Figure 16.12. The charge dq' will create a small electric field $d\vec{E}'$ as illustrated. When we add together $d\vec{E}$ and $d\vec{E}'$, the two y components will cancel, and only the x components will sum together. Similarly, for any dq that we choose, there will always be another dq' such that when we sum together their respective electric fields, the y component will cancel. Thus, by symmetry, we can argue that the net y component of the electric field, E_y, must be identically zero. We thus only need to evaluate the x component of $\vec{E}$:

$$E_x = \int dE\cos\theta = \int k\frac{dq}{R^2}\cos\theta$$

In order to solve this integral, we need to consider which variables change for different choices of the point charge dq. In this case, the distance R is the same anywhere along the semi-circle, so only θ changes with different choices of dq, as k is a constant. We thus need to express dq in terms of $d\theta$ so that we can solve the integral. $d\theta$ corresponds to a small change in the angle θ, and is the angle that is subtended by the charge dq. That is, the charge dq covers a small arc length, ds, of the semi-circle, which is related to $d\theta$ by:

$$ds = Rd\theta$$

The total charge on the wire is given by Q, and the wire has a length πR (half the circumference of a circle). Since the charge is distributed uniformly on the wire, the charge per unit length of any piece of wire must be constant. In particular, dq divided by ds must be equal to Q divided by πR:

$$\begin{aligned}\frac{dq}{ds} &= \frac{Q}{\pi R} \\ \therefore dq &= \frac{Q}{\pi R}ds = \frac{Q}{\pi}d\theta\end{aligned}$$

where in the last equality we used the relation $ds = Rd\theta$. We now have all the ingredients to solve the integral:

$$\begin{aligned}
E_x &= \int k\frac{dq}{R^2}\cos\theta = \int_{-\pi/2}^{+\pi/2} k\frac{Q}{\pi R^2}\cos\theta d\theta \\
&= k\frac{Q}{\pi R^2}\int_{-\pi/2}^{+\pi/2}\cos\theta d\theta = k\frac{Q}{\pi R^2}\left[\sin\theta\right]_{-\pi/2}^{+\pi/2} \\
&= k\frac{2Q}{\pi R^2}
\end{aligned}$$

The total electric field vector at the centre of the circle is thus given by:

$$\vec{E} = k\frac{2Q}{\pi R^2}\hat{x}$$

Note that if we had not realized that we did not need to solve the integral for the y component, we would still find that it is zero:

$$E_y = -k\frac{Q}{\pi R^2}\int_{-\pi/2}^{+\pi/2}\cos\theta d\theta = -k\frac{Q}{\pi R^2}\left[-\cos\theta\right]_{-\pi/2}^{+\pi/2} = 0$$

In order to determine the electric field at some point from any continuous charge distribution, the procedure is generally the same:

1. Make a *good* diagram.
2. Choose a charge element dq.
3. Draw the electric field element, $d\vec{E}$, at the point of interest.
4. Write out the electric field element vector, $d\vec{E}$, in terms of dq and any other relevant variables.
5. Think of symmetry: will any of the component of $d\vec{E}$ sum to zero over all of the dq?
6. Write the total electric field as the sum (integral) of the electric field elements.
7. Identify which variables change as one varies the dq and choose an integration variable to express dq and everything else in terms of that variable and other constants.
8. Do the sum (integral).

Example 16-4

A ring of radius R carries a total charge $+Q$. Determine the electric field a distance a from the centre of the ring, along the axis of symmetry of the ring.

Solution

In order to determine the electric field, we carry out the procedure outlined above, and start by drawing a good diagram, as in Figure 16.13, showing: our coordinate system, our choice of dq, the electric field element vector $d\vec{E}$ that corresponds to dq, and variables (r, θ) to specify the position of dq.

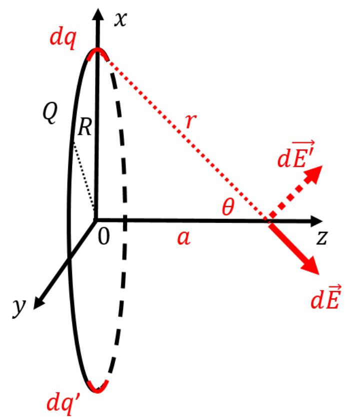


Figure 16.13: Determining the electric field on the axis of a ring of radius R carrying charge Q.

In this case, the figure is challenging to draw and visualize because of the three-dimensional nature of the problem. With the specific dq that we chose, the electric field element vector is given by:

$$d\vec{E} = -dE\sin\theta\hat{x} + 0\hat{y} + dE\cos\theta\hat{z}$$

where $d\vec{E}$ has magnitude:

$$dE = k\frac{dq}{r^2}$$

The x and z components of the total electric field will then be given by:

$$E_x = -\int dE\sin\theta = -\int k\frac{dq}{r^2}\sin\theta$$
$$E_z = \int dE\cos\theta = \int k\frac{dq}{r^2}\cos\theta$$

In general, if we had chosen a dq that is not along one of the axes of the coordinate system, the electric field element vector would have components in all three directions. However, if we consider the symmetry of the ring, we can note that once we sum together all of the electric field elements, only the z components will survive. Indeed, we have shown in Figure 16.13 that for each dq, there will be a dq' located on the opposite side of the ring that will create an electric field element that will cancel all by the z component of the field element from dq. We thus only need to consider the z components of the electric field elements when determining the total electric field:

$$\vec{E} = E_z\hat{z}$$

We now have to evaluate the integral for the z component of the electric field:

$$E_z = \int k\frac{dq}{r^2}\cos\theta$$

and determine which quantities change as we move dq around the ring. In this case, both r^2 and $\cos\theta$ are the same for all elements on the ring, and the integral is trivial:

$$E_z = k\frac{1}{r^2}\cos\theta\int dq = k\frac{Q}{r^2}\cos\theta = kQ\frac{a}{(R^2+a^2)^{\frac{3}{2}}}$$

where the integral $\int dq$ simply means "sum all of the charges dq together", which is equal to Q, the total charge on the ring. In the last equality, we replaced $\cos\theta$ with the variables a and R that are provided in the question.

Example 16-5

You have rubbed a glass rod with a silk cloth such that the glass rod has acquired a positive charge. The rod has a length, L, a negligible cross-section, and has acquired a total charge, Q, that is uniformly distributed along the length of the rod. What is the electric field a distance R from the centre of the rod?

Solution

In order to determine the electric field, we carry out the procedure outlined above, and start by drawing a good diagram, as in Figure 16.14, showing: our coordinate system, our choice of dq at a distance y above the centre of the rod, the electric field element vector $d\vec{E}$ that corresponds to dq, and variables (y, r, θ) to specify the position of dq.

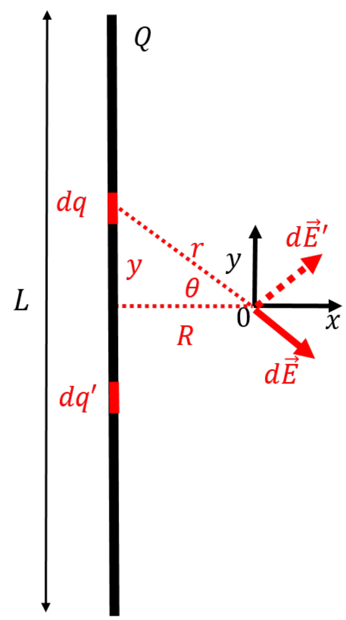


Figure 16.14: Determining the electric field a distance R from the centre of a rod of length L carrying charge Q.

We define the origin to be located at the point where we want to determine the electric field, and the angle θ to be the angle between the horizontal and the position vector of dq. We can write the electric field element vector as:

$$d\vec{E} = dE\cos\theta\hat{x} - dE\sin\theta\hat{y}$$

where $d\vec{E}$ has magnitude:

$$dE = k\frac{dq}{r^2}$$

The x and y components of the total electric field will then be given by:

$$\begin{aligned}E_x &= \int dE\cos\theta = \int k\frac{dq}{r^2}\cos\theta \\ E_y &= -\int dE\sin\theta = -\int k\frac{dq}{r^2}\sin\theta\end{aligned}$$

Again, before proceeding with the integrals, we consider symmetry. Specifically, if we consider a charge dq' located symmetrically about the x axis from dq (as illustrated in Figure 16.14), we see that the y component of the electric field element $d\vec{E}'$ that it creates will cancel the y component of $d\vec{E}$. For each choice of dq, there will exist a corresponding choice dq' which will result in the y component of the net electric field being zero. We thus only need to evaluate the x component of the total electric field:

$$\vec{E} = E_x\hat{x} = \left(\int k\frac{dq}{r^2}\cos\theta\right)\hat{x}$$

Within the integrand, both r and θ will change as we sum over the different charges dq along the rod. A straightforward option to write the integral is to use y as the integration constant, and to write dq, r, and $\cos\theta$ in terms of y. The charge dq covers an infinitesimal length of the rod, dy. Since the rod is uniformly charged, the charge per unit length must be the same over a small length dy as it is over the whole length of the rod:

$$\begin{aligned}\frac{dq}{dy} &= \frac{Q}{L}\\ \therefore dq &= \frac{Q}{L}dy\end{aligned}$$

It is often useful to introduce a constant charge per unit length, $\lambda = \frac{Q}{L}$, so that we can write the charge dq as:

$$dq = \lambda dy$$

We can also express r^2 and $\cos\theta$ in terms of y (and R, which is constant):

$$\begin{aligned}r^2 &= y^2 + R^2\\ \cos\theta &= \frac{R}{r} = \frac{R}{\sqrt{y^2+R^2}}\end{aligned}$$

Finally, we can combine this all into an integral that we can evaluate:

$$\begin{aligned}E_x &= \int k\frac{dq}{r^2}\cos\theta\\ &= k\int_{-L/2}^{L/2}\lambda\frac{1}{y^2+R^2}\frac{R}{\sqrt{y^2+R^2}}dy\\ &= kR\lambda\int_{-L/2}^{L/2}\frac{1}{(y^2+R^2)^{\frac{3}{2}}}dy\\ &= kR\lambda\left[\frac{y}{R^2\sqrt{y^2+R^2}}\right]_{-L/2}^{L/2}\\ \therefore E_x &= \frac{k\lambda}{R}\frac{L}{\sqrt{\left(\frac{L}{2}\right)^2+R^2}}\end{aligned}$$

If the rod were infinitely long (or very long compared to the distance R), the electric field becomes:

$$\lim_{L\to\infty} E_x = \frac{2k\lambda}{R}$$

By using the charge per unit length, λ, we were able to easily generalize our result to that expected for an infinitely long rod with uniform charge density.

Solving the integral above in terms of the integration variable y is difficult without some knowledge of integrals. For this specific integral, the easiest method to use from calculus is "trig substitution". We show below how we can arrive at a much easier integral if we had instead chosen the angle θ as the integration variable instead of y, and we will see that this is a physical illustration of the "trig substitution method" from calculus!

We go back to step 7 in our procedure and choose θ as the integration variable for the integral:

$$E_x = \int k\frac{dq}{r^2}\cos\theta$$

That is, we need to express $1/r^2$ and dq in terms of θ. Referring to Figure 16.14, we have:

$$\begin{aligned}
r &= \frac{R}{\cos\theta}\\
\therefore \frac{1}{r^2} &= \frac{\cos^2\theta}{R^2}\\
y &= R\tan\theta\\
\therefore dy &= \frac{dy}{d\theta}d\theta = \frac{R}{\cos^2\theta}d\theta\\
\therefore dq &= \lambda dy = \lambda\frac{R}{\cos^2\theta}d\theta
\end{aligned}$$

The only difficulty is in determining the angle $d\theta$ subtended by dq, which was determined above by first relating dy and $d\theta$. With these substitutions, the integral becomes trivial:

$$\begin{aligned}
E_x &= \int k\frac{dq}{r^2}\cos\theta\\
&= k\int_{-\theta_0}^{\theta_0}\lambda\frac{R}{\cos^2\theta}\frac{\cos^2\theta}{R^2}\cos\theta d\theta = \frac{k\lambda}{R}\int_{-\theta_0}^{\theta_0}\cos\theta d\theta = \frac{k\lambda}{R}\left[\sin\theta\right]_{-\theta_0}^{\theta_0}\\
&= \frac{2k\lambda}{R}\sin\theta_0
\end{aligned}$$

where θ_0 is the angle subtended by half of the rod. Referring to Figure 16.14, we can easily see that:

$$\sin\theta_0 = \frac{L/2}{\sqrt{\left(\frac{L}{2}\right)^2 + R^2}}$$

So that the total electric field is given by:

$$E_x = \frac{2k\lambda}{R}\sin\theta_0 = \frac{k\lambda}{R}\frac{L}{\sqrt{\left(\frac{L}{2}\right)^2 + R^2}}$$

as found before. Furthermore, in the limit of an infinitely long rod, the angle θ_0 tends to $\frac{\pi}{2}$, so that the electric field becomes:

$$E_x = \lim_{\theta_0 \to \frac{\pi}{2}} \frac{2k\lambda}{R}\sin\theta_0 = \frac{2k\lambda}{R}$$

Discussion: In this example, we saw how to apply the principle of superposition to determine the electric field near a finite and a infinite line of charge with constant charge per unit length. We showed that it was relatively straightforward to set up the integral in terms of dy, but not so easy to solve the integral. We then showed that by using θ as the integration variable, we could arrive at a much easier integral. This change of variable corresponds to a physical variable in our problem, but is also the basis for the more abstract "trig substitution" method used to solve integrals in calculus.

Example 16-6

Calculate the electric field a distance, a, above a infinite plane that carries uniform charge per unit area, σ.

Solution

In this case, we need to determine the field above an object that is two dimensional (a plane). In the previous examples (a ring, a line of charge), we modelled a one dimensional object (e.g. the line), as being made of many point charges (0-dimensional objects). We treated those point charges has having an infinitesimal length along the object so that we could sum them together to obtain the object (e.g. dy was the length of the charge for the rod/line of charge).

In order to model the two-dimensional object (the plane), we model it has being the sum of many one dimensional objects. We can model a plane either as a rectangle of width, W, and length, L, as shown in the left panel of Figure 16.15 or as a disk of radius, R, as shown in the right panel. To model an infinite plane, we can then take the limit of either L and W going to infinity (rectangle), or of R going to infinity (disk).

We can model the rectangle as being the sum of many lines of **finite** length, L, and infinitesimal width, dx. Similarly, we can model the disk as the sum of infinitesimally thin rings of **finite** radius, r, and thickness, dr. In both cases, we know how to model the field from a line of charge (Example 16-5) or from a ring (Example 16-4).

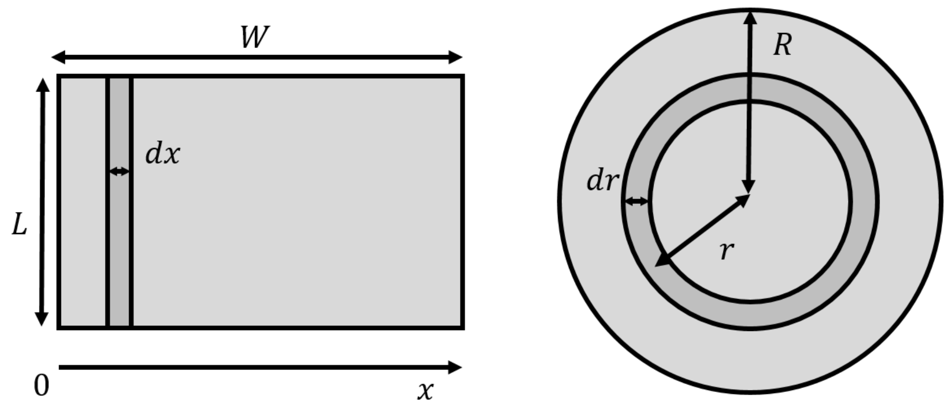


Figure 16.15: A two-dimensional object such as a plane modelled as a the sum of infinitely thin lines (left panel) or as the sum of infinitely thin rings (right panel).

We proceed by modelling the plane as a disk made up of infinitesimal rings. Our infinitesimal charge, dq, is thus that of a ring of radius r and thickness dr, as illustrated in Figure 16.16.

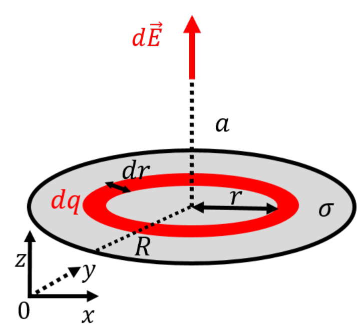


Figure 16.16: Modelling the field from a disk as the sum of fields from concentric thin rings.

We know from Example 16-4 that the magnitude of the electric field a distance a from the centre of the ring, along its axis of symmetry (the z axis in Figure 16.16), is given by:

$$dE = kdq\frac{a}{(r^2+a^2)^{\frac{3}{2}}}$$

By symmetry, for all of the different infinitesimal rings that make up the disk, the field will always point along the z axis. In order to determine the total field, we sum (integrate) the values of dE, over all of the rings, from a radius of $r = 0$ to a radius

$r = R$. For each ring, the value of r will be different, so we need to express dq in terms of dr in order to perform the integral. We know that the plane has a uniform charge per unit area given by σ. The charge dq of an infinitesimal ring is given by:

$$dq = \sigma dA = \sigma 2\pi r dr$$

where $dA = 2\pi r dr$ is the area of the infinitesimal ring of radius r and thickness dr (think of unfolding the ring into a rectangle of height dr and length $2\pi r$, the circumference of the circle, in order to determine the area). We now have all of the ingredients in order to determine the total electric field:

$$\begin{aligned} E &= \int dE = \int_0^R kdq \frac{a}{(r^2+a^2)^{\frac{3}{2}}} = 2\pi k a\sigma \int_0^R \frac{r}{(r^2+a^2)^{\frac{3}{2}}} dr \\ &= 2\pi k a\sigma \left[\frac{-1}{\sqrt{r^2+a^2}} \right]_0^R = 2\pi k\sigma \left(1 - \frac{a}{R^2+a^2} \right) \end{aligned}$$

Finally, we can take the limit of $R \to \infty$ in order to get the electric field above an infinite plane:

$$E = \lim_{R\to\infty} 2\pi k\sigma \left(1 - \frac{a}{R^2+a^2} \right) = 2\pi k\sigma = \frac{\sigma}{2\epsilon_0}$$

where we used ϵ_0 in the last equality as the result is a little cleaner without the factors of π. Note that for an infinite plane of charge, the electric field does not depend on the distance (our variable a) from the plane!

Discussion: In this example, we showed how we can model a two-dimensional charge distribution as the sum of one-dimensional charge distributions. In particular, we showed that an infinite plane of charge can be modelled as the sum of many lines charges or of many rings of charge (we chose the latter in the above). We also found that the electric field above an infinite plane of charge does not depend on the distance from the plane; that is, the electric field is constant above an infinite plane of charge.

16.4 The electric dipole

Electric dipoles are a specific combination of a positive charge $+Q$ held at a fixed distance, l, from an equal and opposite charge, $-Q$, as illustrated in Figure 16.17. Dipoles can be represented by their "electric dipole vector" (or "electric dipole moment"), $\vec{p}$, defined to point in the direction **from the negative charge to the positive charge**, with magnitude:

$$p = Ql$$

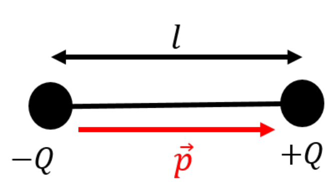


Figure 16.17: An electric dipole.

Dipoles arise often in nature, for example, a water molecule can be modelled as a dipole, because the two hydrogen atoms are not symmetrically arranged around the oxygen atom. The electrons in a water molecule tend to stay closer to the oxygen atom, which acquires an excess of 2 electrons, while each proton has a deficit of 1 electron, resulting in a separation of charge (polarization), which can be modelled as a an electric dipole, as in Figure 16.18.

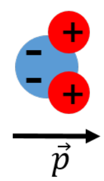


Figure 16.18: A water molecule can be modelled as an electric dipole.

When a dipole is immersed in a uniform electric field, as illustrated in Figure 16.19, the net force on the dipole is zero because the force on the positive charge will always be equal and in the opposite direction from the force on the negative charge.

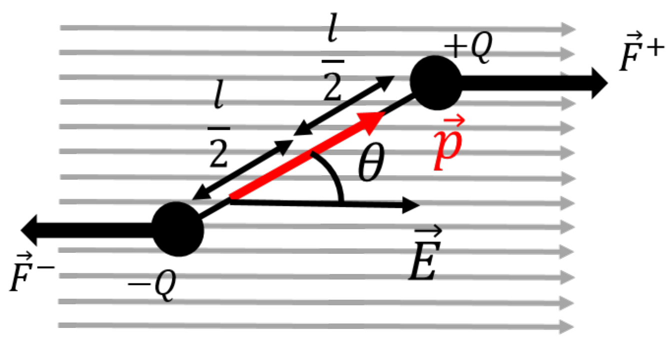


Figure 16.19: An electric dipole in a uniform electric field.

Although the net force on the dipole is zero, there is still a net torque about its centre that will cause the dipole to rotate (unless the dipole vector is already parallel to the electric field vector). If the dipole vector makes an angle of θ with the electric field vector (as in Figure 16.19), the magnitude of the net torque on the dipole about its centre is given by:

$$\tau = \frac{l}{2}F^+ \sin\theta + \frac{l}{2}F^- \sin\theta = \frac{l}{2}QE\sin\theta + \frac{l}{2}QE\sin\theta = QlE\sin\theta = pE\sin\theta$$

In Figure 16.19, the torque vector is into the page (the forces will make it rotate clockwise), which is the same direction as the cross product, $\vec{p}\times\vec{E}$. Note that the magnitude of the torque is also equal to the magnitude of the cross product. Thus, in general, the torque vector on a dipole, $\vec{p}$, from an electric field, $\vec{E}$, is given by:

$$\boxed{\vec{\tau} = \vec{p}\times\vec{E}}$$

In particular, note that the torque is zero when the dipole and electric field vectors are parallel. Thus, a dipole will always experience a torque that tends to align it with the electric field vector. The dipole is thus in a stable equilibrium when it is parallel to the electric field.

Checkpoint 16-5

When an electric dipole is such that its dipole vector is anti-parallel to the electric field vector, the dipole is

A) not in equilibrium.
B) in a stable equilibrium.
C) in an unstable equilibrium.

We can also model the behaviour of the dipole using energy. If a dipole is rotated away from its equilibrium orientation and released, it will gain (rotational) kinetic energy as it tries to return to equilibrium, and will oscillate about the equilibrium position. When the dipole is held out of equilibrium, we can think of it has having potential energy. To determine the functional form of that potential energy function, we consider the work done in rotating the dipole from an angle θ_1 to an angle θ_2 (where the angle is between the dipole and the electric field vectors):

$$\begin{aligned}W &= \int_{\theta_1}^{\theta_2} \tau d\theta = \int_{\theta_1}^{\theta_2} -pE\sin\theta d\theta = -pE\int_{\theta_1}^{\theta_2}\sin\theta d\theta \\ &= pE[\cos\theta]_{\theta_1}^{\theta_2} = pE\cos\theta_2 - pE\cos\theta_1\end{aligned}$$

where the negative sign in the torque is to indicate that the torque is in the opposite direction from increasing θ (in Figure 16.19, the torque is clockwise whereas the angle θ increases counter-clockwise). The net work done in going from position θ_1 to θ_2 is the negative of the change in potential energy in going from θ_1 to θ_2. Thus, we define the potential energy of an electric dipole, $\vec{p}$, in an electric field, $\vec{E}$, as:

$$\boxed{U = -pE\cos\theta = -\vec{p}\cdot\vec{E}}$$

which has a negative sign, and we also recognize that this is equivalent to the scalar product between $\vec{p}$ and $\vec{E}$. Note that the negative sign makes sense because systems experience a force/torque that will decrease their potential energy. When the angle is zero, $\cos\theta$, is maximal. Since we need the position with $\theta = 0$ to have the lowest potential energy, the minus sign guarantees that all values of θ other than zero will give a potential energy that is higher. Remember that only changes in potential energy are relevant, so the minus sign should not bother you, although you should think about whether it makes sense.

16.5 Summary

Key Takeaways

Objects can acquire a net charge if they acquire a net excess or deficit of electrons. Charges are never created, they are only transferred from one object to another. One can charge an object by friction, conduction, or induction. Materials can be classified broadly as conductors, where electrons can move freely in a material, or conductors, in which electrons remain tightly bound to the atoms in the material. If a conducting object acquires a net charge, those charges will migrate to the surface of the conductor.

Coulomb was the first to quantitatively describe the electric force exerted on a point charge, Q_1, by a second point charge, Q_2, located a distance, r, away:

$$\vec{F}_{12} = k\frac{Q_1Q_2}{r^2}\hat{r}_{21} = \frac{1}{4\pi\epsilon_0}\frac{Q_1Q_2}{r^2}\hat{r}_{21}$$

where $\hat{r}_{21}$ is the unit vector from Q_2 to Q_1. One can write the force using either Coulomb's constant, k, or the permittivity of free space, ϵ_0. Coulomb's force is attractive if the product Q_1Q_2 is negative, and repulsive if the product is positive. Thus, charges of the same sign exert a repulsive force on each other, whereas opposite charges exert an attractive for on each other.

Mathematically, Coulomb's Law is identical to the gravitational force in Newton's Universal Theory of Gravity, which implies that it is conservative. The electric field vector at some position in space is defined to be the electric force per unit charge at that position in space. That is, at some position in space where the electric field vector is $\vec{E}$, a charge, q, will experience an electric force:

$$\vec{F} = q\vec{E}$$

much like a mass, m, will experience a gravitational force, $m\vec{g}$, in a position in space where the gravitational field is $\vec{g}$. A positive charge will experience a force in the same direction as the electric field, whereas a negative charge will experience a force in the direction opposite of the electric field. The electric field at position, $\vec{r}$, from a point charge, Q, located at the origin, is given by:

$$\vec{E} = k\frac{Q}{r^2}\hat{r}$$

One can visualize an electric field by using "field lines". The field vector at any point in space has a magnitude that is proportional to the number of field lines at that point, and a direction that is tangent to the field lines at that point.

We can model the electric field from a continuous charged object (i.e. not a point charge) by modelling the object as being made up of many point charges. Often, it is easiest to model an N-dimensional object as being the sum of objects of dimension $N-1$

and an infinitesimal length in the remaining dimension. For example, we modelled a line of charge as the sum of point charges that have an infinitesimal length, and we modelled a disk of charge as the the sum of rings that have an infinitesimal thickness. In general, the strategy to model the electric field from a continuous distribution of charge is the same:

1. Make a *good* diagram.
2. Choose a charge element dq.
3. Draw the electric field element, $d\vec{E}$, at the point of interest.
4. Write out the electric field element vector, $d\vec{E}$, in terms of dq and any other relevant variables.
5. Think of symmetry: will any of the component of $d\vec{E}$ sum to zero over all of the dq?
6. Write the total electric field as the sum (integral) of the electric field elements.
7. Identify which variables change as one varies the dq and choose an integration variable to express dq and everything else in terms of that variable and other constants.
8. Do the sum (integral).

Finally, we introduced the electric dipole, which is an object comprised of two equal and opposite charges, $+Q$ and $-Q$, held at fixed distance, l, from each other. One can model an electric dipole using its dipole vector, $\vec{p}$, defined to point in the direction from $-Q$ to $+Q$, with magnitude:

$$p = Ql$$

When a dipole is immersed in a uniform electric field, $\vec{E}$, it will experience a torque given by:

$$\vec{\tau} = \vec{p} \times \vec{E}$$

The torque will act such as to align the vector $\vec{p}$ with the electric field vector. We can define a potential energy, U, to model the energy that is stored in a dipole when it is not aligned with the electric field:

$$U = -\vec{p} \cdot \vec{E}$$

The point of lowest potential energy corresponds to the case when $\vec{p}$ and $\vec{E}$ are parallel, whereas the point of highest potential energy is when the two vectors are anti-parallel.

Important Equations

Momentum of a point particle:

$$\vec{p} = m\vec{v}$$

$$\frac{d}{dt}\vec{p} = \sum \vec{F} = \vec{F}^{net}$$

Position of the Centre of Mass of a system:

$$\vec{r}_{CM} = \frac{1}{M}\sum_i m_i\vec{r}_i$$

16.6 Thinking about the material

Reflect and research

1. Explain

To try at home

1. Try

To try in the lab

1. Propose an experiment

16.7 Sample problems and solutions

16.7.1 Problems

Problem 16-1:

(Solution)

16.7.2 Solutions

Solution to problem 17-1:

17 Gauss' Law

In this chapter, we take a detailed look at Gauss' Law applied in the context of the electric field. We have already encountered Gauss' Law briefly in Section 9.2.3 when we examined the gravitational field. Since the electric force is mathematically identical to the gravitational force, we can apply the same tools, including Gauss' Law, to model the electric field as we do the gravitational field. Many of the results from this chapter are thus equally applicable to the gravitational force.

Learning Objectives

- Understand the concept of flux for a vector field.
- Understand how to calculate the flux of a vector field through an open and a closed surface.
- Understand how to apply Gauss' Law quantitatively to determine an electric field.
- Understand how to apply Gauss' Law qualitatively to discuss charges on a conductor.

Think About It

A neutral spherical conducting shell encloses a point charge, Q, located at the centre of the shell. Due to separation of charge, the outer surface of the shell will acquire a net positive charge. What is the magnitude of that charge?

A) less than Q.
B) exactly Q.
C) more than Q.

17.1 Flux of the electric field.

Gauss' Law makes use of the concept of "flux". Flux is always defined based on:

- A surface.
- A vector field (e.g. the electric field).

and can be thought of as a measure of the number of field lines from the vector field that cross the given surface. For that reason, one usually refers to the "flux of the electric field through a surface". This is illustrated in Figure 17.1 for a uniform horizontal electric field, and a flat surface, whose normal vector, $\vec{A}$, is shown. If the surface is perpendicular to the field (left panel), and the field vector is thus parallel to the vector, $\vec{A}$, then the flux through that surface is maximal. If the surface is parallel to the field (right panel), then no field lines cross that surface, and the flux through that surface is zero. If the surface is rotated with respect to the electric field, as in the middle panel, then the flux through the surface is between zero and the maximal value.

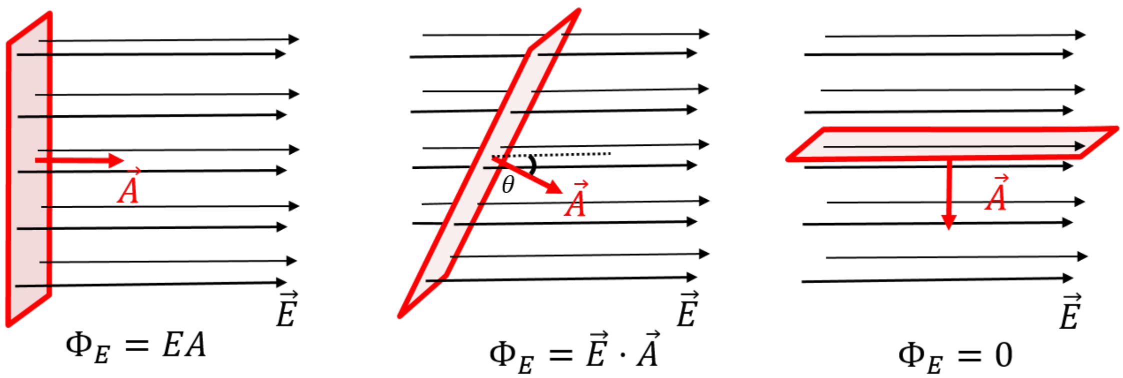


Figure 17.1: Flux of an electric field through a surface that makes different angles with respect to the electric field. In the leftmost panel, the surface is oriented such that the flux through it is maximal. In the rightmost panel, there are no field lines crossing the surface, so the flux through the surface is zero.

We define a vector, $\vec{A}$, associated with the surface such that the magnitude of $\vec{A}$ is equal to the area of the surface, and the direction of $\vec{A}$ is such that it is perpendicular to the surface, as illustrated in Figure 17.1. We define the flux, Φ_E, of the electric field, $\vec{E}$, through the surface represented by vector, $\vec{A}$, as:

$$\Phi_E = \vec{E} \cdot \vec{A} = EA\cos\theta$$

since this will have the same properties that we described above (e.g. no flux when $\vec{E}$ and $\vec{A}$ are perpendicular, flux proportional to number of field lines crossing the surface). Note that the flux is only defined up to an overall sign, as there are two possible choices for the direction of the vector $\vec{A}$, since it is only required to be perpendicular to the surface. By convention, we usually choose $\vec{A}$ so that the flux is positive.

Example 17-1

A uniform electric field is given by: $\vec{E} = E\cos\theta\hat{x} + E\sin\theta\hat{y}$ throughout space. A rectangular surface is defined by the four points $(0,0,0)$, $(0,0,H)$, $(L,0,0)$, $(L,0,H)$. What is the flux of the electric field through the surface?

Solution

The surface that is defined corresponds to a rectangle in the xz plane with area $A = LH$. Since the rectangle lies in the xz plane, a vector perpendicular to the surface will be along the y direction. We choose the positive y direction, since this will give a positive number for the flux (as the electric field has a positive component in the y direction). The vector $\vec{A}$ is given by:

$$\vec{A} = A\hat{y} = LH\hat{y}$$

The flux through the surface is thus given by:

$$\begin{aligned}\Phi_E &= \vec{E}\cdot\vec{A} = (E\cos\theta\hat{x} + E\sin\theta\hat{y})\cdot(LH\hat{y})\\ &= ELH\sin\theta\end{aligned}$$

where one should note that the angle θ, in this case, is not the angle between $\vec{E}$ and $\vec{A}$, but rather the complement of that angle.

Discussion: In this example, we calculated the flux of a uniform electric field through a rectangle of area, $A = LH$. Since we knew the components of both the electric field vector, $\vec{E}$, and the surface vector, $\vec{A}$, we used their scalar product to determine the flux through the surface. In some cases, it is easier to work with the magnitude of the vectors and the angle between them to determine the scalar product (although note that in this example, the angle between $\vec{E}$ and $\vec{A}$ is $90° - \theta$).

17.1.1 Non-uniform fields

So far, we have considered the flux of a uniform electric field, $\vec{E}$, through a surface, S, described by a vector, $\vec{A}$. In this case, the flux, Φ_E, is given by:

$$\Phi_E = \vec{E}\cdot\vec{A}$$

However, if the electric field is not constant in magnitude and/or in direction over the entire surface, then we divide the surface, S, into many infinitesimal surfaces, dS, and sum together (integrate) the fluxes from those infinitesimal surfaces:

$$\boxed{\Phi_E = \int \vec{E}\cdot d\vec{A}}$$

where, $d\vec{A}$, is the normal vector for the infinitesimal surface, dS. This is illustrated in Figure 17.2, which shows, in the left panel, a surface for which the electric field changes magnitude along the surface (as the field lines are closer in the lower left part of the surface), and, in the right panel, a scenario in which the direction (and magnitude) of the electric field vary along the surface.

In order to calculate the flux through the total surface, we first calculate the flux through an infinitesimal surface, dS, over which we assume that $\vec{E}$ is constant in magnitude and direction, and then, we sum (integrate) the fluxes from all of the infinitesimal surfaces together. Remember, the flux through a surface is related to the number of field lines that cross that surface; it thus makes sense to count the lines crossing an infinitesimal surface, dS, and then adding those together over all the infinitesimals surfaces to determine the flux through the total surface, S.

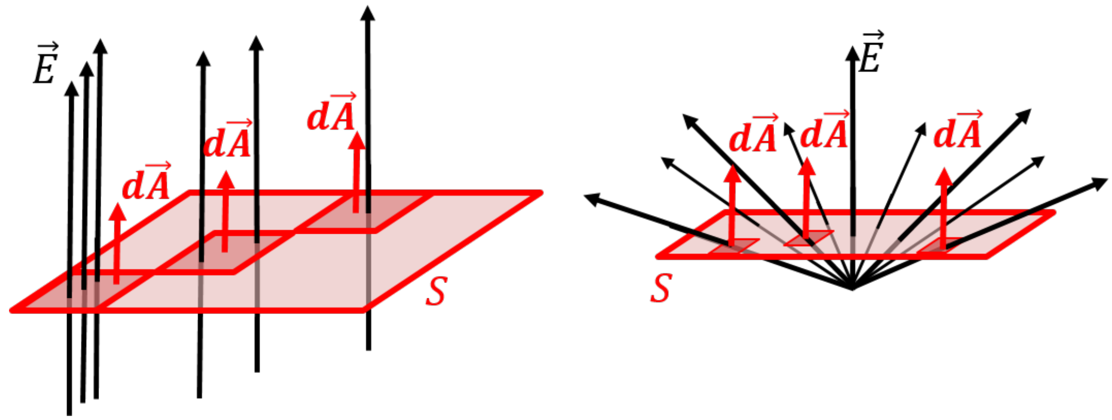


Figure 17.2: Examples of surfaces that need to be sub-divided in order to determine the net flux through them. The surface on the left must be subdivided because the electric field changes magnitude over the surface, whereas the one on the right needs to be subdivided because the angle between $\vec{E}$ and $d\vec{A}$ is not constant (and the magnitude of $\vec{E}$ also changes along the surface).

Example 17-2

An electric field points in the z direction everywhere in space. The magnitude of the electric field depends linearly on the x position in space, so that the electric field vector is given by: $\vec{E} = (a - bx)\hat{z}$, where, a, and, b, are constants. What is the flux of the electric field through a square of side, L, that is located in the xy plane?

Solution

We need to calculate the flux of the electric field through a square of side L in the xy plane. The electric field is always in the z direction, so the angle between $\vec{E}$ and $d\vec{A}$ (the normal vector for any infinitesimal area element) will remain constant.

We can calculate the flux through the square by dividing up the square into thin strips of length L in the y direction and infinitesimal width dx in the x direction, as illustrated in Figure 17.3. In this case, because the electric field does not change with y, the dimension of the infinitesimal area element in the y direction is finite (L). If the electric field varied both as a function of x and y, we would start with area elements that have infinitesimal dimensions in both the x and the y directions.

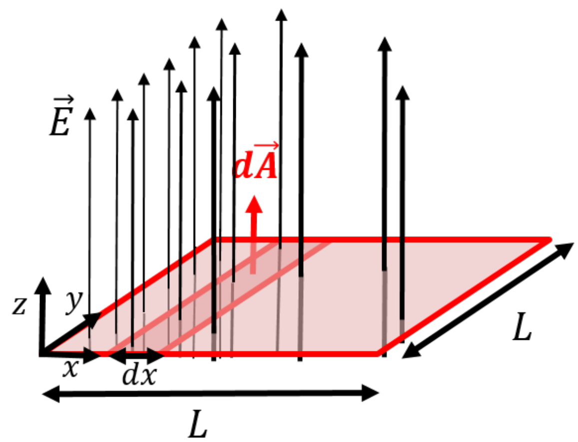


Figure 17.3: Dividing a square in the xy plane into thin strips of length L and width dx.

As illustrated in Figure 17.3, we first calculate the flux through a thin strip of area, $dA = Ldx$, located at position x along the x axis. Choosing, $d\vec{A}$, in the direction to give a positive flux, the flux through the strip that is illustrated is given by:

$$d\Phi_E = \vec{E} \cdot d\vec{A} = EdA = (ax - b)Ldx$$

where $\vec{E} \cdot d\vec{A} = EdA$, since the angle between $\vec{E}$ and $\vec{A}$ is zero. Summing together the fluxes from the strips, from $x = 0$ to $x = L$, the total flux is given by:

$$\Phi_E = \int d\Phi_E = \int_0^L (ax - b)Ldx = \frac{1}{2}aL^3 - bL^2$$

Discussion: In this example, we showed how to calculate the flux from an electric field that changes magnitude with position. We modelled a square of side, L, as being made of many thin strips of length, L, and width, dx. We then calculated the flux through each strip and added those together to obtain the total flux through the square.

17.1.2 Closed surfaces

One can distinguish between a "closed" surface and an "open" surface. A surface is closed if it completely defines a volume that could, for example, be filled with a liquid. A closed surface has a clear "inside" and an "outside". For example, the surface of a sphere, of a cube, or of a cylinder are all examples of closed surfaces. A plane, a triangle, and a disk are, on the other hand, examples of "open surfaces".

For a closed surface, one can unambiguously define the direction of the vector $\vec{A}$ (or $d\vec{A}$) as the direction that it is perpendicular to the surface and **points towards the outside**. Thus, the sign of the flux out of a closed surface is meaningful. The flux will be positive if there is a net number of field lines exiting the surface (since $\vec{E}$ and $\vec{A}$ will be parallel on average) and the flux will be negative if there is a net number of field lines entering the surface (as $\vec{E}$ and $\vec{A}$ will be anti-parallel on average). The flux through a closed surface is thus zero if the number of field lines that enter the surface is the same as the number of field lines that exit the surface.

When calculating the flux over a closed surface, we use a different integration symbol to show that the surface is closed:

$$\Phi_E = \oint \vec{E} \cdot d\vec{A}$$

which is the same integration symbol that we used for indicating a path integral when the initial and final points are the same (see for example Section 8.1).

Example 17-3

A negative electric charge, $-Q$, is located at the origin of a coordinate system. Calculate the flux of the electric field through a spherical surface of radius, R, that is centred at the origin.

Solution

Figure 17.4 shows the spherical surface of radius, R, centred on the origin where the charge $-Q$ is located.

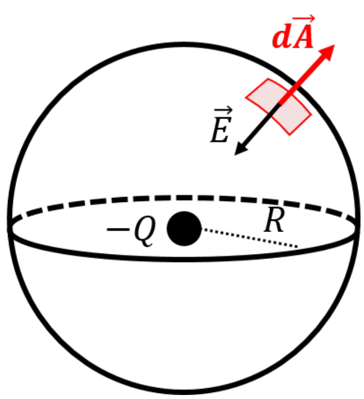


Figure 17.4: Calculating the flux through a spherical surface.

At all points along the surface, the electric field has the same magnitude:

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{R^2}$$

as given by Coulomb's law for a point charge. Although the vector, $\vec{E}$, changes direction everywhere along the surface, it always makes the same angle ($-180°$) with the corresponding vector, $d\vec{A}$, at any particular location. Indeed, for a point charge, the electric field points in the radial direction (inwards for a negative charge) and is thus perpendicular to the spherical surface at all points. Since the surface is closed, the vector, $d\vec{A}$, points outwards anywhere on the surface. Thus, at any point on the surface, we can evaluate the flux through an infinitesimal area element, $d\vec{A}$:

$$d\Phi_E = \vec{E} \cdot d\vec{A} = EdA\cos(-180°) = -EdA$$

where the overall minus sign comes from the fact that, $\vec{E}$, and, $d\vec{A}$, are anti-parallel. The total flux through the spherical surface is obtained by summing together the fluxes through each area element:

$$\Phi_E = \oint d\Phi_E = \oint -EdA = -E\oint dA = -E(4\pi R^2)$$

where we factored, E, out of the integral, since the magnitude of the electric field is constant over the entire surface (a constant distance R from the charge). In the last equality, we recognized that, $\oint dA$, simply means "sum together all of the areas, dA, of the surface elements", which gives the total surface area of the sphere, $4\pi R^2$. The flux through the spherical surface is negative, because the charge is negative, and the field lines point towards $-Q$.

Using the value that we obtained for the magnitude of the electric field from Coulomb's Law, the total flux is given by:

$$\Phi_E = -E(4\pi R^2) = -\frac{1}{4\pi\epsilon_0}\frac{Q}{R^2}(4\pi R^2) = -\frac{Q}{\epsilon_0}$$

which, surprisingly, is independent of the radius of the spherical surface. Note that we used ϵ_0 instead of Coulomb's constant, k, since the result is cleaner without the extra factor of 4π.

Discussion: In this example, we calculated the flux of the electric field from a negative point charge through a spherical surface concentric with the charge. We found the flux to be negative, which makes sense, since the field lines go towards a negative charge, and there is thus a net number of field lines entering the spherical surface. Perhaps surprisingly, we found that the total flux through the surface does not depend on the radius of the surface! In fact, that statement is precisely Gauss' Law: the net flux out of a closed surface depends only on the amount of charge enclosed by that surface (and the constant, ϵ_0). Gauss' Law is of course more general, and applies to surfaces of any shape, as well as charges of any shape (whereas Coulomb's Law only holds for point charges).

17.2 Gauss' Law

Gauss' Law is a relation between the net flux through a closed surface and the amount of charge, Q^{enc}, in the volume enclosed by that surface:

$$\boxed{\oint \vec{E} \cdot d\vec{A} = \frac{Q^{enc}}{\epsilon_0}}$$

In particular, note that Gauss' Law holds true for **any** closed surface, and the shape of that surface is not specified in Gauss' Law. That is, we **can always choose the surface to use** when calculating the flux. For obvious reasons, we often call the surface that we choose a "gaussian surface". But again, this surface is simply a mathematical tool, there is no actual property that makes a surface "gaussian"; it simply means that we chose that surface in order to apply Gauss' Law. In Example 17-3 above, we confirmed that Gauss' Law is compatible with Coulomb's Law for the case of a point charge and a spherical gaussian surface.

Physically, Gauss' Law is a statement that field lines must begin or end on a charge (electric field lines original from positive charges and terminate on negative charges). Recall, flux is a measure of the net number of lines coming out of a surface. If there is a net number of lines coming out of a closed surface (a positive flux), that surface must enclose a positive charge from where those field lines originate. Similarly, if there are the same number of field lines entering a closed surface as there are lines exiting that surface (a flux of zero), then the surface encloses no charge. Gauss' Law simply states that the number of field lines exiting a closed surface is proportional to the amount of charge enclosed by that surface.

Primarily, Gauss' Law is a useful tool to determine the magnitude of the electric field from a given charge, or charge distribution. We usually have to use symmetry to determine the direction of the electric field vector. In general, the integral for the flux is difficult to evaluate, and Gauss' Law can only be used analytically in cases with a high degree of symmetry. Specifically, the integral for the flux is easiest to evaluate if:

1. **The electric field makes a constant angle with the surface**. When this is the case, the scalar product can be written in terms of the cosine of the angle between $\vec{E}$ and $d\vec{A}$, which can be taken out of the integral if it is constant:
$$\oint \vec{E}\cdot d\vec{A} = \oint E\cos\theta dA = \cos\theta \oint EdA$$
Ideally, one has chosen a surface such that this angle is 0 or 180°.
2. **The electric field is constant in magnitude along the surface**. When this is the case, the integral can be simplified further by factoring out, E, and simply becomes an integral over dA (which corresponds to the total area of the surface, A):
$$\oint \vec{E}\cdot d\vec{A} = \cos\theta \oint EdA = E\cos\theta \oint dA = EA\cos\theta$$

Ultimately, the points above should dictate the choice of gaussian surface **so that** the integral for the flux is easy to evaluate. The choice of surface will depend on the symmetry of the problem. For a point (or spherical) charge, a spherical gaussian surface allows the flux to easily be calculated (Example 17-3). For a line of charge, as we will see, a cylindrical surface results is a good choice for the gaussian surface. Broadly, the steps for applying Gauss' Law to determine the electric field are as follows:

1. Make a diagram showing the charge distribution.

2. Use symmetry arguments to determine in which way the electric field vector points.
3. Choose a gaussian surface that goes through the point for which you want to know the electric field. Ideally, the surface is such that the electric field is constant in magnitude and always makes the same angle with the surface, so that the flux integral is straightforward to evaluate.
4. Calculate the flux, $\oint \vec{E} \cdot d\vec{A}$.
5. Calculate the amount of charge located within the volume enclosed by the surface, Q^{enc}.
6. Apply Gauss' Law, $\oint \vec{E} \cdot d\vec{A} = \frac{Q^{enc}}{\epsilon_0}$.

Example 17-4

An insulating sphere of radius, R, contains a total charge, Q, which is uniformly distributed through out its volume. Determine an expression for the electric field as a function of distance, r, from the centre of the sphere.

Solution

Note that this is identical, mathematically, as the derivation that is done in Section 9.2.3 for the case of gravity.

When applying Gauss' Law, we first need to think about symmetry in order to determine the direction of the electric field vector. We also need to think about all possible regions of space in which we need to determine the electric field. In particular, for this case, we need to determine the electric field both inside ($r \leq R$) and outside ($r \geq R$) of the charged sphere.

Figure 17.5 shows the charged sphere of radius R. If we consider the direction of the electric field outside the sphere (where $\vec{E}_{out}$ is drawn), we realize that it can only point in the radial direction (towards or away from the centre of the sphere), as this is the only choice that preserves the symmetry of the sphere. Being a sphere, the charge looks the same from all angles; thus, the electric field must also look the same from all angles, otherwise, there would be a preferred orientation for the sphere. The same argument holds for the electric field vector inside the sphere (drawn as $\vec{E}_{in}$).

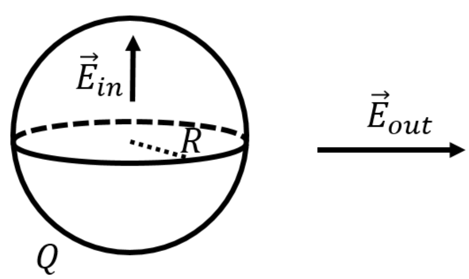


Figure 17.5: For a spherical charge distribution, the electric field inside and outside must point in the radial direction, by symmetry.

We now need to choose a gaussian surface that will make the flux integral easy to evaluate. Ideally, we can find a surface over which the electric field makes the same angle with the surface and over which the electric field is constant in magnitude. Again, based on the symmetry of the charge distribution, it is clear that a spherical surface of radius, r, will satisfy these properties.

We start by applying Gauss' Law outside the charge (with $r \geq R$) to determine the electric field, $\vec{E}_{out}$. Figure 17.6 shows our choice of spherical gaussian surface (labelled S) of radius, r, which is concentric with the spherical charge distribution of radius, R, and total charge, $+Q$.

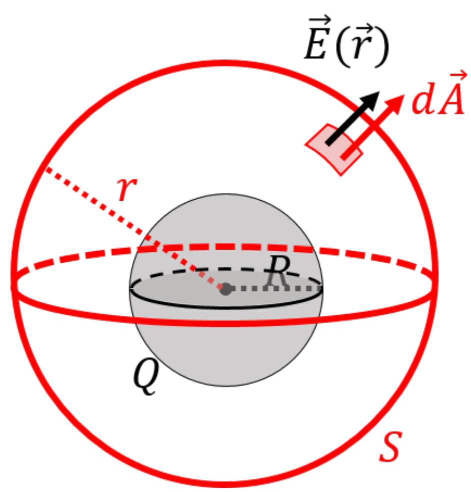


Figure 17.6: A spherical gaussian surface to determine the electric field outside a sphere of radius, R, holding charge, $+Q$.

In order to apply Gauss' Law, we need to calculate:

- the net flux through the surface.
- the charge in the volume enclosed by the surface.

The net flux through the surface is found identically as in Example 17-3, and is given by:

$$\Phi_E = \oint \vec{E} \cdot d\vec{A} = \oint EdA = E \oint dA = E(4\pi r^2)$$

where our choice of spherical surface led to $\vec{E} \cdot d\vec{A} = EdA$, since $\vec{E}$ and $d\vec{A}$ are always parallel. Furthermore, by symmetry, the electric field must be constant in magnitude along the whole surface, or the spherical symmetry would be broken. This allowed us to factor the E out of the integral, leaving us with, $\oint dA$, which is simply the area of our gaussian spherical surface, $4\pi r^2$.

The gaussian surface with $r \geq R$ encloses the whole charged sphere, so the charge enclosed is simply the charge of the sphere, $Q^{inc} = Q$. Applying Gauss' Law allows us to determine the magnitude of the electric field:

$$\begin{aligned}\oint \vec{E} \cdot d\vec{A} &= \frac{Q^{enc}}{\epsilon_0} \\ E(4\pi r^2) &= \frac{Q^{enc}}{\epsilon_0} \\ \therefore E &= \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}\end{aligned}$$

which is the same as the electric field a distance r from a point charge. Thus, from the outside, a spherical charge distribution leads to the same electric field as if the charge were concentrated at the centre of the sphere.

Next, we determine the magnitude of the electric field inside the charged sphere. In this case, we choose a spherical gaussian surface of radius $r \leq R$, that is concentric with the sphere, as illustrated by the surface labelled, S, that is shown in Figure 17.7.

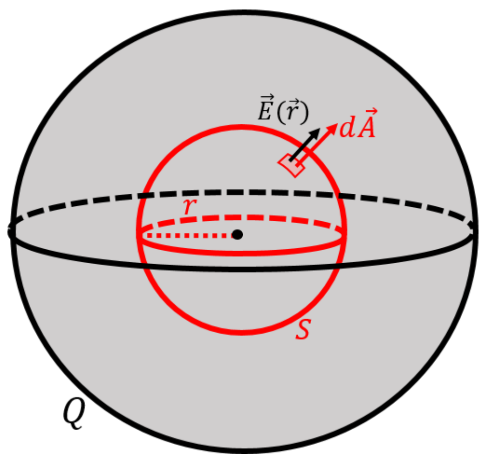


Figure 17.7: A spherical gaussian surface to determine the electric field outside a sphere of radius, R, holding charge, $+Q$.

The flux integral is trivial again, since the electric field always makes the same angle with the gaussian surface, and the magnitude of the electric field is constant in magnitude along the surface:

$$\Phi_E = \oint \vec{E} \cdot d\vec{A} = \oint E dA = E \oint dA = E(4\pi r^2)$$

In this case, however, the charge in the volume enclosed by the gaussian surface is less than Q, since the whole charge is not enclosed. We are told that the charge is distributed uniformly throughout the spherical volume of radius R. We can thus define a volume charge density, ρ, (charge per unit volume) for the sphere:

$$\rho = \frac{Q}{V} = \frac{Q}{\frac{4}{3}\pi R^3}$$

The volume enclosed by the gaussian surface is $\frac{4}{3}\pi r^3$, thus, the charge, Q^{enc}, contained in that volume is given by:

$$Q^{enc} = \frac{4}{3}\pi r^3 \rho = \frac{4}{3}\pi r^3 \frac{Q}{\frac{4}{3}\pi R^3} = Q\frac{r^3}{R^3}$$

Finally, we apply Gauss' Law to find the magnitude of the electric field inside the sphere:

$$\begin{aligned}
\oint \vec{E} \cdot d\vec{A} &= \frac{Q^{enc}}{\epsilon_0} \\
E(4\pi r^2) &= \frac{Q^{enc}}{\epsilon_0} = \frac{Q}{\epsilon_0}\frac{r^3}{R^3} \\
\therefore E &= \frac{Q}{4\pi\epsilon_0 R^3} r
\end{aligned}$$

Note that the electric field increases linearly with radius inside of the charge sphere, and then decreases with radius squared outside of the sphere. Also, note that at the centre of the sphere, the electric field has a magnitude of zero, as expected from symmetry.

Discussion: In this example, we showed how to use Gauss' Law to determine the electric field inside and outside of a uniformly charged sphere. We recognized the spherical symmetry of the charge distribution and chose to use a spherical surface in order to apply Gauss' Law. This, in turn, allowed the flux to be easily calculated. We found that outside the sphere, the electric field decreases in magnitude with radius squared, just as if the entire charge were concentrated at the centre of the sphere. Inside the sphere, we found that the electric field is zero at the centre, and increases linearly with radius.

Example 17-5

An infinitely long straight wire carries a uniform charge per unit length, λ. What is the electric field at a distance, R, from the wire?

Solution

We start by making a diagram of the charge distribution, as in Figure 17.8, so that we can use symmetry arguments to determine the direction of the electric field vector. At any point in space, the electric field vector must be radial (point to/from the centre of the wire) and in the plane perpendicular to the wire. If this were not the case, one would be able to look at the electric field to determine a preferred direction (either around the wire, if the field were not radial, or upwards/downwards, if the field were not perpendicular to the wire).

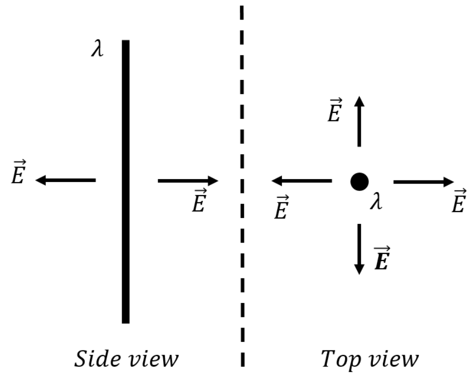


Figure 17.8: An infinite line of charge carrying uniform charge per unit length, λ. The left panel shows a side view and the right panel a view from above. The electric field must be in the radial direction or there would be a preferred direction.

Next, we need to choose a gaussian surface in order to apply Gauss' Law. A convenient choice is a cylinder (a "pill box") of radius, R, and length, L, as shown in Figure 17.9, as this goes through a point that is a distance, R, from the wire (where we are asked for the electric field). At all points on the cylindrical surface, the electric field vector is either perpendicular or parallel to the surface.

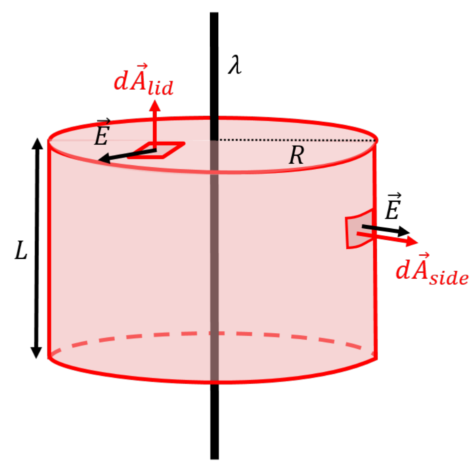


Figure 17.9: A cylindrical gaussian surface is used to calculate the flux from an infinite line of charge.

We can think of the cylindrical surface as being composed of three surfaces: 2 disks on either end (the lids of the pill box), and the curved surface that makes up the side of the cylinder. The flux through the entire cylindrical surface will be the sum of the fluxes through the two lids plus the flux through the side:

$$\oint \vec{E} \cdot d\vec{A} = \int_{side} \vec{E} \cdot d\vec{A} + \int_{lid} \vec{E} \cdot d\vec{A} + \int_{lid} \vec{E} \cdot d\vec{A}$$

where you should note that the closed integral ($\oint$) was separated into three normal integrals ($\int$) corresponding to the three "open" surfaces that make up the closed surface. Again, remember that the flux is proportional to the net number of field lines exiting/entering the closed surface, so it make sense to count those lines over the three open surfaces and add them together to get the total number for the closed surface.

The flux through the lids is identically zero, since the electric field is perpendicular to $d\vec{A}$ everywhere on the lids. The total flux is thus equal to the flux through the curved side surface, for which the electric field vector is always parallel to $d\vec{A}$, and for which the electric field vector is constant in magnitude:

$$\oint \vec{E} \cdot d\vec{A} = \int_{side} \vec{E} \cdot d\vec{A} = \int_{side} EdA = E \int_{side} dA = E(2\pi RL)$$

where we recognized that the side surface can be unfolded into a rectangle of height, L, and width, $2\pi R$, corresponding to the circumference of the cylinder, so that the area is given by $A = 2\pi RL$.

Next, we determine the charge inside the volume enclosed by the surface. Since the cylinder encloses a length, L, of wire, the enclosed charge is given by:

$$Q^{enc} = \lambda L$$

where λ is the charge per unit length on the wire. Putting this altogether into Gauss' Law gives us the electric field at a distance, R, from the wire:

$$\begin{aligned}
\oint \vec{E} \cdot d\vec{A} &= \frac{Q^{enc}}{\epsilon_0} \\
E(2\pi RL) &= \frac{\lambda L}{\epsilon_0} \\
\therefore E &= \frac{\lambda}{2\pi\epsilon_0 R}
\end{aligned}$$

Note that this is the same result that we obtained in Example 16-5 when we took the limit of the finite line of charge having infinite length.

Discussion: In this example, we applied Gauss' Law to determine the electric field at a distance from an infinitely long charged wire. We used symmetry to argue that the field should be radial and in the plane perpendicular to the wire, and recognized that a cylindrical gaussian surface would exploit the symmetry so that the flux can easily be calculated. We obtained the same result as we did from integrating Coulomb's Law in Example 16-5. However, using Gauss' Law was much less work than integrating Coulomb's Law.

Example 17-6

Determine the electric field above an infinitely large plane of charge with uniform surface charge per unit area, σ.

Solution

Figure 17.10 shows a portion of the infinite plane. The electric field vector must be perpendicular to the plane or a preferred direction could otherwise be inferred from the direction of the electric field. We can also argue that the horizontal components of the electric field will cancel everywhere above the plane, since the plane is infinite. The electric field will point away from (towards) the plane, if the charge is positive (negative).

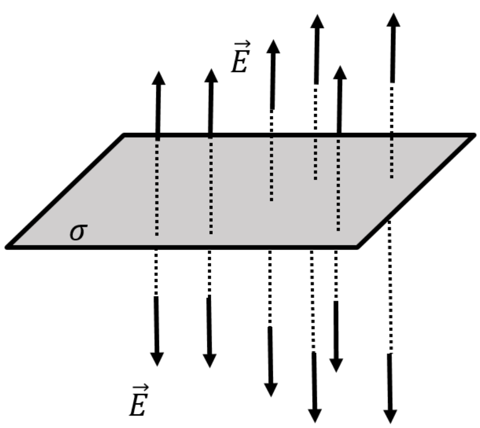


Figure 17.10: The electric field above an infinite plane with uniform charge per unit area, σ, must be perpendicular to the plane.

A cylindrical or box-shaped gaussian surface would both lead to the flux integral being easy to calculate, as illustrated in Figure 17.11. Indeed, since the electric field is perpendicular to the plane, only the parts of the surface that are parallel to the plane (the lids on the cylinder, the two horizontal planes in the box) will have a net flux through them.

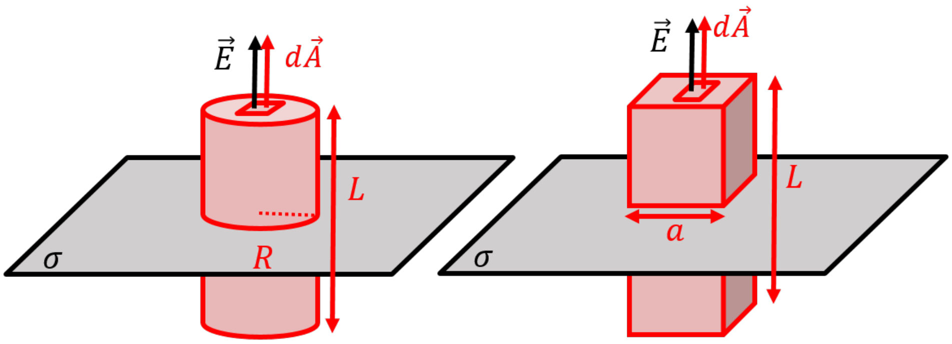


Figure 17.11: A cylindrical surface or a box are both good choices for a gaussian surface above a plane, since only the parts of the surface parallel to the plane will have net flux through them.

Let us choose a box (right panel of Figure 17.11) of length, L, with a square cross-section of side, a. We place the box such that the plane intersects the centre of the box (although this is not required, since we already know that the electric field will not depend on distance from the plane). The flux through the box is simply the flux through the two horizontal planes (of area a^2):

$$\oint \vec{E} \cdot d\vec{A} = \int_{top} EdA + \int_{bottom} EdA = 2Ea^2$$

The box encloses a section of the plane with area a^2, so that the net charge enclosed by the surface is:

$$Q^{enc} = \sigma a^2$$

Applying Gauss' Law allows us to determine the magnitude of the electric field:

$$\begin{aligned}\oint \vec{E} \cdot d\vec{A} &= \frac{Q^{enc}}{\epsilon_0}\\ 2Ea^2 &= \frac{\sigma a^2}{\epsilon_0}\\ \therefore E &= \frac{\sigma}{2\epsilon_0}\end{aligned}$$

which is the same result that we found in Example 16-6.

Discussion: In this example, we used Gauss' Law to determine the electric field above an infinite plane. We found that we had a choice of gaussian surfaces (cylinder, box) that allowed us to apply Gauss' Law. We found the same result that we had found in Example 16-6 where we had integrated Coulomb's Law (twice, once for a ring of charge, then for a disk, then took the limit of the disk radius going to infinity). Again, we see that in configurations with a high degree symmetry, Gauss' Law can be very straightforward to apply.

17.3 Charges in a conductor

We can use Gauss' Law to understand how charges arrange themselves on a conductor. Consider (again) an infinite plane that carries a total charge per unit area, σ, similar to what we considered in Example 17-6. In this case, we explicitly consider the plane to be a conductor and to have a finite thickness. If we zoom into the plane, we will see that the charges will migrate to the surface of the plane, as illustrated in Figure 17.12, where the plane is seen edge on. Thus, the **charge density at the surface is half of the total charge density** of the plane.

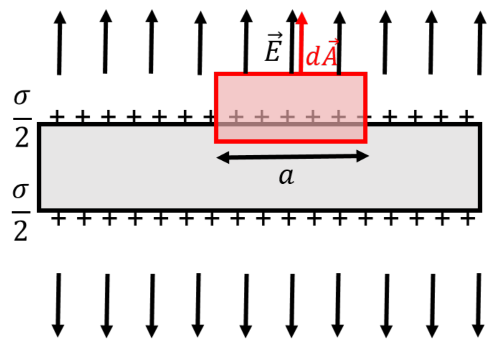


Figure 17.12: Cross-section of a conducting plane where the charges migrate to the surface. A box-shaped gaussian surface is also shown as seen from the side (the third dimension of the box is perpendicular to the plane of the page).

To determine the electric field near the plane, we choose a gaussian surface that is a box (as in Example 17-6), but require the lower end of the box to go through the plane, as illustrated in Figure 17-6. With this choice of gaussian surface, only the top surface (area a^2) will have flux through it, since the **electric field inside a conductor must be zero**[1]. The total flux is given by:

$$\oint \vec{E}\cdot d\vec{A} = \int_{top} EdA = Ea^2$$

The charge enclosed is given by:

$$Q^{enc} = \frac{\sigma}{2}a^2$$

where we used the fact that only half of the charges are inside the volume enclosed by our gaussian surface, so that the charge per unit area is half ($\frac{\sigma}{2}$) of that for the entire plane.

[1]Since charges can freely move in a conductor, they will move until there is no reason to move. Eventually, the charges accumulate in such a way that the net field in the conductor is zero. For a plane, this means that half of the charges will move to each side, as illustrated.

Applying Gauss' Law, we find that the electric field is given by:

$$\begin{aligned}\oint \vec{E}\cdot d\vec{A} &= \frac{Q^{enc}}{\epsilon_0}\\ Ea^2 &= \frac{\sigma a^2}{2\epsilon_0}\\ \therefore E &= \frac{\sigma}{2\epsilon_0} \quad \text{(Field above an infinite plane)}\end{aligned}$$

as before, but the factor of 2 now came from the charge density, rather than from the fact that two of the faces of the box had non-zero flux (as was the case in Example 17-6). We can generalize this result to determine the electric field near the surface of any conductor. Very close to the surface of any object, one can consider the surface as being similar to an infinite plane. If that surface carries charge per unit area, σ, then the electric field just above the surface is given by:

$$E = \frac{\sigma}{\epsilon_0} \quad \text{(Field near a conducting surface)}$$

In this case, there is no factor of 2, because the charge density in this equation is the charge density at the surface of the conductor. In the previous equation, the charge density on the surface of the conducing plane was $\frac{\sigma}{2}$.

Consider, now, a neutral spherical conducting shell, as shown from the side in the left panel of Figure 17.13. When a charge, $+Q$, is placed at the centre of the shell (right panel), charges inside the shell will move until the field in the shell is identically zero. The negative charges will move towards the inner surface (as they are attracted to $+Q$) and positive charges will be repelled onto the outer surface, under the influence of the electric field created by $+Q$ (shown in the diagram as $\vec{E}_Q$). Eventually, the separation of charges will lead to an electric field (shown in the diagram as $\vec{E}_\sigma$) in the opposite direction. The charges will stop moving once the total electric field in the conductor is zero (when the two fields cancel exactly everywhere in the conductor).

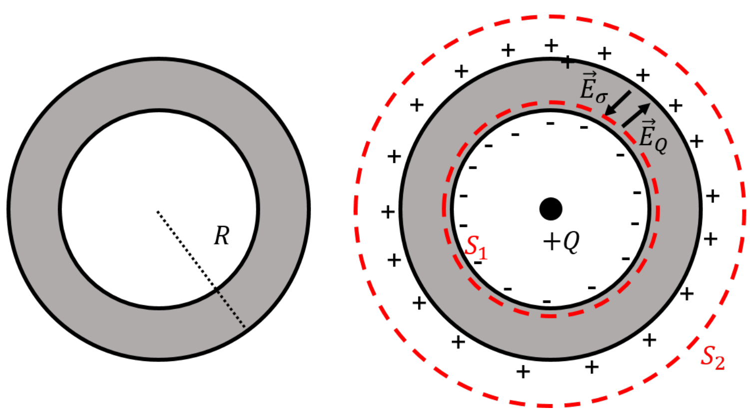


Figure 17.13: Left: a neutral conducting spherical shell (seen edge on). Right: A positive charge, $+Q$, placed at the centre of the shell. Charges in the shell will separate in order to keep the electric field inside the conductor zero.

We can use Gauss' Law to determine the amount of charge that has accumulated on the inner surface. Consider the gaussian spherical surface, S_1, in Figure 17.13, that is concentric with the shell and has a radius such that the surface is just inside the shell. Since the electric field is zero inside the shell, the flux out of the gaussian surface must be zero. By Gauss' Law, the amount of charge enclosed by the surface must also be zero. Thus, a total charge, $-Q$, will have accumulated on the inner surface of the conductor (since $Q^{enc} = -Q + Q = 0$). Because one cannot just create charge from nothing, there must be an equal amount of opposite charge, $+Q$, on the outer surface of the shell. This is true of any conducting material with a cavity inside of it: if you place a charge $+Q$ in the cavity, a charge, $-Q$ will accumulated on the inner surface and a charge, $+Q$, will accumulate on the outer surface.

If we now consider the flux out of the surface, S_2, outside of the shell, the net charge enclosed will be $Q^{enc} = +Q - Q + Q = +Q$. The flux out of the spherical surface of radius, say, r, is then given by:

$$\oint \vec{E} \cdot d\vec{A} = E(4\pi r^2)$$

and the electric field, from Gauss' Law, is simply that of a point charge, $+Q$:

$$E = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}$$

and the shell has no effect on the field in regions where there is no conducting material from the shell. Right at the surface of the shell (outer radius, R), the surface charge density is given by:

$$\sigma = \frac{Q}{4\pi r^2}$$

Above, we found the electric field at the surface of a conductor that carries charge per unit area, σ, to be:

$$E = \frac{\sigma}{\epsilon_0}$$

which is clearly the same result that we obtained using the spherical surface, S_2:

$$E = \frac{\sigma}{\epsilon_0} = \frac{1}{4\pi\epsilon_0}\frac{Q}{r^2}$$

Note that we found the electric field using Gauss' Law only in this last case, and found it to be equal to the electric field that one obtains from Coulomb's law. Thus, Gauss' Law only works if the field has an "inverse square law" dependence. If Gauss' Law does not provide the correct electric field, then the force does not depend on $1/r^2$. Gauss' Law can be used to make extremely stringent tests of whether the force goes as $1/r^2$ or deviates from this model.

17.4 Interpretation of Gauss' Law and vector calculus

In this section, we provide a little more theoretical background and intuition on Gauss' Law, as well as its connection to vector calculus. Very generally, Gauss' Law is a statement that connects a property of a vector field to the "source" of that field. We think of mass as the source for the gravitational field, and we think of charge as the source for the electric field. The property of the field that we considered in this case was its "flux out of a closed surface".

Recall that determining the flux of a field out of a closed surface is equivalent to counting the net number of field lines that exit that closed surface. Field lines must start on a positive charge and must end on a negative charge. Thus, if there is a net number of field lines exiting the surface, there must be a positive charge in the volume defined by the surface (a "source" of field lines). If there is a net number of field lines entering the surface, then the volume defined by the surface must enclose a negative charge (a "sink" of field lines). Gauss' Law is simply a statement that the number of field lines entering/exiting a closed surface is proportional to the amount of charge enclosed in that volume.

The flux out of a closed surface is tightly connected to the vector calculus concept of "divergence", which describes whether field lines are diverging (spreading out or getting closer together). When a point charge is present, field lines will emanate radially from that point charge; in other words, they will diverge. We say that the electric field has non-zero divergence if there is a source of the electric field in that position of space. The key difference between the concept of divergence and that of "flux out of a closed surface", is that divergence is a local property of the field (it is true at a point), whereas the flux out of a surface must be calculated using a finite volume and makes it challenging to define the field at a specific position. Gauss's Law defined using flux is thus not as useful for describing how the field changes at specific positions, and is usually limited to situations with a high degree of symmetry.

The divergence, $\nabla \cdot \vec{E}$, of a vector field, $\vec{E}$, at some position is defined as:

$$\nabla \cdot \vec{E} = \frac{\partial E}{\partial x} + \frac{\partial E}{\partial y} + \frac{\partial E}{\partial z}$$

and corresponds to the sum of three partial derivatives evaluated at that position in space. Gauss' Theorem (also called the Divergence Theorem) states that:

$$\int_V \nabla \cdot \vec{E} = \oint_S \vec{E} \cdot d\vec{A}$$

where the V (S) on the integral indicate whether the sum (integral) should be carried out over a volume, V, or over a closed surface, S, as we have practised in this chapter. While it is not important at this level to understand the theorem in detail, the point is that one can convert a "flux over a closed surface" into an integral of the divergence of the field. In other words, we can convert a global property (flux) to a local property (divergence). Gauss' Law in terms of divergence can be written as:

$$\boxed{\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0}} \quad \text{(Local version of Gauss' Law)}$$

where ρ is the charge per unit volume at a specific position in space. This is the version of Gauss' Law that is usually seen in advanced textbooks and in Maxwell's unified theory of electromagnetism. This version of Gauss's Law relates a local property of the field (its divergence) to a local property of charge at that position in space (the charge per unit volume at that position in space). If we integrate both sides of the equation over volume, we recover the original formulation of Gauss' Law: the left hand side, by the Divergence Theorem, leads to flux when integrated over volume, whereas on the right hand side, the integral over volume of charge per unit volume, ρ, will give the total charge enclosed in that volume, Q^{enc}:

$$\begin{aligned}
\int_V \left(\nabla \cdot \vec{E}\right) dV &= \int_V \left(\frac{\rho}{\epsilon_0}\right) dV \\
\oint_S \vec{E} \cdot d\vec{A} &= \frac{Q^{enc}}{\epsilon_0}
\end{aligned}$$

17.5 Summary

Key Takeaways

We can define the **flux** of a uniform and constant vector field, $\vec{E}$, through a flat surface, as:

$$\Phi_E = \vec{E} \cdot \vec{A} = EA\cos\theta$$

where, $\vec{A}$, is a vector that is perpendicular to the surface with a magnitude equal to the area of that surface, and, θ, is the angle between, $\vec{A}$ and $\vec{E}$. The flux of a field through a surface is proportional to the number of field lines that cross that surface. If the surface is parallel to the field ($\vec{A}$ and $\vec{E}$ are thus perpendicular), the flux through that surface is zero (no field lines cross the surface, the scalar product is zero).

If $\vec{E}$ and $\vec{A}$ change over the surface ($\vec{E}$ and/or $\vec{A}$ change magnitude and/or direction along the surface), then we treat the surface as being made of infinitesimal surface elements over which the two vectors are constant. We define a vector $d\vec{A}$ to be perpendicular to the surface element with an infinitesimal area, dA. The total flux is then obtained by summing the fluxes through each surface element:

$$\Phi_E = \int \vec{E} \cdot d\vec{A} = \int E dA\cos\theta$$

Note that the direction of the vector $d\vec{A}$ (or $\vec{A}$) is ambiguous, as one can choose either of two directions perpendicular to a surface. Usually, one chooses the direction of $\vec{A}$ so that the flux is positive (i.e. $\vec{A}$ has a component parallel to $\vec{E}$). However, if the surface is "closed" (that is, it defines a volume), then we always choose the direction of $d\vec{A}$ so that it points outwards from the surface (since the surface encloses a volume, one can define an "inside" and an "outside").

In the case of the electric field, Gauss' Law relates the flux of the electric field from a closed surface to the amount of charge, Q^{enc}, contained in the volume enclosed by that surface:

$$\oint \vec{E} \cdot d\vec{A} = \frac{Q^{enc}}{\epsilon_0}$$

Physically, Gauss' Law is a statement that field lines must begin or end on a charge (electric field lines original from positive charges and terminate on negative charges). If there is a net number of lines coming out of a closed surface (a positive flux), that surface must enclose a positive charge from where those field lines originate. Similarly, if there are the same number of field lines entering a closed surface as there are lines exiting that surface (a flux of zero), then the surface encloses no charge. Gauss' Law states that the number of field lines exiting a closed surface is proportional to the amount of charge enclosed by that surface.

Gauss' Law is useful to determine the electric field. However, this can only be done analytically for charge distributions with a very high degree of symmetry. This is because the flux integral is not usually easy to evaluate unless:

1. **The electric field makes a constant angle with the surface**. When this is the case, the scalar product can be written in terms of the cosine of the angle between $\vec{E}$ and $d\vec{A}$, which can be taken out of the integral if it is constant:

$$\oint \vec{E} \cdot d\vec{A} = \oint E \cos\theta dA = \cos\theta \oint EdA$$

2. **The electric field is constant in magnitude along the surface**. When this is the case, the integral can be simplified further by factor out E, and simply becomes an integral over dA (which corresponds to the total area of the surface, A):

$$\oint \vec{E} \cdot d\vec{A} = \cos\theta \oint EdA = E\cos\theta \oint dA = EA\cos\theta$$

Note that Gauss' Law does not specify a closed surface over which to calculate the flux; it holds for any surface. We can thus choose a surface that will make the flux integral easy to evaluate - we call this choice a "gaussian surface" (not because it has some special property, but because we chose that surface to apply Gauss' Law). A procedure for applying Gauss' Law to determine the electric field at some point in space can be written as:

1. Make a diagram showing the charge distribution.
2. Use symmetry arguments to determine in which way the electric field vector points.
3. Choose a gaussian surface that goes through the point for which you want to know the electric field. Ideally, the surface is such that the electric field is constant in magnitude and always makes the same angle with the surface, so that the flux integral is straightforward to evaluate.
4. Calculate the flux, $\oint \vec{E} \cdot d\vec{A}$.
5. Calculate the amount of charge in the volume enclosed by the surface, Q^{enc}.
6. Apply Gauss' Law, $\oint \vec{E} \cdot d\vec{A} = \frac{Q^{enc}}{\epsilon_0}$.

We showed how Gauss' Law can be used to understand and quantify how charges arrange themselves on a conductor, in such a way that the electric field is zero everywhere in the conductor. Finally, we briefly introduced a more modern version of Gauss' Law that uses divergence instead of flux:

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0}$$

This last version has the advantage that it relates a local property of the field (divergence) to a local property of charge (charge density at some position in space).

Important Equations

Momentum of a point particle:

$$\vec{p} = m\vec{v}$$

$$\frac{d}{dt}\vec{p} = \sum \vec{F} = \vec{F}^{net}$$

Position of the Centre of Mass of a system:

$$\vec{r}_{CM} = \frac{1}{M}\sum_i m_i\vec{r}_i$$

17.6 Thinking about the material

Reflect and research

1. Explain

To try at home

1. Try

To try in the lab

1. Propose an experiment

17.7 Sample problems and solutions

17.7.1 Problems

Problem 17-1:

(Solution)

17.7.2 Solutions

Solution to problem 17-1:

A Vectors

This appendix gives a very brief introduction to coordinate systems and vectors.

Learning Objectives

- Understand the definition of a coordinate system
- Understand the definition of a vector and of a scalar
- Be able to perform algebra with vectors (addition, scalar products, vector products)

A.1 Coordinate systems

Coordinate systems are used to describe the position of an object in space. A coordinate system is an artificial mathematical tool that we construct in order to describe the position of a real object.

A.1.1 1D Coordinate systems

The easiest coordinate system to construct is one that we can use to describe the location of objects in one dimensional space. For example, we may wish to describe the location of a train along a straight section of track that runs in the East-West direction. In order to do so, we must first define an "origin", which is the reference point of our coordinate system. For example, the origin for our train track may be the Kingston train station (Figure A.1).

We can describe the position of the train by specifying how far it is from the train station (the origin), using a single real number, say x. If the train is at position $x = 0$, then we know that it is at the Kingston station. If the object is not at the origin, then we need to be able to specify on which side (East or West in our train example) of the origin the object is located. We do this by choosing a direction for our one dimensional coordinate x. For example, we may choose that the East side of the track corresponds to positive values of x and that the West side of the track correspond to the negative values of x. Thus, in order to fully specify a one-dimensional coordinate system we need to choose:

- the location of the origin.
- the direction in which the coordinate, x, increases.
- the units in which we wish to express x.

In one dimension, it is common to use the variable x to define the position along the "x-axis". The x-axis *is* our coordinate system in one dimension, and we represent it by drawing a line with an arrow in the direction of increasing x and indicate where the origin is located (as in Figure A.1).

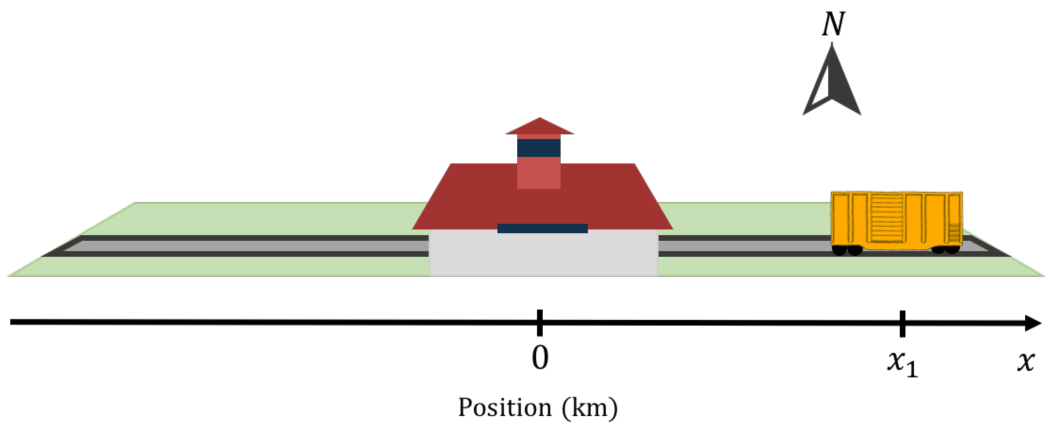


Figure A.1: A 1d coordinate system describing the position of a train. The Kingston train station is the origin and the East side of the track corresponds to positive values of x. The train is located at position x_1.

A.1.2 2D Coordinate systems

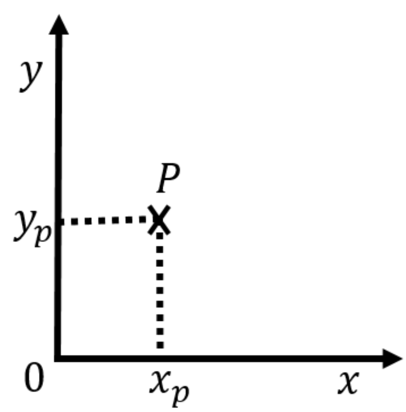


Figure A.2: Example of Cartesian coordinate system and a point P with coordinates (x_p, y_p).

To describe the position of an object in two dimensions (e.g. a marble rolling on a table), we need to specify two numbers. The easiest way to do this is to define two axes, x and

y, whose origin and direction we must define. Figure A.2 shows an example of such a coordinate system. Although it is not necessary to do so, we chose x and y axes that are perpendicular to each other. The origin of the coordinate system is where the two axes intersect. One is free to choose any two directions for the axes (as long as they are not parallel). However, choosing axes that are perpendicular (a "Cartesian" coordinate system) is usually the most convenient.

To fully describe the position of an object, we must specify both its position along the x and y axes. For example, point P in Figure A.2 has two **coordinates**, x_p and y_p, that define its position. The x coordinate is found by drawing a line through P that is parallel to the y axis and is given by the intersection of that line with the x axis. The y coordinate is found by drawing a line through point P that is parallel to the x axis and is given by the intersection of that line with the y axis.

Checkpoint A-1

Figure A.3 shows a coordinate system that is not orthogonal (where the x and y axes are not perpendicular). Which value on the figure correctly indicates the y coordinate of point P?

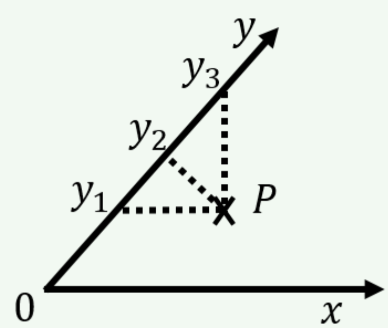


Figure A.3: A non-orthogonal coordinate system (the x and y axes are not perpendicular).

A) y_1
B) y_2
C) y_3

The most common choice of coordinate system in two dimensions is the Cartesian coordinate system that we just described, where the x and y axes are perpendicular and share a common origin, as shown in Figure A.2. When applicable, by convention, we usually choose the y axis to correspond to the vertical direction.

Another common choice is a "polar" coordinate system, where the position of an object is specified by a distance to the origin, r, and an angle, θ, relative to a specified direction, as shown in Figure A.4. Often, a polar coordinate system is defined alongside a Cartesian system, so that r is the distance to the origin of the Cartesian system and θ is the angle with respect to the x axis.

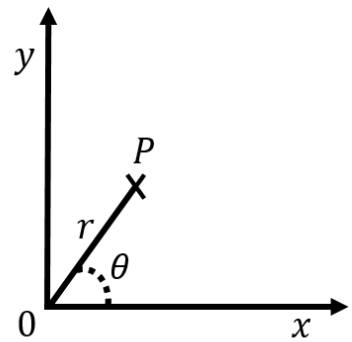


Figure A.4: Example of a polar coordinate system and a point P with coordinates (r, θ).

One can easily convert between the two Cartesian coordinates, x and y, and the two corresponding polar coordinates, r and θ:

$$\begin{aligned} x &= r\cos(\theta) \\ y &= r\sin(\theta) \\ r &= \sqrt{x^2+y^2} \\ \tan(\theta) &= \frac{y}{x} \end{aligned}$$

Polar coordinates are often used to describe the motion of an object moving around a circle, as this means that only one of the coordinates (θ) changes with time (if the origin of the coordinate system is chosen to coincide with the centre of the circle).

A.1.3 3D Coordinate systems

In three dimensions, we need to specify three numbers to describe the position of an object (e.g. a bird flying in the air). In a three dimensional Cartesian coordinate system, we simply add a third axis, z, that is mutually perpendicular to both x and y. The position of an object can then be specified by using the three coordinates, x, y, and z. By convention, we use the z axis to be the vertical direction in three dimensions.

Two additional coordinate systems are common in three dimensions: "cylindrical" and "spherical" coordinates. All three systems are illustrated in Figure A.5 superimposed onto the Cartesian system.

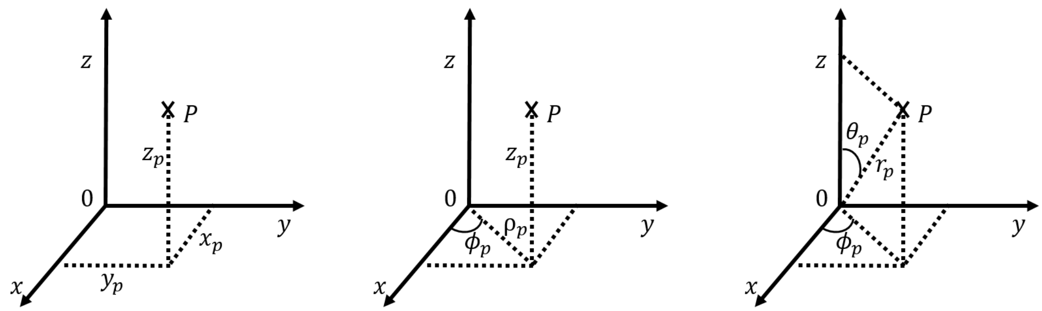


Figure A.5: Cartesian (left), cylindrical (centre) and spherical (right) coordinate systems used in three dimensions. The y and z axes are in the plane of the page, whereas the x axis comes out of the page.

Cylindrical coordinates can be thought of as an extension of the polar coordinates. We keep the same Cartesian coordinate z to indicate the height above the xy plane, however, we use the *azimuthal angle*, ϕ, and the radius, ρ, to describe the position of the projection of a point onto the xy plane. ϕ is the angle between the x axis and the line from the origin to the projection of the point in the xy plane and ρ is the distance between the point and the z axis. Thus, cylindrical coordinates are very similar to the polar coordinate system introduced in two dimensions, except with the addition of the z coordinate. Cylindrical coordinates are useful for describing situations with azimuthal symmetry, such as the motion along the surface of a cylinder. For example, consider point P in Figure A.6. Point P is located a distance ρ from the z axis, as it is located on the surface of the cylinder (the circular end of the cylinder has a radius ρ). Point P is a height z above the xy plane, and a line from the z axis to point P makes an angle ϕ with the x axis.

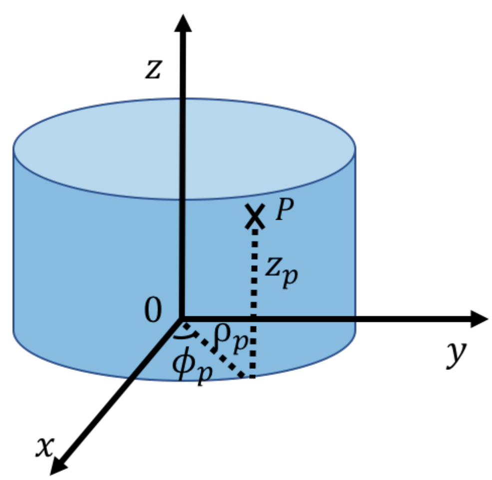


Figure A.6: Describing the position of P, located on the surface of a cylinder, in cylindrical coordinates.

The cylindrical coordinates are related to the Cartesian coordinates by:

$$\begin{aligned}
\rho &= \sqrt{x^2 + y^2} \\
\tan(\phi) &= \frac{y}{x} \\
z &= z
\end{aligned}$$

In spherical coordinates, a point P is described by the radius, r, the *polar angle* θ, and the *azimuthal angle*, ϕ. The radius is the distance between the point and the origin. The polar angle is the angle with the z axis that is made by the line from the origin to the point. The azimuthal angle is defined in the same way as in polar coordinates. Note that the value of ϕ must be between 0 and 2π, whereas the value of θ must be between 0 and π.

Spherical coordinates are useful for describing situations that have spherical symmetry, such as a person walking on the surface of the Earth, since the radial coordinate will not change. For example, this is shown with Point P in Figure A.7, located on the surface of a sphere of radius r.

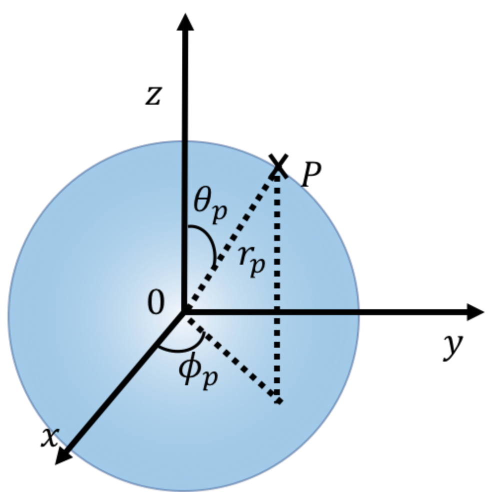


Figure A.7: Describing the position of P, located on the surface of a sphere, in spherical coordinates.

The spherical coordinates are related to the Cartesian coordinates by:

$$\begin{aligned}
r &= \sqrt{x^2 + y^2 + z^2} \\
\cos(\theta) &= \frac{z}{r} = \frac{z}{\sqrt{x^2 + y^2 + z^2}} \\
\tan(\phi) &= \frac{y}{x}
\end{aligned}$$

A.2 Vectors

So far, we have seen how to use a coordinate system to describe the position of a single point in space relative to an origin. In this section, we introduce the notion of a "vector", which allows us to describe quantities that have a **magnitude** and a **direction**. For example, you can use a vector to describe the fact that you walked 5 km in the North direction. A vector can be visualized by an arrow. The length of the arrow is the magnitude that we wish to describe, and the direction of the arrow corresponds to the direction that we would like to describe.

Unlike a point in space, vectors **have no location**. That is, vectors are simply an arrow, and you can choose to draw that arrow anywhere you like. In two dimensional space, one requires two numbers to completely define a vector. In three dimensional space, one requires three numbers to completely define a vector. Figure A.8 shows a two dimensional vector, $\vec{d}$, twice. Because both arrows in the figure have the same magnitude and direction, they represent the *same* vector. When we refer to quantities that are vectors, we usually draw an arrow on top of the quantity ($\vec{d}$) to indicate that they are vectors. We use the word "scalar" to refer to numbers that are not vectors (a regular number is thus also called a scalar to distinguish it from a quantity that is a vector).

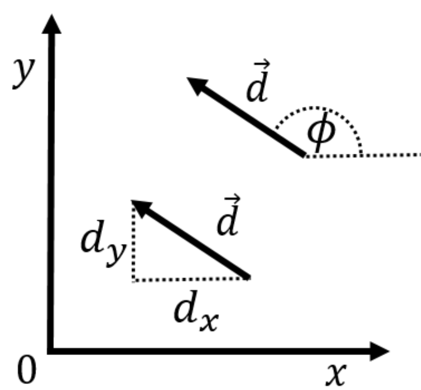


Figure A.8: A vector $\vec{d}$ shown twice, once with its Cartesian components (d_x, d_y) and once with its magnitude and direction (d, ϕ).

In analogy with coordinate systems, we have multiple ways to choose the numbers that we use to describe the vector. The most convenient choice is usually to use the "Cartesian components" of the vector which correspond to the length of the vector when projected onto a Cartesian coordinate system. For example, in Figure A.8, the Cartesian components of the vector $\vec{d}$ are labelled as (d_x, d_y) indicating that the vector has a length of d_x in the x direction and d_y in the y direction. Furthermore, the number d_x is negative, since the vector points in the negative x direction. Another common choice is to use the length of the vector, which we label d (the name of the vector without the arrow on top), and the angle, ϕ that the vector makes with the x-axis, as illustrated in Figure A.8. In terms of the two dimensional Cartesian components, the magnitude of the vector is given by:

$$d = ||\vec{d}|| = \sqrt{d_x^2 + d_y^2}$$

where we also introduced the notation that placing two vertical bars around a vector ($||\vec{d}||$) is used to indicated its magnitude. Note that in three dimensions, it is usually not convenient to specify the direction unless the vector lies in one of the planes defined by the coordinate system (e.g the xy plane). In three dimensions, it is usually most convenient to specify the three Cartesian components.

A.2.1 Unit vectors

A special category of vectors is "unit vectors", which are simply vectors that have a length (magnitude) of 1 (in whichever units the coordinate system is defined). Unit vectors are particularly useful for indicating direction. For example, in Figure A.8, we may be interested in indicating the direction of the vector $\vec{d}$. Unit vectors are denoted by using a "hat" instead of an arrow. Thus, the vector $\hat{d}$, is the vector of length 1 that points in the same direction as $\vec{d}$. The (Cartesian) components of $\hat{d}$ are easily found by dividing the corresponding components of $\vec{d}$ by d (the magnitude):

$$\begin{aligned}
(\hat{d})_x &= \frac{d_x}{d} = \frac{d_x}{\sqrt{d_x^2 + d_y^2}} \\
(\hat{d})_y &= \frac{d_y}{d} = \frac{d_y}{\sqrt{d_x^2 + d_y^2}} \\
\therefore d = ||\hat{d}|| &= \sqrt{(\hat{d})_x^2 + (\hat{d})_y^2} = \sqrt{\frac{d_x^2}{d_x^2 + d_y^2} + \frac{d_y^2}{d_x^2 + d_y^2}} = 1
\end{aligned}$$

A specific type of unit vector is the units vectors that are parallel to the axes of the coordinate system. Those vectors are denoted $\hat{x}$, $\hat{y}$, $\hat{z}$ (and sometimes $\hat{\imath}$, $\hat{\jmath}$, $\hat{k}$ or $\hat{e}_x$, $\hat{e}_y$, $\hat{e}_z$) for the x, y, and z axes, respectively. Thus, the vector $d\hat{x}$, is the vector of length d that points in the positive x direction.

A.2.2 Notations and representation of vectors

There are multiple notations for describing a vector using its components. The following are all equivalent ways to write down the vector $\vec{d}$ in terms of its components d_x and d_y:

$$\begin{aligned}
\vec{d} &= (d_x, d_y) && \text{row vector} \\
&= \begin{pmatrix} d_x \\ d_y \end{pmatrix} && \text{column vector} \\
&= d_x\hat{x} + d_y\hat{y} && \text{using } \hat{x},\ \hat{y} \\
&= d_x\hat{\imath} + d_y\hat{\jmath} && \text{using } \hat{\imath},\ \hat{\jmath}
\end{aligned}$$

The vectors $\hat{x}$ ($\hat{\imath}$) and $\hat{y}$ ($\hat{\jmath}$) are unit vectors in x and y directions respectively. For example, the unit vector $\hat{y}$ can be written down as (0,1) in two dimensions or (0,1,0) in three dimensions, using the row notation.

Checkpoint A-2

What is the magnitude (the length) of the vector $5\hat{x} - 2\hat{y}$?

A) 3.0
B) 5.4
C) 7.0
D) 10.0

Illustrating a vector graphically in two dimensions is straightforward, but difficult in three dimensions. To help remedy this, a notation is introduced in order to draw vectors that point in or out of the page (perpendicular to the plane of the page). The notation comes from imagining that the vector is an archery arrow. If the vector is coming out of the page (at you!), then you would see the head of the arrow, which we represent as a circle with a dot (the dot is the point of the arrow, the circle is the base of the conically shaped arrowhead). If instead, the vector points into the page, then you would see the back of the arrow, which we represent as a cross (the cross being the feathers in the tail of the arrow). This is illustrated in Figure A.9.

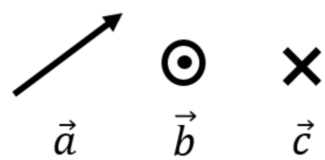


Figure A.9: Geometric representation of three vectors. The vector $\vec{a}$ lies in the plane of the page, the vector $\vec{b}$ is pointing out of the page, and the vector $\vec{c}$ is pointing into the page.

A.3 Vector algebra

In this section, we describe the various algebraic operations that can be performed using vectors.

A.3.1 Multiplication/division of a vector by a scalar

One can multiply (or divide) a vector by a scalar (a number). Suppose that we are given a vector $\vec{v} = (v_x, v_y, v_z)$ and a scalar a. The multiplication $a\vec{v}$ is defined to be a new vector, say $\vec{w}$, whose components are the components of $\vec{v}$ multiplied by a:

$$\vec{w} = a\vec{v} = (av_x, av_y, av_z)$$

Similarly, the division of a vector by a scalar is defined analogously by dividing each Cartesian component by the scalar::

$$\vec{w} = \frac{\vec{v}}{a} = \left(\frac{v_x}{a}, \frac{v_y}{a}, \frac{v_z}{a}\right)$$

Checkpoint A-3

What happens to the length of a vector if the vector is multiplied by 2 (a scalar)?

A) The length doubles
B) The length is halved
C) The length is quadrupled
D) It depends on the direction of the vector

In particular, this makes it easy to determine the unit vector, $\hat{v}$, that points in the same direction as $\vec{v}$:

$$\hat{v} = \frac{\vec{v}}{v}$$

where v is the (scalar) magnitude of $\vec{v}$.

A.3.2 Addition/subtraction of two vectors

The sum of two vectors, $\vec{a}$ and $\vec{b}$, is found by adding the components of the two vectors. Similarly, the difference between two vectors is found by subtracting the components. For example, if $\vec{c} = \vec{a} + \vec{b}$, the components of $\vec{c}$ are given by:

$$\begin{aligned} \vec{c} = \vec{a} + \vec{b} &= \begin{pmatrix} a_x \\ a_y \end{pmatrix} + \begin{pmatrix} b_x \\ b_y \end{pmatrix} \\ \therefore \begin{pmatrix} c_x \\ c_y \end{pmatrix} &= \begin{pmatrix} a_x + b_x \\ a_y + b_y \end{pmatrix} \end{aligned}$$

where we chose to use the "column vector" notation. The column vector notation highlights the fact that the algebra (addition, subtraction) is performed independently on the x and y components. We can thus use write this sum equivalently as two scalar equations, one for each coordinate:

$$\begin{aligned} c_x &= a_x + b_x \\ c_y &= a_y + b_y \end{aligned}$$

Vectors can thus be used as a short-hand notation for representing multiple equations (one equation per component). When we use vectors to write an equation such as:

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

we really mean that there is one scalar equation per component of the vectors:

$$\begin{aligned} F_x &= ma_x \\ F_y &= ma_y \\ F_z &= ma_z \end{aligned}$$

Example A-1

Given two vectors, $\vec{a} = 2\hat{x} + 3\hat{y}$, and $\vec{b} = 5\hat{x} - 2\hat{y}$, calculate the vector $\vec{c} = 2\vec{a} - 3\vec{b}$.

Solution

This can easily be solved algebraically by collecting terms for each component, $\hat{x}$ and $\hat{y}$:

$$\begin{aligned}\vec{c} &= 2\vec{a} - 3\vec{b}\\ &= 2(2\hat{x} + 3\hat{y}) - 3(5\hat{x} - 2\hat{y})\\ &= (4\hat{x} + 6\hat{y}) - (15\hat{x} - 6\hat{y})\\ &= (4 - 15)\hat{x} + (6 + 6)\hat{y}\\ &= -11\hat{x} + 12\hat{y}\end{aligned}$$

We can think of these operations as being performed independently on the components:

$$\begin{aligned}c_x &= 2a_x - 3b_x = -11\\ c_y &= 2a_y - 3b_y = 12\end{aligned}$$

Geometrically, one can easily visualize the addition and subtraction of vectors. This is illustrated in Figure A.10 for the case of adding vectors $\vec{a}$ and $\vec{b}$ to get the vector $\vec{c}$. Geometrically, the sum of the vectors $\vec{a}$ and $\vec{b}$ (sometimes also called the "resultant") can be found by:

1. Placing the "tail" of vector $\vec{b}$ at the "head" of $\vec{a}$ (think of an arrow, the pointy part is the head and the feathery part is the tail)
2. Drawing the vector that goes from the tail of vector $\vec{a}$ to the head of vector $\vec{b}$.

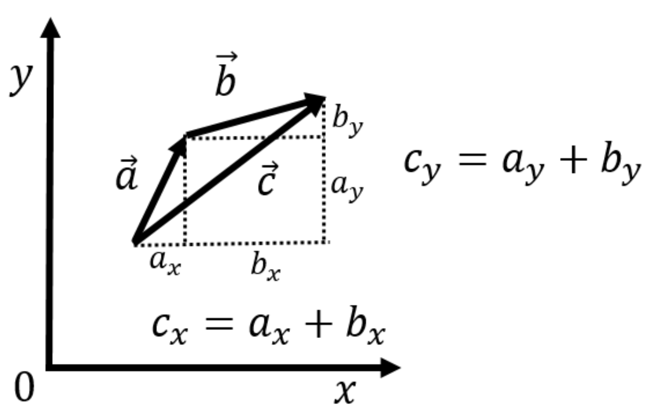


Figure A.10: Geometric addition of the vectors $\vec{a}$ and $\vec{b}$ by placing them "head to tail".

Subtracting two vectors geometrically is done in the same way as addition. For example, the vector $\vec{c}$, given by $\vec{c} = \vec{a} - \vec{b}$ can also be expressed as $\vec{c} = \vec{a} + (-1)\vec{b}$. That is, first multiply the vector $\vec{b}$ by minus 1 (which just reverses its direction), then add that vector, "head to tail", to the vector $\vec{a}$.

Now that we know how to add vectors, we can better understand the notation $\vec{a} = a_x\hat{x} + a_y\hat{y}$. This is not simply a notation, but is in fact algebraically correct. It means: "multiply the vector $\hat{x}$ by a_x (thus giving it a length of a_x) and then add a_y times the vector $\hat{y}$". This is illustrated in Figure A.11, which shows the unit vectors, $\hat{x}$ and $\hat{y}$, which are then multiplied by a_x and a_y, respectively, and then added together "head to tail".

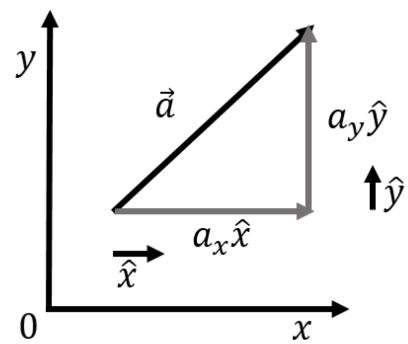


Figure A.11: Illustration that the notation $\vec{a} = a_x\hat{x} + a_y\hat{y}$ is in fact the vector addition of $a_x\hat{x}$ and $a_y\hat{y}$.

A.3.3 The scalar product

There are two ways to "multiply" vectors: the "scalar product" and the "vector product". The scalar product (or "dot product") takes two vectors and results in a scalar (a number). The vector product (or "cross product") takes two vectors and results in a third vector.

The scalar product, $\vec{a} \cdot \vec{b}$, of two vectors $\vec{a}$ and $\vec{b}$, is defined as the following:

$$\vec{a} \cdot \vec{b} = a_x b_x + a_y b_y$$

That is, one multiplies the individual components of the two vectors and then adds those products for each component. This is easily extended to the three dimensional case by adding a term $a_z b_z$ to the sum. The scalar product is also related to the angle between the two vectors when the vectors are placed "tail to tail", as in Figure A.12

$$\vec{a} \cdot \vec{b} = ab\cos\theta$$

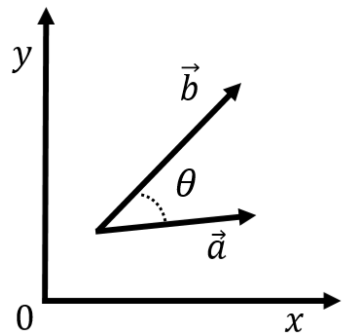


Figure A.12: Illustration of the angle between vectors $\vec{a}$ and $\vec{b}$ when these are placed tail to tail.

The scalar product between two vectors of a fixed length will be maximal when the two vectors are parallel ($\cos\theta = 1$) and zero when the vectors are perpendicular ($\cos\theta = 0$). The scalar product is thus useful when we want to calculate quantities that are maximal when two vectors are parallel.

Checkpoint A-4

The vectors $\vec{a}$ and $\vec{b}$ in the three diagrams below have the same magnitude. Order the diagrams from the one that gives the smallest scalar product $\vec{a} \cdot \vec{b}$ to the largest scalar product.

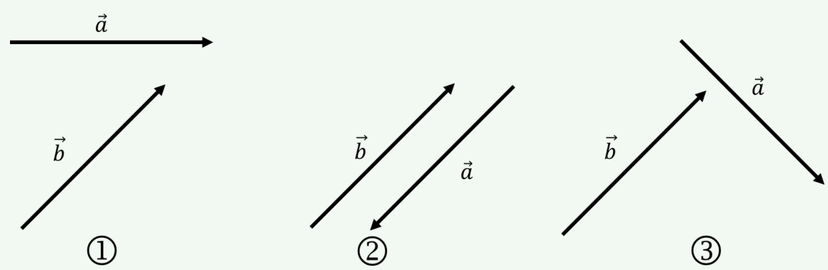


Figure A.13: Put these in order of the magnitude of their scalar product.

A.3.4 The vector product

The vector (or cross) product takes two vectors to produce a third vector that is **mutually perpendicular** to both vectors. The vector product only has meaning in three dimensions. Two vectors that are not co-linear, meaning they can not be arranged so that they lie along the same line, can always be used to define a plane in three dimensions. The cross product of those two vectors will give a third vector that is perpendicular to the plane (making it perpendicular to both vectors).

Algebraically, the three components of the vector product, $\vec{a} \times \vec{b}$, of vectors $\vec{a}$ and $\vec{b}$ are found as follows:

$$\vec{a} \times \vec{b} = \begin{pmatrix} a_y b_z - a_z b_y \\ a_z b_x - a_x b_z \\ a_x b_y - a_y b_x \end{pmatrix} \tag{A.1}$$

One important property to note is that $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$; that is, the cross product is not commutative (the order matters). The magnitude of the vector obtained by a cross product is given by:

$$||\vec{a} \times \vec{b}|| = ab \sin\theta \tag{A.2}$$

where θ is the angle between the vectors $\vec{a}$ and $\vec{b}$ when these are placed tail to tail (Figure A.12). The vector resulting from a cross product will be null (have a zero length) if the vectors $\vec{a}$ and $\vec{b}$ are parallel, and will have a maximal length when these are perpendicular. The cross product is useful to determine quantities that are maximal when two vectors are perpendicular.

Geometrically, one can determine the direction of the cross product of two vectors by using a "right hand rule". To distinguish it from another right hand rule (see Section A.4.3), we will call it "the right hand rule for the cross product"). This is done by using your right hand, aligning your thumb with the first vector and your index with the second vector. The cross product will point in the direction of your middle finger (when you hold your middle finger perpendicular to the other two fingers). This is illustrated in Figure A.14. Thus, you can often avoid using equation A.1 and instead use the right hand rule to determine the direction of the cross product and equation A.2 to find its magnitude.

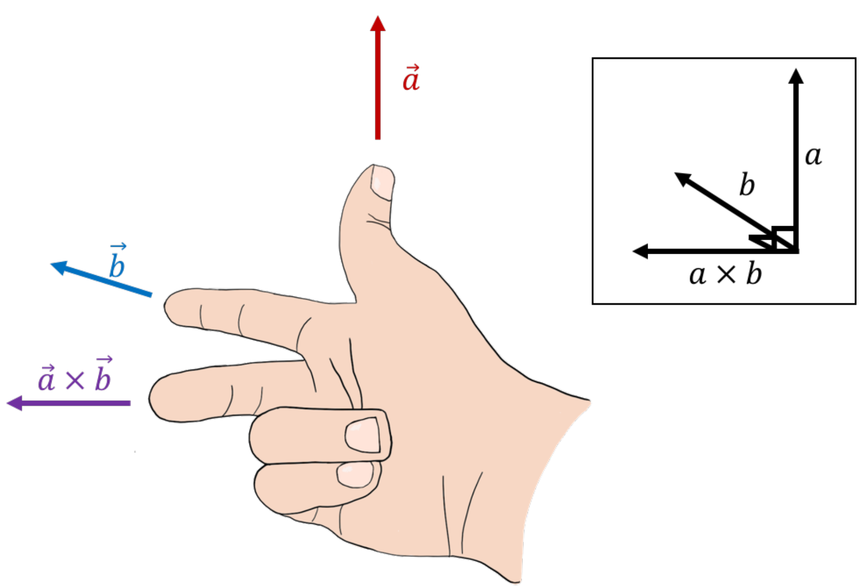


Figure A.14: Using the right hand rule for cross products to find the direction of the cross product of vectors $\vec{a}$ (upwards) and $\vec{b}$ (into the page).

The unit vectors that define a coordinate system have the following properties relative to the cross product:

$$\vec{x} \times \vec{y} = \vec{z}$$
$$\vec{y} \times \vec{z} = \vec{x}$$
$$\vec{z} \times \vec{x} = \vec{y}$$

For these properties to be correct, it should be noted that the direction of the z axis in three dimensions is specified by the choice of x and y axes. That is, one can freely choose the direction of the x and y axes, which then define a plane to which the z axis will be perpendicular. The direction of the z axis must be chosen so that $\vec{x} \times \vec{y} = \vec{z}$ (this guarantees that the coordinate system is "right handed"), as in Figure A.15.

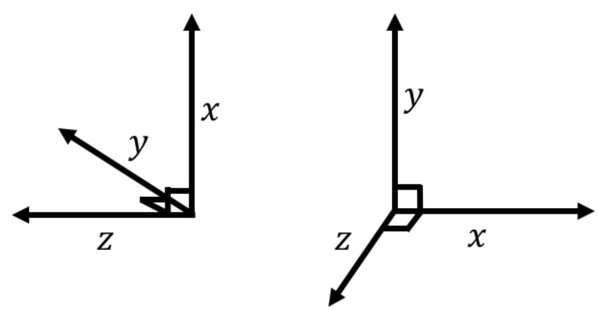


Figure A.15: Two possible orientations for a three dimensional coordinate system. You can confirm using the right hand rule that the z axis is the cross product $\vec{x} \times \vec{y}$.

A.4 Example uses of vectors in physics

This section gives a quick overview of some applications of vectors in physics.

A.4.1 Kinematics and vector equations

Kinematics is the description of the position and motion of an object (Chapters 3 and 4). The laws of physics are the principles that ultimately allow us to determine how the position of an object changes with time. For example, Newton's Laws are a mathematical framework that introduce the concepts of force and mass in order to model and determine how an object will move through space.

We often use a **position vector**, $\vec{r}(t)$, to describe the position of an object as a function of time. Because the object can move, the position vector is a function of time. A position vector is a special vector in the sense that it should be considered to be fixed in space; the position vector for an object points from the origin of a coordinate system to the location of the object.

The three components of the position vector in Cartesian coordinates, are the x, y, and z coordinates of the object:

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

where the three coordinates of the object are functions of time if the object can move. Suppose that the object was initially at position $\vec{r}_1 = (x_1, y_1, z_1)$ at some time $t = t_1$, and that later, at time $t = t_2$, the object was at as second position, $\vec{r}_2 = (x_1, y_1, z_1)$. We can define the **displacement vector**, $\vec{d}$, as the vector from position $\vec{r}_1$ to position $\vec{r}_2$:

$$\vec{d} = \vec{r}_2 - \vec{r}_1 = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{pmatrix} = \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}$$

The displacement vector is such that one can add the vector $\vec{d}$ to the vector $\vec{r}_1$ to describe the new position of the object at time t_2:

$$\begin{aligned} \vec{d} &= \vec{r}_2 - \vec{r}_1 \\ \therefore \vec{r}_2 &= \vec{r}_1 + \vec{d} \end{aligned}$$

The components of the displacement vector, Δx, Δy, and Δz correspond to the displacements (the distance travelled) along the x, y, and z axes, respectively. This is illustrated for the two dimensional case in Figure A.16.

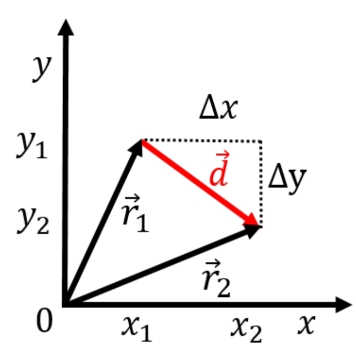


Figure A.16: Illustration of a displacement vector, $\vec{d} = \vec{r}_2 - \vec{r}_1$, for an object that was located at position $\vec{r}_1$ at time t_1 and at position $\vec{r}_2$ at time t_2.

The velocity vector of the object, $\vec{v} = (v_x, v_y, v_z)$, is defined to be the displacement vector, $\vec{d}$, divided by the amount of time (a scalar) that elapsed, $\Delta t = t_2 - t_1$, while the object moved by the corresponding displacement:

$$\vec{v} = \frac{\vec{d}}{\Delta t} = \begin{pmatrix} \frac{\Delta x}{\Delta t} \\ \frac{\Delta y}{\Delta t} \\ \frac{\Delta z}{\Delta t} \end{pmatrix}$$

We used the property that dividing a vector by a scalar (Δt) is defined as dividing each component by the scalar. If we write the components of the velocity vector out explicitly, we have:

$$\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = \begin{pmatrix} \frac{\Delta x}{\Delta t} \\ \frac{\Delta y}{\Delta t} \\ \frac{\Delta z}{\Delta t} \end{pmatrix}$$

That is, we can think of each row in this "vector equation" as an independent equation. That is, when we write the vector equation:

$$\vec{v} = \frac{\vec{d}}{\Delta t}$$

we are really just using a shorthand notation for writing the three **independent** equations that are true for each individual component of the vectors:

$$v_x = \frac{\Delta x}{\Delta t}$$
$$v_y = \frac{\Delta y}{\Delta t}$$
$$v_z = \frac{\Delta z}{\Delta t}$$

Whenever we write an equation using vectors, we are really writing out multiple equations all at once, one for each component. Newton's Second Law:

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

thus corresponds to the three (scalar) equations:

$$F_x = ma_x$$
$$F_y = ma_y$$
$$F_z = ma_z$$

A.4.2 Work and scalar products

As we will see, "work" is a scalar quantity that allows us to determine the change in the speed (squared) of an object that results from a force exerted over a particular displacement (Chapter 7). Both force and the displacement are vector quantities (a force has a magnitude and is exerted in a particular direction). The work, W, done by a force, $\vec{F}$, over a displacements, $\vec{d}$, is defined as:

$$W = \vec{F} \cdot \vec{d}$$

The work energy theorem tells us that this work is related to the change in speed squared of the object as it moves along the displacement vector d. If the work is zero, the object has the same speed at the beginning and end of the displacement. If the work is positive, the object is moving faster at the end of the displacement (and slower if the work is negative). A one dimensional example is shown in Figure A.17, which shows a force $\vec{F}$ being applied to a block as it slides along the ground over a distance d (represented by the displacement vector $\vec{d}$).

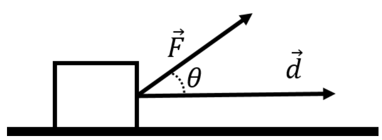


Figure A.17: Example of a force $\vec{F}$ being applied on an object as it moves along the displacement vector $\vec{d}$.

Intuitively, it makes sense that only the horizontal component of the force would contribute to changing the speed of the object as it moves along the horizontal trajectory defined by the vector $\vec{d}$. The vertical component of the force does not contribute to changing the speed of the object. Thus, the work (the change in speed), should only depend on the component of the force that is parallel to the displacement vector. The scalar product allows us to formalize this in an equation. The scalar product is given by:

$$\vec{F} \cdot \vec{d} = Fd\cos\theta = F_{\parallel}d$$

where we introduced $F_{\parallel} = F\cos\theta$ as the component of $\vec{F}$ that is parallel to $\vec{d}$ (see Figure A.17). The scalar product thus "picks out" the component of $\vec{F}$ that is parallel to $\vec{d}$, which is exactly what we need to in order for work to make sense.

A.4.3 Using vectors to describe rotational motion

Often, we need to describe rotational motion in physics. If an object is rotating, one must specify:

1. The axis about which the object is rotating
2. The direction about that axis in which the object is rotating (e.g. clockwise or counter-clockwise)
3. How fast the object is rotating

We introduce a new type of vector, an "axial vector", to describe this kind of rotational motion. We choose the direction of the vector to be co-linear with the axis of rotation and the magnitude of the vector to represent the speed with which the object is rotating. We are thus left with two choices for the direction of the vector. For example, consider the wheels on a car that is moving away from you (Figure A.18, the car is moving into the page). The axis of rotation is the axis of the wheel, so we know that the vector describing the wheel's rotation (the angular velocity vector) must point either to the left or to the right.

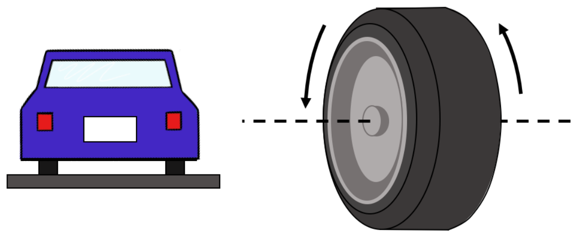


Figure A.18: The wheels on a car that is driving away from you.

We choose the direction of the vector by using another right hand rule. We will refer to this as "the right hand rule for axial vectors" to distinguish it from the right hand rule for the cross product. When using the right hand rule for axial vectors, the vector points in the direction of your thumb when you curl your fingers in the direction of rotation, as in Figure A.19. For the car moving away from you, the wheels will be turning such that the closest point to you is moving up and the furthest point is moving down. Using the right hand rule, we find that the rotation vector points to the left.

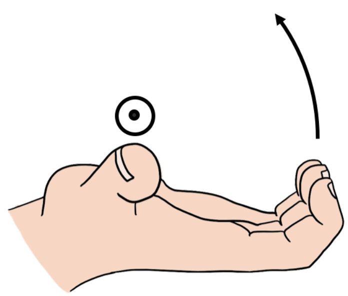


Figure A.19: Using the right hand rule for axial quantities. In this case, the direction of rotation is counter clockwise when looking at the page (the direction that the fingers curl), so the rotation vector points out of the page (the direction of the thumb).

We have to distinguish axial vectors from "true" vectors because they do not behave like true vectors in all cases. For instance, imagine that there is a giant mirror that runs parallel to the road (Figure A.20). When the car is reflected in the mirror, the reflected car will also be moving away from you. This means that the wheels will be turning in the same direction as before, so the rotation vector still points to the left. Now consider a true vector, like a velocity vector. If the velocity vector initially pointed to the left (i.e. if the car was moving to the left), the reflected car would be moving to the *right*. So, we expect a true vector to change directions when it is reflected in this way. Since the rotation vector does not always behave like a true vector, we call it an axial vector or a "pseudovector."

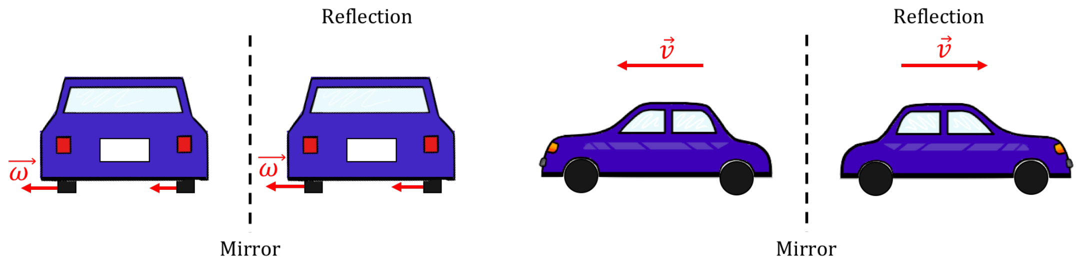


Figure A.20: Left: The angular velocity vector for the rotation of the wheels, $\vec{\omega}$, which points to the left, also points left in the reflection. Right: The velocity vector, pointing to the left, points to the right in the reflection of the car. The angular velocity vector is called an "axial" or "pseudo" vector because it does not change direction under a reflection.

A.4.4 Torque and vector products

We will introduce the concept of a torque in order to describe how a force can cause an object to rotate. Consider the disk illustrated in Figure A.21 that is free to rotate about an axis that goes through its centre and that is perpendicular to the plane of the page. A force $\vec{F}$ is applied at the edge of the disk (imagine pulling on a string attached to the edge of the disk), at a position that is displaced from the axis of rotation by the vector $\vec{r}$. The torque, $\vec{\tau}$, of the force about the centre of the disk is defined to be:

$$\vec{\tau} = \vec{r} \times \vec{F}$$

and represents how much the force $\vec{F}$ will contribute to making the disk rotate about its axis. If the force vector were parallel to the vector $\vec{r}$, the disk would not rotate; if you pull outwards on a disk, it will not rotate about its centre. However, if the force is perpendicular to the vector $\vec{r}$ (i.e. tangent to the circumference of the disk), then it will maximally cause the disk to rotate. The magnitude of the torque (cross-product) is given by:

$$\tau = rF\sin\theta = F_{\perp}r = Fr_{\perp}$$

where θ is the angle between the vectors when placed tail to tail, as in the right side of Figure A.21. In the last two equalities, we have defined $F_{\perp} = F\sin\theta$ or $r_{\perp} = r\sin\theta$ to refer to the part of the vector $\vec{F}$ that is perpendicular to the vector $\vec{r}$ or the part of the vector $\vec{r}$ that is perpendicular to the vector $\vec{F}$. That is, the vector product "picks out" the part of a vector that is perpendicular to the other, which is exactly the property that we need for the physical quantity of torque.

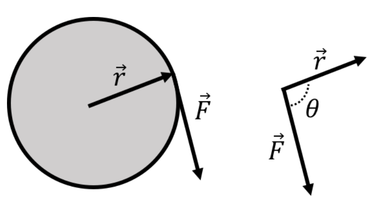


Figure A.21: A force, $\vec{F}$, is exerted in the plane of a disk at a position given by the vector $\vec{r}$ relative to the centre of the disk.

Checkpoint A-5

Referring to Figure A.21, in which direction does the torque vector point?

A) to the right
B) to the left
C) out of the page
D) into the page

A.5 Summary

Key Takeaways

Cartesian coordinate systems can be defined using an origin, and mutually perpendicular axes that specify a direction in which each corresponding coordinate increases. The position of a point is described by the coordinates of the point (one coordinate per axis). Polar, cylindrical and spherical coordinate systems can be defined relative to a Cartesian coordinate system and sometimes facilitate the description of situations with cylindrical (azimuthal) or spherical symmetry.

Vectors can be represented by arrows and are quantities that have both a magnitude and a direction, as opposed to "scalars", which are simply numbers. Vectors are not fixed in space, so two vectors are equal if they have the same magnitude and direction, regardless of where they are drawn. We place a little arrow above a variable, $\vec d$, to indicate that it is a vector. There are several, equivalent, notations to indicate the components of a vector:

$$\begin{aligned}
\vec d &= (d_x, d_y, d_z) && \text{row vector}\\
&= \begin{pmatrix} d_x \\ d_y \\ d_z \end{pmatrix} && \text{column vector}\\
&= d_x\hat x + d_y\hat y + d_z\hat z && \text{using } \hat x,\ \hat y,\ \hat z\\
&= d_x\hat\imath + d_y\hat\jmath + d_z\hat\jmath && \text{using } \hat\imath,\ \hat\jmath,\ \hat k
\end{aligned}$$

If we multiply (divide) a vector by a scalar, we multiply (divide) each component of the vector individually by that quantity. As a result, the magnitude of the vector will also be multiplied (divided) by that quantity:

$$a\vec d = \begin{pmatrix} ad_x \\ ad_y \\ ad_z \end{pmatrix}$$

In particular, we can define a unit vector, $\hat d$, to be a vector of length 1 in the same direction as $\vec d$, by simply dividing $\vec d$ by its magnitude, d:

$$\hat d = \frac{\vec d}{d}$$

where the magnitude of the vector, $||\vec d|| = d$, expressed in Cartesian coordinates, is given by:

$$||\vec{d}|| = d = \sqrt{d_x^2 + d_y^2 + d_z^2}$$

We can add two vectors by independently adding the individual components of the vectors:

$$\begin{aligned}\vec{c} &= \vec{a} + \vec{b}\\ \therefore c_x &= a_x + b_x\\ \therefore c_y &= a_y + b_y\\ \therefore c_z &= a_z + b_z\end{aligned}$$

Graphically, this corresponds to adding vectors "head to tail". This also highlights that an equation written using vectors (as the first line above) really represents three independent equations, one for each coordinate of the vectors (or two in two dimensions). Subtraction of vectors is treated in the same way as addition (but using minus signs where appropriate).

One can define the scalar (or dot) product between two vectors, as a scalar quantity that is obtained from the two vectors:

$$\vec{a} \cdot \vec{b} = a_x b_x + a_y b_y + a_z b_z$$

The scalar product is also related to the angle, θ, between the two vectors when these are placed "tail to tail":

$$\vec{a} \cdot \vec{b} = ab\cos\theta$$

In particular, the scalar product between two vectors is zero if the two vectors are perpendicular to each other ($\cos\theta = 0$), and maximal when these are parallel to each other.

The vector (or cross) product between two vectors is a vector that is mutually perpendicular to both vectors and is defined as the following:

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

The vector product can only be defined in three dimensions, since it must be mutually perpendicular to the vectors. The magnitude of the vector product is given by:

$$||\vec{a} \times \vec{b}|| = ab\sin\theta$$

where θ is the angle between the two vectors when these are placed tail to tail. In particular, the vector product between two vectors is zero if the two vectors are parallel to each other (and maximal when these are perpendicular). The direction of the vector product is given by the right-hand rule for the cross product.

An axial vector can be used to describe a quantity that is related to rotation. The direction of the axial vector is co-linear with the axis of rotation, its magnitude is given by the magnitude of the rotational quantity (e.g. angular speed), and its direction is defined using the right-hand rule for axial vectors.

A.6 Thinking about the Material

Reflect and research

1. What are some quantities that need to be represented by a vector?
2. Can a vector in three dimensions be represented using spherical coordinates? How would you calculate the scalar product between two vectors represented in spherical coordinates?

A.7 Sample problems and solutions

A.7.1 Problems

Problem A-1: (Solution)

a) What is the displacement vector from position $(1, 2, 3)$ to position $(4, 5, 6)$?
b) What angle does that displacement vector make with the x axis?

A.7.2 Solutions

Solution to problem A-1:

a) The displacement vector is given by:

$$\vec{d} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}$$

b) We can find the angle that this vector makes with the x axis by taking the scalar product of the displacement vector and the unit vector in the x direction (1,0,0):

$$\hat{x} \cdot \vec{d} = (1)(3) + (0)(3) + (0)(3) = 3$$

This is equal to the product of the magnitude of $\hat{x}$ and $\vec{d}$ multiplied by the cosine of the angle between them:

$$\begin{aligned} \hat{x} \cdot \vec{d} &= ||\hat{x}||||\vec{d}|| \cos\theta = (1)(\sqrt{3^2 + 3^2 + 3^2}) \cos\theta = \sqrt{27} \cos\theta \\ 3 &= \sqrt{27} \cos\theta \\ \therefore \cos\theta &= \frac{3}{\sqrt{27}} = \frac{1}{\sqrt{3}} \\ \theta &= 54.7^\circ \end{aligned}$$

B Calculus

This appendix gives a very brief introduction to calculus with a focus on the tools needed in physics.

Learning Objectives

- Understand how to determine a derivative and that it measures a rate of change.
- Understand how to determine partial derivatives and gradients.
- Understand how to determine anti-derivatives and that integrals are sums.

B.1 Functions of real numbers

In calculus, we work with functions and their properties, rather than with variables as we do in algebra. We are usually concerned with describing functions in terms of their slope, the area (or volumes) that they enclose, their curvature, their roots (when they have a value of zero) and their continuity. The functions that we will examine are a mapping from one or more *independent* real numbers to one real number. By convention, we will use x, y, z to indicate independent variables, and $f()$ and $g()$, to denote functions. For example, if we say:

$$\begin{aligned} f(x) &= x^2 \\ \therefore f(2) &= 4 \end{aligned}$$

we mean that $f(x)$ is a function that can be evaluated for any real number, x, and the result of evaluating the function is to square the number x. In the second line, we evaluated the function with $x = 2$. Similarly, we can have a function, $g(x, y)$ of multiple variables:

$$\begin{aligned} g(x, y) &= x^2 + 2y^2 \\ \therefore g(2, 3) &= 22 \end{aligned}$$

We can easily visualize a function of 1 variable, for example by plotting it in python (see Appendix D):

Python Code B.1: Plotting a function of 1 variable

```
#import pacakges for creating arrays of values and for plotting
import numpy as np #arrays
import pylab as pl #plotting

#define the function:
def f(x):
    return x*x

#create 100 values of x between -5 and +5
xvals = np.linspace(-5,5,100)

#Plot the function evaluated at the values of x against the values of x:
pl.plot(xvals,f(xvals))
pl.xlabel('x')
pl.ylabel('f(x)')
pl.title('f(x)=$x^2$')
pl.grid()
pl.show()
```

Output B.1:

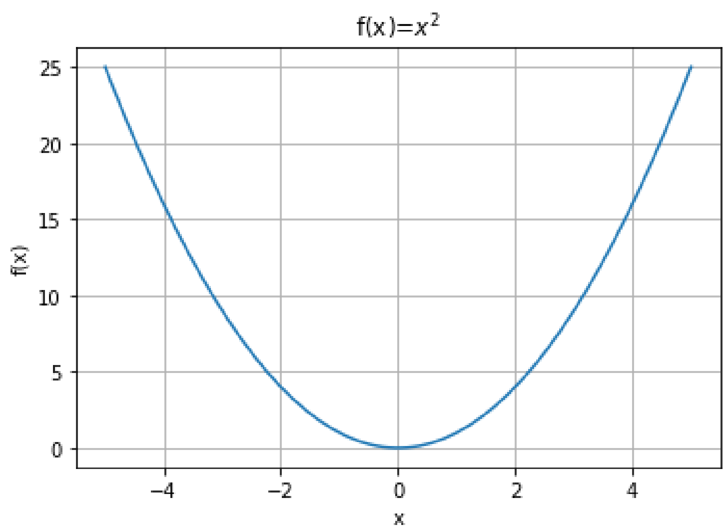


Figure B.1: $f(x) = x^2$ plotted between $x = -5$ and $= +5$.

Plotting a function of 2 variables is a little trickier, since we need to do it in three dimensions (one axis for x, one axis for y, and one axis for $g(x, y)$). This can be done in python with a little more work:

Python Code B.2: Plotting a function of 2 variables

```
#import pacakges for creating arrays of values and for plotting
import numpy as np #arrays
import pylab as pl #plotting
```

```python
#import package for handling 3D graphs:
from mpl_toolkits.mplot3d import Axes3D

#define the function:
def g(x,y):
    return x*x+2*y*y

#create 100 values of x and y between -5 and +5
xvals = np.linspace(-5,5,100)
yvals = np.linspace(-5,5,100)
#create a grid with the values of x and y:
X,Y = np.meshgrid(xvals,yvals)
#evaluate the function everywhere on the grid
gvals = g(X,Y)

#Plot the function as a surface (create a figure, add 3D, plot it):
fig = pl.figure(figsize=(10,10))
ax = fig.add_subplot(111, projection='3d')
ax.plot_surface(X,Y,gvals,cmap="Blues")
#show contours for the surface, projected on xy plane:
ax.contour(X, Y, gvals,offset=-1,cmap="Blues")
#add some labels
ax.set_xlabel('x')
ax.set_ylabel('y')
ax.set_zlabel('g(x,y)')
ax.set_title("$g(x,y)=x^2+2y^2$")
#choose the view point:
ax.view_init(elev=30, azim=-25)
pl.show()
```

Output B.2:

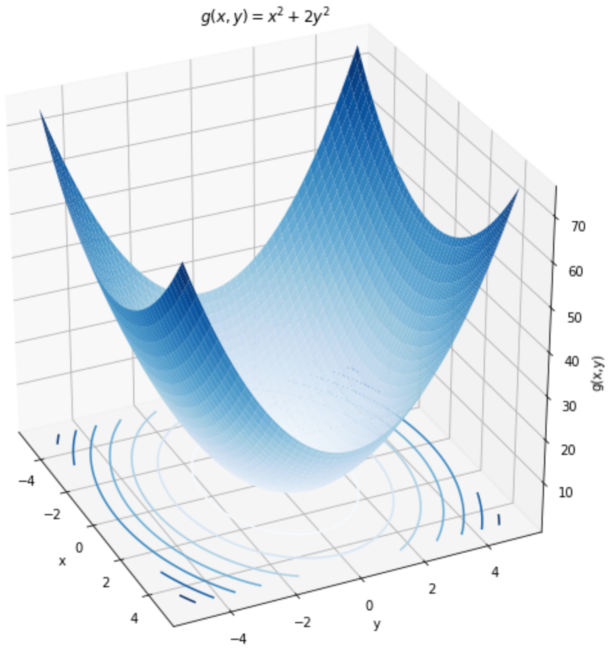


Figure B.2: $g(x, y) = x^2 + 2y^2$ plotted for x between -5 and +5 and for y between -5 and +5. A function of two variables can be visualized as a surface in three dimensions. One can also visualize the function by look at its "contours" (the lines drawn in the xy plane).

Unfortunately, it becomes difficult to visualize functions of more than 2 variables, although one can usually look at projections of those functions to try and visualize some of the features (for example, contour maps are 2D projections of 3D surfaces, as shown in the xy plane of Figure B.2). When you encounter a function, it is good practice to try and visualize it if you can. For example, ask yourself the following questions:

- Does the function have one or more maxima and/or minima?
- Does the function cross zero?
- Is the function continuous everywhere?
- Is the function always defined for any value of the independent variables?

B.2 Derivatives

Consider the function $f(x) = x^2$ that is plotted in Figure B.1. For any value of x, we can define the slope of the function as the "steepness of the curve". For values of $x > 0$ the function increases as x increases, so we say that the slope is positive. For values of $x < 0$, the function decreases as x increases, so we say that the slope is negative. A synonym for the word slope is "derivative", which is the word that we prefer to use in calculus. The derivative of a function $f(x)$ is given the symbol $\frac{df}{dx}$ to indicate that we are referring to the slope of $f(x)$ when plotted as a function of x.

We need to specify which variable we are taking the derivative with respect to when the function has more than one variable but only one of them should be considered *independent*.

For example, the function $f(x) = ax^2 + b$ will have different values if a and b are changed, so we have to be precise in specifying that we are taking the derivative with respect to x. The following notations are equivalent ways to say that we are taking the derivative of $f(x)$ with respect to x:

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

The notation with the prime ($f'(x), f'$) can be useful to indicate that the derivative itself is *also* a function of x.

The slope (derivative) of a function tells us how rapidly the value of the function is changing when the independent variable is changing. For $f(x) = x^2$, as x gets more and more positive, the function gets steeper and steeper; the derivative is thus increasing with x. The sign of the derivative tells us if the function is increasing or decreasing, whereas its absolute value tells how quickly the function is changing (how steep it is).

We can approximate the derivative by evaluating how much $f(x)$ changes when x changes by a small amount, say, Δx. In the limit of $\Delta x \to 0$, we get the derivative. In fact, this is the formal definition of the derivative:

$$\boxed{\frac{df}{dx} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}} \tag{B.1}$$

where Δf is the small change in $f(x)$ that corresponds to the small change, Δx, in x. This makes the notation for the derivative more clear, dx is Δx in the limit where $\Delta x \to 0$, and df is Δf, in the same limit of $\Delta x \to 0$.

As an example, let us determine the function $f'(x)$ that is the derivative of $f(x) = x^2$. We start by calculating Δf:

$$\begin{aligned}
\Delta f &= f(x + \Delta x) - f(x) \\
&= (x + \Delta x)^2 - x^2 \\
&= x^2 + 2x\Delta x + \Delta x^2 - x^2 \\
&= 2x\Delta x + \Delta x^2
\end{aligned}$$

We now calculate $\frac{\Delta f}{\Delta x}$:

$$\begin{aligned}
\frac{\Delta f}{\Delta x} &= \frac{2x\Delta x + \Delta x^2}{\Delta x} \\
&= 2x + \Delta x
\end{aligned}$$

and take the limit $\Delta x \to 0$:

$$\begin{aligned}
\frac{df}{dx} &= \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} \\
&= \lim_{\Delta x \to 0} (2x + \Delta x) \\
&= 2x
\end{aligned}$$

We have thus found that the function, $f'(x) = 2x$, is the derivative of the function $f(x) = x^2$. This is illustrated in Figure B.3. Note that:

- For $x > 0$, $f'(x)$ is positive and increasing with increasing x, just as we described earlier (the function $f(x)$ is increasing and getting steeper).
- For $x < 0$, $f'(x)$ is negative and decreasing in magnitude as x increases. Thus $f(x)$ decreases and gets less steep as x increases.
- At $x = 0$, $f'(x) = 0$ indicating that, at the origin, the function $f(x)$ is (momentarily) flat.

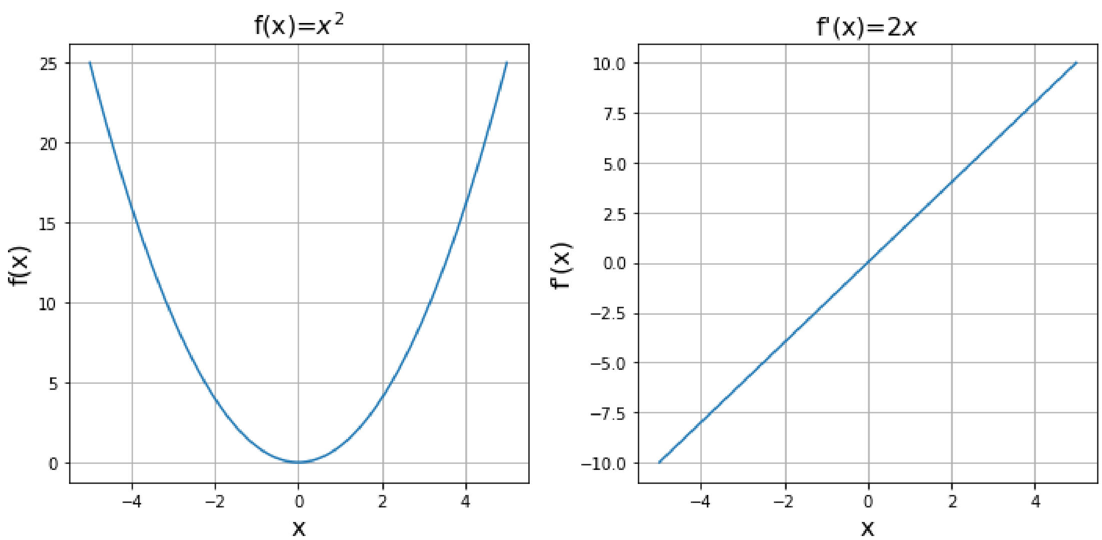


Figure B.3: $f(x) = x^2$ and its derivative, $f'(x) = 2x$ plotted for x between -5 and +5.

Checkpoint B-1

When a function has a maximum, its derivative at that point

A) also has a maximum
B) is zero
C) has a minimum
D) is infinite

B.2.1 Common derivatives and properties

It is beyond the scope of this document to derive the functional form of the derivative for any function using equation B.1. Table B.1 below gives the derivatives for common functions. In all cases, x is the independent variable, and all other variables should be thought of as constants:

Function, $f(x)$	**Derivative,** $f'(x)$
$f(x) = a$	$f'(x) = 0$
$f(x) = x^n$	$f'(x) = nx^{n-1}$
$f(x) = \sin(x)$	$f'(x) = \cos(x)$
$f(x) = \cos(x)$	$f'(x) = -\sin(x)$
$f(x) = \tan(x)$	$f'(x) = \frac{1}{\cos^2(x)}$
$f(x) = e^x$	$f'(x) = e^x$
$f(x) = \ln(x)$	$f'(x) = \frac{1}{x}$

Table B.1: Common derivatives of functions.

If two functions of 1 variable, $f(x)$ and $g(x)$, are combined into a third function, $h(x)$, then there are simple rules for finding the derivative, $h'(x)$, based on the derivatives $f'(x)$ and $g'(x)$. These are summarized in Table B.2 below.

Function, $h(x)$	**Derivative,** $h'(x)$
$h(x) = f(x) + g(x)$	$h'(x) = f'(x) + g'(x)$
$h(x) = f(x) - g(x)$	$h'(x) = f'(x) - g'(x)$
$h(x) = f(x)g(x)$	$h'(x) = f'(x)g(x) + f(x)g'(x)$ (The product rule)
$h(x) = \frac{f(x)}{g(x)}$	$h'(x) = \frac{f'(x)g(x)-f(x)g'(x)}{g^2(x)}$ (The quotient rule)
$h(x) = f(g(x))$	$h'(x) = f'(g(x))g'(x)$ (The Chain Rule)

Table B.2: Derivatives of combined functions.

Example B-1

Use the properties from Table B.2 to show that the derivative of $\tan(x)$ is $\frac{1}{\cos^2(x)}$

Solution

Since $\tan(x) = \frac{\sin(x)}{\cos(x)}$, we can write:

$$h(x) = \frac{f(x)}{g(x)}$$
$$f(x) = \sin(x)$$
$$g(x) = \cos(x)$$

Using the fourth row in Table B.2, and the common derivatives from Table B.1, we have:

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

as required.

Example B-2

Use the properties from Table B.2 to calculate the derivative of $h(x) = \sin^2(x)$

Solution

To calculate the derivative of $h(x)$, we need to use the Chain Rule. $h(x)$ is found by first taking $\sin(x)$ and then taking that result squared. We can thus identify:

$$\begin{aligned}
h(x) &= \sin^2(x) = f(g(x)) \\
f(x) &= x^2 \\
g(x) &= \sin(x)
\end{aligned}$$

Using the common derivatives from Table B.1, we have:

$$\begin{aligned}
f'(x) &= 2x \\
g'(x) &= \cos(x)
\end{aligned}$$

Applying the Chain Rule, we have:

$$\begin{aligned}
h'(x) &= f'(g(x))g'(x) \\
&= 2\sin(x)g'(x) \\
&= 2\sin(x)\cos(x)
\end{aligned}$$

where $f'(g(x))$ means apply the derivative of $f(x)$ to the function $g(x)$. Since the derivative of $f(x)$ is $f'(x) = 2x$, when we apply it to $g(x)$ instead of $2x$, we get $2g(x) = 2\cos(x)$.

B.2.2 Partial derivatives and gradients

So far, we have only looked at the derivative of a function of a single independent variable and used it to quantify how much the function changes when the independent variable changes. We can proceed analogously for a function of multiple variables, $f(x, y)$, by quantifying how much the function changes along the direction associated with a particular variable. This is illustrated in Figure B.4 for the function $f(x, y) = x^2 - 2y^2$, which looks somewhat like a saddle.

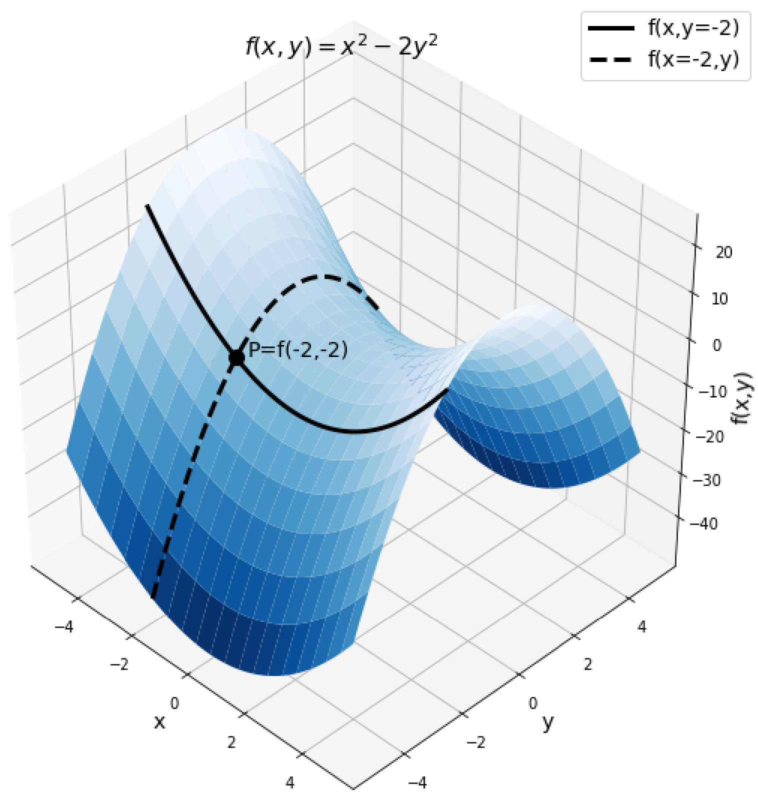


Figure B.4: $f(x, y) = x^2 - 2y^2$ plotted for x between -5 and +5 and for y between -5 and +5. The point P labelled on the figure shows the value of the function at $f(-2, -2)$. The two lines show the function evaluated when one of x or y is held constant.

Suppose that we wish to determine the derivative of the function $f(x)$ at $x = -2$ and $y = -2$. In this case, it does not make sense to simply determine the "derivative", but rather, we must specify *in which direction* we want the derivative. That is, we need to specify in which direction we are interested in quantifying the rate of change of the function.

One possibility is to quantify the rate of change in the x direction. The solid line in Figure B.4 shows the part of the function surface where y is fixed at -2, that is, the function evaluated as $f(x, y = -2)$. The point P on the figure shows the value of the function when $x = -2$ and $y = -2$. By looking at the solid line at point P, we can see that as x increases, the value of the function is gently decreasing. The derivative of $f(x, y)$ with respect to x when y is held constant and evaluated at $x = -2$ and $y = -2$ is thus negative. Rather than saying "The derivative of $f(x, y)$ with respect to x when y is held constant" we say "The **partial derivative** of $f(x, y)$ with respect to x".

Since the partial derivative is different than the ordinary derivative (as it implies that we are holding independent variables fixed), we give it a different symbol, namely, we use ∂ instead of d:

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x, y) \text{ (Partial derivative of f with respect to x)}$$

Calculating the partial derivative is very easy, as we just treat all variables as constants except for the variable with respect to which we are differentiating[1]. For the function $f(x, y) = x^2 - 2y^2$, we have:

$$\begin{aligned}
\frac{\partial f}{\partial x} &= \frac{\partial}{\partial x}(x^2 - 2y^2) = 2x \\
\frac{\partial f}{\partial y} &= \frac{\partial}{\partial y}(x^2 - 2y^2) = -4y
\end{aligned}$$

At $x = -2$, the partial derivative of $f(x, y)$ is indeed negative, consistent with our observation that, along the solid line, at point P, the function is decreasing.

A function will have as many partial derivatives as it has independent variables. Also note that, just like a normal derivative, a partial derivative is still a function. The partial derivative with respect to a variable tells us how steep the function is in the direction in which that variable increases and whether it is increasing or decreasing.

Example B-3

Determine the partial derivatives of $f(x, y, z) = ax^2 + byz - \sin(z)$.

Solution

In this case, we have three partial derivatives to evaluate. Note that a are b constants

[1] To take the derivative is to "differentiate"!

and can be thought of as numbers that we do not know.

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}(ax^2 + byz - \sin(z)) = 2ax$$
$$\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}(ax^2 + byz - \sin(z)) = bz$$
$$\frac{\partial f}{\partial z} = \frac{\partial}{\partial z}(ax^2 + byz - \sin(z)) = by - \cos(z)$$

Since the partial derivatives tell us how the function changes in a particular direction, we can use them to find the direction in which the function changes *the most rapidly*. For example, suppose that the surface from Figure B.4 corresponds to a real physical surface and that we place a ball at point P. We wish to know in which direction the ball will roll. The direction that it will roll in is the opposite of the direction where $f(x, y)$ increases the most rapidly (i.e. it will roll in the direction where $f(x, y)$ decreases the most rapidly). The direction in which the function increases the most rapidly is called the "gradient" and denoted by $\nabla f(x, y)$.

Since the gradient is a direction, it cannot be represented by a single number. Rather, we use a "vector" to indicate this direction. Since $f(x, y)$ has two independent variables, the gradient will be a vector with two components. The components of the gradient are given by the partial derivatives:

$$\nabla f(x, y) = \frac{\partial f}{\partial x}\hat{x} + \frac{\partial f}{\partial y}\hat{y}$$

where $\hat{x}$ and $\hat{y}$ are the unit vectors in the x and y directions, respectively (sometimes, the unit vectors are denoted $\hat{i}$ and $\hat{j}$). The direction of the gradient tells us in which direction the function increases the fastest, and the magnitude of the gradient tells us how much the function increases in that direction.

Example B-4

Determine the gradient of the function $f(x, y) = x^2 - 2y^2$ at the point $x = -2$ and $y = -2$.

Solution

We have already found the partial derivatives that we need to evaluate at $x = -2$ and

$y = -2$:

$$\frac{\partial f}{\partial x} = 2x$$
$$\frac{\partial f}{\partial y} = -4y$$
$$\therefore \nabla f(x, y) = \frac{\partial f}{\partial x}\hat{x} + \frac{\partial f}{\partial y}\hat{y}$$
$$= 2x\hat{x} - 4y\hat{y}$$

Evaluating the gradient at $x = -2$ and $y = -2$:

$$\begin{aligned}\nabla f(x, y) &= 2x\hat{x} - 4y\hat{y}\\ &= -4\hat{x} + 8\hat{y}\\ &= 4(-\hat{x} + 2\hat{y})\end{aligned}$$

The gradient vector points in the direction $(-1, 2)$. That is, the function increases the most in the direction where you would take 1 pace in the negative x direction and 2 paces in the positive y direction. You can confirm this by looking at point P in Figure B.4 and imagining in which direction you would have to go to climb the surface to get the steepest climb.

The gradient is itself a function, but it is not a real function (in the sense of a real number), since it evaluates to a vector. It is a mapping from real numbers x, y to a vector. As you take more advanced calculus courses, you will eventually encounter "vector calculus", which is just the calculus for functions of multiple variables to which you were just introduced. The key point to remember here is that the gradient can be used to find the vector that points in the direction of maximal increase of the corresponding multi-variate function. This is precisely the quantity that we need in physics to determine in which direction a ball will roll when placed on a surface (it will roll in the direction opposite to the gradient vector).

Checkpoint B-2

The gradient of a function of one variable, $f(x)$, is

A) undefined
B) zero
C) equal to its derivative
D) infinite

B.2.3 Common uses of derivatives in physics

The simplest case of using a derivative is to describe the speed of an object. If an object covers a distance Δx in a period of time Δt, it's "average speed", v_{avg}, is defined as the distance covered by the object divided by the amount of time it took to cover that distance:

$$v_{avg} = \frac{\Delta x}{\Delta t}$$

If the object changes speed (for example it is slowing down) over the distance Δx, we can still define its "instantaneous speed", v, by measuring the amount of time, Δt, that it takes the object to cover a *very small distance*, Δx. The instantaneous speed is defined in the limit where $\Delta x \to 0$:

$$v = \lim_{\Delta x \to 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt}$$

which is precisely the derivative of $x(t)$ with respect to t. $x(t)$ is a function that gives the position, x, of the object along some x axis as a function of time. The speed of the object is thus the rate of change of its position.

Similarly, if the speed is changing with time, then we can define the "acceleration", a, of an object as the rate of change of its speed:

$$a = \frac{dv}{dt}$$

B.3 Anti-derivatives and integrals

In the previous section, we were concerned with determining the derivative of a function $f(x)$. The derivative is useful because it tells us how the function $f(x)$ varies as a function of x. In physics, we often know how a function varies, but we do not know the actual function. In other words, we often have the opposite problem: we are given the derivative of a function, and wish to determine the actual function. For this case, we will limit our discussion to functions of a single independent variable.

Suppose that we are given a function $f(x)$ and we know that this is the derivative of some other function, $F(x)$, which we do not know. We call $F(x)$ the **anti-derivative** of $f(x)$. The anti-derivative of a function $f(x)$, written $F(x)$, thus satisfies the property:

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

Since we have a symbol for indicating that we take the derivative with respect to x ($\frac{d}{dx}$), we also have a symbol, $\int dx$, for indicating that we take the anti-derivative with respect to x:

$$\int f(x)dx = F(x)$$
$$\therefore \frac{d}{dx}\left(\int f(x)dx\right) = \frac{dF}{dx} = f(x)$$

Earlier, we justified the symbol for the derivative by pointing out that it is like $\frac{\Delta f}{\Delta x}$ but for the case when $\Delta x \to 0$. Similarly, we will justify the anti-derivative sign, $\int f(x)dx$, by showing that it is related to a sum of $f(x)\Delta x$, in the limit $\Delta x \to 0$. The $\int$ sign looks like an "S" for sum.

While it is possible to exactly determine the derivative of a function $f(x)$, the anti-derivative can only be determined up to a constant. Consider for example a different function, $\tilde{F}(x) = F(x) + C$, where C is a constant. The derivative of $\tilde{F}(x)$ with respect to x is given by:

$$\begin{aligned}
\frac{d\tilde{F}}{dx} &= \frac{d}{dx}\left(F(x) + C\right) \\
&= \frac{dF}{dx} + \frac{dC}{dx} \\
&= \frac{dF}{dx} + 0 \\
&= f(x)
\end{aligned}$$

Hence, the function $\tilde{F}(x) = F(x) + C$ is also an anti-derivative of $f(x)$. The constant C can often be determined using additional information (sometimes called "initial conditions"). Recall the function, $f(x) = x^2$, shown in Figure B.3 (left panel). If you imagine shifting the whole function up or down, the derivative would not change. In other words, if the origin of the axes were not drawn on the left panel, you would still be able to determine the derivative of the function (how steep it is). Adding a constant, C, to a function is exactly the same as shifting the function up or down, which does not change its derivative. Thus, when you know the derivative, you cannot know the value of C, unless you are also told that the function must go through a specific point (a so-called initial condition).

In order to determine the derivative of a function, we used equation B.1. We now need to derive an equivalent prescription for determining the anti-derivative. Suppose that we have the two pieces of information required to determine $F(x)$ completely, namely:

1. the function $f(x) = \frac{dF}{dx}$ (its derivative).
2. the condition that $F(x)$ must pass through a specific point, $F(x_0) = F_0$.

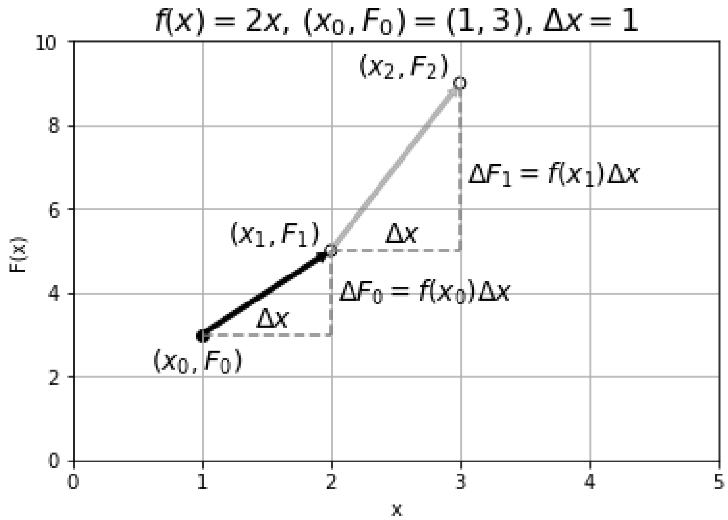


Figure B.5: Determining the anti-derivative, $F(x)$, given the function $f(x) = 2x$ and the initial condition that $F(x)$ passes through the point $(x_0, F_0) = (1, 3)$.

The procedure for determining the anti-derivative $F(x)$ is illustrated above in Figure B.5. We start by drawing the point that we know the function $F(x)$ must go through, (x_0, F_0). We then choose a value of Δx and use the derivative, $f(x)$, to calculate ΔF_0, the amount by which $F(x)$ changes when x changes by Δx. Using the derivative $f(x)$ evaluated at x_0, we have:

$$\begin{aligned}\frac{\Delta F_0}{\Delta x} &\approx f(x_0) \quad (\text{in the limit} \Delta x \to 0)\\ \therefore \Delta F_0 &= f(x_0)\Delta x\end{aligned}$$

We can then estimate the value of the function $F_1 = F(x_1)$ at the next point, $x_1 = x_0 + \Delta x$, as illustrated by the black arrow in Figure B.5

$$\begin{aligned}F_1 &= F(x_1)\\ &= F(x + \Delta x)\\ &\approx F_0 + \Delta F_0\\ &\approx F_0 + f(x_0)\Delta x\end{aligned}$$

Now that we have determined the value of the function $F(x)$ at $x = x_1$, we can repeat the procedure to determine the value of the function $F(x)$ at the next point, $x_2 = x_1 + \Delta x$. Again, we use the derivative evaluated at x_1, $f(x_1)$, to determine ΔF_1, and add that to F_1 to get $F_2 = F(x_2)$, as illustrated by the grey arrow in Figure B.5:

$$\begin{aligned}F_2 &= F(x_1 + \Delta x)\\ &\approx F_1 + \Delta F_1\\ &\approx F_1 + f(x_1)\Delta x\\ &\approx F_0 + f(x_0)\Delta x + f(x_1)\Delta x\end{aligned}$$

Using the summation notation, we can generalize the result and write the function $F(x)$ evaluated at any point, $x_N = x_0 + N\Delta x$:

$$F(x_N) \approx F_0 + \sum_{i=1}^{i=N} f(x_{i-1})\Delta x$$

The result above will become exactly correct in the limit $\Delta x \to 0$:

$$F(x_N) = F(x_0) + \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} f(x_{i-1})\Delta x \tag{B.2}$$

Let us take a closer look at the sum. Each term in the sum is of the form $f(x_{i-1})\Delta x$, and is illustrated in Figure B.6 for the same case as in Figure B.5 (that is, Figure B.6 shows $f(x)$ that we know, and Figure B.5 shows $F(x)$ that we are trying to find).

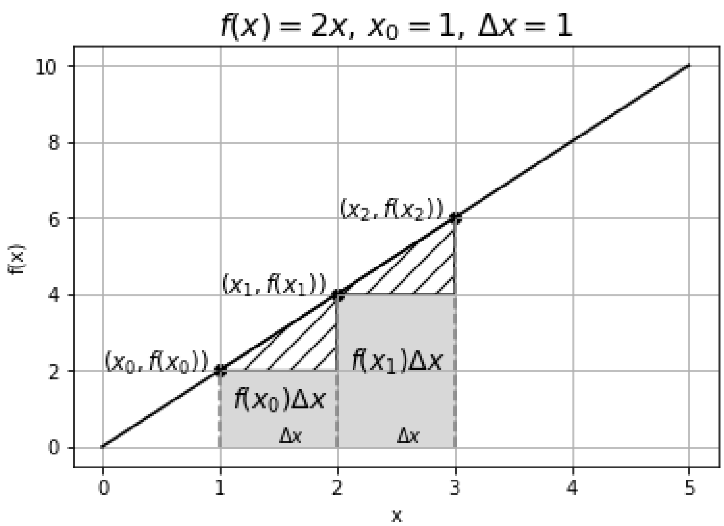


Figure B.6: The function $f(x) = 2x$ and illustration of the terms $f(x_0)\Delta x$ and $f(x_1)\Delta x$ as the area between the curve $f(x)$ and the x axis when $\Delta x \to 0$.

As you can see, each term in the sum corresponds to the area of a rectangle between the function $f(x)$ and the x axis (with a piece missing). In the limit where $\Delta x \to 0$, the missing pieces (shown by the hashed areas in Figure B.6) will vanish and $f(x_i)\Delta x$ will become exactly the area between $f(x)$ and the x axis over a length Δx. The sum of the rectangular areas will thus approach the area between $f(x)$ and the x axis between x_0 and x_N:

$$\lim_{\Delta x \to 0} \sum_{i=1}^{i=N} f(x_{i-1})\Delta x = \text{Area between f(x) and x axis from } x_0 \text{ to } x_N$$

Re-arranging equation B.2 gives us a prescription for determining the anti-derivative:

$$F(x_N) - F(x_0) = \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} f(x_{i-1})\Delta x$$

We see that if we determine the area between $f(x)$ and the x axis from x_0 to x_N, we can obtain the difference between the anti-derivative at two points, $F(x_N) - F(x_0)$

The difference between the anti-derivative, $F(x)$, evaluated at two different values of x is called the **integral** of $f(x)$ and has the following notation:

$$\boxed{\int_{x_0}^{x_N} f(x)dx = F(x_N) - F(x_0) = \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} f(x_{i-1})\Delta x} \tag{B.3}$$

As you can see, the integral has labels that specify the range over which we calculate the area between $f(x)$ and the x axis. A common notation to express the difference $F(x_N) - F(x_0)$ is to use brackets:

$$\int_{x_0}^{x_N} f(x)dx = F(x_N) - F(x_0) = \Big[F(x)\Big]_{x_0}^{x_N}$$

Recall that we wrote the anti-derivative with the same $\int$ symbol earlier:

$$\int f(x)dx = F(x)$$

The symbol $\int f(x)dx$ without the limits is called the **indefinite integral**. You can also see that when you take the (definite) integral (i.e. the difference between $F(x)$ evaluated at two points), any constant that is added to $F(x)$ will cancel. Physical quantities are always based on definite integrals, so when we write the constant C it is primarily for completeness and to emphasize that we have an indefinite integral.

As an example, let us determine the integral of $f(x) = 2x$ between $x = 1$ and $x = 4$, as well as the indefinite integral of $f(x)$, which is the case that we illustrated in Figures B.5 and B.6. Using equation B.3, we have:

$$\begin{aligned}
\int_{x_0}^{x_N} f(x)dx &= \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} f(x_{i-1})\Delta x \\
&= \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} 2x_{i-1}\Delta x
\end{aligned}$$

where we have:

$$\begin{aligned}
x_0 &= 1 \\
x_N &= 4 \\
\Delta x &= \frac{x_N - x_0}{N}
\end{aligned}$$

Note that N is the number of times we have Δx in the interval between x_0 and x_N. Thus, taking the limit of $\Delta x \to 0$ is the same as taking the limit $N \to \infty$. Let us illustrate the sum for the case where $N = 3$, and thus when $\Delta x = 1$, corresponding to the illustration in Figure B.6:

$$\begin{aligned}\sum_{i=1}^{i=N=3} 2x_{i-1}\Delta x &= 2x_0\Delta x + 2x_1\Delta x + 2x_2\Delta x \\ &= 2\Delta x(x_0 + x_1 + x_2) \\ &= 2\frac{x_3 - x_0}{N}(x_0 + x_1 + x_2) \\ &= 2\frac{(4) - (1)}{(3)}(1 + 2 + 3) \\ &= 12\end{aligned}$$

where in the second line, we noticed that we could factor out the $2\Delta x$ because it appears in each term. Since we only used 4 points, this is a pretty coarse approximation of the integral, and we expect it to be an underestimate (as the missing area represented by the hashed lines in Figure B.6 is quite large).

If we repeat this for a larger value of N, $N = 6$ ($\Delta x = 0.5$), we should obtain a more accurate answer:

$$\begin{aligned}\sum_{i=1}^{i=6} 2x_{i-1}\Delta x &= 2\frac{x_6 - x_0}{N}(x_0 + x_1 + x_2 + x_3 + x_4 + x_5) \\ &= 2\frac{4-1}{6}(1 + 1.5 + 2 + 2.5 + 3 + 3.5) \\ &= 13.5\end{aligned}$$

Writing this out again for the general case so that we can take the limit $N \to \infty$, and factoring out the $2\Delta x$:

$$\begin{aligned}\sum_{i=1}^{i=N} 2x_{i-1}\Delta x &= 2\Delta x \sum_{i=1}^{i=N} x_{i-1} \\ &= 2\frac{x_N - x_0}{N}\sum_{i=1}^{i=N} x_{i-1}\end{aligned}$$

Now, consider the combination:

$$\frac{1}{N}\sum_{i=1}^{i=N} x_{i-1}$$

that appears above. This corresponds to the arithmetic average of the values from x_0 to x_{N-1} (sum the values and divide by the number of values). In the limit where $N \to \infty$, then the value $x_{N-1} \approx x_N$. The average value of x in the interval between x_0 and x_N is simply given by the value of x at the midpoint of the interval:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{i=N} x_{i-1} = \frac{1}{2}(x_N + x_0)$$

Putting everything together:

$$\begin{aligned}\lim_{N\to\infty}\sum_{i=1}^{i=N} 2x_{i-1}\Delta x &= 2(x_N + x_0)\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{i=N} x_{i-1}\\ &= 2(x_N - x_0)\frac{1}{2}(x_N + x_0)\\ &= x_N^2 - x_0^2\\ &= (4)^2 - (1)^2 = 15\end{aligned}$$

where in the last line, we substituted in the values of $x_0 = 1$ and $x_N = 4$. Writing this as the integral:

$$\int_{x_0}^{x_N} 2x dx = F(x_N) - F(x_0) = x_N^2 - x_0^2$$

we can immediately identify the anti-derivative and the indefinite integral:

$$\begin{aligned} F(x) &= x^2 + C\\ \int 2x dx &= x^2 + C\end{aligned}$$

This is of course the result that we expected, and we can check our answer by taking the derivative of $F(x)$:

$$\frac{dF}{dx} = \frac{d}{dx}(x^2 + C) = 2x$$

We have thus confirmed that $F(x) = x^2 + C$ is the anti-derivative of $f(x) = 2x$.

Checkpoint B-3

The quantity $\int_a^b f(t)dt$ is equal to

A) the area between the function $f(t)$ and the f axis between $t = a$ and $t = b$
B) the sum of $f(t)\Delta t$ in the limit $\Delta t \to 0$ between $t = a$ and $t = b$
C) the difference $f(b) - f(a)$.

B.3.1 Common anti-derivative and properties

Table B.3 below gives the anti-derivatives (indefinite integrals) for common functions. In all cases, x, is the independent variable, and all other variables should be thought of as constants:

Function, $f(x)$	**Anti-derivative,** $F(x)$
$f(x) = a$	$F(x) = ax + C$
$f(x) = x^n$	$F(x) = \frac{1}{n+1}x^{n+1} + C$
$f(x) = \frac{1}{x}$	$F(x) = \ln(\lvert x\rvert) + C$
$f(x) = \sin(x)$	$F(x) = -\cos(x) + C$
$f(x) = \cos(x)$	$F(x) = \sin(x) + C$
$f(x) = \tan(x)$	$F(x) = -\ln(\lvert\cos(x)\rvert) + C$
$f(x) = e^x$	$F(x) = e^x + C$
$f(x) = \ln(x)$	$F(x) = x\ln(x) - x + C$

Table B.3: Common indefinite integrals of functions.

Note that, in general, it is much more difficult to obtain the anti-derivative of a function than it is to take its derivative. A few common properties to help evaluate indefinite integrals are shown in Table B.4 below.

Anti-derivative	**Equivalent anti-derivative**
$\int(f(x) + g(x))dx$	$\int f(x)dx + \int g(x)dx$ (sum)
$\int(f(x) - g(x))dx$	$\int f(x)dx - \int g(x)dx$ (subtraction)
$\int af(x)dx$	$a\int f(x)dx$ (multiplication by constant)
$\int f'(x)g(x)dx$	$f(x)g(x) - \int f(x)g'(x)dx$ (integration by parts)

Table B.4: Some properties of indefinite integrals.

B.3.2 Common uses of integrals in Physics - from a sum to an integral

Integrals are extremely useful in physics because they are related to sums. If we assume that our mathematician friends (or computers) can determine anti-derivatives for us, using integrals is not that complicated.

The key idea in physics is that **integrals are a tool to easily performing sums**. As we saw above, integrals correspond to the area underneath a curve, which is found by *summing* the (different) areas of an infinite number of infinitely small rectangles. In physics, it is often the case that we need to take the sum of an infinite number of small things that keep varying, just as the areas of the rectangles.

Consider, for example, a rod of length, L, and total mass M, as shown in Figure B.7. If the rod is uniform in density, then if we cut it into, say, two equal pieces, those two pieces will weigh the same. We can define a "linear mass density", μ, for the rod, as the mass per unit length of the rod:

$$\mu = \frac{M}{L}$$

The linear mass density has dimensions of mass over length and can be used to find the mass of any length of rod. For example, if the rod has a mass of $M = 5\,\text{kg}$ and a length of $L = 2\,\text{m}$, then the mass density is:

$$\mu = \frac{M}{L} = \frac{(5\,\text{kg})}{(2\,\text{m})} = 2.5\,\text{kg/m}$$

Knowing the mass density, we can now easily find the mass, m, of a piece of rod that has a length of, say, $l = 10\,\text{cm}$. Using the mass density, the mass of the 10 cm rod is given by:

$$m = \mu l = (2.5\,\text{kg/m})(0.1\,\text{m}) = 0.25\,\text{kg}$$

Now suppose that we have a rod of length L that is not uniform, as in Figure B.7, and that does not have a constant linear mass density. Perhaps the rod gets wider and wider, or it has a holes in it that make it not uniform. Imagine that the mass density of the rod is instead given by a function, $\mu(x)$, that depends on the position along the rod, where x is the distance measured from one side of the rod.

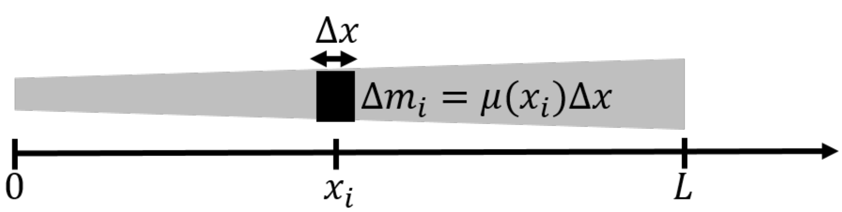


Figure B.7: A rod with a varying linear density. To calculate the mass of the rod, we consider a small mass element Δm_i of length Δx at position x_i. The total mass of the rod is found by summing the mass of the small mass elements.

Now, we cannot simply determine the mass of the rod by multiplying $\mu(x)$ and L, since we do not know which value of x to use. In fact, we have to use all of the values of x, between $x = 0$ and $x = L$.

The strategy is to divide the rod up into N pieces of length Δx. If we label our pieces of rod with an index i, we can say that the piece that is at position x_i has a tiny mass, Δm_i. We assume that Δx is small enough so that $\mu(x)$ can be taken as constant over the length of that tiny piece of rod. Then, the tiny piece of rod at $x = x_i$, has a mass, Δm_i, given by:

$$\Delta m_i = \mu(x_i)\Delta x$$

where $\mu(x_i)$ is evaluated at the position, x_i, of our tiny piece of rod. The total mass, M, of the rod is then the sum of the masses of the tiny rods, in the limit where $\Delta x \to 0$:

$$\begin{aligned} M &= \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} \Delta m_i \\ &= \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} \mu(x_i) \Delta x \end{aligned}$$

But this is precisely the definition of the integral (equation B.2), which we can easily evaluate with an anti-derivative:

$$\begin{aligned} M &= \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} \mu(x_i) \Delta x \\ &= \int_0^L \mu(x) dx \\ &= G(L) - G(0) \end{aligned}$$

where $G(x)$ is the anti-derivative of $\mu(x)$.

Suppose that the mass density is given by the function:

$$\mu(x) = ax^3$$

with anti-derivative (Table B.3):

$$G(x) = a\frac{1}{4}x^4 + C$$

Let $a = 5\,\mathrm{kg/m^4}$ and let's say that the length of the rod is $L = 0.5\,\mathrm{m}$. The total mass of the rod is then:

$$\begin{aligned} M &= \int_0^L \mu(x) dx \\ &= \int_0^L ax^3 dx \\ &= G(L) - G(0) \\ &= \left[a\frac{1}{4}L^4\right] - \left[a\frac{1}{4}0^4\right] \\ &= 5\,\mathrm{kg/m^4}\frac{1}{4}(0.5\,\mathrm{m})^4 \\ &= 78\,\mathrm{g} \end{aligned}$$

With a little practice, you can solve this type of problem without writing out the sum explicitly. Picture an *infinitesimal* piece of the rod of length dx at position x. It will have an *infinitesimal* mass, dm, given by:

$$dm = \mu(x) dx$$

The total mass of the rod is the then the sum (i.e. the integral) of the mass *elements*

$$M = \int dm$$

and we really can think of the $\int$ sign as a sum, when the things being summed are *infinitesimally* small. In the above equation, we still have not specified the range in x over which we want to take the sum; that is, we need some sort of index for the mass elements to make this a meaningful definite integral. Since we already know how to express dm in terms of dx, we can substitute our expression for dm using one with dx:

$$M = \int dm = \int_0^L \mu(x)dx$$

where we have made the integral definite by specifying the range over which to sum, since we can use x to "label" the mass elements.

One should note that coming up with the above integral is physics. Solving it is math. We will worry much more about writing out the integral than evaluating its value. Evaluating the integral can always be done by a mathematician friend or a computer, but determining which integral to write down is the physicist's job!

B.4 Summary

Key Takeaways

The derivative of a function, $f(x)$, with respect to x can be written as:

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

and measures the rate of change of the function with respect to x. The derivative of a function is generally itself a function. The derivative is defined as:

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

Graphically, the derivative of a function represents the slope of the function, and it is positive if the function is increasing, negative if the function is decreasing and zero if the function is flat. Derivatives can always be determined analytically for any continuous function.

A partial derivative measures the rate of change of a multi-variate function, $f(x, y)$, with respect to one of its independent variables. The partial derivative with respect to one of the variables is evaluated by taking the derivative of the function with respect to that variable while treating all other independent variables as if they were constant. The partial derivative of a function (with respect to x) is written as:

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

The gradient of a function, $\nabla f(x, y)$, is a vector in the direction in which that function is increasing most rapidly. It is given by:

$$\nabla f(x, y) = \frac{\partial f}{\partial x}\hat{x} + \frac{\partial f}{\partial y}\hat{y}$$

Given a function, $f(x)$, its anti-derivative with respect to x, $F(x)$, is written:

$$F(x) = \int f(x)dx$$

$F(x)$ is such that its derivative with respect to x is $f(x)$:

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

The anti-derivative of a function is only ever defined up to a constant, C. We usually write this as:

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

since the derivative of $F(x) + C$ will also be equal to $f(x)$. The anti-derivative is also called the "indefinite integral" of $f(x)$.

The definite integral of a function $f(x)$, between $x = a$ and $x = b$, is written:

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

and is equal to the difference in the anti-derivative evaluated at $x = a$ and $x = b$:

$$\int_a^b f(x)dx = F(b) - F(a)$$

where the constant C no longer matters, since it cancels out. Physical quantities only ever depend on definite integrals, since they must be determined without an arbitrary constant.

Definite integrals are very useful in physics because they are related to a sum. Given a function $f(x)$, one can relate the sum of terms of the form $f(x_i)\Delta x$ over a range of values from $x = a$ to $x = b$ to the integral of $f(x)$ over that range:

$$\lim_{\Delta x \to 0} \sum_{i=1}^{i=N} f(x_{i-1})\Delta x = \int_{x_0}^{x_N} f(x)dx = F(x_N) - F(x_0) =$$

B.5 Thinking about the Material

Reflect and research

1. When was calculus first discovered, and by whom?
2. What is an example of a physical quantity that is given by a derivative (other than speed or acceleration)?
3. What is a case when you would need to perform an integral to evaluate a physical quantity?

B.6 Sample problems and solutions

B.6.1 Problems

Problem B-1: You find that the number of customers in your store as a function of time is given by:

$$N(t) = a + bt - ct^2$$

where a, b and c are constants. At what time does your store have the most customers, and what will the number of customers be? (Give the answer in terms of a, b and c). (Solution)

Problem B-2: You measure the speed, $v(t)$, of an accelerating train as function of time, t, to be given by:

$$v(t) = at + bt^2$$

where a and b are constants. How far does the train move between $t = t_0$ and $t = t_1$? (Solution)

B.6.2 Solutions

Solution to problem B-1: We need to find the value of t for which the function $N(t)$ is maximal. This will occur when its derivative with respect to t is zero:

$$\begin{aligned}\frac{dN}{dt} &= b - 2ct = 0\\ \therefore t &= \frac{b}{2c}\end{aligned}$$

At that time, the number of customers will be:

$$\begin{aligned}N\left(t = \frac{b}{2c}\right) &= a + bt - ct^2\\ &= a + \frac{b^2}{2c} - \frac{b^2}{4c} = a + \frac{3b^2}{4c}\end{aligned}$$

Solution to problem B-2: We are given the speed of the train as a function of time, which is the rate of change of its position:

$$v(t) = \frac{dx}{dt}$$

We need to find how its position, $x(t)$, changes with time, given the speed. In other words, we need to find the anti-derivative of $v(t)$ to get the function for the position as a function of time, $x(t)$:

$$\begin{aligned}x(t) &= \int v(t)dt = \int (at + bt^2)dt\\ &= \frac{1}{2}at^2 + \frac{1}{3}bt^3 + C\end{aligned}$$

where C is an arbitrary constant. The distance covered, Δx, between time t_0 and time t_1 is simply the difference in position at those two times:

$$\begin{aligned}\Delta x &= x(t_1) - x(t_0)\\ &= \frac{1}{2}at_1^2 + \frac{1}{3}bt_1^3 + C - \frac{1}{2}at_0^2 + \frac{1}{3}bt_0^3 - C\\ &= \frac{1}{2}a(t_1^2 - t_0^2) + \frac{1}{3}b(t_1^3 - t_0^3)\end{aligned}$$

C Guidelines for lab related activities

This chapter introduces the skills that are necessary for thinking about how to design an experiment and to report on its results.

Learning Objectives

- Develop skills in general scientific writing.
- Learn to write scientific proposals and experimental reports.
- Learn to review others' scientific proposals and experimental reports.

C.1 The process of science and the need for scientific writing

Conducting experiments that test a scientific theory is integral to the advancement of science and to the refining of scientific theories. In practice, scientists do not have a lab full of equipment ready to go and to be used for testing whichever theory suits their fancy. Instead, they need to write a "proposal" for conducting a particular experiment to a funding source (e.g. a funding agency). That funding source will then select a panel of experts in the field to review whether the proposal is feasible and useful in advancing science, to decide whether it should be funded. If the scientist is awarded with funds, they are then expected to carry out their experiment and report on the results in a peer-reviewed scientific journal. Again, before the results are published, the scientific journal will ask a panel of experts to review the results to ensure that they are scientifically valid and interesting.

In order for a proposal to be funded, it must thus propose an experiment that is well-thought out and feasible. For example, the reviewers will want to make sure that the proposed experiment is designed in the best possible way to test a theory. Often, this means that thought has been put into designing an experiment that minimizes the uncertainty on the result, so that the test of the theory is as stringent as possible.

A proposal needs to be well-written and precise. We generally call this type of writing "scientific writing", and it is a style of writing that takes some practice. Similarly, when reporting on the results of an experiment, the report will need to be clear and precise as well. For example, in scientific writing, one avoids giving opinions or using sentences that do not add necessary information or that are not factual.

This chapter provides some guidelines for scientific writing, writing proposals, and writing reports. In addition to this, guidelines for reviewing others' proposals and reports are also presented. Not only is it important to develop the ability to critically evaluate others' work, but it is also helpful in learning to reflect and improve on one's own work.

C.2 Scientific writing

Scientific writing is important in communicating with other scientists. Think of scientific writing as a style of writing where **every word counts**. It makes for rather "dry" reading, but it is important for clearly and precisely communicating factual information. The main guidelines for scientific writing are **be concise, precise, factual, and clear**. Below are some tips to help with scientific writing:

- Avoid subjective/imprecise terms: avoid using subjective and imprecise terms, stick to factual statements and avoid opinions. Instead of saying "our calculated value of g was much greater than the expected value", say "our calculated value of g was greater than the expected value". Your opinion that it was "much greater" does not communicate anything and is imprecise (much greater in relation to what?).
- Definitive statements: avoid attributing definitive causes to your experimental outcomes. You can never prove a theory to be correct, so at most, your results will be consistent with a theory. For example, instead of saying "as the data exhibit, we have detected the Purple Particle", you should state that "the data are consistent with the detection of the Purple Particle".
- Data is the plural of datum. "This data shows" is incorrect, rather, "these data show", or "this set of data shows".
- Active vs. passive voice: when writing scientific papers, it is recommended to use the third person, passive voice. For example, this would mean saying "the drop time for balls at various heights was measured" rather than "we measured the drop time for balls at various heights". However, both passive and active voices are acceptable in scientific writing, as long as it is consistent throughout the text.
- Tense. Generally, for a proposal, you would use the future tense, and you would use the past tense for reporting on your results.

Emma's Thoughts

Writing and editing - how can I be more concise? We've all felt that our writing was lacking at some point or another. Here are some general tips to avoid overall "wordiness" and to increase ease of reading when writing scientifically:

- What would you want to read? Let's say that you wanted to know the strength of Earth's magnetic field, and how it was found, so you decide to do a literature search. Would you choose a brief, succinct article, or a wordy Magnetic Field Manifesto?
- The kindergarten test: If you had to explain your concept to a six year old cousin, how would you break it down in a way that they could understand it? If you can't break it down enough to explain to a six year old, perhaps you need to revisit your own understanding of the concept before writing about it scientifically.
- Avoid unnecessary adjectives: while this might be ok in a creative writing class, in scientific writing, the goal is to get your point across as succinctly as possible. Using "big" words might be ok (as long as they properly describe what you are trying to say), but it is important to communicate your message in the simplest manner.
- Think about it: every time you use a comma, dash or even an "and", you should reconsider the brevity of your statement. In scientific writing, commas are carefully placed, and semicolons are rare.
- Cut it in half: For every word you read, think of another that you can cut. For every sentence that you read, think of three sentences that communicate the same idea. Pick the sentence that is the shortest and most concise.
- Proofread - the more, the better.

The following sections provide basic outlines for writing a proposal and a lab report, as well as rubrics for evaluating/reviewing proposals and reports. Additionally, samples of a proposal, proposal review, report, and report review for the experiment "Measuring g using a pendulum" are provided. In the sample proposal and lab report, errors are purposefully included and addressed in the reviews. It is important to entirely read the rest of this section to capture the common proposal/lab mistakes and their corresponding corrections. That is, do not take the sample proposal as a "perfect proposal", but rather, consider it in the light of the corresponding review.

C.3 Guide for writing a proposal

Summary and Goal

Write a few short sentences briefly summarizing the aim of your experiment, how it will be conducted, and how precise of a result you expect to obtain.

Method and equipment

Clearly describe, in as much detail as required, the method/procedure that you will use to carry out your experiment, and how you will analyse the results. Justify the choices that you made (no need to say you chose to use a ruler because you will need to measure a distance, but perhaps say why you need to measure a given distance, or that you chose to measure something in a particular way as it would reduce the corresponding uncertainty). Provide a list of the equipment that you will need. Also, propose a method of assessing whether or not your project was successful.

Consider the following questions:

- What theory are you testing and through what model?
- How precisely do you estimate that you will be able to make your measurement? Estimate the uncertainty that you will obtain with the proposed experiment. Use this in guiding the design of your experiment.
- What materials, equipment and/or tools are necessary in making your measurements?
- What are the cost of these materials? Can they be easily obtained?
- Where should this experiment be conducted?
- Are there any safety concerns?
- How will you make your measurements? How many times will you make them?
- How will you record your measurements?
- How will you maximize the precision of your experiments?
- How will you determine uncertainties?
- How will you analyse the data?
- What issues could arise in your experiment? How do you plan to resolve these issues?

Timeline and Team

Provide the names of team members, and assign relevant duties to each member. Give a rough outline of the timeline to conduct the experiment, to analyse the data, and to report on the results.

C.4 Guide for reviewing a proposal

Summary

Summarize your overall evaluation of the proposal in 2-3 sentences. Focus on the experiment's methods and goals. For example, "The authors wish to drop balls from different heights to determine the value of g". You don't need to go into the specific details, just give a high level summary of the proposal and your opinion on whether this is a strong proposal. If the proposal is unclear, specify this.

Review

This is where you give your detailed review of the proposal. Consider the following questions:

- Is the proposed experiment well thought-out and feasible?
- Is the experimental procedure clear and concise? Could you could carry out the experiment without asking the authors for additional information? Do the authors specify what instruments to use to measure different quantities and how to determine the associated uncertainties?
- Does the experimental design minimize uncertainties?
- Is it possible to complete the experiment in a reasonable period of time?
- Is it possible to obtain the equipment/materials to conduct the experiment?
- Do the authors describe how to analyse the data (correctly)?
- Does the plan incorporate a mechanism to assess success?
- Is a troubleshooting plan in place, in case of unexpected difficulties?

Overall Rating of the Experiment

Give the proposal an overall score, based on the criteria described above. Use one of the following to rate the proposal and include a sentence to justify your choice.

- Excellent
- Good
- Satisfactory
- Needs work
- Incomplete

C.5 Guide for writing a lab report

Abstract

Write a few short sentences briefly summarizing what you did, how you did it, what you found and whether anything went wrong in your experiment.

Procedure

Describe relevant theories that relate to your experiment here, and the steps to carry out your procedure.

Consider the following questions:

- What are the relevant theories/principles that you used?
- What equations did you use? Show how you modelled your experiment.
- What materials, equipment and/or tools were necessary in making your measurements?
- Where was this experiment conducted?
- How did you make your measurements? How many times did you make them?
- How did you record your measurements?
- How did you determine and minimize the uncertainties in your measurements? Why did you choose to measure a specific quantity in a certain way?

Prediction It can be useful to predict the value (and uncertainty) that you expect to measure before conducting the measurement. You should report on this initial prediction in order to help you better understand the data from your experiment.

Consider the following questions:

- Predict your measured values and uncertainties. How precise do you expect your measurements to be?
- What assumptions did you have to make to predict your results?
- Have these predictions influenced how you should approach your procedure? Make relevant adjustments to the procedure based on your predictions.

Data and Analysis

Present your data. Include relevant tables/graphs. Describe in detail how you analysed the data, including how you propagated uncertainties. If the data do not agree with your model prediction (or the prediction from your proposal), examine whether you can improve your model.

Consider the following questions:

- How did you obtain the “final” measurement/value from your collected data?
- How did you propagate uncertainties? Why did you do it that way?
- What is the relative uncertainty on your value(s)?

Discussion and Conclusion

Summarize your findings, and address whether or not your model described the data. Discuss possible reasons why your measured value is not consisted with your model expectation (is it the model? is it the data?).

Consider the following questions:

- Were there any systematic errors that you didn’t consider?
- Did you learn anything that you didn’t previously know? (eg. about the subject of your experiment, about the scientific method in general)
- If you could redo this experiment, what would you change (if anything)?

C.5.1 Guide for reviewing a lab report

Summary

Summarize your overall evaluation of the report in 2-3 sentences. Focus on the experiment's method and its result. For example, "The authors dropped balls from different heights to determine the value of g". You don't need to go into the specific details, just give a high level summary of the report. If the report is unclear, specify this.

Review

Consider the following questions:

- Is the the procedure well thought-out, clearly and concisely described?
- Do you have sufficient information that you could repeat this experiment?
- Does the report clearly describe how different quantities were measured and how the uncertainties were determined?
- Does the report motivate why the specific procedure was chosen? (e.g. to minimize uncertainties).
- Does the experiment clearly state how uncertainties were propagated and how the data were analysed?
- Do you believe their result to be scientifically valid?

Overall Rating of the Experiment

Give the report an overall score, based on the criteria described above. Use one of the following to rate the proposal and include a sentence to justify your choice.

- Excellent
- Good
- Satisfactory
- Needs work
- Incomplete

C.6 Sample proposal (Measuring g using a pendulum)

Summary and Goal

One can measure the gravitational constant, g, by measuring the period of a pendulum of a known length, requiring only a string, mass, ruler and timer. Because the experimental design can be easily adjusted and the experiment is simple, the experiment has a high chance of success.

Method and equipment

The period of a pendulum of length L is easily shown to be given by:

$$T = 2\pi\sqrt{\frac{L}{g}}$$

Thus, by measuring the period, T, of a pendulum as well as its length, one can determine the value of g:

$$g = \frac{4\pi^2 L}{T^2}$$

One can carry out the experiment using the following materials:

- a mass
- inextensible string
- a metre stick
- stand to attach string
- cell-phone with timer and slow-motion camera

The materials listed above are all inexpensive and can be easily obtained. It is recommended that the experiment be completed indoors at room temperature, in order to minimize any environmental effects.

One should tie the string to the mass at one end and the stand at the other, and measure the length, L, of the string from the point on the stand to the centre of mass of the mass.

The period of the pendulum is measured by timing how long it takes the pendulum to complete 20 oscillations and dividing that time by 20. This will be more precise than trying to time the period of a single oscillation.

The pendulum should be released from 90°. When releasing the pendulum, the string should be pulled taught, and the team member's eye that is measuring the angle should be situated parallel to the measuring device.

A slow-motion video will be taken of the pendulum to track the time of the oscillation in order to minimize error due to reaction time. The team member in charge of taking the video will start the video shortly before the pendulum is released. After releasing the pendulum, the team should record 20 oscillations before stopping the pendulum and the video. Data from the video should be entered into a Jupyter Notebook. It is recommended that this measurement be repeated at least 5 times.

The uncertainty in the time should be taken as half of the smallest division of the cell-phone timer, and the uncertainty in the length of the pendulum as half the smallest division of the metre stick used to measure the length of the pendulum.

Foreseeable issues in this experiment may arise when trying to find a string that is optimally inextensible, as any extensibility will cause error in the results. Additionally, being able to measure exactly 90° as the drop-angle for the pendulum could be difficult. In order to correct for this, the team member who is dropping the pendulum must stand directly parallel to the measuring device, minimizing parallax error.

The measure of success will be determined by the uncertainty and precision of the measured value of g. If the measured value of g has a relative uncertainty that is less than 10 %, and is consistent with the accepted value, then one can consider the experiment to have been carried out successfully.

Team and timeline

One should be able to complete the experiment and analysis in approximately 1 hour and 30 minutes with the data being collected in the first 30 minutes. The remainder of the time should be spent processing the data and writing the experimental report. Following the strengths of the members of the team, the following people should be responsible for leading the following tasks, while everyone participates:

- Alice: building the pendulum
- Brice: taking the measurements
- Chloë: analysing the data
- Dennis: writing and formatting

C.7 Sample proposal review (Measuring g using a pendulum)

Summary and Goal

The authors propose to measure the value of g to within 10% by measuring the period of a simple pendulum, using the SHM equations and theory. The proposal is reasonably clear, but lacks some details in how to measure the initial angle of the pendulum. The authors propose to use a an amplitude of 90° for the pendulum, but at such a large angle, the motion is not expected to be SHM, since it is only so at small angles. By using a smaller angle, the experiment has a good chance of being successful in the proposed timeline.

Review

The experimental methods are described clearly and succinctly, with most information clearly stated. For the materials list, it is stated that “a mass” must be used. Here, it should be stated that a small, solid, non-deformable mass should be used to minimize drag and to act as a point mass. The authors refer to a “measuring device” when determining the amplitude of the pendulum, but this is not described. Anyhow, the amplitude of the oscillations in irrelevant for a pendulum in SHM, as long as the amplitude is small.

Most equations are described in the theory section, but it is incorrectly assumed that the period of a pendulum is independent of the drop angle for all angles. The small angle approximation is not expected to apply with an oscillation amplitude of 90°.

No justification is provided for the use of 20 oscillations prior to measuring the period - it may be necessary to iterate on the reason why 20 oscillations was chosen.

The equipment can be easily obtained and is fairly inexpensive. Adequate resources are available to the group to perform this experiment. A clear troubleshooting plan is described and a method for evaluating success is included.

Timeline and team

This experiment is fairly simple and the equipment/setup is not difficult to handle. The proposed team should be qualified to perform this experiment in the proposed amount of time, although I worry a little bit about Dennis, as he seems to be a bit of a menace.

Overall Rating of the Proposal

Good - this proposal was clearly explained and is scientifically sound, apart from the use of a large angle for the oscillations. It was succinctly written, and most components of the experiment were clearly described. A little more detail in the justification for using 20 oscillations is necessary.

C.8 Sample lab report (Measuring g using a pendulum)

Abstract

In this experiment, we measured g by measuring the period of a pendulum of a known length. We measured $g = (7.650 \pm 0.378)\,\text{m/s}^2$. This correspond to a relative difference of 22% with the accepted value ($9.8\,\text{m/s}^2$), and our result is not consistent with the accepted value.

Theory

A pendulum exhibits simple harmonic motion (SHM), which allowed us to measure the gravitational constant by measuring the period of the pendulum. The period, T, of a pendulum of length L undergoing simple harmonic motion is given by:

$$T = 2\pi\sqrt{\frac{L}{g}}$$

Thus, by measuring the period of a pendulum as well as its length, we can determine the value of g:

$$g = \frac{4\pi^2 L}{T^2}$$

We assumed that the frequency and period of the pendulum depend on the length of the pendulum string, rather than the angle from which it was dropped.

Predictions

We built the pendulum with a length $L = (1.0000 \pm 0.0005)\,\text{m}$ that was measured with a ruler with 1 mm graduations (thus a negligible uncertainty in L). We plan to measure the period of one oscillation by measuring the time to it takes the pendulum to go through 20 oscillations and dividing that by 20. The period for one oscillation, based on our value of L and the accepted value for g, is expected to be $T = 2.0\,\text{s}$. We expect that we can measure the time for 20 oscillations with an uncertainty of 0.5 s. We thus expect to measure one oscillation with an uncertainty of 0.025 s (about 1% relative uncertainty on the period). We thus expect that we should be able to measure g with a relative uncertainty of the order of 1%

Procedure

The experiment was conducted in a laboratory indoors.

1. Construction of the pendulum

We constructed the pendulum by attaching a inextensible string to a stand on one end and to a mass on the other end. The mass, string and stand were attached together with knots. We adjusted the knots so that the length of the pendulum was (1.0000 ± 0.0005) m. The uncertainty is given by half of the smallest division of the ruler that we used.

2. Measurement of the period

The pendulum was released from 90° and its period was measured by filming the pendulum with a cell-phone camera and using the phone's built-in time. In order to minimize the uncertainty in the period, we measured the time for the pendulum to make 20 oscillations, and divided that time by 20. We repeated this measurement five times. We transcribed the measurements from the cell-phone into a Jupyter Notebook.

Data and Analysis

Using a 100 g mass and 1.0 m ruler stick, the period of 20 oscillations was measured over 5 trials. The corresponding value of g for each of these trials was calculated. The following data for each trial and corresponding value of g are shown in the table below.

Trial	Angle (Degrees)	Measured Period (s)	Value of g (m/s^2)
1	90	2.24	7.87
2	90	2.37	7.03
3	90	2.28	7.59
4	90	2.26	7.73
5	90	2.22	8.01

Our final measured value of g is $(7.650 \pm 0.378)\,\text{m/s}^2$. This was calculated using the mean of the values of g from the last column and the corresponding standard deviation. The relative uncertainty on our measured value of g is 4.9% and the relative difference with the accepted value of $9.8\,\text{m/s}^2$ is 22%, well above our relative uncertainty.

Discussion and Conclusion

In this experiment, we measured $g = (7.650 \pm 0.378)\,\text{m/s}^2$. This has a relative difference of 22% with the accepted value and our measured value is not consistent with the accepted value. All of our measured values were systematically lower than expected, as our measured periods were all systematically higher than the §2.0s that we expected from our prediction. We also found that our measurement of g had a much larger uncertainty (as determined from the spread in values that we obtained), compared to the 1% relative uncertainty that we predicted.

We suspect that by using 20 oscillations, the pendulum slowed down due to friction, and this resulted in a deviation from simple harmonic motion. This is consistent with the fact that our measured periods are systematically higher. We also worry that we were not able to accurately measure the angle from which the pendulum was released, as we did not use a protractor.

If this experiment could be redone, measuring 10 oscillations of the pendulum, rather than 20 oscillations, could provide a more precise value of g. Additionally, a protractor could be taped to the top of the pendulum stand, with the ruler taped to the protractor. This way, the pendulum could be dropped from a near-perfect 90° rather than a rough estimate.

C.9 Sample lab report review (Measuring g using a pendulum)

Summary

The authors measured the period of a pendulum to determine g. They measured g to be $(7.650 \pm 0.378)\,\text{m/s}^2$ which is inconsistent with the accepted value. The authors were incorrect in assuming that the pendulum would undergo simple harmonic motion in the conditions that they used.

Review

The experimental procedure was clearly written and one could mostly reproduce this experiment with the given description.

The authors thought about minimizing uncertainties by measuring the period over several oscillations, although it appears that 20 was perhaps too large, as friction was likely to have an effect. The authors should have taken more care in determining the number of oscillations to use so that the uncertainty in the time is minimized while also keeping the effects of friction negligible. Ultimately, the authors did not specify the uncertainty in the time that they measured.

The authors also claim to have measured the length of the pendulum with a precision of 0.5 mm, but did not specify the length of the ruler that they used. I would not expect the measurement to be that precise unless they used a very precise ruler that is longer than 1 m. However, the authors made the length of the pendulum as long as possible so as to minimize the uncertainty in the length.

The authors did not describe the mass that was attached at the end of the pendulum, and whether its size would be expected to cause significant air drag.

The authors made a mistake in assuming that a pendulum would undergo simple harmonic motion with an amplitude of $90°$, as the small angle approximation used to determine the period does not apply in this case.

The experimental procedure was scientifically sound, other than the choices for the number of oscillations and their amplitude.

Overall rating of the Experiment

Satisfactory - The experiment was well described, but the authors should have paid more attention to their choice of 20 oscillations, and they made a mistake in assuming that their pendulum would exhibit simple harmonic oscillation with a large amplitude.

D The Python programming language

This appendix gives a very brief introduction to programming in python and is primarily aimed at introducing tools that are useful for the experimental side of physics.

Learning Objectives

- Be able to perform simple algebra using python.
- Be able to plot a function in python.
- Be able to propagate uncertainties in python.
- Be able to plot and fit data to a straight line.
- Understand how to use Python to numerically calculate *any* integral.

In this textbook, we will encourage you to use computers to facilitate making calculations and displaying data. We will make use of a popular programming language called Python, as well as several "modules" from Python that facilitate working with numbers and data. Do not worry if you do not have any programming experience; we assume that you have none and hope that by the end of this book, you will have some capability to decrease your workload by using computer programming.

The only way to become proficient at programming is through practice. If you want to effectively learn from this chapter, it is important that you take the time to actually type the commands into a Python environment rather than simply reading through the chapter. Reading through the chapter will at least give you a sense of what is possible and some terminology, but it will not teach you programming!

D.1 A quick intro to programming

In Python, as in other programming languages, the equal sign is called the **assignment operator**. Its role is to *assign* the value on its right to the variable on its left. The following code does the following:

- *assigns* the value of `2` to the variable `a`
- *assigns* the values of `2*a` to the variable `b`
- prints out the value of the variable `b`

Python Code D.1: Declaring variables in Python

```
#This is a comment, and is ignored by Python
a = 2
b = 2*a
print(b)
```

Output D.1:

```
4
```

Note that any text that follows a pound sign (#) is intended as a comment and will be ignored by Python. Inserting comments in your code is very important for being able to understand your computer program in the future or if you are sharing your code with someone who would like to understand it. In the above example, we called the `print()` **function** and passed to it the variable `b` as an **argument**; this allowed us to print (display) the value of the variable `b` and verify that it was indeed equal to the number 4.

In Python, if you want to have access to "functions", which are a more complex series of operations, then you typically need to load the *module* that defines those operations.

A large number of functions are provided in Python. Most of these functions need to be "imported" from "modules". For example, if you want to be able to take the square root of a number, then you need to load (import) the "math module" which contains the square root function, as in the following example:

Python Code D.2: Using functions from modules

```
#First, we load (import) the math module
import math as m
a = 9
b = m.sqrt(a)
print(b)
```

Output D.2:

```
3
```

In the above code, we loaded the math module (and renamed it `m`); this then allows us to use the functions that are part of that module, including the square root function (`m.sqrt()`).

D.2 Arrays

It is often the case that we need to represent a series of numbers. For example, imagine that you have measured the position of an object as a function of time. **Arrays** are a convenient way to hold a series of numbers that are all alike, for example, all of the values of the position and corresponding time values for the trajectory of the object. In Python, we can define variables that hold arrays instead of a single value (arrays are called "lists" in Python):

Python Code D.3: Arrays in python

```
#define an array of values for the position of the object
position = [0,1,4,9,16,25]
#define an array of values for the corresponding times
time = [0,1,2,3,4,5]
```

D.3 Plotting

Several modules are available in python for plotting. We will show here how to use the `pylab` module (which is equivalent to the `matplotlib` module). For example, we can easily plot the data in the two arrays from the previous section in order to plot the position versus time for the object:

Python Code D.4: Plotting two arrays

```
#import the pylab module
import pylab as pl

#define an array of values for the position of the object
position = [0,1,4,9,16,25]
#define an array of values for the corresponding times
time = [0,1,2,3,4,5]

#make the plot showing points and the line (.-)
pl.plot(time, position, '.-')
#add some labels:
pl.xlabel("time") #label for x-axis
pl.ylabel("position") #label for y-axis
#show the plot
pl.show()
```

Output D.4:

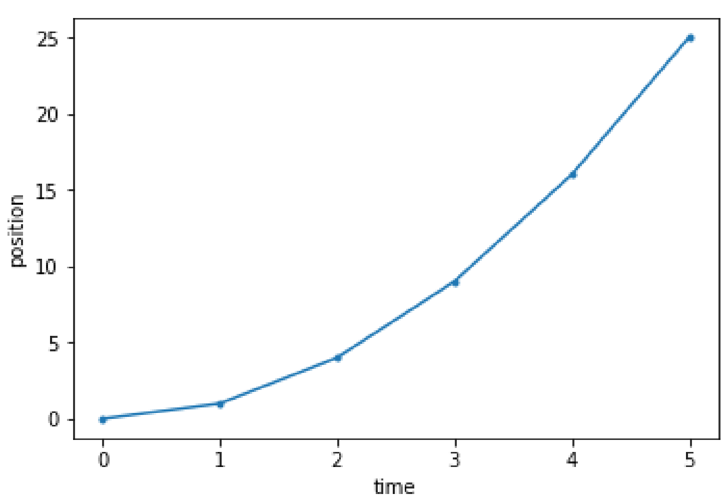


Figure D.1: Using two arrays and plotting them.

Checkpoint D-1

How would you modify the Python code above to show only the points, and not the line?

We can use Python to plot any mathematical function that we like. It is important to realize that computers do not have a representation of a continuous function. Thus, if we would like to plot a continuous function, we first need to evaluate that function at many points, and then plot those points. The `numpy` module provides many useful features for working with arrays of numbers and applying functions directly to those arrays.

Suppose that we would like to plot the function $f(x) = cos(x^2)$ between $x = -3$ and $x = 5$. In order to do this in Python, we will first generate an array of many values of x between -3 and 5 using the `numpy` package and the function `linspace(min,max,N)` which generates N linearly spaced points between min and max. We will then evaluate the function at all of those points to create a second array. Finally, we will plot the two arrays against each other:

Python Code D.5: Plotting a function of 1 variable

```
#import the pylab and numpy modules
import pylab as pl
import numpy as np

#Use numpy to generate 1000 values of x between -3 and 5.
#xvals is an array with 1000 values in it:
xvals =np.linspace(-3,5,1000)

#Now, evaluate the function for all of those values of x.
#We use the numpy version of cos, since it allows us to take the cos
#of all values in the array.
#fvals will be an array with the 1000 corresponding cosines of the xvals
    squared
fvals = np.cos(xvals**2)

#make the plot showing only a line, and color it
pl.plot(xvals, fvals, color='red')
#show the plot
pl.show()
```

Output D.5:

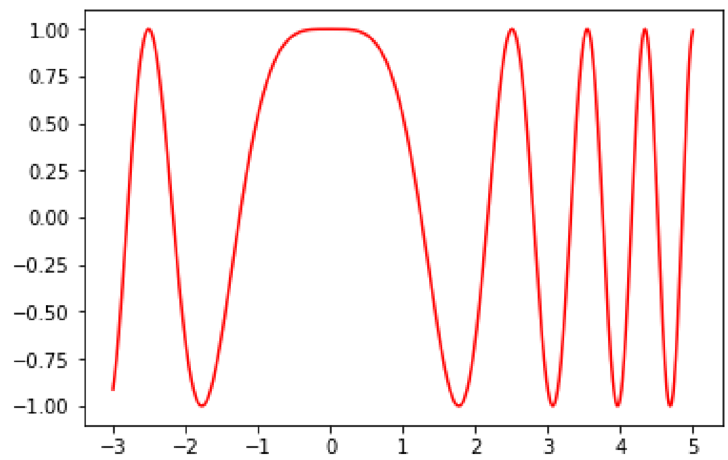


Figure D.2: Plotting a function using arrays.

D.4 The QExpy python package for experimental physics

QExpy is a Python module that was developed with students from Queen's University to handle all aspects of undergraduate physics laboratories. In this section, we look at how to use QExpy to propagate uncertainties and to plot experimental data.

D.4.1 Propagating uncertainties

In Chapter 2, we saw how to use the "derivative method" to propagate the uncertainty from measurements into the uncertainty in a value that depended on those measurements. In Example 2-7, we propagated the uncertainties $x = (3.00 \pm 0.01)\,\text{m}$ and $t = (0.76 \pm 0.15)\,\text{s}$ to the quantity $k = \frac{t}{\sqrt{x}}$. We show below how easily this can be done with QExpy:

Python Code D.6: QExpy to propagate uncertainties

```
#First, we load the QExpy module
import qexpy as q
#Now define our measurements with uncertainties:
t = q.Measurement(0.76, 0.15) # 0.76 +/- 0.15
x = q.Measurement(3,0.1) # 3 +/- 0.1
#Now define k, which depends on t and x:
k = t/q.sqrt(x) # use the QExpy version of sqrt() since x is of type
    Measurement
#Print the result:
print(k)
```

Output D.6:

```
0.44 +/- 0.09
```

which is the result that we obtained when manually applying the derivative method. Note that we used the square root function from the QExpy module, as it "knows" how to take the square root of a value with uncertainty (a "Measurement" in the language of QExpy).

We also saw that when we had repeated measurements of the same quantity (Section 2.3.1), one could define a central value and uncertainty for that quantity by using the mean and standard deviations of the measurements. QExpy can easily take a set of measurements (an array of values) and convert them into a single quantity (a "Measurement") with a central value and uncertainty that correspond to the mean and standard deviation of the set of measurements:

Python Code D.7: QExpy to calculate mean and standard deviation

```
#First, we load the QExpy module
import qexpy as q
#We define $t$ as an array of values (note the square brackets):
t = q.Measurement([1.01, 0.76, 0.64, 0.73, 0.66])
#Choose the number of significant figures to print:
q.set_sigfigs(2)
#Print the result:
print("t = ",t)
```

Output D.7:

```
t = 0.76 +/- 0.15
```

By using QExpy, we do not need to tediously calculate the mean and standard deviation, as we had in Example 2-6.

D.4.2 Plotting experimental data with uncertainties

In Chapter 2 we had presented the data in Table D.1 which corresponded to our measurements of how long it took (t) for an object to drop a certain distance, x. We had also introduced Chloë's Theory of gravity that predicted that the data should be described by the following model:

$$t = k\sqrt{x}$$

where k was an undetermined constant of proportionality.

x [m]	**t** [s]	$\sqrt{x}$ [$\mathrm{m}^{\frac{1}{2}}$]	**k** [$\mathrm{s\,m}^{-\frac{1}{2}}$]
1.00	0.33	1.00	0.33
2.00	0.74	1.41	0.52
3.00	0.67	1.73	0.39
4.00	1.07	2.00	0.54
5.00	1.10	2.24	0.49

Table D.1: Measurements of the drop times, t, for a bowling ball to fall different distances, x. We have also computed $\sqrt{x}$ and the corresponding value of k.

The easiest way to visualize and analyse those data is to plot them. In particular, if we plot (graph) t versus $\sqrt{x}$, we expect that the points will fall on a straight line that goes through zero, with a slope of k (if the data are described by Chloë's Theory). We can use QExpy to graph the data as well as determine ("fit") for the slope of the line that best describes the data, since we expect that the slope will correspond to the value of k. When plotting data and fitting them to a line (or other function), it is important to make sure that the values have at least an uncertainty in the quantity that is being plotted on the y axis. In this case, we have assumed that all of the measurements of time have an uncertainty of 0.15 s and that the measurements of the distance have no (or negligible) uncertainties. The python code below shows how to use QExpy to plot and fit the data to a straight line.

Python Code D.8: Using QExPy to plot and fit linear data

```
#First, we load the QExpy module:
import qexpy as q

#Use matplotlib as the plot engine (try using 'bokeh' instead of 'mpl')
q.plot_engine = 'mpl'

#Set the number of significant figures to 2:
q.set_sigfigs(2)

#Then we enter the data:
#start with the values for the square root of height:
sqx = [1.  , 1.41, 1.73, 2., 2.24]
#and then, the corresponding times:
t = [ 0.33,  0.74,  0.67,  1.07,  1.1 ]

#Let us attribute an uncertainty of 0.15 to each measured values of t:
terr = 0.15

#We now make the plot. First, we create the plot object with the data
#Note that x and y refer to the x and y axes
fig = q.MakePlot( xdata = sqx, xname = "sqrt(distance) [m^0.5]",
                  ydata = t, yerr = terr, yname = "time [s]",
                  data_name = "My data")

#Ask QExpy to also determine the line of best fit
fig.fit("linear")

#Then, we show it:
fig.show()
```

Output D.8:

```
-----------------Fit results-------------------
Fit of  My data  to  linear
Fit parameters:
My data_linear_fit0_fitpars_intercept = -0.24 +/- 0.22,
My data_linear_fit0_fitpars_slope = 0.61 +/- 0.13

Correlation matrix:
[[ 1.     -0.968]
 [-0.968  1.    ]]

chi2/ndof = 2.04/2
---------------End fit results----------------
```

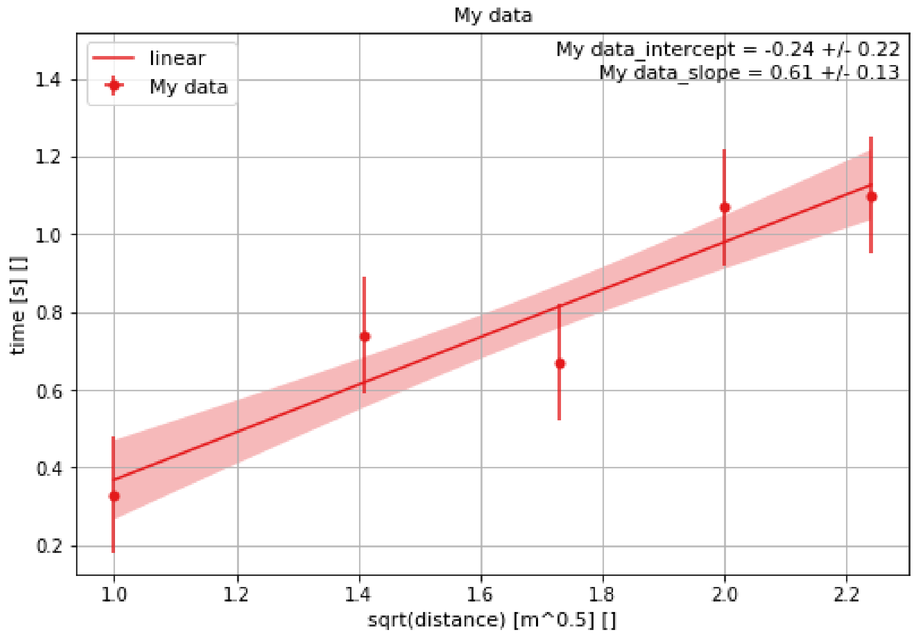


Figure D.3: QExpy plot of t versus $\sqrt{x}$ and line of best fit.

The plot in Figure D.3 shows that the data points are consistent with falling on a straight line, when their error bars are taken into account. We've also asked QExpy to show us the line of best fit to the data, represented by the line with the shaded area. When we asked for the line of best fit, QExpy not only drew the line, but also gave us the values and uncertainties for the slope and the intercept of the line. The shaded area around the line corresponds to other possible lines that one would obtain using different values of the slope and intercept within their corresponding uncertainties. The output also provides a line that tells us that `chi2/ndof = 2.04/2`; although you do not need to understand the details, this is a measure of how well the data are described by the line of best fit. Generally, the fit is assumed to be "good" if this ratio is close to 1 (the ratio is called "the reduced chi-squared"). The "correlation matrix" tells us how the best fit value of the slope is linked to the best fit value of the intercept, which you do not need to worry about here.

Since we expect the slope of the data to be k, this provides us a method to determine k from the data as $(0.61 \pm 0.13)\,\mathrm{s}\,\mathrm{m}^{-\frac{1}{2}}$. **Performing a linear fit of the data is the best way to determine a constant of proportionality between the measurements**. Finally, we expect the intercept to be equal to zero according to our model. The best fit line from QExpy has an intercept of $(-0.24 \pm 0.22)\,\mathrm{s}$, which is slightly below, but consistent, with zero. From these data, we would conclude that the measurements are consistent with Chloë's Theory.

D.5 Advanced topics

This section introduces a few more advanced topics that allow you to use computer programming to simplifying many tasks. In this section, we will show you how you can write your own program to numerically estimate the value of an integral of any function.

D.5.1 Defining your own functions

Although Python provides many modules and functions, it is often useful to be able to define your own functions. For example, suppose that you would like to define a function that calculates $\frac{1}{3}x^2 + \frac{1}{4}x^3 + \cos(2x)$, for a given value of x. This is done easily using the `def` keyword in Python:

Python Code D.9: Defining a function

```
#import the math module in order to use cos
import math as m

#define our function and call it myfunction:
def myfunction(x):
  return x**2 / 3 + x**3 / 4 + m.cos(2*x)

#Test our function by printing out the result of evaluating it at x = 3
print( myfunction(3) )
```

Output D.9:

```
10.710170286650365
```

A few things to note about the code above:

- Functions are defined using the `def` keyword followed by the name that we choose for the function (in our case, `myfunction`)
- If functions take arguments, those are specified in parenthesis after the name of the function (in our case, we have one argument that we chose to call `x`)
- After the name of the function and the arguments, we place a colon
- The code that belongs to the function, after the colon, must be indented (this allows Python to know where the code for the function ends)
- The function can "return" a value; this is done by using the `return` keyword.
- We used the "operator" `**` to take the power of a number (`x**2`), and the operator `*`, to multiply numbers. Python would not understand something like `2x`; you need to use the multiplication operator, i.e. `2*x`.

In the example above, we wrote a Python function to represent a mathematical function. However, one can write a function to execute any set of tasks, not just to apply a mathematical function. Python functions are very useful in order to avoid having to repeatedly type the same code.

Recall that the `numpy` module allows us to apply functions to arrays of numbers, instead of a single number. We can modify the code above slightly so that, if the argument to the function, `x`, is an array, the function will gracefully return an array of numbers to which the function has been applied. This is done by simply replacing the call to the `math` version of the `cos` function by using the `numpy` version:

Python Code D.10: Defining a function that works on an array

```
#import the numpy module in order to use cos to an array
```

```
import numpy as np

#define our function and call it myfunction:
def myfunction(x):
  return x**2 / 3 + x**3 / 4 + np.cos(2*x)

#Test our function by printing out the result of evaluating it at x = 3 (same
    as before)
print( myfunction(3) )

#Test it with an array
xvals = np.array([1,2,3])
print ( myfunction(xvals) )
```

Output D.10:

```
10.710170286650365
[ 0.1671865   2.67968971 10.71017029]
```

where we created the array `xvals` using the `numpy` module.

D.5.2 Using a loop to calculate an integral

The ability to define our own functions in Python allows us to easily simplify complex tasks. Using "loops" is another way that computer programming can greatly simplify calculations that would otherwise be very tedious. In a loop, one is able to repeat the same task many times. The example below simply prints out a statement five times:

Python Code D.11: A simple loop

```
#A loop to print out a statement 5 times:

for i in range(5):
  print("The value of i is ",i)
```

Output D.11:

```
The value of i is  0
The value of i is  1
The value of i is  2
The value of i is  3
The value of i is  4
```

A few notes on the code above:

- The loop is defined by using the keywords `for ... in`
- The value after the keyword `for` is the "iterator" variable and will have a different value each time that the code inside of the loop is run (in our case, we called the variable `i`)
- The value after the keyword `in` is an array of values that the iterator will take
- The `range(N)` function returns an array of `N` integer values between 0 and `N-1` (in our case, this returns the five values 0,1,2,3,4)
- The code to be executed at each "iteration" of the loop is preceded by a colon and indented (in the same way as the code for a function also follows a colon and is

indented)

We now have all of the tools to evaluate an integral numerically. Recall that the integral of the function $f(x)$ between x_a and x_b is simply a sum:

$$\int_{x_a}^{x_b} f(x)dx = \lim_{\Delta x \to 0} \sum_{i=0}^{i=N-1} f(x_i)\Delta x$$
$$\Delta x = \frac{x_b - x_a}{N}$$
$$x_i = x_a + i\Delta x$$

The limit of $\Delta x \to 0$ is equivalent to the limit $N \to \infty$. Our strategy for evaluating the integral is:

1. Define a Python function for $f(x)$.
2. Create an array, `xvals`, of N values of x between x_a and x_b.
3. Evaluate the function for all those values and store those into an array, `fvals`.
4. Loop over all of the values in the array `fvals`, multiply them by Δx, and sum them together.

Let's use Python to evaluate the integral of the function $f(x) = 4x^3 + 3x^2 + 5$ between $x = 1$ and $x = 5$:

Python Code D.12: Numerical integration of a function

```
#import numpy to work with arrays:
import numpy as np

#define our function
def f(x):
  return 4*x**3 + 3*x**2 + 5

#Make N and the range of integration variables:
N = 1000
xmin = 1
xmax = 5

#create the array of values of x between xmin and xmax
xvals = np.linspace(xmin, xmax, N)

#evaluate the function at all those values of x
fvals = f(xvals)

#calculate delta x
deltax = (xmax - xmin) / N

#initialize the sum to be zero:
sum = 0
```

```python
#loop over the values fvals and add them to the sum
for fi in fvals:
  sum = sum + fi*deltax

#print the result:
print("The integral between {} and {} using {} steps is {:.2f} ".format(xmin,
    xmax, N, sum))
```

Output D.12:

```
The integral between 1 and 5 using 1000 steps is 768.42
```

One can easily integrate the above function analytically and obtain the exact result of 768. The numerical answer will approach the exact answer as we make N bigger. Of course, the power of numerical integration is to use it when the function cannot be integrated analytically.

Checkpoint D-2

What value of N should you use above in order to get within 0.01 of the exact analytic answer?