

The Art of Modelling: Introduction to Physics

Ryan D. Martin

May 24, 2018

Contents

1	The Scientific Method and Physics	3
1.1	Science and the Scientific Method	3
1.2	Theories and models	5
1.3	Fighting intuition	6
1.4	The scope of Physics	7
1.4.1	Classical Physics	7
	Mechanics	7
	Electromagnetism	8
1.4.2	Modern Physics	8
	Quantum mechanics and particle physics	9
	The Special and General Theories of Relativity	9
	Cosmology and astrophysics	9
	Particle astrophysics	9
1.5	Thinking like a physicist	10
1.6	Summary	11
2	Comparing Model and Experiment	13
2.1	Orders of magnitude	13
2.2	Units and dimensions	14
2.2.1	Base dimensions and their SI units	15
2.2.2	Dimensional analysis	16
2.3	Making measurements	18
2.3.1	Measurement uncertainties	20
	Determining the central value and uncertainty	20
	Random and systematic sources of error/uncertainty	22
	Propagating uncertainties	23
2.3.2	Reporting measured values	26
2.3.3	Comparing model and measurement - discussing a result	27
2.4	Summary	29
3	Describing motion in one dimension	31
3.1	Motion with constant speed	32
3.2	Motion with constant acceleration	36
3.2.1	Visualizing motion with constant acceleration	40

3.2.2	Speed versus velocity	42
3.3	Using calculus to describe motion	42
3.3.1	Instantaneous and average velocity	42
3.3.2	Using calculus to obtain acceleration from position	44
3.3.3	Using calculus to obtain position from acceleration	45
3.4	Relative motion	47
3.5	Summary	51
3.5.1	Important equations	51
3.6	Problems	52
4	Describing motion in multiple dimensions	57
4.1	Motion in two dimensions	57
4.1.1	Using vectors to describe motion in two dimensions	57
4.1.2	Accelerated motion when the velocity vector changes direction	63
4.1.3	Relative motion	66
4.2	Motion in three dimensions	68
4.3	Circular motion	70
4.3.1	Period and frequency	75
4.4	Summary	77
A	Calculus	79
A.1	Functions of real numbers	79
A.2	Derivatives	82
A.2.1	Common derivatives and properties	83
A.2.2	Partial derivatives and gradients	86
A.3	Anti-derivatives and integrals	89
A.3.1	Common anti-derivative and properties	94
A.3.2	Common uses of integrals in Physics - from a sum to an integral	95
B	Vectors	99
B.1	Coordinate systems	99
B.1.1	1D Coordinate systems	99
B.1.2	2D Coordinate systems	100
B.1.3	3D Coordinate systems	102
B.2	Vectors	103
B.2.1	Unit vectors	104
B.2.2	Notations and representation of vectors	105
B.3	Vector algebra	106
B.3.1	Multiplication/division of a vector by a scalar	106
B.3.2	Addition/subtraction of two vectors	107
B.3.3	The scalar product	108
B.3.4	The vector product	109
B.4	Example uses of vectors in physics	110
B.4.1	Kinematics and vector equations	110
B.4.2	Work and scalar products	112

B.4.3	Using vectors to describe rotational motion	113
B.4.4	Torque and vector products	113
C	The Python programming language	115
C.1	A quick intro to programming	115
C.2	Arrays	116
C.3	Plotting	117
C.4	The QExpy python package for experimental physics	119
C.4.1	Propagating uncertainties	119
C.4.2	Plotting experimental data with uncertainties	120
C.5	Advanced topics	122
C.5.1	Defining your own functions	123
C.5.2	Using a loop to calculate an integral	124

1

The Scientific Method and Physics

Checkpoint 1-1: A good scientific theory...

- A) Must explain the physical world; may or may not be experimentally verifiable.
- B) Prove our models to be correct; must be experimentally verifiable.
- C) Describe the physical world; must be experimentally verifiable.
- D) Must disprove other theories; may or may not be experimentally verifiable.

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

1.1 Science and the Scientific Method

Science is an attempt to *describe* 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 1 m or 2 m and see if their measurement is consistent 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:

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 :

$$\begin{aligned} t &\propto \sqrt{h} \\ \rightarrow t &= k\sqrt{h} \end{aligned}$$

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

$$\begin{aligned} t_1 &= k\sqrt{h_1} = k\sqrt{(1 \text{ m})} \\ t_2 &= k\sqrt{h_2} = k\sqrt{(2 \text{ m})} \end{aligned}$$

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

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

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 scientists to find the most general and succinct description of the physical world.

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 1-2: Fill in the blanks:

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

- A) explains, "Why?"
- B) describes, "How?"

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 prediction. We will consider a "model" to be a situation-specific description of a phenomenon *based on a theory*. 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 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 1-3: 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 repulsion 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 1-4: 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, it is possible for an object to be 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. Both of these theories have however 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.

“Explanations aren't satisfying because they're beautiful, they're beautiful because they're satisfying. - Rebecca Goldstein”

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 are described.

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 a particular observation. 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 Newton's 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 1-5: Albert Einstein's postulation of General Relativity was useful in explaining the precession of Mercury because:

- A) It showed that Newton's model of Mercury was correct. (wrong)
- B) It showed that Newton's theory correctly described the motion of Uranus, but did not correctly describe the motion of more massive objects. (right)
- C) It proved that Verrier's theory was correct, but did not correctly describe the motion of more massive objects. (wrong)
- D) It proved Einstein's theory of General Relativity to be correct. (wrong)

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, leading to a theory that could not predict how particular situations evolve in time, but only 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 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 moving fast, which has been experimentally verified by flying, precise, atomic clocks on air planes 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.

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.

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



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

example, the 2015 Nobel Prize in Physics was awarded to Art McDonald (a Canadian physicist from Queen's University) for using neutrinos¹ 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.

¹Neutrinos are the lightest subatomic particles that we know of

1.6 Summary

- Science attempts to *describe* the physical world (answers the question “How?”, not “Why?”).
- The Scientific Method provides a prescription for arriving at theories that describe the physical world, that 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.
- Theories are typically valid only in well-defined situations.
- 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.

Reflect and Research 1-6: Research the following topics and try to answer these questions:

Which of the following is a branch of modern physics?

- A Newtonian mechanics
- B Classical electrodynamics
- C Quantum chromodynamics

Checkpoint 1-7: What particle helps to give mass to all of the massive elementary particles?

- A) Up quark
- B) Neutrino
- C) Photon
- D) Higgs Boson

Checkpoint 1-8: In the Double Slit Experiment, the behaviour of an electron travelling through a double slit depends on whether or not ...

- A) The electron was in an excitable state.
- B) An observer was measuring the system.
- C) the electron had a certain critical speed.

Checkpoint 1-9: Name that physicist! Who was the first to propose that the universe is expanding?

- A) Richard Feynman
- B) Stephen Hawking
- C) Edwin Hubble
- D) Erwin Schrödinger

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
- Be able to use a computer for simple data analysis

2.1 Orders of magnitude

Although one should try to fight intuition when building a model to describe a particular phenomenon, one should not abandon critical thinking and should always ask if a model (or a prediction of the model) makes sense. One of the most straightforward ways to verify 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.

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?

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 m (3 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 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.

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 Google these!

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 Earth	
Radius of the Sun	
Distance to the Moon	
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 talk about 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 jiffies¹ 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 unit.

Checkpoint 2-2: The speed limit on a highway:

- A) has dimension of length over time and can be expressed in units of kilometers per hour
- B) has dimension of length can be expressed in units of kilometers
- C) has dimension of time over length and can be expressed in units of meters per second
- D) has 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 expressing 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 ?? and thus in terms of the corresponding base SI units.

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 $\frac{L}{T}$ (length over time) and the corresponding SI unit is m/s (meters per second)² and corresponds to a measure of how much distance an object can cover per unit time (the higher the speed, the larger the distance covered per unit time). Table ?? shows a few derived dimensions and their corresponding SI units.

By convention, we can indicate the dimension of a quantity, X , by writing it in square

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

²Note 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.

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 [0]	unitless []

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

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.

brackets, $[X]$. For example, $[X] = I$, would mean that the quantity X has dimensions of electric current. Similarly, we can indicate the SI units of X with $SI[X]$; since X has dimensions 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. 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. You can only add or subtract two quantities if they have the same dimension: $[X+Y] = [X] = [Y]$

2. The dimension of the product is the product of the dimensions: $[XY] = [X] \cdot [Y]$
3. The dimension of the ratio 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 dimensions of mass multiplied by acceleration. What are the dimensions and SI units of force, expressed in terms of the base dimensions and units?

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

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

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

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

and the SI units of force are thus:

$$SI[\text{force}] = \frac{\text{kg} \cdot \text{m}}{\text{s}^2}$$

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

According to Table ??, voltage has dimensions:

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

while force, speed and current have dimensions:

$$[\text{force}] = M \cdot L \cdot T^{-2}$$

$$[\text{speed}] = L \cdot T^{-1}$$

$$[\text{current}] = I$$

The dimension of force multiplied by speed divided by electric charge

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

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 describe a situation, your model will typically provide a value for a quantity that you are interested in modelling. You should always use dimensional analysis to ensure that the dimension of the quantity your model predicts has the correct dimension. For example, suppose that you model the speed, v , that an object has after falling from a height of 100 m on the surface of the planet Mars. Presumably, v will depend on the mass and radius of the planet. You can be guaranteed that your model for v is incorrect if the dimension of v is not speed. Dimensional analysis should always be used to check that your model is not incorrect (note that getting the correct dimension is not a guarantee of the model being correct, only that it is “not definitely wrong”). Similarly, you should also use order of magnitude estimates to evaluate whether your model gives a reasonable prediction.

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

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 , from 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 ??, 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 $k = \frac{t}{\sqrt{x}}$:

x [m]	t [s]	\sqrt{x} [m $^{\frac{1}{2}}$]	k [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 ??, 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 ??.

x [m]	t [s]	\sqrt{x} [m $^{\frac{1}{2}}$]	k [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 $x = 3$ 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 ?? 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, $x = 3$ m, we obtained a spread in values of the drop time, t , between roughly 0.6 s and 1.0 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 ?? show that for a fixed experimental setup (a drop height of 3 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”:

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

where 0.76 s is called the “central value” and 0.15 s the “uncertainty” or the “error”³. 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 s and 0.91 s (given by 0.76 s - 0.15 s and 0.76 s + 0.15 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 which you can learn about on the internet or in a more advanced course.

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 ???

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 ?. 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:

$$\bar{t} = \frac{1}{N} \sum_{i=1}^{i=N} t_i = \frac{t_1 + t_2 + t_3 + \dots + t_N}{N} \quad (2.1)$$

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

$$\sigma_t = \sqrt{\sigma_t^2} \quad (2.3)$$

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.

³We use the word error as a synonym for uncertainty, not “mistake”.

Example 2-4: Calculate the mean and standard deviation of the values for k from Table ??.

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:

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

We can now calculate the variance:

$$\sigma_k^2 = \frac{1}{4}[(0.58 - 0.44)^2 + (0.44 - 0.44)^2 + (0.37 - 0.44)^2 + (0.42 - 0.44)^2 + (0.38 - 0.44)^2] = 7.3 \times 10^{-3} \text{ s}^2 \text{ m}$$

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

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

Using the mean and standard deviation, we would quote our value of k as $k = (0.44 \pm 0.09) \text{ s m}^{-\frac{1}{2}}$.

Any value that we measure should 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 value⁴. 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.

For example, we would quote the length of the grey object in Figure 2.1 to be $L = (2.8 \pm 0.5) \text{ cm}$ based on the rules in Table 2.6, since 2.8 cm is the closest value on the ruler

⁴In 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.

that matches the length of the object and 0.5 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.

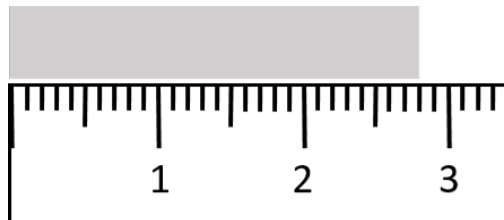


Figure 2.1: The length of the grey rectangle would be quoted as $L = (2.8 \pm 0.5)$ cm using the rule of “half the smallest division”.

The **relative uncertainty** in a measured value is given by dividing the uncertainty by the central value, and expressing the result in percent. For example, the relative uncertainty in $t = (0.76 \pm 0.15)$ s is given by $\frac{0.15}{0.76} = 20\%$. 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 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.

In this section, we want to further emphasize the difference between “error” and “mistake”. “Uncertainty” or “error” in a measurement comes from an unavoidable source that affects experimental results. A “mistake” also affects experimental results, but is preventable. If a “mistake” occurs in physics, the experiment is generally re-done and the previous data is discarded. The broad term “human error” may refer to a variety of errors or mistakes, so it is best to avoid this term in experimental physics (especially in your lab reports)!

The following table refers to examples of error that may be confused with "human error" that could be more accurately explained:

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 random 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 error - a type of random error (in this situation).
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 - a type of random error.

Table 2.6: Different situations in the lab and type of error.

Propagating uncertainties

Going back to the data in Table ??, we found that for a known drop height of $x = 3$ m, we measured different values of the drop time, which we found to be $t = (0.76 \pm 0.15)$ s (using the mean and standard deviation). We also calculated a value of k corresponding to each value of t , and found $k = (0.44 \pm 0.09) \text{ s m}^{-\frac{1}{2}}$. Suppose that we did not have access to the individual values of t , but only to the value of $t = (0.76 \pm 0.15)$ 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 from the measurements. 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 on 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:

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

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

uncertainty is given by:

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

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

2. The Min-Max method

A pedagogical way to determine k and its uncertainty is to use the “Min-Max method”. Since $k = \frac{t}{\sqrt{x}}$, 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 , $t^{\min} = 0.61 \text{ s}$ and the maximum value of t , $t^{\max} = 0.91 \text{ s}$:

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

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:

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

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 m) which of course is not realistic. Let us suppose that we have measured x to within 1 cm so that $x = (3.00 \pm 0.01) \text{ m}$. 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. At the end of the chapter, 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 2-5: Use the derivative method to evaluate $k = \frac{t}{\sqrt{x}}$ for $x = (3.00 \pm 0.01) \text{ m}$ and $t = (0.76 \pm 0.15) \text{ s}$.

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

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

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

$$\frac{\partial k}{\partial t} = \frac{1}{\sqrt{x}} \frac{d}{dt} t = \frac{1}{\sqrt{x}} = \frac{1}{\sqrt{(3 \text{ m})}} = 0.58 \text{ m}^{-\frac{1}{2}}$$

$$\frac{\partial k}{\partial x} = t \frac{d}{dx} x^{-\frac{1}{2}} = -\frac{1}{2} t x^{-\frac{3}{2}} = -\frac{1}{2} (0.76 \text{ s}) (3.00 \text{ m})^{-\frac{3}{2}} = -0.073 \text{ s m}^{-\frac{3}{2}}$$

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

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

So we find that:

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

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

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 2.7. 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 \frac{df}{dx} \sigma_x \right $

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 ± 0.5) m in (4.0 ± 0.5) s. What is the speed (with uncertainty) of the llama?

2.3.2 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 sentence describing 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”⁵ 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. $(10\,123 \pm 310)$ m would be $(10.12 \pm 0.31) \times 10^3$ m or $(101.2 \pm 3.1) \times 10^2$ m)

⁵Significant digits are those excluding leading and trailing zeroes.

Checkpoint 2-5: Someone has measured the average height of tables in the laboratory to be 1.0535 m with a standard deviation of 0.0525 m. What is the best way to present this measurement?

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

2.3.3 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 $k = (0.44 \pm 0.09) \text{ s m}^{-\frac{1}{2}}$. 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 Theory of Gravity does predict a value for k of $0.45 \text{ s m}^{-\frac{1}{2}}$ 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 $k = (0.55 \pm 0.08) \text{ s m}^{-\frac{1}{2}}$ so that the lowest value compatible with our measurement, $k = 0.55 \text{ s m}^{-\frac{1}{2}} - 0.08 \text{ s m}^{-\frac{1}{2}} = 0.47 \text{ s m}^{-\frac{1}{2}}$ is not compatible with Newton's prediction. Would we conclude that our measurement invalidates Newton's theory? The answer is: it depends... And 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?
- 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?
- What assumptions were made in calculating your measured value?
- What assumptions were made in determining the model prediction?

In the above, our value of $k = (0.55 \pm 0.08) \text{ s m}^{-\frac{1}{2}}$ 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 $\sigma_k = 0.08 \text{ s m}^{-\frac{1}{2}}$, 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 un-

certainty, 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.

Example 2-6: During a physics lecture, you look under your seat and find a sheet containing data from what seems to be a juggling experiment.

Attempting to model the situation with math, the students have written this equation at the bottom of the sheet: $= \frac{v_2^2 - v_1^2}{2a}$

where

- 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 (m/s^2) - $9.8m/s^2$

Unfortunately, the students spilled ketchup on the left hand side of their equation, making it illegible. Luckily, you're proficient in unit analysis. What variable were the students trying to calculate, based on this model?

2.4 Summary

- 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.
- 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.
- 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!

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).

Think about it: You are taking your 6 year old cousin, Lily, to see the aquarium in Toronto. You are sitting on the train in Kingston waiting to leave the station. Lily exclaims that your train must be moving, and is excited that you are on your way to Toronto. You wonder why she said this, so you look out the window and notice that train beside you is moving backwards. You conclude that your train is not moving after all. How do you explain this to Lily? Do you know for certain that you aren’t moving?

Learning Objectives:

- Describe motion in 1D using functions and defining an axis.
- Define position, velocity, speed, and acceleration.
- Use calculus to describe motion.
- Define the meaning of an inertial frame of reference.
- Use Galilean and Lorentz transformations to convert the description of an object’s position from one inertial frame to another.

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.

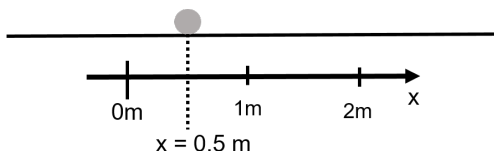


Figure 3.1: In order to describe the motion of the grey ball along a straight line, we introduce the x -axis, represented by an arrow to indicate the direction of increasing x , and the location of the origin, where $x = 0$ m. Given our choice of origin, the ball is currently at a position of $x = 0.5$ 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). The easiest way to visualize the values in the table is to plot them on a graph. 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. We can easily plot these values in Python:

Python Code 3.1: Plotting position versus time

```
#First, we load pylab module for plotting
import pylab as pl
#We define t as a list of values (note the square brackets):
t = [0.0, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0]
#Similarly, we define the corresponding positions:
```

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.

```

x = [0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0]
#Define the plot:
pl.plot(t,x, '. ')# the '.' means that it will show the actual points instead
    of a line
#Set the range of the axes, add some labels and a grid
pl.ylim(0,6)
pl.xlim(0,10)
pl.xlabel('time [s]')
pl.ylabel('position [m]')
pl.grid()
#Show the plot
pl.show()

```

Output 3.1:

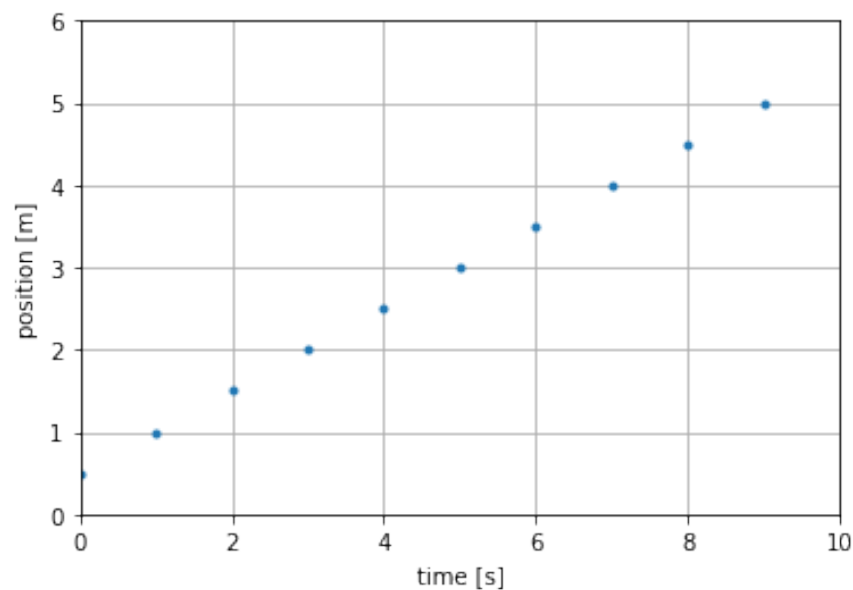


Figure 3.2: Plot of position as a function of time using the values from Table 3.1.

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). 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 $x = 0.5$ 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 $t = 9.5$ s, if it continues its motion undisturbed?

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:

$$x(t) = a + bt$$

The constant a is the “offset” of the function, the value that the function has at $t = 0$ s. The constant b is the slope and gives the rate of change of the position as a function of time. We can determine the values for the constants a and b by choosing any two rows from Table 3.1 (to determine 2 unknown quantities, you need 2 equations), and obtain 2 equations and 2 unknowns. For example, choosing the points where $t = 0$ s and $t = 2.0$ s:

$$\begin{aligned} x(t = 0 \text{ s}) &= 0.5 \text{ m} = a + b(0 \text{ s}) \\ x(t = 2.0 \text{ s}) &= 1.5 \text{ m} = a + b(2.0 \text{ s}) \end{aligned}$$

The first equation immediately gives $a = 0.5$ m, which we can substitute into the second equation to get b :

$$\begin{aligned} 1.5 \text{ m} &= a + b(2.0 \text{ s}) = 0.5 \text{ m} + b(2.0 \text{ s}) \\ \therefore b &= \frac{(1.5 \text{ m}) - (0.5 \text{ m})}{(2.0 \text{ s})} = 0.5 \text{ m/s} \end{aligned}$$

which gives us the functional form for $x(t)$:

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

where you should note that a and b have different dimensions. Since a is added to something that must then give dimensions of length (for position, x), a has dimensions of length. b is multiplied by time, and that product must have dimensions of length as well; b thus has dimensions of length over time, or “speed” (with S.I. units of m/s).

We can generalize the description of an object whose position increases linearly with time as:

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

where x_0 is the position of the object at time $t = 0$ s (a from above), and v_x corresponds to the distance that the object covers per unit time (b from above) along the x -axis. We call v_x the “velocity” of the object. 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.

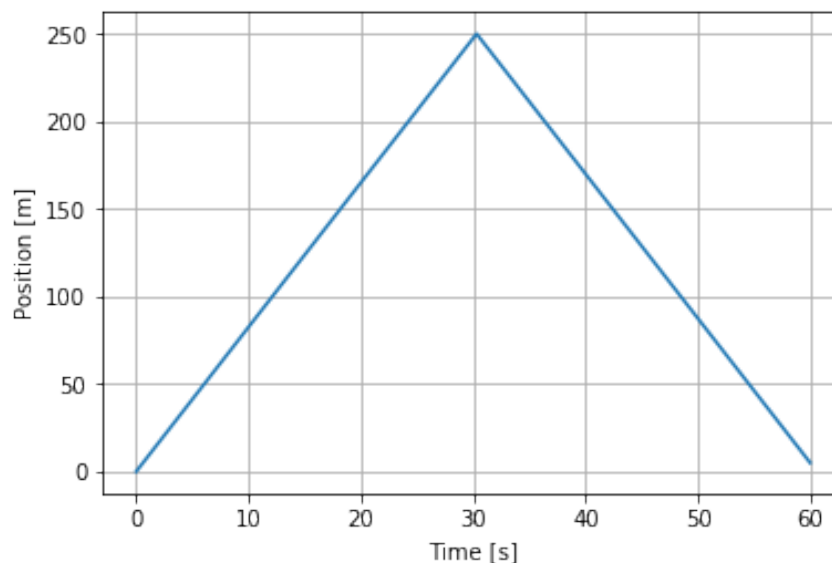


Figure 3.3: Position as a function of time for an object.

Checkpoint 3-2: Referring to Figure 3.3, what can you say about the motion of the object?

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

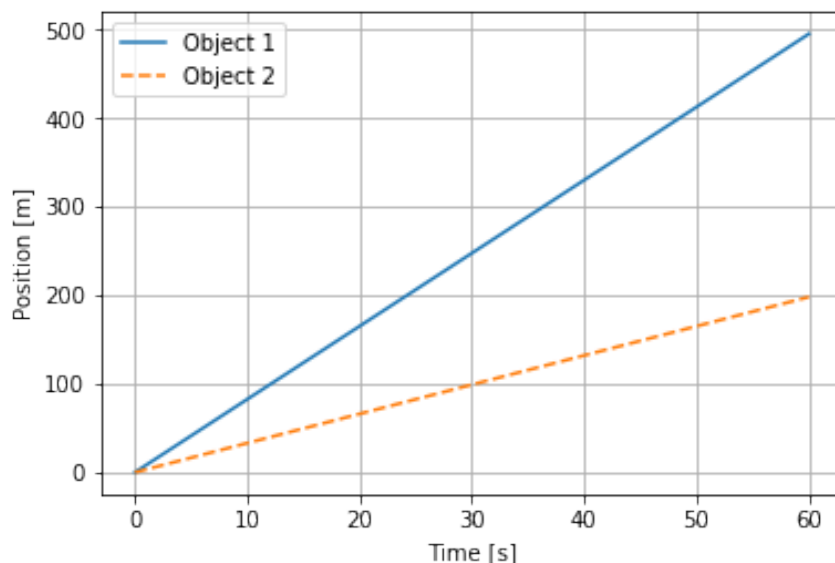


Figure 3.4: Positions as a function of time for two objects.

Checkpoint 3-3: 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). Suppose that we wish to describe the position of a falling object that we released from rest at time $t = 0$ s. The object will start with a velocity of 0 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:

$$x(t) = x_0 + v_x t \quad (3.1)$$

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

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

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:

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

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).

Now that we have an understanding of acceleration, how do we 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.

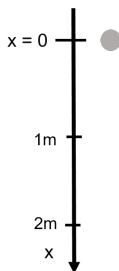


Figure 3.5: *X-axis for an object that starts at rest at $x = 0\text{ m}$ when $t = 0\text{ s}$ and falls downwards (in the direction of increasing x).*

Let us work out the corresponding equation for position as a function of time for accelerated motion using the x -axis depicted in Figure 3.5. We will determine $x(t)$ for the grey ball that starts at rest ($v_{0x} = 0\text{ m/s}$) at the position $x = 0\text{ m}$ at time $t = 0\text{ s}$ with a constant positive acceleration $a_x = 10\text{ m/s}^2$. We would like to use equation 3.1, but we cannot because it only applies if the velocity is constant. To remedy this, we pretend (we “approximate”) that for a very small amount of time, the velocity is almost constant. Let us take a very small interval in time, say $\Delta t = 0.001\text{ s}$, and approximate that the velocity is constant during that interval.

At $t = 0\text{ s}$, we have $x = 0\text{ m}$, $v_{0x} = 0\text{ m/s}$ and $a_x = 10\text{ m/s}^2$. We can use equation 3.2 to find the velocity at $t = \Delta t$ (at the end of the first interval):

$$\begin{aligned}v_x(t = \Delta t) &= v_{0x} + a_x \Delta t \\ &= (0\text{ m/s}) + a_x \Delta t \\ &= a_x \Delta t\end{aligned}$$

The average velocity during the first interval, v_1^{avg} is then given by averaging the velocity at the beginning and at the end of the interval:

$$\begin{aligned}
 v_1^{avg}(t = \Delta t) &= \frac{1}{2} [v(t = 0) + v(t = \Delta t)] \\
 &= \frac{1}{2} (v_{0x} + a_x \Delta t) \\
 &= \frac{1}{2} ((0 \text{ m/s}) + a_x \Delta t) \\
 &= \frac{1}{2} (10 \text{ m/s}^2)(0.001 \text{ s}) \\
 &= 0.005 \text{ m/s}
 \end{aligned}$$

Using the average velocity during the interval, we can use equation 3.1 to find the position at $t = \Delta t$:

$$\begin{aligned}
 x(t = \Delta t) &= x_0 + v_1^{avg} \Delta t \\
 &= (0 \text{ m}) + \frac{1}{2} a_x (\Delta t)^2 \\
 &= \frac{1}{2} (10 \text{ m/s}^2)(0.001 \text{ s})^2 \\
 &= 0.000\,005 \text{ m}
 \end{aligned}$$

Thus, at time $t = 0.001 \text{ s}$, the object will have a velocity of $v = 0.005 \text{ m/s}$ and be at a position $x = 0.000\,005 \text{ m}$. We can now use these values as the starting velocity and position for the next interval in time. Using variables, at the beginning of the second interval, the velocity is $v(t = \Delta t) = a_x \Delta t$ and at the end of the second interval, it will be $v(t = 2\Delta t) = 2a_x \Delta t$. The average velocity during the second interval is thus given by:

$$\begin{aligned}
 v_2^{avg}(t = 2\Delta t) &= \frac{1}{2} [v(t = \Delta t) + v(t = 2\Delta t)] \\
 &= \frac{1}{2} (a_x \Delta t + 2a_x \Delta t) \\
 &= \frac{3}{2} a_x \Delta t \\
 &= \frac{3}{2} (10 \text{ m/s}^2)(0.001 \text{ s}) \\
 &= 0.015 \text{ m/s}
 \end{aligned}$$

To find the position at the end of the second time interval, when $t = 2\Delta t$, we use equation 3.1 again, but with a different starting position and the average velocity that we just found:

$$\begin{aligned}
 x(t = 2\Delta t) &= x(t = \Delta t) + v_2^{avg} \Delta t \\
 &= \frac{1}{2} a_x (\Delta t)^2 + \frac{3}{2} a_x (\Delta t)^2 \\
 &= \frac{1}{2} a_x (2\Delta t)^2 \\
 &= \frac{1}{2} (10 \text{ m/s}^2)(2 \times 0.001 \text{ s})^2 = 0.000\,02 \text{ m}
 \end{aligned}$$

You can carry out this exercise to ultimately find the position at any time. However, if you carry it out over a few more intervals, you may notice the following pattern: For the N th interval when $t = N\Delta t$ at the end of the interval, we have:

$$\begin{aligned} v(t = (N - 1)\Delta t) &= a_x(N - 1)\Delta t && \text{(at beginning of interval } N) \\ v(t = N\Delta t) &= a_x N\Delta t && \text{(at end of interval } N) \\ v_N^{avg} &= \frac{1}{2}a_x(2N - 1)\Delta t && \text{(average during interval)} \\ x(t = N\Delta t) &= \frac{1}{2}a_x(N\Delta t)^2 && \text{(position at end of interval)} \end{aligned}$$

The last line gives us exactly what we were after, namely the position as a function of time for a constant acceleration, a_x , when the object started at rest at a position of $x = 0$ m:

$$x(t) = \frac{1}{2}a_x t^2 \quad (3.3)$$

If at $t = 0$, the object had an initial position along the x -axis of x_0 , then the position $x(t)$ would be shifted by an amount x_0 :

$$x(t) = x_0 + \frac{1}{2}a_x t^2 \quad (3.4)$$

Finally, if the object had an initial speed v_{0x} at $t = 0$, one can easily reproduce the iterations above to find that we need to add an additional term to account for this. We arrive at the general description of the position of an object moving in a straight line with acceleration, a_x :

$$\boxed{x(t) = x_0 + v_{0x}t + \frac{1}{2}a_x t^2} \quad (3.5)$$

Note that equation 3.1 is just a special case of the above when $a = 0$.

Example 3-1: A ball is thrown upwards with a velocity of 10 m/s. After which distance will the ball stop before falling back down? Assume that gravity causes a constant downwards acceleration of 9.8 m/s².

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 3.5, we have:

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

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:

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

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

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

and we find that the ball will rise by 5.10 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:

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

where the velocity changes linearly with time, and the position changes quadratically with time (it goes as t^2). Figure 3.6 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.

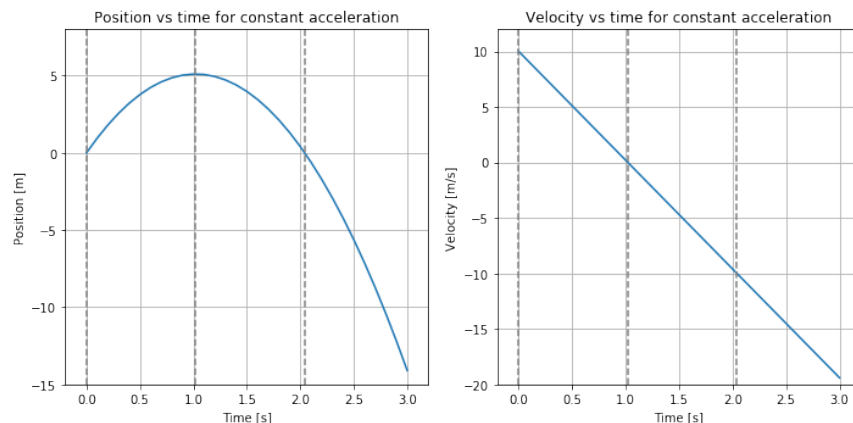


Figure 3.6: Position and speed as a function of time for the ball in example 3-1.

We can divide the motion into three parts (shown by the vertical dashed lines in Figure 3.6):

1) Between $t = 0$ s and $t = 1.02$ s

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

2) Between $t = 1.02$ s and $t = 2.04$ s

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

3) After $t = 2.04$ s

If nothing is there to stop the ball, it continues to move downwards with ever increasing velocity. The position continues to become more negative and the velocity continues to become larger in magnitude and more negative.

Checkpoint 3-4: Make a sketch of the acceleration as a function of time corresponding to the position and velocity shown in Figure 3.6.

3.2.2 Speed versus velocity

In the previous example, our language was not quite as precise as it should be when conducting science. Specifically, we need a way to distinguish the situation when the velocity is decreasing (becoming more negative), while the object is actually going faster and faster (after $t = 1.02$ s in Figure 3.6). We will use the term **speed** to refer to how fast an object is moving (how much distance it covers per unit time), and we will use the term **velocity** to also indicate the direction of the motion. In other words, the speed is the absolute value of the velocity¹. The speed is thus always positive, whereas the velocity can also be negative.

With this vocabulary, the speed of the ball in Figure 3.6 decreases between $t = 0$ s and $t = 1.02$ s, and increases thereafter. On the other hand, the velocity continuously decreases (it is always becoming more and more negative). Velocity is thus the more general term since it tells us both the speed and the direction of the motion.

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} , over the time interval. That average velocity will be an average over the interval, between some time t and $t + \Delta t$. If we shrink the time interval, and take the limit $\Delta t \rightarrow 0$, we can define the **instantaneous velocity**:

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

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:

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

¹This is true for one-dimensional motion, whereas in two or more dimensions, velocity is a vector and speed is the magnitude of that vector.

If we go back to Figure 3.6, 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 $t < 1.02$ s, the slope of the $x(t)$ graph is positive but decreasing (as is $v(t)$). At $t = 1.02$ s, the slope of $x(t)$ is instantaneously 0 m/s (as is the velocity). Finally, for $t > 1.02$ 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 $\lim_{\Delta t \rightarrow 0}$:

$$\boxed{v(t) = \lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt} = \frac{d}{dt}x(t)} \quad (3.6)$$

where we can really think of dt as $\lim_{\Delta t \rightarrow 0} \Delta t$, 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 $v(t)$:

$$\boxed{a_x(t) = \frac{dv}{dt} = \frac{d}{dt}v(t)} \quad (3.7)$$

Olivia's Thoughts 3-5: Speeding up or slowing down?

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 slope. 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.

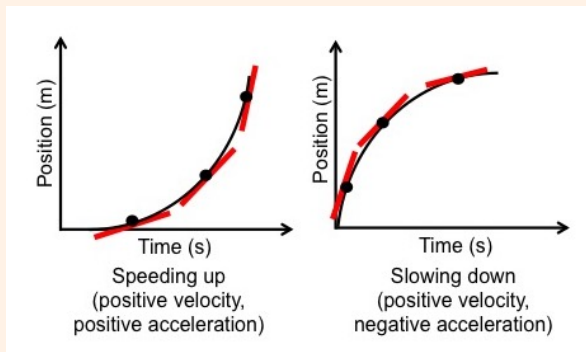


Figure 3.7: Two graphs of $x(t)$ showing tangent lines.

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 3.7 are describing the motion of a person in heavy wind. In the graph on the left, the person is running with the wind ($v(t)$ and $a(t)$ positive), and in the second graph the person is running against the wind ($v(t)$ 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):

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

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

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

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

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

as expected.

Checkpoint 3-6: 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 $x(t) = (40 \text{ m/s}^2)t^2 + (20 \text{ m/s}^3)t^3$. What is the acceleration of the vicuñas?

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

^aNever 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:

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

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:

$$v(t) = \frac{dx}{dt}$$

$$\therefore x(t) = \int v(t) dt$$

In the case where acceleration is constant, this gives:

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

$$= \int (v_{0x} + a_x t) dt$$

$$= v_{0x} t + \frac{1}{2} a_x t^2 + C'$$

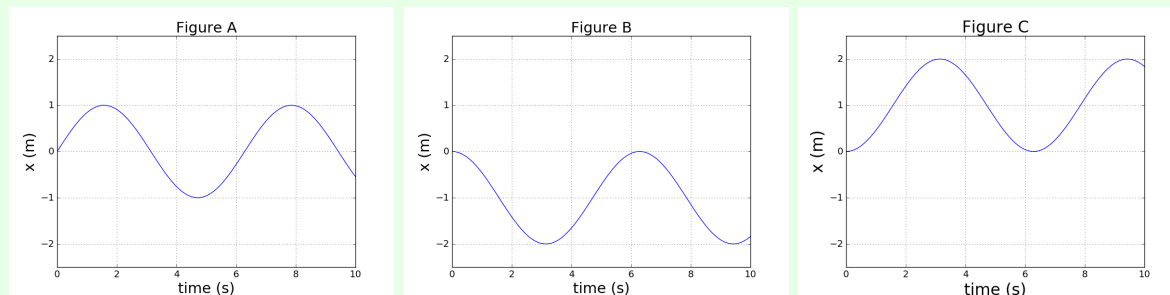
where C' is a different constant than the one we had when determining velocity. The constant is given by our initial conditions. If the object was located at position $x = x_0$ at time $t = 0$, then $C' = x_0$ and we recover the equation for position as a function of time for constant acceleration:

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

Checkpoint 3-7: 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 ($v(t) \propto t^3$)

Checkpoint 3-8: Choose the graph of $x(t)$ for the case when acceleration is given by $\cos(\omega t)$, where ω is a constant. The velocity and position are zero at $t = 0$



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

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, 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 surfer, 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.

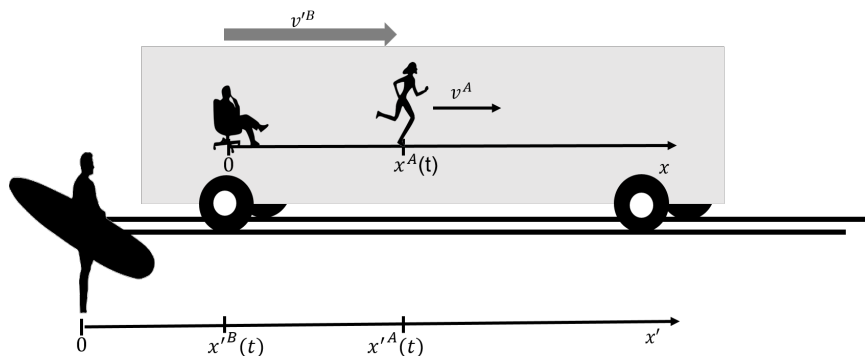


Figure 3.8: Alice is walking in the train and her position is described by both Brice, who is sitting in the train (using the x axis), and Igor, who is at rest on the ground (using the x' axis).

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.

Not only is it impossible to define a frame of reference that is truly at rest, the rules from transforming from one frame to the other depend on the speed between the reference frames. Our common experience is described by what we call “Galilean Relativity”, but if the speed between trains is very large, close to the speed of light, then we need to use Einstein's Special Theory of Relativity.

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 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 Brice), then the position of Brice's origin 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:

$$\boxed{x'^A(t) = x'^B(t) + x^A(t) = v'^B t + x^A(t)} \quad (3.8)$$

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:

$$v'^A(t) = \frac{d}{dt}x'^A(t) \quad (3.9)$$

$$= \frac{d}{dt}(v'^B t + x^A(t)) \quad (3.10)$$

$$= v'^B + \frac{d}{dt}x^A(t) \quad (3.11)$$

$$= v'^B + v^A(t) \quad (3.12)$$

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:

$$a'^A(t) = \frac{d}{dt}v'^A(t) \quad (3.13)$$

$$= \frac{d}{dt}(v'^B + v^A) \quad (3.14)$$

$$= 0 + \frac{d}{dt}v^A(t) \quad (3.15)$$

$$= a^A \quad (3.16)$$

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 $v'^B = 15$ 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 $v^A = 3$ 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?

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 origin of the boat's coordinate system was $x' = 0$. The origin of the boat's coordinate system 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:

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

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 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

- 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).
- 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” and is a function given by the time-derivative of position.
- The rate of change of velocity is called “acceleration” and is a function given by the time-derivative of velocity.
- Given a function for acceleration, $a_x(t)$, one can use its anti-derivative to determine velocity.
- Given a function for velocity, $v_x(t)$, one can use its anti-derivative to determine position.
- 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 a position $x^A(t)$ in a given inertial frame of reference, x , that is moving with a velocity v'^B compared to a different inertial frame of reference, x' , then the position of the object in the x' frame of reference is given by $x'(t) = v'^B t + x^A(t)$.

3.5.1 Important equations

If the position of an object is described by a function $x(t)$, then, its velocity, $v_x(t)$, and acceleration, $a_x(t)$, are given by:

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

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

Conversely, given the acceleration, $a_x(t)$, 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 $t = 0$:

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

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

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:

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

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

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

3.6 Problems

Problem 3-1: Rob is riding his bike at a speed of 8 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 m/s².

- Assuming it takes 3 seconds for the velociraptor to react, how long does it take from the moment Rob passes by for the velociraptor to catch up to him?
- 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?

Solution:

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 set $t = 0$ to be the moment the velociraptor starts running.

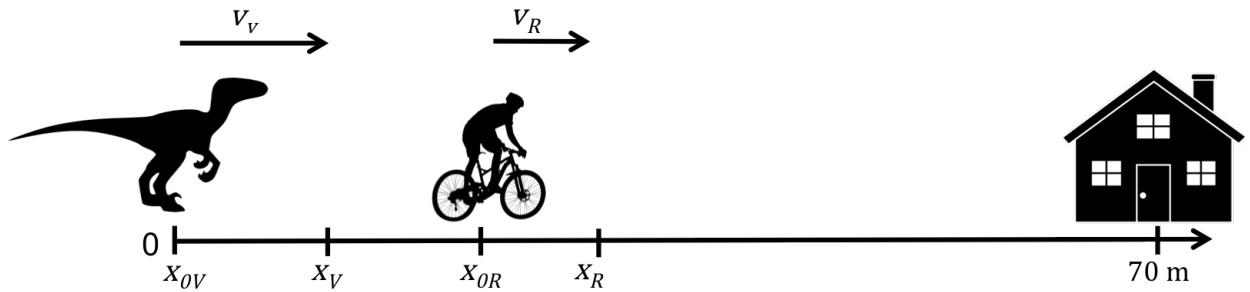


Figure 3.9: Rob is being chased by a velociraptor. At $t = 0$, Rob is a distance x_{0R} from the velociraptor. Safety is 70 m away from the origin.

Since we are dealing with constant accelerations, we can use the equation:

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

We can use a table to take stock of our known values and variables:

The value of x_{0R} is unknown but can be easily solved for. We defined $t = 0$ as the moment the velociraptor starts running. It takes 3 seconds for the velociraptor to react, so at $t = 0$, Rob has moved $(8\text{ m/s}) \times (3\text{ s}) = 24\text{ m}$.

Rob	Velociraptor
$x_{0R} = ?$	$x_{0V} = 0 \text{ m}$
$v_R = 8 \text{ m/s}$	$v_{0V} = 0 \text{ m/s}$
$a_R = 0 \text{ m/s}^2$	$a_V = 4 \text{ m/s}^2$
t	t
x_R	x_V

Next, we can write out our equations as,

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

Now what do we mean by “catch up”? It means that Rob and the velociraptor are in the same place at the same time. So, we are interested in the value of t when $x_R = x_V$. Setting the above equations equal to one another gives

$$x_{0R} + v_R t = \frac{1}{2} a_V t^2$$

that we can rearrange to get this quadratic:

$$0 = \frac{1}{2} a_V t^2 - v_R t - x_{0R}$$

By solving the quadratic, we find that $t = 6 \text{ s}$. This doesn’t quite give us the answer we want. We want to know how long it takes the velociraptor to catch up *from the moment Rob passes by*, so we have to add on the 3 s reaction time, giving a total time of 9 s.

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

Problem 3-2: Figure 3.10 shows a graph of the acceleration, $a(t)$, of a particle moving in one dimension. Draw the corresponding velocity and position graphs. Assume that $v(0) = x(0) = 0$.

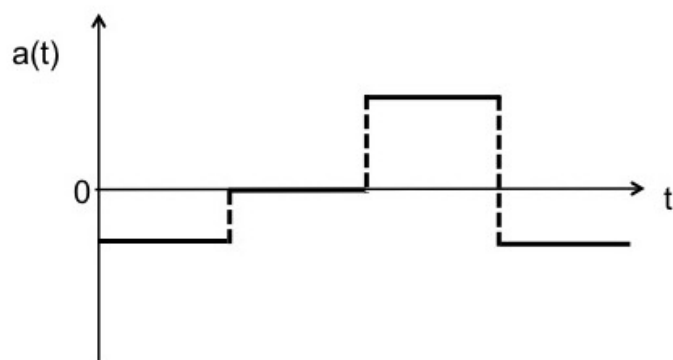
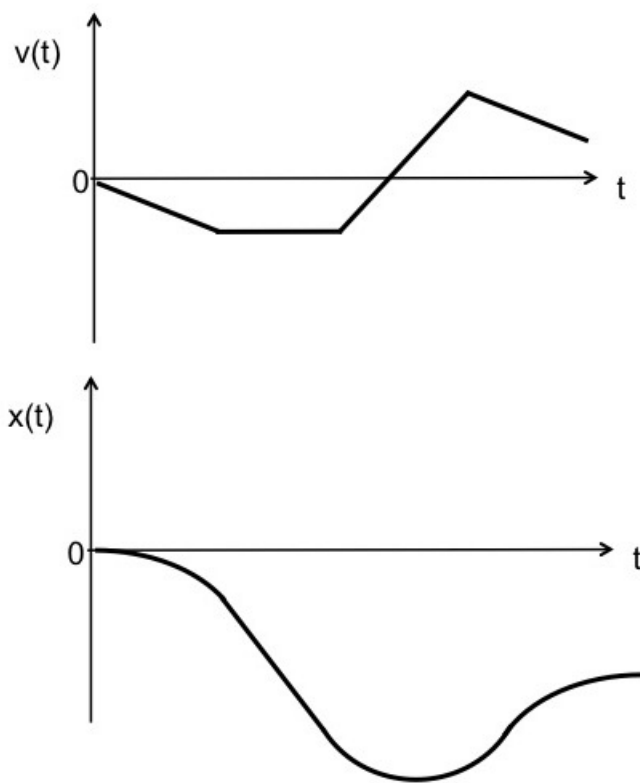


Figure 3.10: A graph of acceleration as a function of time. The scale and units are arbitrary.

Solution:



Solutions may vary, however there are a few key features that must be present. For the graph of $v(t)$:

- Velocity is zero at $t = 0$.
- When acceleration is negative, the velocity is (negative slope). It follows that when the acceleration is positive, the velocity is increasing.

- When the acceleration is zero, the velocity should be a flat line

For the graph of $x(t)$:

- Position is zero at $x = 0$
- When velocity is negative, $x(t)$ is decreasing. When velocity is positive, $x(t)$ is increasing
- Take note of the turning points on the graph of $x(t)$. The particle turns around (the position goes from increasing or decreasing, or from decreasing to increasing) when the velocity changes sign.
- When the velocity and acceleration are in the same direction, the slope of $x(t)$ increases in magnitude. When the acceleration and velocity are in opposite directions, the slope of $x(t)$ decreases in magnitude.

Reflect and Research 3-9:

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?

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.

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 quantify any displacement by a vector. First consider the case of an object moving at a constant velocity in a particular direction. We can describe the object at any time, t , using its position vector, $\vec{r}(t)$, which is a function of time:

$$\begin{aligned}\vec{r}(t = t_0) &= \vec{r}_1 \\ \vec{r}(t = t_0 + \Delta t) &= \vec{r}_2\end{aligned}$$

More generally, we can describe the x and y components of the position vector with independent functions, $x(t)$, and $y(t)$, 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. 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:

$$\vec{v} = \frac{\Delta\vec{r}}{\Delta t} \quad (4.1)$$

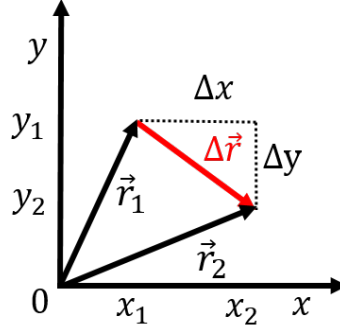


Figure 4.1: Illustration of a displacement vector, $\Delta\vec{r} = \vec{r}_2 - \vec{r}_1$, for an object that was located at position \vec{r}_1 at time t_1 and at position \vec{r}_2 at time $t_2 = t_1 + \Delta t$.

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:

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

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 m/s
- B) 5 m/s
- C) 10 m/s
- D) 15 m/s

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

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

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:

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

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 having two

independent components: 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:

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

For example, 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, $\vec{a} = a_x\hat{x} + a_y\hat{y}$, 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:

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

which really can be considered as two independent equations for the components of the velocity vector:

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

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:

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

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 $t = 0$ 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 m/s^2 and the acceleration in the y direction is -10 m/s^2 .

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

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:

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

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 $t = 10$ 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:

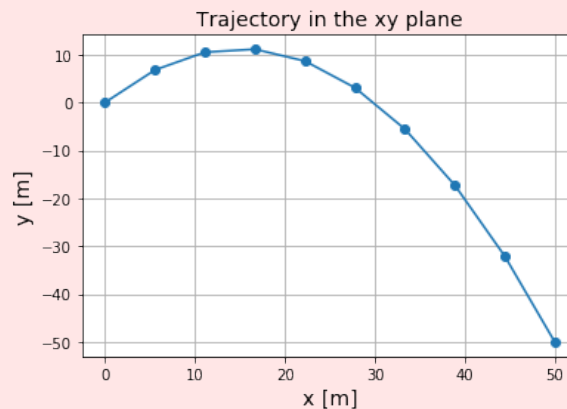


Figure 4.2: Parabolic trajectory of an object with no acceleration in the x direction and a negative acceleration in the y direction.

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

4.1.2 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 a non-zero 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. The length of the velocity vector is called the speed. 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. In this case, the acceleration would not result in a change of speed, but rather in a change of the direction of motion. This is exactly what happens when an object goes around in a circle with a constant speed (the direction of the velocity vector changes).

Figure 4.3 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:

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

and will thus 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

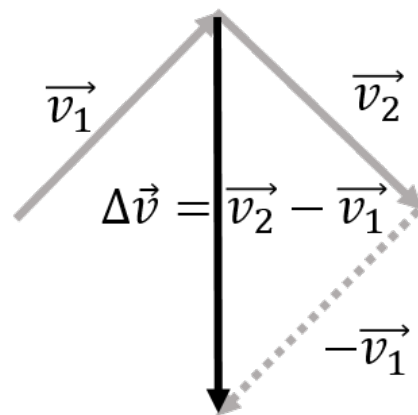


Figure 4.3: Illustration of how the direction of the velocity vector can change when speed is constant.

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:

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

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:

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

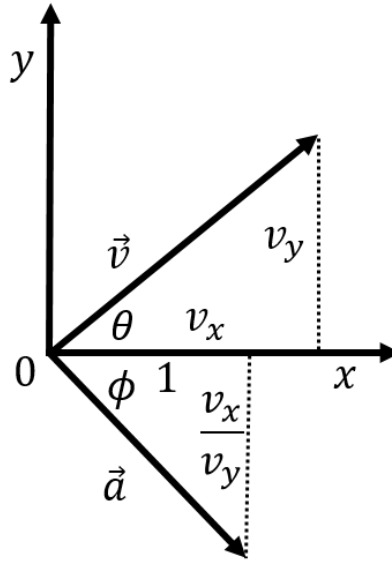


Figure 4.4: Illustration that the acceleration vector is perpendicular to the velocity vector if speed is constant.

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:

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

where most of the algebra that we did was to separate out the x and y components of the acceleration vector. The resulting acceleration vector is illustrated in Figure 4.4 along with

the velocity vector. Rather, a vector parallel to the acceleration vector 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). 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, $\vec{a}_{\parallel}(t)$ 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 4-3: 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.

4.1.3 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. $\vec{r}'^A(t)$ 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 4.5. 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 $\vec{r}^A(t)$.

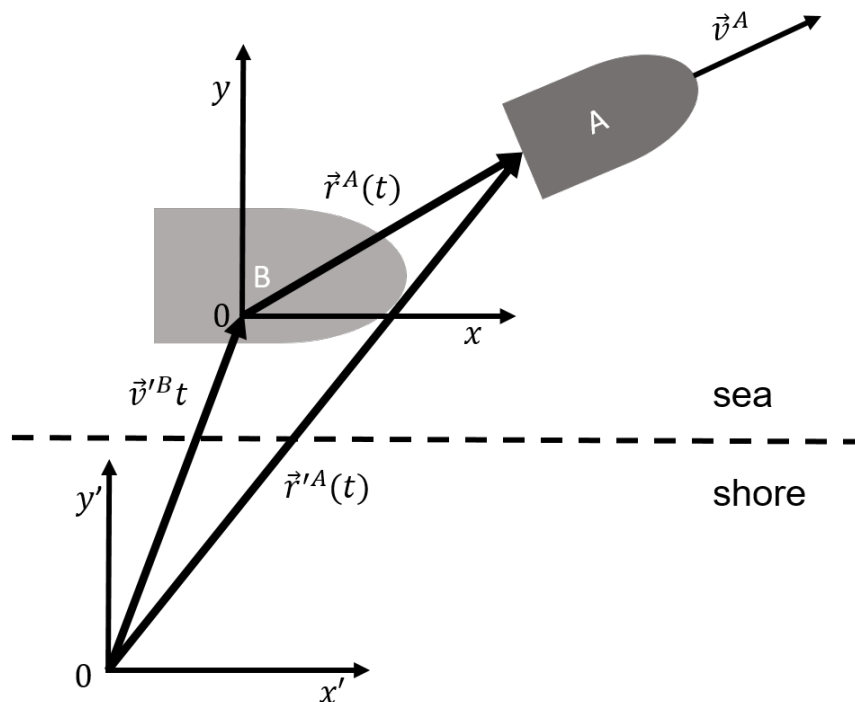


Figure 4.5: Example of converting from one reference frame to another in two dimensions using vector addition.

Writing this out by coordinate, we have:

$$\begin{aligned}x'^A(t) &= v'_x{}^B t + x^A(t) \\ y'^A(t) &= v'_y{}^B t + y^A(t)\end{aligned}$$

and for the velocities:

$$\begin{aligned}v'_x{}^A(t) &= v'_x{}^B + v_x^A(t) \\ v'_y{}^A(t) &= v'_y{}^B + v_y^A(t)\end{aligned}$$

Checkpoint 4-4: 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 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.

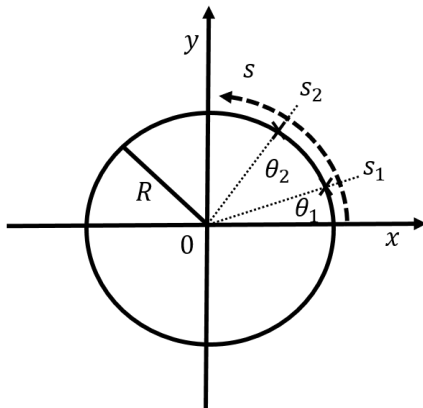


Figure 4.6: Describing the motion of an object around a circle of radius R .

Figure 4.6 shows how we can describe motion on a circle. 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 4.6, 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 4.6. 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:

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

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:

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

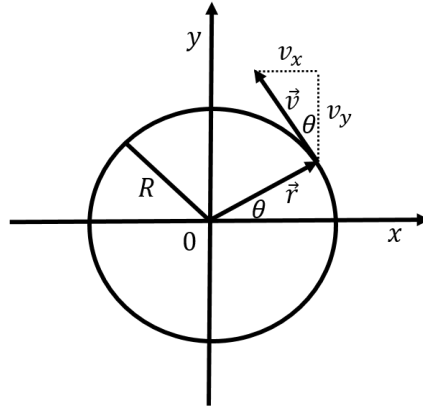


Figure 4.7: The position vector, $\vec{r}(t)$ is always perpendicular to the velocity vector, $\vec{v}(t)$, for motion on a circle.

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). 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 4.7 for an angle θ in the first quadrant ($0 < \theta < \frac{\pi}{2}$). In this case, you can note that the x component of the velocity is negative (in the equation above, and in the Figure). From the equation above, you can also see that $\frac{|v_x|}{|v_y|} = \tan(\theta)$, which is illustrated in Figure 4.7, 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 $\theta = \frac{s}{R}$ 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 as:

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

or, in terms of θ :

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

$$\boxed{\therefore \omega = \frac{d\theta}{dt}}$$

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 acceleration:

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

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

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

$$\boxed{a_d(t) = R\alpha}$$

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

$$s = R\theta \tag{4.8}$$

$$v_s = R\omega \tag{4.9}$$

$$a_s = R\alpha \tag{4.10}$$

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 as:

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

which corresponds to an object that is going around the circle faster and faster.

As you recall from section 4.1.2, 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 $\hat{v} = -\sin(\theta)\hat{x} + \cos(\theta)\hat{y}$:

$$\begin{aligned}\vec{a}_{\perp}(t) &= v\frac{d\hat{v}}{dt} \\ &= (\omega R)\frac{d}{dt}[-\sin(\theta)\hat{x} + \cos(\theta)\hat{y}] \\ &= \omega R\left[-\frac{d}{dt}\sin(\theta)\hat{x} + \frac{d}{dt}\cos(\theta)\hat{y}\right] \\ &= \omega R\left[-\cos(\theta)\frac{d\theta}{dt}\hat{x} - \sin(\theta)\frac{d\theta}{dt}\hat{y}\right] \\ &= \omega R[-\cos(\theta)\omega\hat{x} - \sin(\theta)\omega\hat{y}] \\ \boxed{\vec{a}_{\perp}(t) = \omega^2 R[-\cos(\theta)\hat{x} - \sin(\theta)\hat{y}]} &\quad (4.11)\end{aligned}$$

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:

$$a_c(t) = ||\vec{a}_{\perp}(t)|| = \omega^2(t)R = \frac{v^2(t)}{R} \quad (4.12)$$

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

4.3.1 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 $\theta = 2\pi$:

$$\omega = \frac{\Delta\theta}{T} = \frac{2\pi}{T}$$

$$\boxed{\therefore T = \frac{2\pi}{\omega}} \quad (4.13)$$

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 :

$$v = \frac{2\pi R}{T}$$

$$T = \frac{2\pi R}{v} = \frac{2\pi R}{\omega R} = \frac{2\pi}{\omega}$$

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.

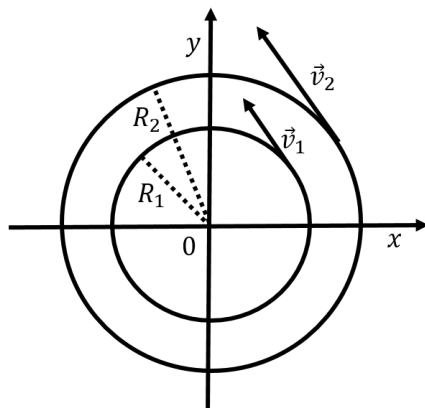


Figure 4.8: For a given angular velocity, the linear velocity will be larger on a larger circle ($v = \omega R$).

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 4.8, 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.

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

4.4 Summary

- Something interesting

A

Calculus

This appendix gives a very brief introduction to calculus with a focus on the tools needed in physics.

Learning Objectives:

- Understand that derivatives measure the rate of change
- Understand partial derivatives and gradients
- Understand anti-derivatives and that integrals are sums

A.1 Functions of real numbers

In calculus, we work with functions and their properties, rather than variables. 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 a single 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 that number. 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:

Python Code A.1: Plotting a function of 1 variable

```

#import packages 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()

```

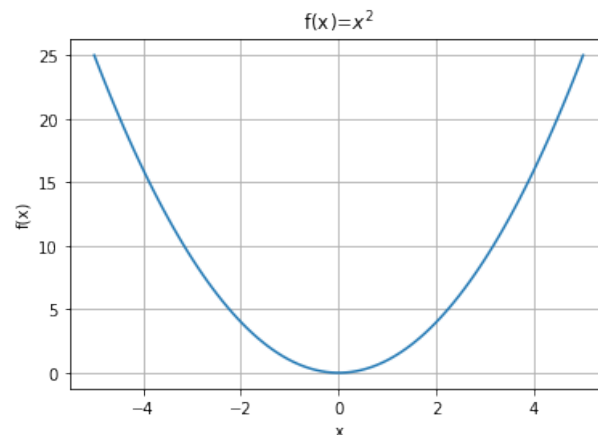
Output A.1:

Figure A.1: $f(x) = x^2$ plotted between $x = -5$ and $x = +5$.

Plotting a function of 2 variables is a little trickier, since we need to do it in three dimensions (1 axis for x , 1 axis for y , and one axis for $g(x, y)$). This can be done in python with a little more work:

Python Code A.2: Plotting a function of 2 variables

```

#import packages for creating arrays of values and for plotting
import numpy as np #arrays
import pylab as pl #plotting
#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 A.2:

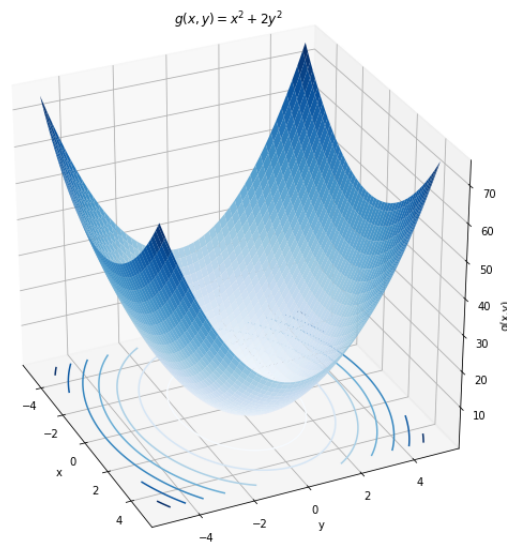


Figure A.2: $g(x,y) = x^2 + 2y^2$ plotted for x between -5 and $+5$ and for y between -5 and $+5$.

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 A.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?

A.2 Derivatives

Consider the function $f(x) = x^2$ that is plotted in Figure ???. 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 $f(x) = ax^2 + c$ 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 function value 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 (its magnitude) tells how steeply the function is changing.

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 $\Delta x \rightarrow 0$, we get the derivative. In fact, this is the formal definition of the derivative:

$$\boxed{\frac{df}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}} \quad (\text{A.1})$$

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 $\Delta x \rightarrow 0$, and df is Δf , in the same limit of $\Delta x \rightarrow 0$.

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 $\Delta x \rightarrow 0$:

$$\begin{aligned}\frac{df}{dx} &= \lim_{\Delta x \rightarrow 0} \frac{\Delta f}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} (2x + \Delta x) \\ &= 2x\end{aligned}$$

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 A.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.

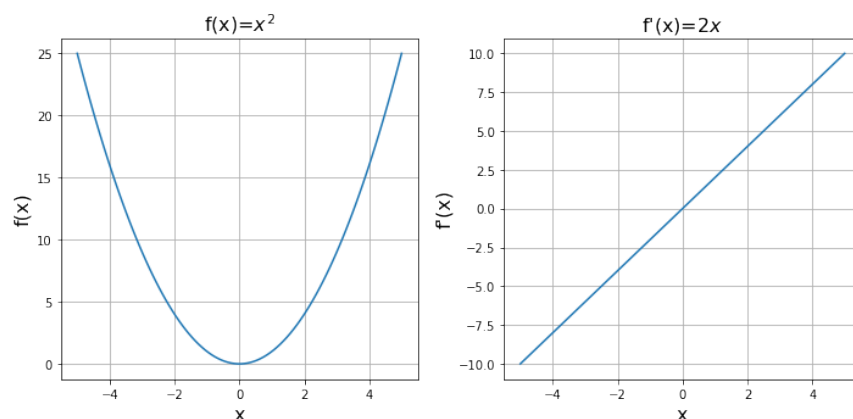


Figure A.3: $f(x) = x^2$ and its derivative, $f'(x) = 2x$ plotted for x between -5 and $+5$.

A.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 A.1. Table A.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 A.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 of $h(x)$, $h'(x)$. These are summarized in Table A.2 below.

Function, $h(x)$	Derivative, $h'(x)$
$h(x) = f(x) + g(x)$	$h'(x) = f'(x) + g'(x)$
$h(x) = f(x) - g(x)$	$h'(x) = f'(x) - g'(x)$
$h(x) = f(x)g(x)$	$h'(x) = f'(x)g(x) + f(x)g'(x)$
$h(x) = \frac{f(x)}{g(x)}$	$h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$
$h(x) = f(g(x))$	$h'(x) = f'(g(x))g'(x)$ (The Chain Rule)

Table A.2: Derivatives of combined functions.

Example A-1: Use the properties from Table A.2 to show that the derivative of $\tan(x)$ is $\frac{1}{\cos^2(x)}$

Since $\tan(x) = \frac{\sin(x)}{\cos(x)}$, we can write:

$$\begin{aligned}
 h(x) &= \frac{f(x)}{g(x)} \\
 f(x) &= \sin(x) \\
 g(x) &= \cos(x)
 \end{aligned}$$

Using the fourth row in Table A.2, and the common derivatives from Table A.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 A-2: Use the properties from Table A.2 to calculate the derivative of $h(x) = \sin^2(x)$

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 A.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 $2g(x) = 2\cos(x)$.

A.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 of a particular variable. This is illustrated in Figure A.4 for the function $f(x, y) = x^2 - 2y^2$, which looks somewhat like a saddle.

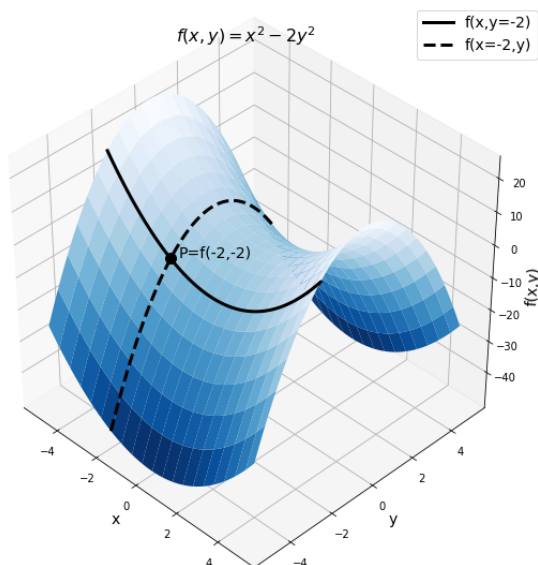


Figure A.4: $f(x, y) = x^2 - 2y^2$ plotted for x between -5 and $+5$ and for y between -5 and $+5$. The point P labelled on the figure shows the value of the function at $f(-2, -2)$. The two lines show the function evaluated when one of x or y is held constant.

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 A.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 :

$$\frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x, y) \text{ (Partial derivative of } f \text{ with respect to } x)$$

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 differentiating¹. For the function $f(x, y) = x^2 - 2y^2$, we have:

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{\partial}{\partial x} (x^2 - 2y^2) = 2x \\ \frac{\partial f}{\partial y} &= \frac{\partial}{\partial y} (x^2 - 2y^2) = -4y \end{aligned}$$

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 of that variable and whether it is increasing or decreasing.

Example A-3: Determine the partial derivatives of $f(x, y, z) = ax^2 + byz - \sin(z)$.

In this case, we have three partial derivatives to evaluate. Note that a and b are constants and can be thought of as numbers that we do not know.

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{\partial}{\partial x} (ax^2 + byz - \sin(z)) = 2ax \\ \frac{\partial f}{\partial y} &= \frac{\partial}{\partial y} (ax^2 + byz - \sin(z)) = bz \\ \frac{\partial f}{\partial z} &= \frac{\partial}{\partial z} (ax^2 + byz - \sin(z)) = by - \cos(z) \end{aligned}$$

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 A.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

¹To take the derivative is to “differentiate”!

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} , but this is just a matter of notation. 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 A-4: Determine the gradient of the function $f(x, y) = x^2 - 2y^2$ at the point $x = -2$ and $y = -2$.

We have already found the partial derivatives that we need to evaluate at $x = -2$ and $y = -2$:

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2x \\ \frac{\partial f}{\partial y} &= -4y \\ \therefore \nabla f(x, y) &= \frac{\partial f}{\partial x} \hat{x} + \frac{\partial f}{\partial y} \hat{y} \\ &= 2x\hat{x} - 4y\hat{y} \end{aligned}$$

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 A.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.

A.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 :

$$\begin{aligned} \int f(x)dx &= F(x) \\ \therefore \frac{d}{dx} \left(\int f(x)dx \right) &= \frac{dF}{dx} = f(x) \end{aligned}$$

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 $\Delta x \rightarrow 0$. 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 $\Delta x \rightarrow 0$. 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 $\tilde{F}(x)$ is given by:

$$\begin{aligned} \frac{d\tilde{F}}{dx} &= \frac{d}{dx} (F(x) + C) \\ &= \frac{dF}{dx} + \frac{dC}{dx} \\ &= \frac{dF}{dx} + 0 \\ &= f(x) \end{aligned}$$

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 A.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 A.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 $f(x) = \frac{dF}{dx}$
2. the condition that $F(x)$ must pass through a specific point, $F(x_0) = F_0$.

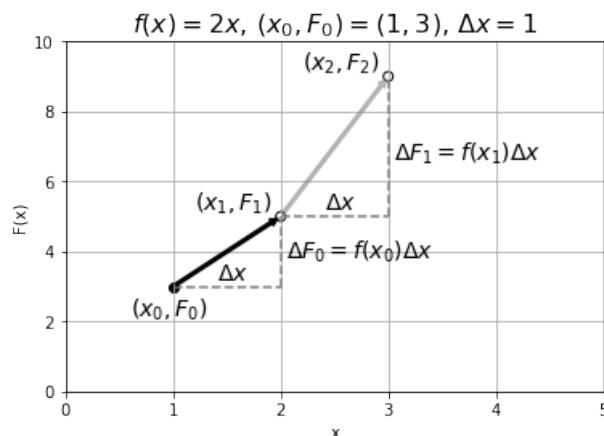


Figure A.5: Determining the anti-derivative, $F(x)$, given the function $f(x) = 2x$ and the initial condition that $F(x)$ passes through the point $(x_0, F_0) = (1, 3)$.

The procedure for determining the anti-derivative $F(x)$ is illustrated above in Figure A.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:

$$\begin{aligned} \frac{\Delta F_0}{\Delta x} &\approx f(x_0) \quad (\text{in the limit } \Delta x \rightarrow 0) \\ \therefore \Delta F_0 &= f(x_0)\Delta x \end{aligned}$$

We can then estimate the value of the function $F(x)$ at the next point, $x_1 = x_0 + \Delta x$, as illustrated by the black arrow in Figure A.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 A.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 $\Delta x \rightarrow 0$:

$$F(x_N) = F(x_0) + \lim_{\Delta x \rightarrow 0} \sum_{i=1}^{i=N} f(x_{i-1})\Delta x \quad (\text{A.2})$$

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 A.6 for the same case as in Figure A.5 (that is, Figure A.6 shows $f(x)$ that we know, and Figure A.5 shows $F(x)$ that we are trying to find).

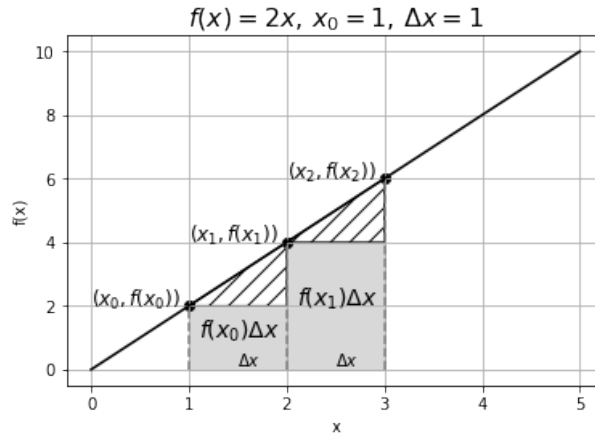


Figure A.6: The function $f(x) = 2x$ and illustration of the terms $f(x_0)\Delta x$ and $f(x_1)\Delta x$ as the area between the curve $f(x)$ and the x axis when $\Delta x \rightarrow 0$.

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 $\Delta x \rightarrow 0$, the missing pieces (shown by the hashed areas in Figure A.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 :

$$\lim_{\Delta x \rightarrow 0} \sum_{i=1}^{i=N} f(x_{i-1})\Delta x = \text{Area between } f(x) \text{ and } x \text{ axis from } x_0 \text{ to } x_N$$

Re-arranging equation A.2 gives us a prescription for determining the anti-derivative:

$$F(x_N) - F(x_0) = \lim_{\Delta x \rightarrow 0} \sum_{i=1}^{i=N} f(x_{i-1}) \Delta x$$

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:

$$\boxed{\int_{x_0}^{x_N} f(x) dx = F(x_N) - F(x_0) = \lim_{\Delta x \rightarrow 0} \sum_{i=1}^{i=N} f(x_{i-1}) \Delta x} \quad (\text{A.3})$$

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:

$$\int_{x_0}^{x_N} f(x) dx = F(x_N) - F(x_0) = \left[F(x) \right]_{x_0}^{x_N}$$

Recall that we wrote the anti-derivative with the same f 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 A.5 and A.6. Using equation A.3, we have:

$$\begin{aligned} \int_{x_0}^{x_N} f(x) dx &= \lim_{\Delta x \rightarrow 0} \sum_{i=1}^{i=N} f(x_{i-1}) \Delta x \\ &= \lim_{\Delta x \rightarrow 0} \sum_{i=1}^{i=N} 2x_{i-1} \Delta x \end{aligned}$$

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 $\Delta x \rightarrow 0$ is the same as taking the limit $N \rightarrow \infty$. Let us illustrate the sum for the case where $N = 3$, and thus when $\Delta x = 1$, corresponding to the illustration in Figure A.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 A.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 $N \rightarrow \infty$, 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 $N \rightarrow \infty$,

then the value $x_{N-1} \approx x_N$. The average of the values between x_0 and x_N is simply given by:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{i=N} x_{i-1} = \frac{1}{2}(x_N - x_0)$$

Putting everything together:

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{i=1}^{i=N} 2x_{i-1} \Delta x &= 2(x_N - x_0) \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^{i=N} x_{i-1} \\ &= 2(x_N - x_0) \frac{1}{2}(x_N - x_0) \\ &= x_N^2 - x_0^2 \\ &= (4)^2 - (1)^2 = 15 \end{aligned}$$

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 $f(x) = 2x$.

A.3.1 Common anti-derivative and properties

Table [A.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, $F(x)$
$f(x) = a$	$F(x) = ax + C$
$f(x) = x^n$	$F(x) = \frac{1}{n+1}x^{n+1} + C$
$f(x) = \frac{1}{x}$	$F(x) = \ln(x) + C$
$f(x) = \sin(x)$	$F(x) = -\cos(x) + C$
$f(x) = \cos(x)$	$F(x) = \sin(x) + C$
$f(x) = \tan(x)$	$F(x) = -\ln(\cos(x)) + C$
$f(x) = e^x$	$F(x) = e^x + C$
$f(x) = \ln(x)$	$F(x) = x \ln(x) - x + C$

Table A.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 the derivative. A few common properties to help evaluate indefinite integrals are shown in Table A.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 A.4: Some properties of indefinite integrals.

A.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 that are each different. In physics, it is often the case that we need to take the sum of an infinite number of small things that keep varying.

Consider, for example, a rod of length, L , and total mass M , as shown in Figure ???. 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:

$$\mu = \frac{M}{L}$$

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 A.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, $\mu(x)$, where x is the distance measured from one side of the rod.

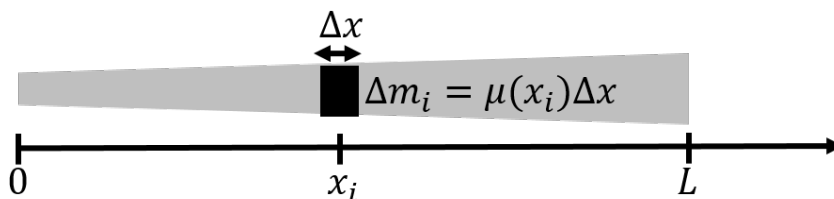


Figure A.7: A rod with a varying linear density. To calculate the mass of the rod, we consider a small mass element Δm_i of length Δx at position x_i . The total mass of the rod is found by summing the mass elements.

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 for that tiny piece of rod. Then, the tiny piece of rod, has a mass 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 $\Delta x \rightarrow 0$:

$$\begin{aligned} M &= \lim_{\Delta x \rightarrow 0} \sum_{i=1}^{i=N} \Delta m_i \\ &= \lim_{\Delta x \rightarrow 0} \sum_{i=1}^{i=N} \mu(x_i) \Delta x \end{aligned}$$

But this is precisely the definition of the integral (equation A.2), which we can easily evaluate with an anti-derivative:

$$\begin{aligned} M &= \lim_{\Delta x \rightarrow 0} \sum_{i=1}^{i=N} \mu(x_i) \Delta x \\ &= \int_0^L \mu(x) dx \\ &= G(L) - G(0) \end{aligned}$$

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 A.3):

$$G(x) = a \frac{1}{4} x^4 + C$$

Let $a = 5 \text{ kg/m}^4$ and let's say that the length of the rod is $L = 0.5 \text{ m}$. The total mass of the rod is then:

$$\begin{aligned} M &= \int_0^L \mu(x) dx \\ &= \int_0^L ax^3 dx \\ &= G(L) - G(0) \\ &= \left[a \frac{1}{4} L^4 \right] - \left[a \frac{1}{4} 0^4 \right] \\ &= 5 \text{ kg/m}^4 \frac{1}{4} (0.5 \text{ m})^4 \\ &= 78 \text{ g} \end{aligned}$$

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 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

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)

B.1 Coordinate systems

Coordinate systems are used in order to be able 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.

B.1.1 1D Coordinate systems

The easiest coordinate system to construct is one where we need 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. 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_T . If the train is at position $x_T = 0$, 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_T and that the West side of the track correspond to the negative values of x_T . Thus, in order to fully specify a 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

TODO: make figure to illustrate 1D x-axis

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.

B.1.2 2D Coordinate systems

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 B.1 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.

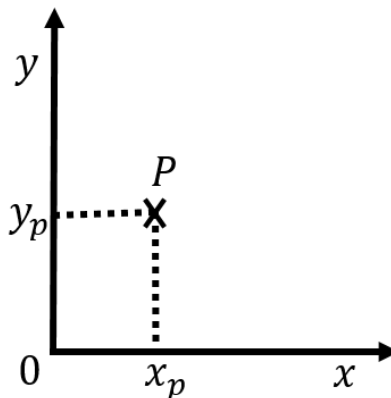


Figure B.1: Example of Cartesian coordinate system and a point P with coordinates (x_p, y_p) .

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 B.1 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 B-1: Figure B.2 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 ?

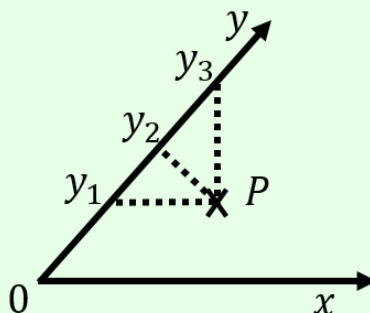


Figure B.2: A non-orthogonal coordinate system (the x and y axes are not perpendicular).

- A) y_1
- B) y_2
- C) y_3

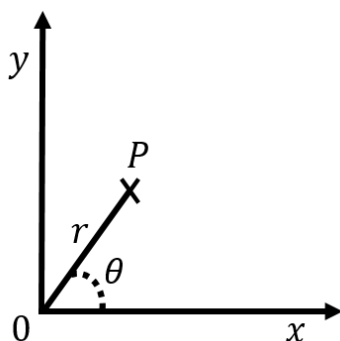


Figure B.3: Example of a polar coordinate system and a point P with coordinates (r, θ) .

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 ???. 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 ???. 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.

One can easily convert between the two Cartesian coordinates, x, y , and the two correspond-

ing polar coordinates, r, 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).

B.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 specified using the three coordinates, x , y , and z .

Two additional coordinate systems are common in three dimensions: “cylindrical” and “spherical coordinates”. All three systems are illustrated in Figure B.4 superimposed onto the Cartesian system.

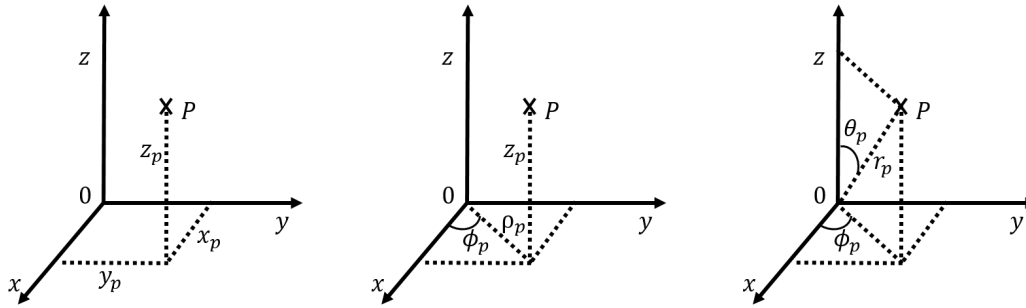


Figure B.4: Cartesian (left), cylindrical (centre) and spherical (right) coordinate systems used in three dimensions. The y and z axes are in the plane of the page, whereas the x axis comes out of the page.

By convention, we use the z axis to be the vertical direction in three dimensions. In cylindrical 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 that the projected point makes with the x axis and ρ is the distance of that projected point to the origin. 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 motion along the surface

of a cylinder. 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. Spherical coordinates are useful for describing situations that have spherical symmetry, such as a person walking on the surface of the Earth. The spherical coordinates are related to the Cartesian coordinates by:

$$\begin{aligned}r &= \sqrt{x^2 + y^2 + z^2} \\ \tan(\theta) &= \frac{z}{r} = \frac{z}{\sqrt{x^2 + y^2 + z^2}} \\ \tan(\phi) &= \frac{y}{x}\end{aligned}$$

B.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 B.5 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).

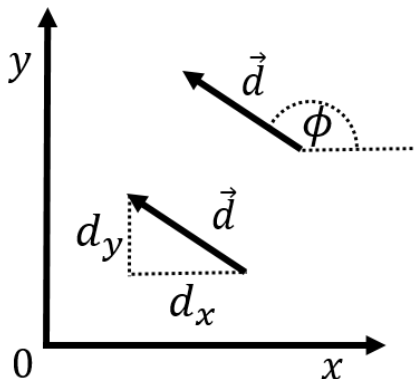


Figure B.5: A vector \vec{d} shown twice, once with its Cartesian components (d_x, d_y) and once with its magnitude and direction (d, ϕ) .

In analogy with coordinate systems, we have multiple ways to choose the numbers that we use to define 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 B.5, 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 B.5. In terms of the 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 indicate its magnitude.

B.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 B.5, 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):

$$\begin{aligned}(\hat{d})_x &= \frac{d_x}{d} = \frac{d_x}{\sqrt{d_x^2 + d_y^2}} \\(\hat{d})_y &= \frac{d_y}{d} = \frac{d_y}{\sqrt{d_x^2 + d_y^2}} \\ \therefore \|\hat{d}\| &= \sqrt{(\hat{d})_x^2 + (\hat{d})_y^2} = \sqrt{\frac{d_x^2}{d_x^2 + d_y^2} + \frac{d_y^2}{d_x^2 + d_y^2}} = 1\end{aligned}$$

A specific type of unit vectors are those unit vectors that are parallel to the axes of the coordinate system. Those vectors are denoted \hat{x} , \hat{y} , \hat{z} (and sometimes \hat{i} , \hat{j} , \hat{k} or \hat{e}_x , \hat{e}_y , \hat{e}_z) for the x , y , and z axes, respectively.

B.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 :

$$\begin{aligned}\vec{d} &= (d_x, d_y) && \text{row vector} \\ &= \begin{pmatrix} d_x \\ d_y \end{pmatrix} && \text{column vector} \\ &= d_x \hat{x} + d_y \hat{y} && \text{using } \hat{x}, \hat{y} \\ &= d_x \hat{i} + d_y \hat{j} && \text{using } \hat{i}, \hat{j}\end{aligned}$$

For example, the unit vector \hat{y} can be written down as $(0,1,0)$ in three dimensions.

Checkpoint B-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 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 B.6.

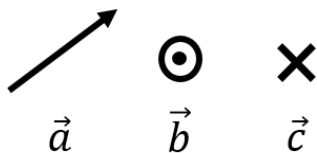


Figure B.6: Geometric representation of three vectors. The vector \vec{a} lies in the plane of the page, the vector \vec{b} is pointing out of the page, and the vector \vec{c} is pointing into the page.

B.3 Vector algebra

In this section, we describe the various algebraic operations that can be performed using vectors.

B.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 :

$$\vec{w} = a\vec{v} = (av_x, av_y)$$

Similarly, the division of a vector by a scalar is defined analogously:

$$\vec{w} = \frac{\vec{v}}{a} = \left(\frac{v_x}{a}, \frac{v_y}{a} \right)$$

Checkpoint B-3: What happens to the length of a vector if the vector is multiplied by 2?

- 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 magnitude of \vec{v} .

B.3.2 Addition/subtraction of two vectors

The addition (subtraction) of two vectors, \vec{a} and \vec{b} , is found by adding (subtracting) the components of the two vectors. For example, if $\vec{c} = \vec{a} + \vec{b}$, the components of \vec{c} are given by:

$$\begin{aligned}\vec{c} = \vec{a} + \vec{b} &= \begin{pmatrix} a_x \\ a_y \end{pmatrix} + \begin{pmatrix} b_x \\ b_y \end{pmatrix} \\ &= \begin{pmatrix} a_x + b_x \\ a_y + b_y \end{pmatrix}\end{aligned}$$

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.

Example B-1: Given two vectors, $\vec{a} = 2\hat{x} + 3\hat{y}$, and $\vec{b} = 5\hat{x} - 2\hat{y}$, calculate the vector $\vec{c} = 2\vec{a} - 3\vec{b}$.

This can easily be solved algebraically:

$$\begin{aligned}\vec{c} &= 2\vec{a} - 3\vec{b} \\ &= 2(2\hat{x} + 3\hat{y}) - 3(5\hat{x} - 2\hat{y}) \\ &= (4\hat{x} + 6\hat{y}) - (15\hat{x} - 6\hat{y}) \\ &= (4 - 15)\hat{x} + (6 + 6)\hat{y} \\ &= -11\hat{x} + 12\hat{y}\end{aligned}$$

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 B.7 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 goes from the tail of vector \vec{a} to the head of vector \vec{b} .

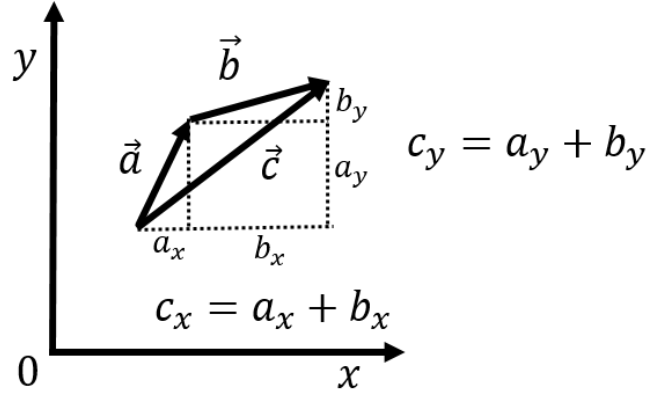


Figure B.7: Geometric addition of the vectors \vec{a} and \vec{b} by placing them “head to tail”.

Subtracting two vectors geometrically is done in the same way as addition. For example, the vector *vecc*, 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 B.8.

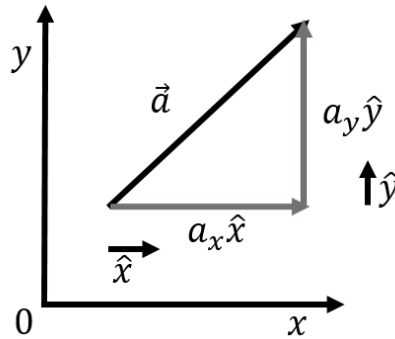


Figure B.8: Illustration that the notation $\vec{a} = a_x\hat{x} + a_y\hat{y}$ is in fact the vector addition of $a_x\hat{x}$ and $a_y\hat{y}$.

B.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. One can easily show that the scalar product is also related to the angle between the two vectors when these are placed “tail to tail”, as in Figure B.9

$$\vec{a} \cdot \vec{b} = ab \cos \theta$$

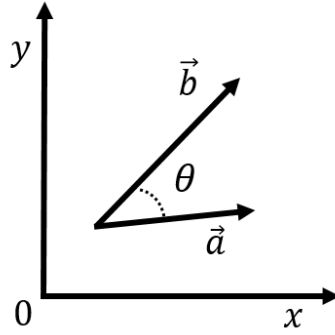


Figure B.9: Illustration of the angle between vectors \vec{a} and \vec{b} when these are placed tail to tail.

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.

B.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 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 thus perpendicular to the plane (thus 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:

$$\vec{a} \times \vec{b} = \begin{pmatrix} a_y b_z - a_z b_y \\ a_z b_x - a_x b_z \\ a_x b_y - a_y b_x \end{pmatrix} \quad (\text{B.1})$$

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:

$$||\vec{a} \times \vec{b}|| = ab \sin \theta \quad (\text{B.2})$$

where θ is the angle between the vectors \vec{a} and \vec{b} when these are placed tail to tail (Figure B.9). 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 thus useful to determine quantities that are maximal when two vectors are perpendicular (the opposite use case from the scalar product).

Geometrically, one can determine the direction of the cross product of two vectors by using the “right hand rule”. This is done by using your right hand, aligning your thumb with the first vector, your index with the second vector, and 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 TODO. Thus, you can often avoid using equation B.1 and instead use the right hand rule and equation B.2 to find the vector resulting from a cross product.

TODO: make a figure for the right hand rule

The unit vectors that define a coordinate system have the following properties relative to the cross product:

$$\begin{aligned}\vec{x} \times \vec{y} &= \vec{z} \\ \vec{y} \times \vec{z} &= \vec{x} \\ \vec{z} \times \vec{x} &= \vec{y}\end{aligned}$$

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”).

B.4 Example uses of vectors in physics

This section gives a quick overview of some applications of vectors in physics.

B.4.1 Kinematics and vector equations

Kinematics is the description of the position and motion of an object (TODO: Chapter reference). 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, that 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 in general if the object is moving relative to the origin of the coordinate system. 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 a second position, $\vec{r}_2 = (x_2, y_2, z_2)$. We can define the **displacement vector**, \vec{d} :

$$\vec{d} = \vec{r}_2 - \vec{r}_1 = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{pmatrix} = \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}$$

where the components of the displacement vector, Δx , Δy , and Δz correspond to the displacements along the x , y , and z axes, respectively. This is illustrated for the two dimensional case in Figure B.10.

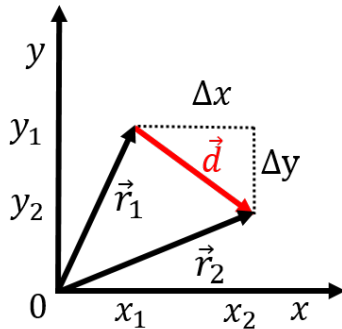


Figure B.10: Illustration of a displacement vector, $\vec{d} = \vec{r}_2 - \vec{r}_1$, for an object that was located at position \vec{r}_1 at time t_1 and at position \vec{r}_2 at time t_2 .

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 that elapsed, $\Delta t = t_2 - t_1$:

$$\vec{v} = \frac{\vec{d}}{\Delta t} = \begin{pmatrix} \frac{\Delta x}{\Delta t} \\ \frac{\Delta y}{\Delta t} \\ \frac{\Delta z}{\Delta t} \end{pmatrix}$$

where 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:

$$\begin{aligned} v_x &= \frac{\Delta x}{\Delta t} \\ v_y &= \frac{\Delta y}{\Delta t} \\ v_z &= \frac{\Delta z}{\Delta t} \end{aligned}$$

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:

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

B.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. 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 (TODO: chapter reference) 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 B.11, 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}).

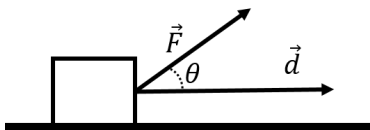


Figure B.11: Example of a force \vec{F} being applied on an object as it moves along the displacement vector \vec{d} .

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. 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 B.11). 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 to calculate work.

B.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 around that axis in which the object is rotating (e.g. clockwise or counter-clockwise)
3. How fast the object is rotating

We can also use a vector to describe this type 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 (it is co-linear with the axis of rotation, but the specific choice of direction has not been made). We choose the direction of the vector by using our right hand in such a way that the vector points in the direction of your thumb when curling your fingers corresponds to the rotational direction, as illustrated in Figure TODO: Figure for rotational right hand rule.

B.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 B.12 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, as 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 vectors when placed tail to tail, as in the right side of Figure B.12. 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.

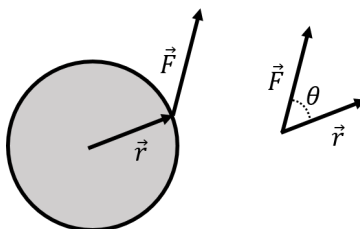


Figure B.12: A force, \vec{F} , is exerted in the plane of a disk at a position given by the vector \vec{r} relative to the centre of the disk.

Checkpoint B-4: Referring to Figure B.12, 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

C

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!

C.1 A quick intro to programming

In Python, as in other programming language, 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 C.1: Declaring variables in Python

```
#This is a comment, and is ignored by Python
a = 2
b = 2*a
print(b)
```

Output C.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 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 C.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 C.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()**).

C.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 C.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]
```

C.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 C.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 C.4:

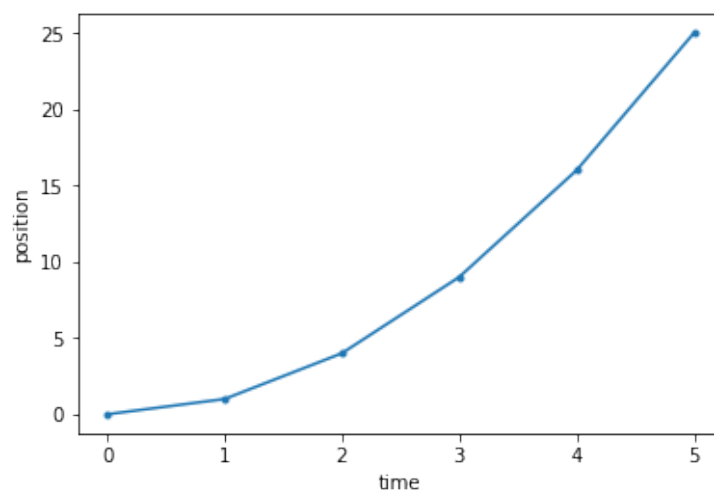


Figure C.1: Plotting two arrays.

Checkpoint C-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 -10 and 25 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 C.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 =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 = np.cos(xvals**2)

#make the plot showing only a line , and color it
pl.plot(xvals , fvals , color='red')
#add some labels:
pl.xlabel("time") #label for x-axis
pl.ylabel("position") #label for y-axis
#show the plot
pl.show()
```

Output C.5:

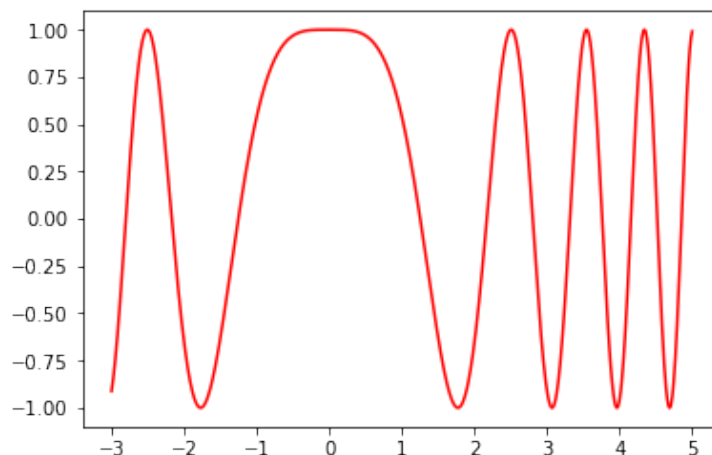


Figure C.2: Plotting a function using arrays.

C.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.

C.4.1 Propagating uncertainties

In Chapter ??, 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 2-5, we propagated the uncertainties $x = (3.00 \pm 0.01)$ m and $t = (0.76 \pm 0.15)$ s to the quantity $k = \frac{t}{\sqrt{x}}$. We show below how easily this can be done with QExpy:

Python Code C.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 C.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 ??), 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 C.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 C.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 ??.

C.4.2 Plotting experimental data with uncertainties

In Chapter ?? we had presented the data in Table C.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} [m ^½]	k [s m ^{-½}]
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 C.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 0.15s and that the measurements of the distance have no (or negligible) uncertainties:

Python Code C.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'

#Then we enter the data in arrays for the x and y axes.
#The values for the square root of height (x axis):
sqx = [1. , 1.41, 1.73, 2. , 2.24]
#and then, the corresponding times (y-axis):
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 C.8:

```
-----Fit results-----
Fit of My data to linear
Fit parameters:
My data_linear_fit0_fitpars.intercept = -0.2 +/- 0.2,
My data_linear_fit0_fitpars.slope = 0.6 +/- 0.1

Correlation matrix:
[[ 1.      -0.968]
 [-0.968   1.    ]]

chi2/ndof = 2.04/2
-----End fit results-----
```

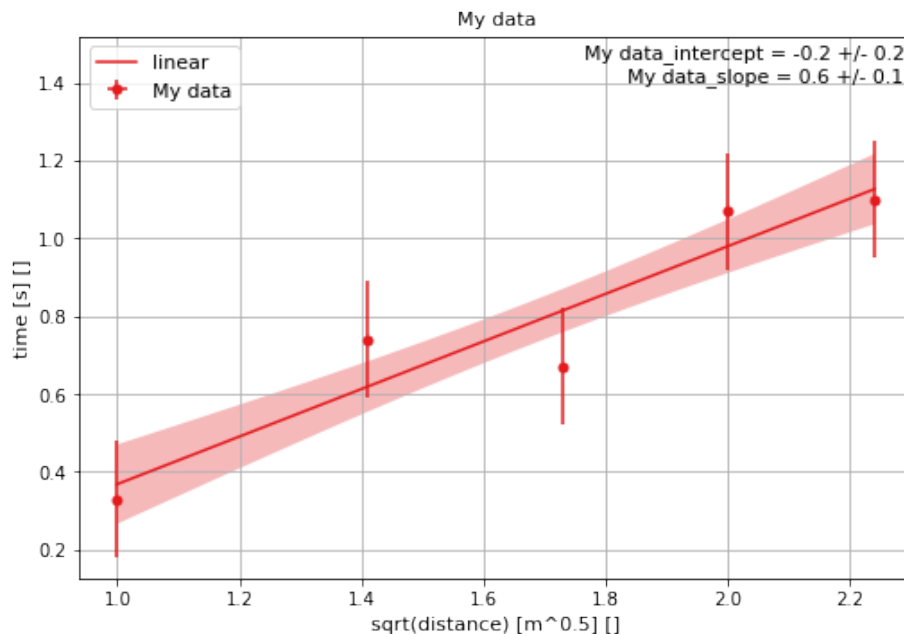


Figure C.3: QExpy plot of t versus \sqrt{x} and line of best fit.

The plot in Figure C.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 $\text{chi2}/\text{ndof} = 2.04/2$; 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 $(0.61 \pm 0.13) \text{ s m}^{-\frac{1}{2}}$. **Performing a linear fit of the data is the best way to determine a constant of proportionality between the measurements.** Finally, we expect the intercept to be equal to zero according to our model. The best fit line from QExpy has an intercept of $(-0.24 \pm 0.22) \text{ s}$, which is slightly below, but consistent, with zero. From these data, we would conclude that our measurements are consistent with Chloé’s Theory. Again, remember that we can never confirm a theory, we can only exclude it; in this case, we cannot exclude Chloé’s Theory.

C.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.

C.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 C.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 C.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. In particular, Python would not understand `2x` which needs to explicitly have the multiplication operator, `2*x` (inside of the cosine function).

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 slight 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 C.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 C.10:

```
10.710170286650365
[ 0.1671865   2.67968971 10.71017029]
```

where we created the array `xvals` using the `numpy` module.

C.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 C.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 C.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 value 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:

$$\int_{x_a}^{x_b} f(x)dx = \lim_{\Delta x \rightarrow 0} \sum_{i=0}^{i=N-1} f(x_i)\Delta x$$

$$\Delta x = \frac{x_b - x_a}{N}$$

$$x_i = x_a + i\Delta x$$

The limit of $\Delta x \rightarrow 0$ is thus equivalent to the limit $N \rightarrow \infty$. Our strategy for evaluating the integral is thus:

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 us thus use Python to evaluate the integral of the function $f(x) = 4x^3 + 3x^2 + 5$ between $x = 1$ and $x = 5$:

Python Code C.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
```

```
#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 C.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 C-2: What value of N should you use above in order to get within 0.01 of the exact analytic answer?