The Art of Modelling: Introduction to Physics

Ryan D. Martin

October 6, 2017

Contents

1

The Scientific Method and Physics

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

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. 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. One could generate a giant table showing how long it takes a tennis ball to drop for 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 $2\,\mathrm{m}$ than it would to drop by $1\,\mathrm{m}$:

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:

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

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

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

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

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

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

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 $2\,\mathrm{m}$ than it will from $1\,\mathrm{m}$, 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.

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 $4.2\,\mathrm{m}$.

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.

> **Checkpoint 1-1:** 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 one event preceded another. 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 Shrödinger's cat, which can be modelled as being both alive and dead at the same time. However, just because we model it that way does not mean that it really is alive and dead at the same time.

1.4   The scope of Physics

Physics describes a wide range of phenomena within the physical sciences, ranging from behaviour of the microscopic particles that make up matter to the evolution of the entire Universe. We often speak of 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.

1.4.1   Classical Physics

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

Mechanics

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

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

In 1859, Urbain Le Verrier also noted that Mercury's orbit around the Sun is different than that predicted by Newton's theory. Again, a new planet was proposed, "Vulcan", but that planet was never discovered and the deviation of Mercury's orbit from Newton's prediction remained unexplained until 1915, when Albert Einstein introduced a new, more complete, theory of gravity, called "General Relativity". This is a good example of the scientific method; although the discovery of Neptune was consistent with Newton's theory, it did not prove that the theory is correct, only that it correctly described the motion of Uranus. The

discrepancy that arose when looking at Mercury ultimately showed that Newtons' theory of gravity fails to provided a proper description of planetary orbits in the proximity of very massive objects (Mercury is the closest planet to the Sun).

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.

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.

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

Particle astrophysics

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

1.5   Thinking like a physicist

In a sense, physics can be thought of as the most fundamental of the sciences, as it describes the interactions of the smallest constituents of matter. In principle, if one can precisely describe how protons, neutrons, and electrons interact, then one can completely describe how a human brain thinks. In practice, the theories of particle physics lead to equations that are too difficult to solve for system 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

---

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

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

1.6 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 and 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.

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 $20\,\text{m}\times2\,\text{m}\times2\,\text{m}$*, with a volume of* $80\,\text{m}^3\sim100\,\text{m}^3$*. We can model a ping pong ball as a sphere with a diameter of* $0.03\,\text{m}$ *(*$3\,\text{cm}$*). When stacking the ping pong balls, we can model them as little cubes with a side given by their diameter, so the volume of a ping pong ball, for stacking, is* $\sim 0.000\,03\,\text{m}^3 = 3 \times 10^{-5}\,\text{m}^3$*. If we divide* $100\,\text{m}^3$ *by* $3 \times 10^{-5}\,\text{m}^3$*, using scientific notation:*

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

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

Checkpoint 2-1: Fill in the following table giving the order of magnitude in meters of the size 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 jiffys[1] are units for the dimension of time.

When we compare two numbers, for example a prediction from a model and a measurement, it is important that both quantities have the same dimension *and* be expressed in the same 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 2.1 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 faster the object, the larger the distance covered per unit time). Table 2.2 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

---

[1]A jiffy is a unit used in electronics and generally corresponds to either 1/50 or 1/60 seconds.
[2]Note that we can also write meters per second as m·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 $[M \cdot L \cdot T^{-2}]$	newton [N]	$[kg \cdot m \cdot s^{-2}]$
Energy $[M \cdot L^2 \cdot T^{-2}]$	joule [J]	$[N \cdot m = kg \cdot m^2 \cdot s^{-2}]$
Power $[M \cdot L^2 \cdot T^{-3}]$	watt [W]	$[J/s = kg \cdot m^2 \cdot s^{-3}]$
Electric Charge $[I \cdot T]$	coulomb [C]	$[A \cdot s]$
Voltage $[M \cdot L^2 \cdot T^{-3} \cdot I^{-1}]$	volt [V]	$[J/C = kg \cdot m^2 \cdot s^{-3} \cdot A^{-1}]$

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.

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

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

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

and the SI units of force are thus:

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

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

Example 2-3: Use Table 2.2 to show that voltage has the same dimension as force multiplied by speed and divided by electric current.

According to Table 2.2, voltage has dimensions:

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

while force, speed and current have dimensions:

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

The dimension of force multiplied by speed divided by electric charge

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

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

When you build a model to 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 $100\,\mathrm{m}$ 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) $\mathrm{T\,L^{\frac{1}{2}}}$
    B) $\mathrm{T\,L^{-\frac{1}{2}}}$
    C) $\mathrm{s\,m^{\frac{1}{2}}}$
    D) $\mathrm{s\,m^{-\frac{1}{2}}}$

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 2.3, where the time, t, was measured for a bowling ball to fall distances of x between $1\,\mathrm{m}$ and $5\,\mathrm{m}$. The table also shows the values computed for $\sqrt{x}$ and the corresponding value of $k = \frac{t}{\sqrt{x}}$:

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

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

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

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

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

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

2.3.1   Measurement uncertainties

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

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

where $0.76\,\mathrm{s}$ is called the "central value" and $0.15\,\mathrm{s}$ the "uncertainty" or the "error"[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 $0.61\,\mathrm{s}$ and $0.91\,\mathrm{s}$ (given by $0.76\,\mathrm{s}$ - $0.15\,\mathrm{s}$ and $0.76\,\mathrm{s}$ + $0.15\,\mathrm{s}$). When we say that we are "pretty sure" that the value is within the quoted range, we usually mean that there is a 68% chance of this being true and allow for the possibility that the true value is actually outside the range that we quoted. The value of 68% comes from statistics and the normal distribution 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 out how to assign a central value and an uncertainty. For example, how did we come up with $t = (0.76 \pm 0.15)\,\mathrm{s}$ from the values in Table 2.4?

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

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

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

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

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

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

---

[3]We 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 2.4.

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 value[4]. 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 2.5 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 2.5, since $2.8 \, \text{cm}$ is the closet value on the ruler

---

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

that matches the length of the object and $0.5\,\mathrm{mm}$ is half of the smallest division on the ruler. Using half of the smallest division of the ruler means that our uncertainty range covers one full division. Note that it is usually better to reproduce a measurement to evaluate the uncertainty instead of using half of the smallest division, although half of the smallest division should be the lower limit on the uncertainty. That is, by repeating the measurements and obtaining the standard deviation, you should see if the uncertainty is *larger* than half of the of the smallest division, not smaller.
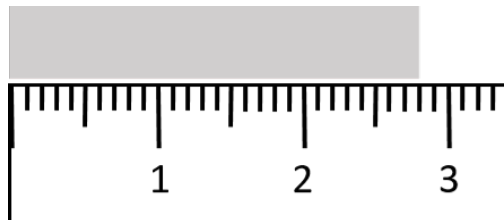

Figure 2.1:   The length of the grey rectangle would be quoted as $L = (2.8 \pm 0.5)\,\mathrm{cm}$ using the rule of "half the smallest division".

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

Propagating uncertainties

Going back to the data in Table 2.4, we found that for a known drop height of $x = 3\,\mathrm{m}$, we measured different values of the drop time, which we found to be $t = (0.76 \pm 0.15)\,\mathrm{s}$ (using the mean and standard deviation). We also calculated a value of k corresponding to each value of t, and found $k = (0.44 \pm 0.09)\,\mathrm{s\,m}^{-\frac{1}{2}}$. Suppose that we did not have access to the individual values of t, but only to the value of $t = (0.76 \pm 0.15)\,\mathrm{s}$ 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

valued derived from the measurements. We now look at different methods for propagating uncertainties.

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:

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

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

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

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

2. The Min-Max method
A pedagogical way to determine k and its uncertainty is to use the "Min-Max method". Since $k = \frac{t}{\sqrt{x}}$, 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, $t^{max} = 0.91\,\text{s}$:

$$k^{min} = \frac{t^{min}}{\sqrt{x}} = \frac{0.61\,s}{\sqrt{(3\,m)}} = 0.35\,\text{s}\,\text{m}^{-\frac{1}{2}}$$

$$k^{max} = \frac{t^{max}}{\sqrt{x}} = \frac{0.91\,s}{\sqrt{(3\,m)}} = 0.53\,\text{s}\,\text{m}^{-\frac{1}{2}}$$

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:

$$\bar{k} = \frac{1}{2}(k^{min} + k^{max}) = 0.44\,\text{s}\,\text{m}^{-\frac{1}{2}}$$

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

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

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 true uncertainties in a measurement.

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

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

$$\bar{F} = F(\bar{x}, \bar{y})$$

$$\sigma_F = \sqrt{\left(\frac{\partial F}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial F}{\partial y}\sigma_y\right)^2} \tag{2.4}$$

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 $x = (3.00 \pm 0.01)\,\mathrm{m}$ and $t = (0.76 \pm 0.15)\,\mathrm{s}$.

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

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

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

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

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

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

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

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

So we find that:

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

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

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 methods leads to a few simple short cuts when propagating the uncertainties for simple operations, as shown in Table 2.6. 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 2.6: *How to propagate uncertainties from measured values $x \pm \sigma_x$ and $y \pm \sigma_y$ to a quantity $z(x, y)$ for common operations.*

Checkpoint 2-4: We have measured that a llama can cover a distance of $(20.0 \pm 0.5)\,\mathrm{m}$ in $(4.0 \pm 0.5)\,\mathrm{s}$. What is the speed (with uncertainty) of the llama?

2.3.2   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. $(10\,123 \pm 310)\,\text{m}$ would be $(10.12 \pm 0.31) \times 10^3\,\text{m}$ or $(101.2 \pm 3.1) \times 10^2\,\text{m}$

> **Checkpoint 2-5:** Someone has measured the average height of tables in the laboratory to be $1.0535\,\text{m}$ with a standard deviation of $0.0525\,\text{m}$. 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 make science advance, 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 $k = (0.44 \pm 0.09)\,\text{s}\,\text{m}^{-\frac{1}{2}}$. 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 $0.45\,\text{s}\,\text{m}^{-\frac{1}{2}}$ 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 $k = (0.55 \pm 0.08)\,\text{s}\,\text{m}^{-\frac{1}{2}}$ so that the lowest value compatible with our measurement, $k = 0.55\,\text{s}\,\text{m}^{-\frac{1}{2}} - 0.08\,\text{s}\,\text{m}^{-\frac{1}{2}} = 0.47\,\text{s}\,\text{m}^{-\frac{1}{2}}$ 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

---

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

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

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

2.4 Using computers to facilitate data analysis

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 computers.

2.4.1 Python: a quick intro to programming

In Python, as in other programming language, the equal sign is called the **assignment operator**. Its role is assign the value on its right to the variable on tits 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 2.1: Declaring variables in Python

```python
#This is a comment, and is ignored by Python
a = 2
```

```
b = 2*a
print(b)
```

Output 2.1:

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 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. 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", as in the following example:

Python Code 2.2: Using functions from modules

```
#First, we load (import) the math module
import math as m
a = 9
b = m.sqrt(a)
print(b)
```

Output 2.2:

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()`).

2.4.2   Using QExpy for propagating uncertainties

QExpy is a Python module that was developed at Queen's University to handle all aspects of undergraduate physics laboratories. In this section, we look at how to use it to propagate uncertainties. Recall Example 2-5, where we propagated the uncertainties in t and x to $k = \frac{t}{\sqrt{x}}$. We show below how easily this can be done with QExpy:

Python Code 2.3: 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)
#Print the result:
print(k)
```

Output 2.3:

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).

Also recall that in Table 2.4, we have 5 different measurements of time that we used to calculate the mean and standard deviation of t to use as central value and uncertainty. We can do this very easily in QExpy, by setting our value of t to be equal to a list of measurements:

Python Code 2.4: QExpy to calculate mean and standard deviation

```
#First, we load the QExpy module
import qexpy as q
#We define $t$ as a list 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 2.4:

```
t = 0.76 +/- 0.15
```

thus, we do not need to tediously calculate the mean and standard deviation, as we did in Example 2-4.

2.4.3 Using QExpy for plotting

Recall Table 2.3, where we measured the time for an object to drop from different heights. One of the easiest ways to look at the data is to visualize them on a graph. In this case, we measured the time, t, that it took to drop different heights, x. Chloë's Theory stated that the time, t, is proportional to the square root of the distance fallen, x, and we introduced a constant of proportionality k:

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

This means that if we make a graph of t versus $\sqrt{x}$, we should expect that the points fall on a straight line that goes through zero, with a slope of k. We can easily use QExpy to make this plot of the data in Table 2.3.

Python Code 2.5: Using QExPy for plotting

```
#First, we load the QExpy module:
import qexpy as q

#Then we enter the data:
#start with the values for the square root of height:
sqx = [1. , 1.41, 1.73, 2., 2.24]
#and then, the corresponding times:
t = [ 0.33, 0.74, 0.67, 1.07, 1.1 ]

#Let us attribute an uncertainty of 0.15 to each measured values of t:
```

```
terr = 0.15

#We now make the plot. First, we create the plot object with the data.
fig = q.MakePlot( xdata = sqx, xname = "sqrt(distance)", xunits = "sqrt(m)",
                  ydata = t, yerr = terr, yname = "time", yunits ="s",
                  data_name = "Data1")

#Ask QExpy to also show the line of best fit
fig.fit("linear")

#Then, we show it:
fig.show()
```

Output 2.5:

```
——————————————————Fit results——————————————————
Fit of  Data1   to   linear
Fit parameters:
Data1_linear_fit0_fitpars_intercept = −0.24 +/− 0.22,
Data1_linear_fit0_fitpars_slope = 0.61 +/− 0.13

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

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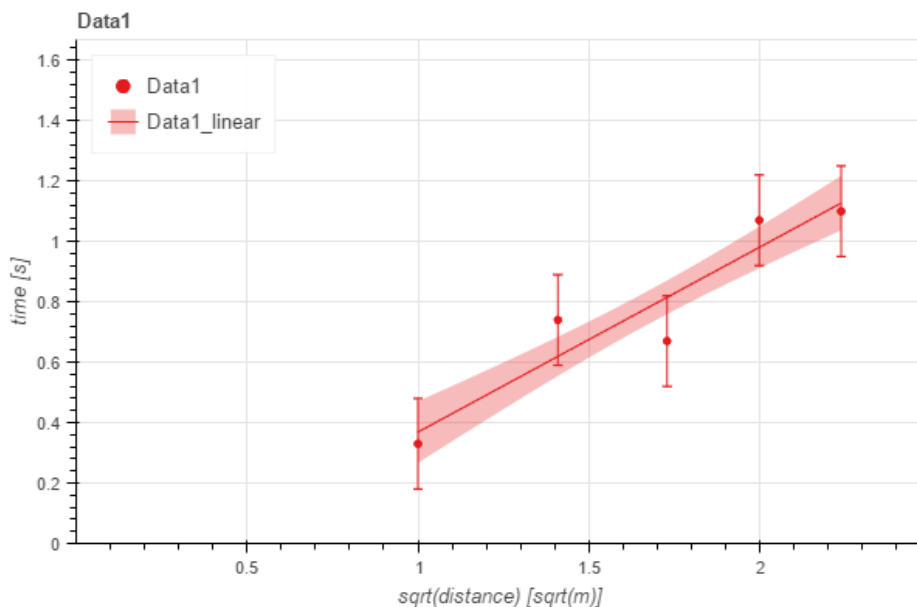

Figure 2.2:   *QExpy plot of t versus $\sqrt{x}$.*

The plot in Figure 2.2 shows that the data points do indeed appear to fall near a straight line. 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 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 offset within their uncertainties. The output also provides a line that tells us that `chi2/ndof = 2.04/2`; although you do not need to understand the details, this is a measure of how well the data are described by the line of best fit. Generally, the fit is assumed to be "good" if this ratio is close to 1 (the ratio is called "the reduced chi-squared"). The "correlation matrix" tells us how the best fit value of the slope is linked to the best fit value of the intercept, which you do not need to worry about here.

Since we expect the slope of the data to be k, this provides us a method to determine k from the data as $(0.61 \pm 0.13)\,\mathrm{s\,m}^{-\frac{1}{2}}$. When we used Table 2.4 to determine k using repeated measurements at a drop height of $3.0\,\mathrm{m}$, we obtained $k = (0.44 \pm 0.09)\,\mathrm{s\,m}^{-\frac{1}{2}}$, which is consistent with what we get from the slope of the best fit line. Finally, we expect the intercept to be equal to zero. The best fit line from QExpy has an intercept of $(-0.24 \pm 0.22)\,\mathrm{s}$, which is slightly below, but consistent, with zero. From these data, we would conclude that our measurements are consistent with Chloë's Theory.

2.5   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.
- When propagating uncertainties, use a computer #becauseits2015
- 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).

> **Learning Objectives:**
> - Describe motion in 1D using functions and defining an axis.
> - Define position, velocity, speed, and acceleration.
> - Use derivatives to determine the rate of change of a quantity.
> - Use integrals to calculate sums.
> - 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 axis along the direction of motion
- we specify an origin where $x = 0$
- we specify a direction along the axis of motion corresponding to increasing values of x
- we specify the units for the quantity, x.

31

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.
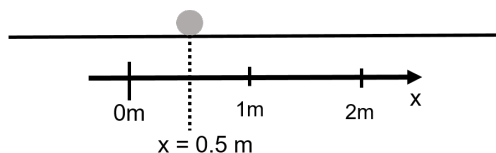

Figure 3.1: In order to describe the motion of the grey ball along a straight line, we introduce the x-axis, represented by an arrow to indicate the direction of increasing x, and the location of the origin, where $x = 0$ m. Given our choice of origin, the ball is currently at a position of $x = 0.5$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

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.

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: QExpy to calculate mean and standard deviation

```
#First, we load the QExpy module
import qexpy as q
#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:
position = [0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0]
#Define the plot, and show it:
fig = q.MakePlot(xdata=t, xunits="s", xname="time",
                 ydata=position, yunits="m", yname="position",
                 data_name="position vs time")
fig.show()
```

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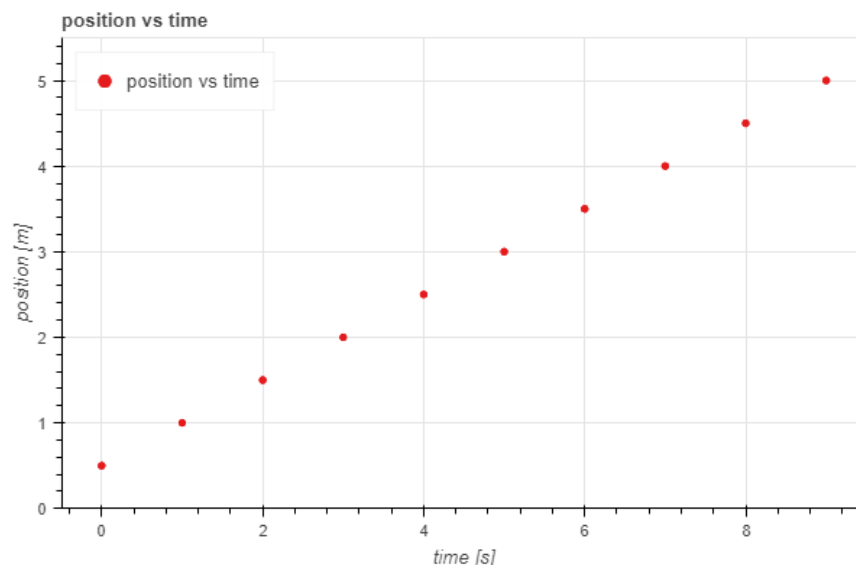
Figure 3.2: Plot of position as a function of time using the values from Table 3.1.

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 $x = 0.5\,\text{m}$.

Checkpoint 3-1: Using the data from Table 3.1, at what position along the x-axis will the ball be when time is $t = 9.5\,\text{s}$, if it continues its motion undisturbed?

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:

$$x(t) = a + b \times t$$

The constant a is the "offset" of the function, the value that the function has at $t = 0\,\text{s}$. 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 $t = 0\,\text{s}$ and $t = 2.0\,\text{s}$:

$$x(t = 0\,\text{s}) = 0.5\,\text{m} = a + b \times (0\,\text{s})$$
$$x(t = 2.0\,\text{s}) = 1.5\,\text{m} = a + b \times (2.0\,\text{s})$$

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

$$1.5\,\text{m} = a + b \times (2.0\,\text{s}) = 0.5\,\text{m} + b \times (2.0\,\text{s})$$
$$\therefore b = \frac{(1.5\,\text{m}) - (0.5\,\text{m})}{(2.0\,\text{s})} = 0.5\,\text{m/s}$$

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

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

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 $t = 0\,\text{s}$ (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 "speed" 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.
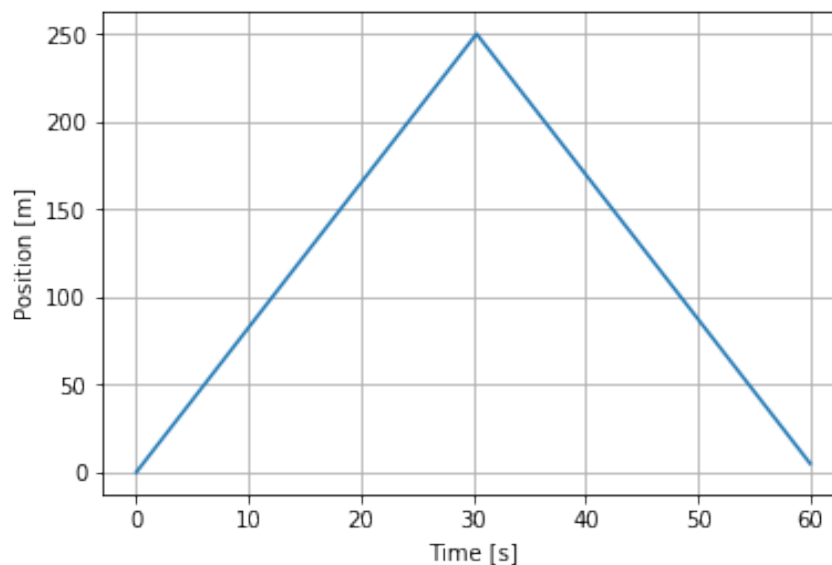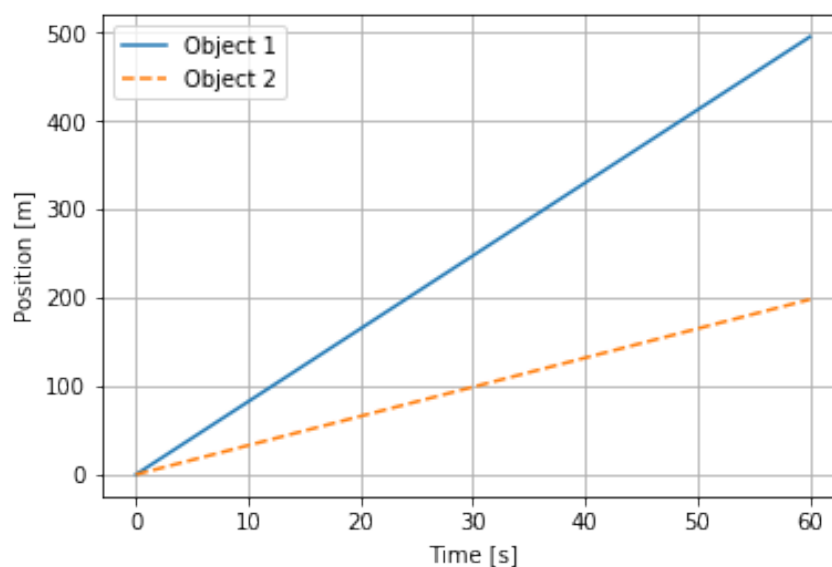
Figure 3.3: Position as a function of time for an object.

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 $t = 0\,\text{s}$ and $t = 30\,\text{s}$, then slowed down to a stop at $t = 60\,\text{s}$.

B) The object moved in the positive x-direction between $t = 0\,\text{s}$ and $t = 30\,\text{s}$, and then turned around and moved in the negative x-direction between $t = 30\,\text{s}$ and $t = 60\,\text{s}$.

C) The object moved with a higher speed between $t = 0\,\text{s}$ and $t = 30\,\text{s}$ than it did between $t = 30\,\text{s}$ and $t = 60\,\text{s}$.

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

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

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

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

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

where a_x is the "acceleration" and v_{0x} is the speed 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 speed:

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

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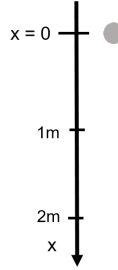
Figure 3.5: X-axis for an object that starts at rest at $x = 0\,\mathrm{m}$ when $t = 0\,\mathrm{s}$ and falls downwards (in the direction of increasing x).

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 ($v_{0x} = 0\,\mathrm{m/s}$) at the position $x = 0\,\mathrm{m}$ at time $t = 0\,\mathrm{s}$ with a constant positive acceleration $a_x = 10\,\mathrm{m/s^2}$. We would like to use equation 3.1, but we cannot because it only applies if the speed is constant. To remedy this, we pretend (we "approximate") that for a a very small amount of time, the speed is almost constant. Let us take a very small interval in time, say $\Delta t = 0.001\,\mathrm{s}$, and approximate that the speed is constant during that interval.

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

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

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

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

With the average speed during the interval, we can use equation 3.1 to find the position at

$t = \Delta t$:

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

Thus, at time $t = 0.001\,\mathrm{s}$, the object will have a speed of $v = 0.005\,\mathrm{m/s}$ and will have covered a distance of $0.000\,005\,\mathrm{m}$. We can now use these values as the starting speed and position for the next interval in time. Using variables, at the beginning of the second interval, the speed 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 speed during the second interval is thus given by:

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

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 speed that we just found:

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

where we kept the $\frac{1}{2}$ factored out and brought a factor of 2 inside the parenthesis with the Δt. 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 Nth 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:

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

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:

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

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:

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

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 speed 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/s^2.

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 speeed 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:

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

where the initial speed is in the positive x-direction, and the acceleration, a_x, is in the negative direction (the speed 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 speed will be zero (that is what it means to stop). We can use equation 3.2 to find what time that corresponds to:

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

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

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

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

3.2.1   Visualizing motion with constant acceleration

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

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

where the speed changed 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.
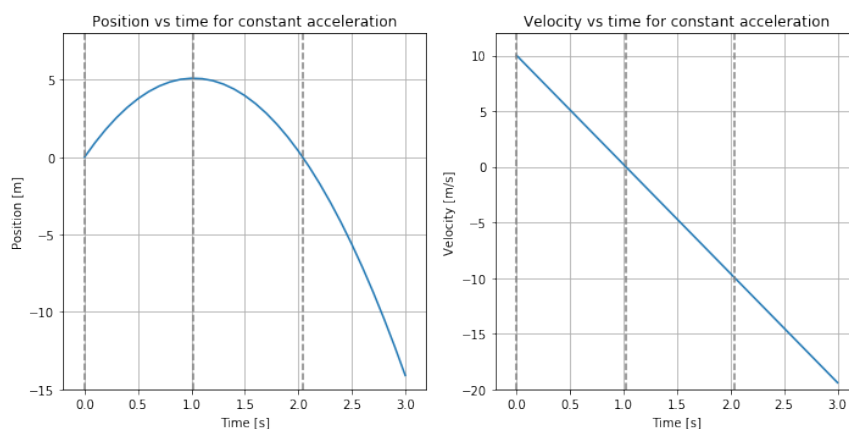

Figure 3.6:   *Position and speed as a function of time for the ball in example 3-1.*

We can divide the motion into three parts:

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

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 continues to increase (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 speed decreases linearly from 10 m/s to 0 m/s due to the constant negative acceleration from gravity. At $t = 1.02$ s, the speed is instantaneously 0 m/s and the ball at 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 speed continues to decrease linearly (it becomes more and more negative) as the ball start to fall back down faster and faster. The position also starts decreasing just after $t = 1.02$ 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 speed (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 speed. The position continues to become more negative and the speed 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 speed 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 velocity[1]. 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. This is why we used the letter 'v' in all of the equations.

---

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

3.3 Using calculus to describe motion

Most objects do not 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 the limit that Δt tends to zero, this will be an accurate description.

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 $\Delta t \to 0$, we can define the **instantaneous velocity**:

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

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

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

If we go back to Figure 3.6, where speed was no longer constant, we can indeed see that the graph of the velocity versus time $(v(t))$ corresponds to the instantaneous slope of the graph of position versus time $(x(t))$. For $t < 1.02\,\text{s}$, the slope of the $x(t)$ graph is positive but decreasing (as is $v(t)$). At $t = 1\,02s$, the slope of $x(t)$ is instantaneously $0\,\text{m/s}$ (as is the velocity). Finally, for $t > 1.02\,\text{s}$, the slope of $x(t)$ is negative and increasing in magnitude, as is $v(t)$.

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

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

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

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

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

3.3.1   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:

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

With our calculus-based definitions above, we should be able to recover that:

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

$$a_x(t) = a_x$$

Let us start by determining $v(t)$:

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

Knowing our function $x(t)$, we can can re-write this as:

$$\boxed{\frac{dx}{dt} = \lim_{\Delta t \to 0} \frac{x(t + \Delta t) - x(t)}{\Delta t}} \quad (3.8)$$

which is the formal definition of the deriva-tive $\frac{dx}{dt}$ and gives us a prescription to evalu-ate it. We will proceed as illustrated in Figure 3.7:
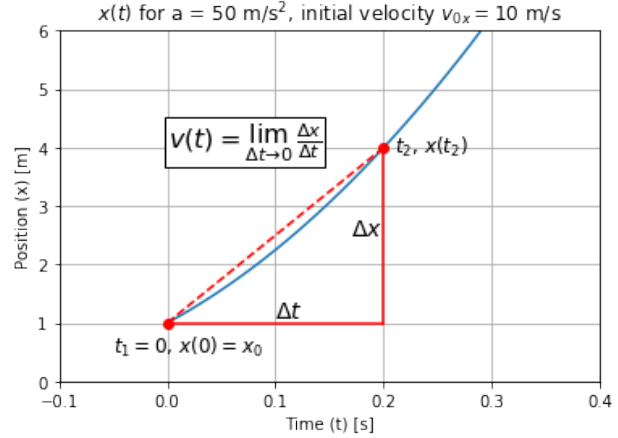

Figure 3.7: Obtaining the instantaneous velocity from the graph of position as a function of time. As $\Delta t \to 0$, $\frac{\Delta x}{\Delta t}$ approaches the instantaneous velocity $v(t)$, which is the slope of the curve $x(t)$ at time t.

1. Use $x(t)$ to determine Δx for a small interval Δt
2. Divide Δx by Δt
3. Take the limit, $\lim_{\Delta t \to 0}$

To obtain Δx, we introduce a start time, t_1, and an end time t_2 for our time interval, such that $t_2 - t_1 = \Delta t$, centred about a time t. The change in position is then given by:

$$\Delta x = x(t_2) - x(t_1)$$
$$= \left(x_0 + v_{0x}t_2 + \frac{1}{2}a_x t_2^2\right) - \left(x_0 + v_{0x}t_1 + \frac{1}{2}a_x t_1^2\right)$$
$$= v_{0x}(t_2 - t_1) + \frac{1}{2}a_x(t_2^2 - t_1^2)$$
$$= v_{0x}\Delta t + \frac{1}{2}a_x(t_2 - t_1)(t_2 + t_1)$$
$$= v_{0x}\Delta t + \frac{1}{2}a_x\Delta t(t_2 + t_1)$$

which we divide by Δt to get $v(t)$:

$$v(t) = \frac{\Delta x}{\Delta t}$$
$$= \frac{v_{0x}\Delta t + \frac{1}{2}a_x\Delta t(t_2 + t_1)}{\Delta t}$$
$$= v_{0x} + \frac{1}{2}a_x(t_2 + t_1)$$

We now take the limit where $\Delta t \to 0$, namely when $t_2 - t_1$ is very small. As we make the interval small, t_2 and t_1 will both approach the same value of time, say t, corresponding to the time at the centre of the interval. In particular, the average of t_1 and t_2, given by $\frac{1}{2}(t_1 + t_2)$, will approach the time at the centre of the interval, t. We thus recover the equation for instantaneous velocity:

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

Of course, once you become more familiar with calculus, you will be able to directly use the formulas for derivatives to recover the answer:

$$v(t) = \frac{d}{dt}\left(x_0 + v_{0x}t + \frac{1}{2}a_x t^2\right)$$
$$= v_{0x} + at$$

Similarly, we can now confirm that the acceleration is a constant, independent of time:

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

You can easily verify that you obtain this result by first calculating $v(t)$ at two different times, t_1 and t_2, taking the difference, $\Delta v = v(t_2) - v(t_1)$, and then taking the limit of $\lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t}$ to get $a_x(t)$.

Checkpoint 3-5: Chloë has been working on a detailed study of how vicuñas[a] run, and found that their position as a function of time when they start running is well modelled by the function $x(t) = (40\,\mathrm{m/s^2})t^2 + (20\,\mathrm{m/s^3})t^3$. What is the acceleration of the vicuñas?

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

---
[a]Never heard of vicuñas? Internet!

3.3.2   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:

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

where as we will see later, the constant C is required in order for the function $v(t)$ to go through the point $v(t = 0) = v_{0x}$. All we need now is to determine how to calculate the anti-derivative, $\int a_x(t)dt$. Since in this case, $a_x(t)$ is a constant, a_x, we can determine the anti-derivative quite easily.

For a small interval in time, Δt, the velocity will change by a small amount, Δv, such that:

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

If we label the start time of the interval as t_1 and the end of the interval as t_2, we have:

$$\Delta t = t_2 - t_1$$
$$\Delta v = v(t_2) - v(t_1) = a_x \Delta t$$
$$\therefore v(t_2) = v(t_1) + a_x \Delta t$$

If we set $t_1 = 0\,\text{s}$ to correspond to the point where $v(t) = v_{0x}$, then we can write the velocity at t_2 as:

$$v(t_2) = v(t = \Delta t) = v_{0x} + a_x \Delta t$$

and we see that the velocity changed by an amount $a\Delta t$ over a period of time Δt. Since a_x is the same at all times, this is always true, and after a period of time $t = N\Delta t$, the

velocity will have changed by $Na_x\Delta t$, and we recover the original equation for velocity as a function of time when acceleration is constant:

$$v(t = N\Delta t) = v_{0x} + Na_x\Delta t$$
$$\therefore v(t) = v_{0x} + a_x t$$

We can identify the anti-derivative for the case where $a(t)$ is constant:

$$\frac{d}{dt}v(t) = a$$
$$\int a\,dt = a_x t + C$$

where the constant, C, is given by v_{0x}.

To now obtain position as a function of time, we proceed in the same manner, namely:

1. Define a small interval in time Δt
2. Calculate the corresponding change in position Δx
3. Add Δx to our original position, and repeat.

Again, we have the time-derivative of the position equal to a function of time:

$$\frac{d}{dt}x(t) = v(t) = v_{0x} + a_x t$$

and we need to find the anti-derivative:

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

given that, at time $t = 0$, the position was $x = x_0$. After a small interval in time, Δt, the position will have changed by AN amount Δx:

$$\Delta x = v(t)\Delta t$$

so that the position at time $t = \Delta t$ will be given by:

$$x(t = \Delta t) = x_0 + \Delta x$$
$$= x_0 + v(t)\Delta t$$

Figure 3.8: *Determining Δx, given $v(t)$ and Δt. Three different choices of $v(t)$ are shown, depending on whether $v(t)$ is evaluated at the start of the interval, in the middle, or at the end. As $\Delta t \to 0$, these all become equal.*

The problem here is to evaluate $v(t)$ since the velocity changes throughout the interval. One possible choice is to evaluate the velocity at $t = \frac{1}{2}\Delta t$, midway in the interval, as we did
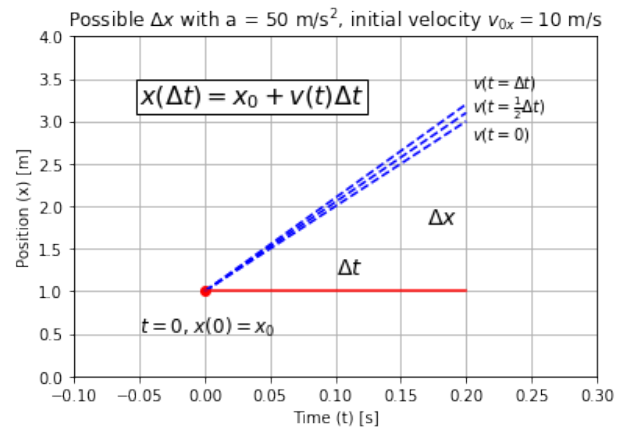
before. We could also choose to use the velocity at the beginning or at the end of the time interval, as all three choices will converge to the same value when $\Delta t \to 0$, as illustrated in Figure 3.8. For now, we will leave the choice open and simply call the velocity that we use v_1 to indicate that it is the velocity in the first interval. We thus write the position, $x(t = \Delta t)$, after a time interval Δt as:

$$x(t = \Delta t) = x_0 + v_1 \Delta t$$

The position, $x(t = 2\Delta t)$, after another interval in time Δt will then be given by:

$$\begin{aligned} x(t = 2\Delta t) &= x(t = \Delta t) + v_2 \Delta t \\ &= x_0 + v_1 \Delta t + v_2 \Delta t \end{aligned}$$

where v_2 is the velocity over the second interval in time (different than v_1, since velocity changes with time). For the Nth interval, we label the position $x_N = x(t = N\Delta t)$:

$$x_N = x(t = N\Delta t) = x_0 + v_1 \Delta t + v_2 \Delta t + \cdots + v_N \Delta t$$
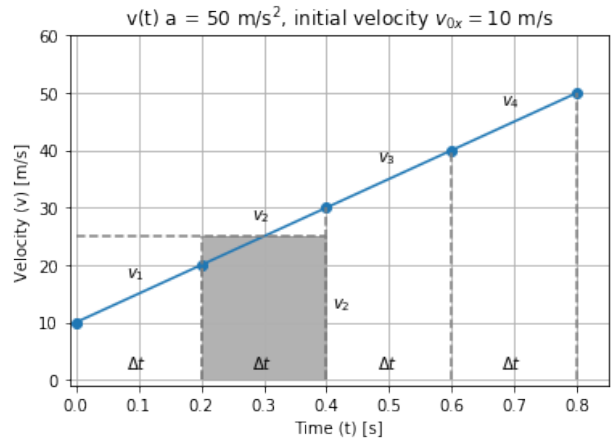
$$= x_0 + \sum_{i=1}^{N} v_i \Delta t$$

where we made use of the summation notation ($\sum$) to avoid writing out every term. The above equation is only correct in the limit of $\Delta t \to 0$, in which case it must be the anti-derivative of $v(t)$:

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

$$= x_0 + \lim_{\Delta t \to 0} \sum_{i=1}^{N} v_i \Delta t$$

so that we can identify:

$$\boxed{\int v(t)dt = \lim_{\Delta t \to 0} \sum_{i=1}^{N} v_i \Delta t + C} \quad (3.9)$$

Figure 3.9: Illustration of the anti-derivative $\int v(t)dt$ as a sum.

where it is understood that v_i is the "average" velocity in the ith interval, and the constant $C = x_0$ ensures that $x(t = 0) = x_0$. This is now a general definition for the anti-derivative, as we have made no specific assumption about the function $v(t)$. Equation 3.9 tells us that the anti-derivative of a function (in this case, $v(t)$) can be obtained from a sum.

The sum is illustrated in Figure 3.9, which shows the function $v(t)$ and several intervals Δt. Over each interval, i, we labeled the average velocity, v_i. As the intervals shrink, $\Delta t \to 0$, the average velocity v_i approaches the instantaneous velocity, $v(t)$, at the centre for the interval. Since the function $v(t)$ is linear, the speed at the middle of an interval is exactly

equal to the average speed in the interval. Taking v_i as the speed in the middle of the interval, we then see that each term in the sum, $v_i \Delta t$, is equal to the area between the curve $v(t)$ and the t-axis. This is illustrated for the second term in the sum, $v_2 \Delta t$, with the grey rectangle in Figure 3.9.

The anti-derivative of a function is thus related to the area between the function and the horizontal axis. If we specify limits on the horizontal axis between which we calculate the area, then the anti-derivative is called an **integral**. For example, if we wish to calculate the sum of $v_i \Delta t$ for values between t_a and t_b, we would write the integral as:

$$\int_{t_a}^{t_b} v(t) dt = \lim_{\Delta t \to 0} \sum_{i=1}^{N} v_i \Delta t$$

where the sum is such that for $i = 1$, v_i is close to $v(t = t_a)$ and for $i = N$, v_N is close to $v(t = t_b)$. An illustration of taking the integral of $v(t)$ is shown in Figure 3.10 where the sum is shown for two different values of Δt. It is clear that as $\Delta t \to 0$, the sum becomes equal to the area between the curve and the horizontal axis.
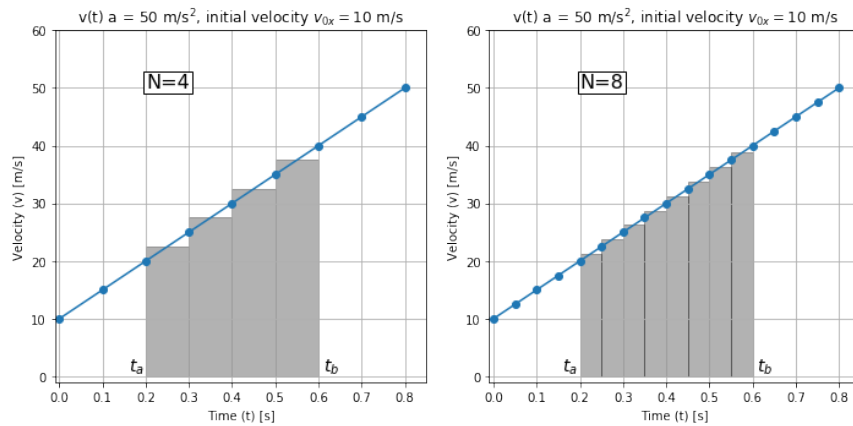

Figure 3.10: Integral of $v(t)$ between $t_a = 0.2$ s and $t_b = 0.6$ s illustrated as a sum with of 4 terms when $\Delta t = 0.1$ s or of 8 terms when $\Delta t = 0.05$ s.

Since $v(t)$ is a linear function when acceleration is constant, we can easily calculate the area between the curve and the horizontal axis. In the case of a linear function, $v(t) = v_{0x} + at$, the area is a trapezoid, and we have:

$$\int_{t_a}^{t_b} v(t) dt = \text{base} \times \text{average height}$$

$$= (t_b - t_a) \times \frac{1}{2} (v(t_a) + v(t_b))$$

$$= (t_b - t_a) \frac{1}{2} (v_{0x} + a_x t_a + v_{0x} + a_x t_b)$$

$$= \left(\frac{1}{2} (v_{0x} t_b + a_x t_a t_b + v_{0x} t_b + a_x t_b^2) \right) - \left(\frac{1}{2} (v_{0x} t_a + a_x t_a^2 + v_{0x} t_a + a_x t_b t_a) \right)$$

$$= \left(v_{0x} t_b + \frac{1}{2} a_x t_b^2 \right) - \left(v_{0x} t_a + \frac{1}{2} a_x t_a^2 \right)$$

If we define a new function, $V(t) = v_{0x}t + \frac{1}{2}a_x t^2$, then we have:

$$\int_{t_a}^{t_b} v(t)dt = V(t_b) - V(t_a)$$

In other words, given a function, $v(t)$, the integral of that function between two values t_a and t_b can be found by evaluating a different function, $V(t)$, at the end points t_a and t_b[2]. As you will see in your calculus course, the function $V(t)$ is precisely what we call the anti-derivative:

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

which has derivative:

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

Note that when taking the integral, the constant C always cancels.

> **Checkpoint 3-6:** The acceleration of a cricket jumping sideways is observed to increase linearly with time, that is, $a_x(t) = a_0 + jt$, where a_0 and j are constants. What can you say about the velocity of the cricket as a function of time?
>
>   A) it is constant
>   B) it increases linearly with time ($v(t) \propto t$)
>   C) it increases quadratically with time ($v(t) \propto t^2$)
>   D) it increases with the cube of time ($v(t) \propto t^3$)

3.4   Relative motion

In order to describe the motion of an object confined to a straight line, we introduced a 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 3.11 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.

---

[2]Note that we only explicitly showed that this true if $v(t)$ is linear, but the result is in fact general.
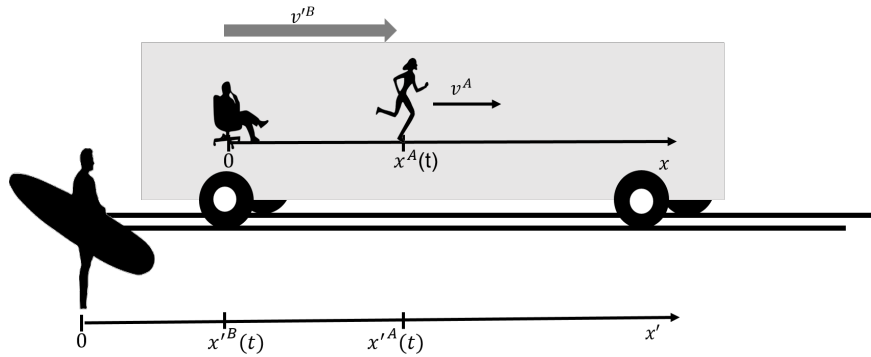
Figure 3.11: Alice is walking in the train and her position is described by both Brice, who is sitting in the train (using the x axis), and Igor, who is at rest on the ground (using the x′ axis).

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 $108\,000\,\mathrm{km/h}$.

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.

3.4.1 Galilean Relativity

Referring to Figure 3.11, 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:

$$\boxed{x'^A(t) = x'^B(t) + x^A(t) = v'^B t + x^A(t)} \tag{3.10}$$

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:

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

$$= \frac{d}{dt}(v'^B t + x^A(t)) \tag{3.12}$$

$$= v'^B + \frac{d}{dt}x^A(t) \tag{3.13}$$

$$= v'^B + v^A(t) \tag{3.14}$$

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:

$$a'^A(t) = \frac{d}{dt}v'^A(t) \tag{3.15}$$

$$= \frac{d}{dt}(v'^B + v^A) \tag{3.16}$$

$$= 0 + \frac{d}{dt}v^A(t) \tag{3.17}$$

$$= a^A \tag{3.18}$$

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 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 a object are directly related to the acceleration experienced by that object. Thus, the forces on a object do not depend on the choice of inertial reference frame.

Example 3-2: A large boat is sailing North at a speed of $v'^B = 15\,\text{m/s}$ and a restless passenger is walking about on the deck. Chloë, another passenger on the boat, finds that the passenger is walking at a constant speed of $v^A = 3\,\text{m/s}$ towards the South (opposite the direction of the boat's motion). Marcel is watching the boat pass by from 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 the in 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:

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

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 $12\,\mathrm{m/s}$, 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.

3.4.2 Special Relativity

In the previous section, we found that the position of an object, x^A, in one inertial reference frame can be converted to a position, x'^A, in a different inertial reference frame using the equation:

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

where v'^B is the velocity of the x frame of reference as measured in the x' frame of reference. We found this by arguing that we could sum the distance from the x' origin to the x origin and the distance from the x origin to the object's position, x^A, in the x reference frame. We call this transformation, from x to x', a "Galilean transformation". In this section, we briefly introduce the more general way to transform descriptions between reference frames, mostly for completeness.

Although most of our common experience is well described by this transformation, Albert Einstein found that it is no longer accurate if the speeds involved in the situation are close to the speed of light. Einstein's Theory of Special Relativity gives the following transformation instead:

$$x'^A = \frac{1}{\sqrt{1 - \left(\frac{v'^B}{c}\right)^2}}(v'^B t + x^A) \tag{3.19}$$

$$t' = \frac{1}{\sqrt{1 - \left(\frac{v'^B}{c}\right)^2}}\left(t + \frac{v'^B x^A}{c^2}\right) \tag{3.20}$$

where c is the speed of light ($3 \times 10^8\,\mathrm{m/s}$). The first thing to note is that this a considerably more complicated expression. The most bizarre aspect is that the two reference frames have different measures of time (t and t'). The above transformation, from x, t to x', t', is called a "Lorentz transformation". The fact that we have to include time in the transformation is an aspect related to the fact that Einstein's Theory of Special Relativity tells us that we must consider Space-Time as a single entity (rather than thinking of space and time as separate things).

Let us examine a few aspects of the Lorentz transformation. First, if the velocity, v'^B,

between reference frames is small when compared to the speed of light, $v'^B << c$, we have:

$$\left(\frac{v'^B}{c}\right)^2 << 1$$

$$\therefore \frac{1}{\sqrt{1 - \left(\frac{v'^B}{c}\right)^2}} \sim 1$$

In this case, we recover the original Galilean transformation:

$$x'^A = \lim_{v_B << c} \frac{1}{\sqrt{1 - \left(\frac{v'^B}{c}\right)^2}} (v'^B t + x^A) = v'^B t + x^A$$

$$t' = \lim_{v_B << c} \frac{1}{\sqrt{1 - \left(\frac{v'^B}{c}\right)^2}} \left(t + \frac{v'^B x^A}{c^2}\right) = t$$

So *a priori*, this aspect of Einstein's Theory of Special Relativity is difficult to test, as we only expect it to deviate from common experience when v_B is comparable to the speed of light. However, in particle accelerators, subatomic particles are regularly accelerated to speeds close to the speed of light, and their position has to be described using Special Relativity.

Suppose that a space ship is carrying Chloë and moving with a speed of $v'^B = 0.7c = 2.1\,\text{m/s}$. Suppose that Chloë fires an arrow in the same direction as the space ship's motion, and that she measures the speed of the arrow to be $v^A = 0.4c = 0.9 \times 10^8\,\text{m/s}$. Suppose that the space ship goes by Marcel, who is stationary on Earth, right at the moment that Chloë fires the arrow. Let us define the x axis on the ship so that it is positive in the direction of motion with its origin where Chloë fired the arrow. Let us define $t = t' = 0$ to be the moment when the origin of the space ship x-axis is aligned with the origin of Marcel's x' axis on Earth. The x' axis is also set up to be positive in the direction of the space ship's motion and has the origin such that at $t' = 0$

Using Galilean Relativity, the speed of the arrow as measured by Marcel would be found by adding the two velocities, $v'^A = v^A + v'^B = 1.1c = 3.3 \times 10^8\,\text{m/s}$. Using Special Relativity, the situation is a little more difficult. The position as measured in Marcel's reference frame is given by:

$$x'^A = \frac{1}{\sqrt{1 - \left(\frac{v'^B}{c}\right)^2}} (v'^B t + x^A)$$

To obtain the arrow's velocity in Marcel's frame, we need the time derivative of x'^A with respect to t', the time measured in Marcel's frame:

$$v'^A = \frac{d}{dt'} x'^A$$

However, x'^A is written in terms of t, not t'. However, equation 3.19, tells us how t depends on t'; we thus need to treat t as a function of t' and use the Chain Rule:

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

$$= \frac{d}{dt}x'^A(t)\left(\frac{dt}{dt'}\right)$$

$$= \frac{d}{dt}\left(\frac{1}{\sqrt{1-\left(\frac{v'^B}{c}\right)^2}}(v'^B t + x^A)\right)\left(\frac{dt}{dt'}\right)$$

$$= \frac{1}{\sqrt{1-\left(\frac{v'^B}{c}\right)^2}}\frac{d}{dt}\left((v'^B t + x^A)\right)\left(\frac{dt}{dt'}\right)$$

$$= \frac{1}{\sqrt{1-\left(\frac{v'^B}{c}\right)^2}}\left(v'^B + \frac{d}{dt}x^A\right)\left(\frac{dt}{dt'}\right)$$

$$= \frac{1}{\sqrt{1-\left(\frac{v'^B}{c}\right)^2}}\left(v'^B + v^A\right)\left(\frac{dt}{dt'}\right)$$

We can differentiate equation 3.19 to get $\frac{dt'}{dt}$:

$$\frac{dt'}{dt} = \frac{d}{dt'}\left(\frac{1}{\sqrt{1+\left(\frac{v'^B}{c}\right)^2}}\left(t + \frac{v'^B x^A}{c^2}\right)\right)$$

$$= \frac{1}{\sqrt{1+\left(\frac{v'^B}{c}\right)^2}}\left(1 + \frac{v'^B v^A}{c^2}\right)$$

where we identified $v^A = \frac{dx^A}{dt}$. We can take the inverse to get $\frac{dt}{dt'}$:

$$\frac{dt}{dt'} = \frac{1}{\frac{dt'}{dt}}$$

$$= \frac{\sqrt{1+\left(\frac{v'^B}{c}\right)^2}}{1+\frac{v'^B v^A}{c^2}}$$

Combining everything, we get the velocity of the arrow as measured in Earth's reference frame by Marcel:

$$v'^A = \frac{1}{\sqrt{1-\left(\frac{v'^B}{c}\right)^2}}\left(v'^B + v^A\right)\left(\frac{dt}{dt'}\right)$$

$$\boxed{v'^A = \frac{v'^B + v^A}{1 + \frac{v'^B v^A}{c^2}}} \tag{3.21}$$

The above equation tells us how to convert the velocity of the arrow as measured in the space ship (v^A) into the velocity as measured on Earth, v'^A, when the space ship moves with velocity v'^B as measured on Earth. When v^A and v'^B are small compared to c, then this equation reduces to the Galilean version $(v'^A = v'^B + v^A)$. However, when the velocities are large this is no longer the case. For the above example, with $v^A = 0.4c$, $v'^B = 0.7c$, we get:

$$
\begin{aligned}
v'^A &= \frac{v'^B + v^A}{1 + \frac{v'^B v^A}{c^2}} \\
&= \frac{(0.7c) + (0.4c)}{1 + \frac{(0.7c)(0.4c)}{c^2}} \\
&= \frac{1.1}{1 + 0.28} c \\
&= 0.86c = 2.57 \times 10^8 \,\mathrm{m/s}
\end{aligned}
$$

which is well short of the $1.1c$ that you would expect from Galilean relativity. Suppose instead that $v^A = c$, that is, Chloë fires the arrow at the speed of light (she sees the arrow move away at the speed of light). Marcel will measure the speed of the arrow to be:

$$
\begin{aligned}
v'^A &= \frac{v'^B + v^A}{1 + \frac{v'^B v^A}{c^2}} \\
&= \frac{v'^B + (1.0c)}{1 + \frac{v'^B (1.0c)}{c^2}} \\
&= \frac{v'^B + c}{1 + \frac{v'^B}{c}} \\
&= c \frac{v'^B + c}{c + v'^B} \\
&= c
\end{aligned}
$$

regardless of the speed of the space ship! This is a truly strange result from Einstein's theory. If Chloë fires an arrow at the speed of light, Marcel will also measure the arrow moving at the speed of light, regardless of whether or not the space ship is moving. Indeed, all observers in inertial frames of references would agree that the arrow is moving at the same speed, c. Although we showed this result as a consequence of the Lorentz transformation, Einstein developed his Theory of Special Relativity by postulating that the speed of light is the same in all inertial reference frames. The Lorentz transformation is then just a consequence of this postulate. This postulate has several strange consequence that have been *experimentally* verified:

- Moving objects contract. That is, fast moving objects become shorter.
- Moving clocks slow down. That is, people moving very fast age slower.
- People in different inertial reference frames cannot agree on the order in which events took place.

We will explore this in more detail when we take a closer look at Einstein's Special Theory of Relativity.

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 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 found by:
    1. Galilean transformation, if all speeds are small
    2. Lorentz transformation, if any of the speeds are comparable to the speed of light

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:

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

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

Conversely, given the acceleration, $a_x(t)$, on can find the velocity and position:

$$v_x(t) = \lim_{\Delta t \to 0} \sum_i a_x(t_i)\Delta t + C = \int a_x(t)dt + C$$

$$x(t) = \lim_{\Delta t \to 0} \sum_i v_x(t_i)\Delta t + C = \int v_x(t)dt + C$$

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

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

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

If an object has position x^A as measured in a frame of reference x that is moving at constant speed v'^B (small compared to the speed of light) as measured in a second frame of reference x', then in the x' reference frame, the object is described by the Galilean transformation:

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

If the x frame of reference (or the object) is moving very fast compared to the speed of light, then in the x' reference frame, it is described by the Lorentz transformation:

$$x'^A = \frac{1}{\sqrt{1 - \left(\frac{v'^B}{c}\right)^2}}(v'^Bt + x^A)$$

$$t' = \frac{1}{\sqrt{1 - \left(\frac{v'^B}{c}\right)^2}}\left(t + \frac{v'^Bx^A}{c^2}\right)$$

$$v'^A = \frac{v'^B + v^A}{1 + \frac{v'^Bv^A}{c^2}}$$

where you should note that one must also transform the meaning of time, t.

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. 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.
- Know how to use polar and spherical coordinates.
- Know how to use vectors.

4.1 Motion in two dimensions

4.1.1 2D Coordinate systems

In an analogous manner to describing motion in one dimension, to describe motion in two dimensions, we first need to define a coordinate system. In two dimensions, we will use two axes, x and y, whose origin and direction we must define. Figure 4.1 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 and that have a common origin. One is free to choose any two directions as axes (as long as they are not parallel) and the origins of the two axes need not coincide. However, the choice that we have made in Figure 4.1 greatly simplifies the description of the motion of an object.

In two dimension, to fully describe the position of an object, we must specify both its position along the x axis
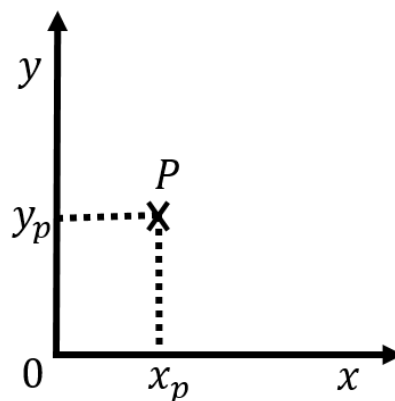
Figure 4.1: Example of Cartesian coordinate system and a point P with coordinates (x_p, y_p).

and its position along the y axis. For example, point P in Figure 4.1 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 4-1: Figure 4.2 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?
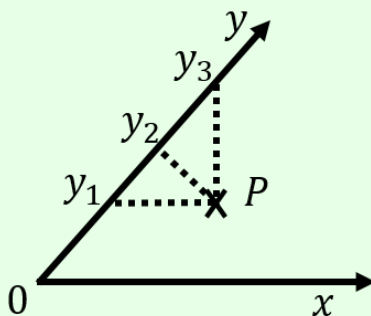
Figure 4.2: A non-orthogonal coordinate system (the x and y axes are not perpendicular).

A) y_1
B) y_2
C) y_3

The most common choice of coordinate system in two dimensions is a "Cartesian" coordinate system, where the x and y axes are perpendicular and share a common origin, as shown in Figure 4.1. When applicable, by convention, we usually choose the y axis to correspond to the vertical direction.
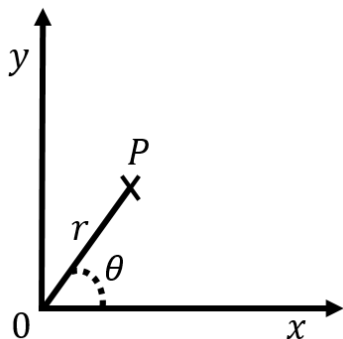

Figure 4.3:   Example of a polar coordinate system and a point P with coordinates (r, θ).

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, as shown in Figure 4.3. Often, a polar coordinate system is defined along side a Cartesian system, so that r is the distance to the origin and θ is the angle with respect to the x axis.

One can easily convert between Cartesian and polar coordina-

tes:

$$x = r\cos(\theta)$$
$$y = r\sin(\theta)$$
$$r = \sqrt{(x^2 + y^2)}$$
$$\tan(\theta) = \frac{y}{x}$$

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).

4.1.2   Position and displacement vectors

When we describe the motion of an object in two dimensions, we must specify two coordinates that can change with time. In Cartesian coordinates, this means that we must specify two functions, $x(t)$ and $y(t)$, to fully describe the motion of an object. When an object moves from one position to another, we need a way to describe that *displacement.* To do this, we use **vectors**, which can be represented as arrows.

Figure 4.4 shows an example of a displacement vector, $\vec{d}$, to represent the motion from P_1 to P_2. A vector has two important features:

Figure 4.4: Example of a displacement vector, $\vec{d}$, from point P_1 to point P_2.
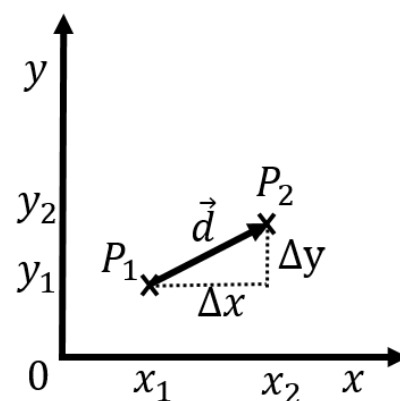
1. a direction (represented by the direction of the arrow)
2. a magnitude (represented by the length of the arrow)

We specify that a variable is a vector by drawing a little arrow on top of the variable. An important feature is that a vector is not fixed anywhere in space. In Figure 4.4, we drew the vector between points P_1 and P_2 for illustration. The vector $\vec{d}$ only represents a change in position in a certain direction and of a certain distance and is not tied to the specific points P_1 and P_2. One could have the *same* displacement vector, $\vec{d}$, between two other points, P_3 and P_4, if the displacement is in the same direction and over the same distance.

All we need to specify a vector in two dimension are two **components**. In Cartesian coordinates, we can specify the length of the vector projected onto the x and y axes (Δx and Δy in Figure 4.4), its x and y components. In polar coordinates, we could specify the total magnitude of the vector (its length, r) and the angle that it makes with the x axis, θ.

Several, equivalent, notations can be used for specifying a vector in Cartesian coordinates

is:

$$\vec{d} = \begin{pmatrix} d_x \\ d_y \end{pmatrix}$$
$$= d_x \hat{x} + d_y \hat{y}$$
$$= d_x \hat{i} + d_y \hat{j}$$

where d_x and d_y are the x and y components of the vector $\vec{d}$, respectively. In Figure 4.4, we have $d_x = \Delta x$ and $d_y = \Delta y$. $\hat{x}$ and $\hat{y}$ are called "unit vectors" for the x and y axes, respectively. $\hat{i}$ and $\hat{j}$ are equivalent to $\hat{x}$ and $\hat{y}$. Unit vectors are vectors that have a length of 1, and are usually denoted with a hat[1] (ˆ) instead of an arrow. The $\hat{x}$ and $\hat{y}$ are special unit vectors as they are parallel to the x and y axes, respectively. In vector notation, they are written:

$$\hat{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\hat{y} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Any vector can be made to have a unit length, if you divide its components by its magnitude. For example, given a vector with components (d_x, d_y), the corresponding unit vector is:

$$\hat{d} = \frac{\vec{d}}{||\vec{d}||}$$
$$= \frac{\vec{d}}{\sqrt{d_x^2 + d_y^2}}$$
$$= \frac{d_x}{\sqrt{d_x^2 + d_y^2}} \hat{x} + \frac{d_y}{\sqrt{d_x^2 + d_y^2}} \hat{y}$$

where $||\vec{d}||$ is the magnitude of $\vec{d}$ and is given by Pythagoras' theorem. If it is clear that $\vec{d}$ is a vector, then one can also denote its magnitude by d (without the arrow), although this notation should only be used if it is clear that you are referring to the magnitude of a vector, and not some other variable.

Note that in the first line above, we implied that we **can divide a vector by a number**. Dividing (or multiplying) a vector by a number means individually dividing (or multiplying) the components of the vector by that number and gives a new vector. If you multiply a vector by (-1), then you get a new vector of the same magnitude but that points in the

---

[1]It's actually called a "caret" not a hat!

opposite direction. When working with vectors, we use the term "scalar" to refer to a quantity that is not a vector (i.e. a regular number is called a scalar). Thus, we have just examined how we can multiply a vector by a scalar to obtain a vector [2]. Also note that multiplying a vector by a scalar ultimately multiplies the length of the vector by that scalar. Thus, multiplying a vector by 2 makes it twice as long and in the same direction.

Referring to Figure 4.4, we have introduced the displacement vector, $\vec{d}$, from the point P_1 to the point P_2. We can also define "position vectors" corresponding to the points P_1 and P_2. The position vector for a point is the displacement vector corresponding to going from the origin to that point (and can be drawn as an arrow that goes from the origin to the point). The components of the position vector for a point are the same as the coordinates of that point. Often, the letter r is used for position vectors. For P_1 and P_2, the position vectors, $\vec{r}_1$ and $\vec{r}_2$, respectively, are given by:

$$\vec{r}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

$$\vec{r}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

The displacement vector, $\vec{d}$, from P_1 to P_2, has the following components:

$$\vec{d} = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix}$$

This suggests that **we can define addition (and subtraction) of two vectors**:

Figure 4.5: Example of adding the displacement vector $\vec{d}$ to the position vector $\vec{r}_1$ to obtain the position vector $\vec{r}_2 = \vec{r}_1 + \vec{d}$.

$$\vec{d} = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} - \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

$$\vec{d} = \vec{r}_2 - \vec{r}_1$$
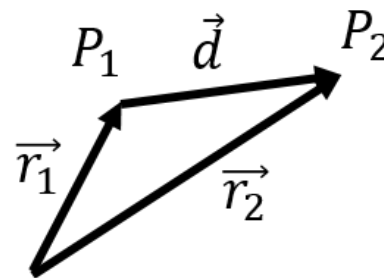
$$\therefore \vec{r}_2 = \vec{r}_1 + \vec{d}$$

where adding (or subtracting) two vectors means to add (or subtract) each component to obtain the corresponding components of the new vector. In the last line, we re-arranged the equation to show that the position vector $\vec{r}_2$ is the sum of $\vec{r}_1$ and $\vec{d}$. This highlights a geometrical aspect of vector addition, shown in Figure 4.5. When you want to add vectors, simply draw the two vectors "head to tail" to get the sum vector (from the tail of the first vector to the head of the last vector).

---

[2]Note that this is **not** called a scalar product.

Example 4-1: 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:

$$\begin{aligned}
\vec{c} &= 2\vec{a} - 3\vec{b} \\
&= 2(2\hat{x} + 3\hat{y}) - 3(5\hat{x} - 2\hat{y}) \\
&= (4\hat{x} + 6\hat{y}) - (15\hat{x} - 6\hat{y}) \\
&= (4 - 15)\hat{x} + (6 + 6)\hat{y} \\
&= -9\hat{x} + 12\hat{y}
\end{aligned}$$

Checkpoint 4-2: 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

4.1.3   Using vectors to describe motion in two dimensions

We now have all of the required elements 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. Suppose that in a period of time Δt, the object goes from position P_1 to position P_2, with position vectors $\vec{r}_1$ and $\vec{r}_2$, respectively. We can describe the object at any time, t, using its position vector, $\vec{r}(t)$, which is a function of time:

$$\vec{r}(t = t_0) = \vec{r}_1$$
$$\vec{r}(t = t_0 + \Delta t) = \vec{r}_2$$

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

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

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

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

The average velocity vector will have the same direction as $\Delta\vec{r}$, since it the displacement vector divided by a scalar (Δt). The velocity vector thus points in the same direction as the displacement. 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 when from point $P_1(x_1, y_1)$ to point $P_2(x_2, y_2)$ in an amount of time Δt, the average velocity vector is given by:

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

$$= \frac{1}{\Delta t} \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix}$$

$$= \frac{1}{\Delta t} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\Delta x}{\Delta t} \\ \frac{\Delta y}{\Delta t} \end{pmatrix}$$

$$= \begin{pmatrix} v_x \\ v_y \end{pmatrix}$$

$$\therefore \vec{v} = v_x \hat{x} + v_y \hat{y}$$

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 4-3: A llama runs in a field from a position $(x_1, y_1) = (2\,\text{m}, 5\,\text{m})$ to a position $(x_2, y_2) = (6\,\text{m}, 8\,\text{m})$ in a time $\Delta t = 0.5\,\text{s}$, as measured by Marcel, a llama farmer standing at the origin of the Cartesian coordinate system. What is the average speed of the llama?

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

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

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

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

$$\boxed{\vec{v}(t) = \frac{d}{dt}\vec{r}(t)} \tag{4.3}$$

$$= \frac{d}{dt}\begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix}$$

$$= \begin{pmatrix} v_x(t) \\ v_y(t) \end{pmatrix}$$

$$\therefore \vec{v}(t) = v_x(t)\hat{x} + v_y(t)\hat{y}$$

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.

Similarly the acceleration vector is given by:

$$\boxed{\vec{a}(t) = \frac{d}{dt}\vec{v}(t)} \tag{4.4}$$

$$= \begin{pmatrix} \frac{dv_x}{dt} \\ \frac{dv_y}{dt} \end{pmatrix}$$

$$= \begin{pmatrix} a_x(t) \\ a_y(t) \end{pmatrix}$$

$$\therefore \vec{a}(t) = a_x(t)\hat{x} + a_y(t)\hat{y}$$

For example, if an object, at time $t = 0$ had a position vector $\vec{r}_0 = (x_0, y_0)$, a velocity vector $\vec{v}_0 = v_{0x}\hat{x} + v_{0y}\hat{y}$, and has a constant acceleration vector, $\vec{a} = a_x\hat{x} + a_y\hat{y}$, then the velocity vector as a function of time, $\vec{v}(t)$, is given by:

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

Or in vector notation:

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

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

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

where you will recognize the first equation from chapter 3. The position vector is given by:

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

with components:

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

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

Example 4-2: An object starts at the origin of a coordinate system at time $t = 0\,\text{s}$, with an initial velocity vector $\vec{v}_0 = (10\,\text{m/s})\hat{x} + (15\,\text{m/s})\hat{y}$. The acceleration in the x direction is $0\,\text{m/s}^2$ and the acceleration in the y direction is $-10\,\text{m/s}^2$.

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

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:

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

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

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

$$= (0\,\mathrm{m}) + (15\,\mathrm{m/s})t + \frac{1}{2}(-10\,\mathrm{m/s^2})t^2$$

$$= (15\,\mathrm{m/s})t - \frac{1}{2}(10\,\mathrm{m/s^2})t^2$$

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

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

$$= \begin{pmatrix} (10\,\mathrm{m/s})t \\ (15\,\mathrm{m/s})t - \frac{1}{2}(10\,\mathrm{m/s^2})t^2 \end{pmatrix}$$

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

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

$$= \begin{pmatrix} (100\,\mathrm{m}) \\ (-350\,\mathrm{m}) \end{pmatrix}$$

c) *We can plot the trajectory using python:*

Python Code 4.1: Trajectory in xy plane

```python
#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()
```
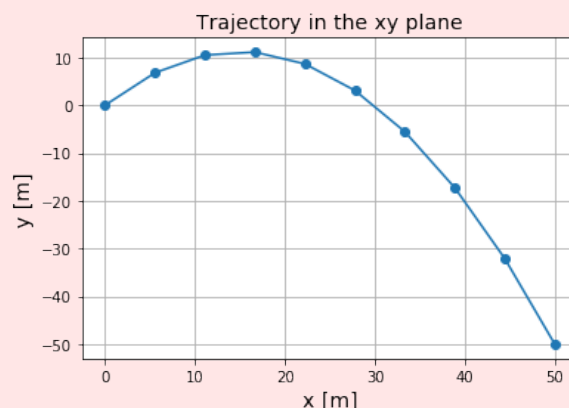
Output 4.1:

Figure 4.6: 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 4-4: In example 4-2, what is the velocity vector exactly at the top of the parabola in Figure 4.6?

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

4.1.4   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. In addition, recall that a vector has two key properties: a length and a direction. 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 there is an acceleration. 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.

Figure 4.7 shows an illustration of the velocity vector, $\vec{v}$, at two different times, $\vec{v}_1$ and $\vec{v}_2$, as well as the vector difference, $\Delta\vec{v} = \vec{v}_2 - \vec{v}_1$, between the two. In this case, the length of the velocity vector did not change with time ($||\vec{v}_1|| = ||\vec{v}_2||$). The acceleration vector is given by:

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

and will thus have a direction parallel to Δv, 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*.

Figure 4.7: Illustration of how the direction of the velocity vector can change when speed is constant.

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

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

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

$$\vec{v}(t) = v(t)\hat{v}(t)$$
$$= v_x(t)\hat{x} + v_y(t)\hat{y}$$

When we take the derivative of $\vec{v}(t)$, we thus have to take the derivative of a product of functions of time, $v(t)\hat{v}(t)$, using the formulas in Table A.2. The acceleration vector is thus
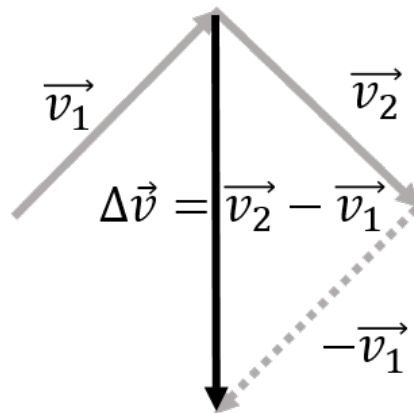
given by:

$$\vec{a} = \frac{d}{dt}\vec{v}(t) = \frac{d}{dt}v(t)\hat{v}(t)$$

$$\boxed{\vec{a} = \frac{dv}{dt}\hat{v}(t) + v(t)\frac{d\hat{v}}{dt}} \tag{4.5}$$

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.

In general, 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:

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

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:

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

$$= 0 + v\frac{d}{dt}\hat{v}(t)$$

$$= v\frac{d}{dt}\left(\frac{v_x(t)}{v}\hat{x} + \frac{v_y(t)}{v}\hat{y}\right)$$

$$= \frac{dv_x}{dt}\hat{x} + \frac{d}{dt}\sqrt{v^2 - v_x(t)^2}\hat{y}$$

$$= \frac{dv_x}{dt}\hat{x} + \frac{1}{2\sqrt{v^2 - v_x(t)^2}}(-2v_x(t))\frac{dv_x}{dt}\hat{y}$$

$$= \frac{dv_x}{dt}\hat{x} - \frac{v_x(t)}{\sqrt{v^2 - v_x(t)^2}}\frac{dv_x}{dt}\hat{y}$$

$$= \frac{dv_x}{dt}\hat{x} - \frac{v_x(t)}{v_y(t)}\frac{dv_x}{dt}\hat{y}$$

$$\boxed{\vec{a} = \frac{dv_x}{dt}\left(\hat{x} - \frac{v_x(t)}{v_y(t)}\hat{y}\right)} \tag{4.6}$$
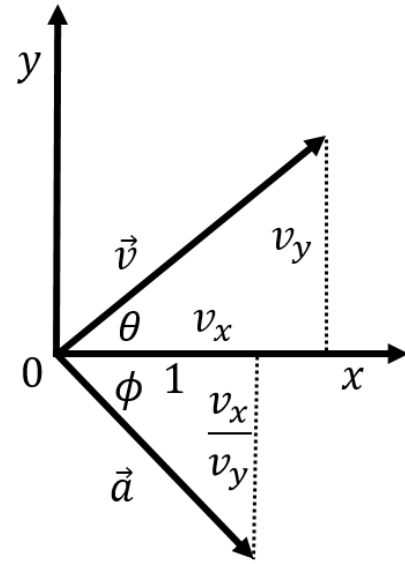

Figure 4.8: Illustration that the acceleration vector is perpendicular to the velocity vector if speed is constant.

The resulting acceleration vector is illustrated in Figure 4.8 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:

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

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 $\vec{a}_\perp(t)$, which is perpendicular to the velocity vector and is only responsible for changing the direction of the motion.

Checkpoint 4-5: 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.

4.1.5 Relative motion

In the previous chapter, we examined how to convert the description of motion from one reference frame to another. Effectively, we found that we only needed to add the velocity of one frame with respect to the other in order to convert between reference frames. We saw that if the speeds are small compared to the speed of light, the addition of the velocity

was trivial, but that it was a little more complicated when the frames of reference or the objects are moving with speeds close to the speed of light. We will start by assuming that speeds are small and that we can use Galilean relativity here.

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 4.9. 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 $\vec{r}^A(t)$.
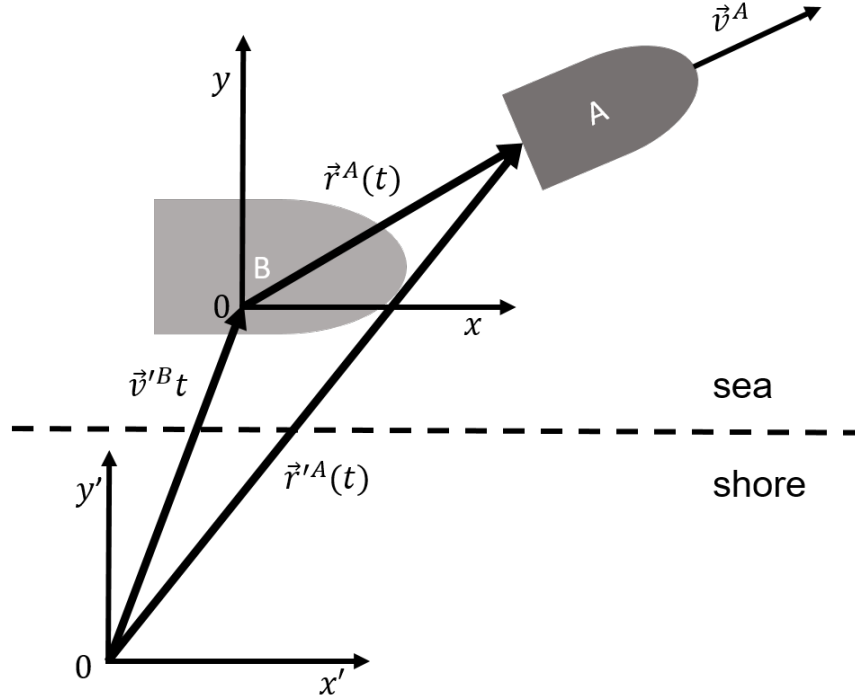
Figure 4.9:   Example of converting from one reference frame to another in two dimensions using vector addition.

Writing this out by coordinate, we have:

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

and for the velocities:

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

It turns out that extending this to Special Relativity is "trivial", if one specifically chooses the x and x' directions to be in the same direction as v'^B. In this case, only the x coordinates are affected by the Lorentz transformation:

$$x'^A = \frac{1}{\sqrt{1 - \left(\frac{v'^B}{c}\right)^2}}(v'^B t + x^A)$$

$$t' = \frac{1}{\sqrt{1 - \left(\frac{v'^B}{c}\right)^2}}\left(t + \frac{v'^B x^A}{c^2}\right)$$

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

where of course, one also has to transform the time "coordinate".

Checkpoint 4-6: 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.

4.2.1 3D Coordinate systems

In a Cartesian coordinate system, we use the variables, x, y, and z, and simply add a third axis, z, that is perpendicular to both x and y. Two additional coordinate systems are common in three dimensions, namely "cylindrical" and "spherical coordinates". All three systems are illustrated in Figure 4.10 superimposed onto the Cartesian system.
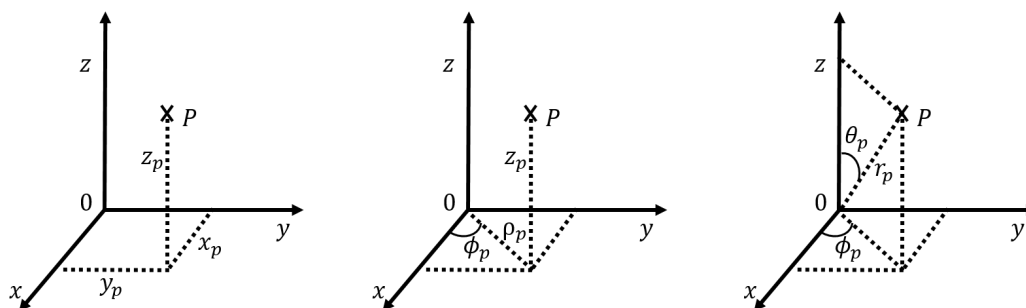

Figure 4.10: Cartesian (left), cylindrical (centre) and spherical (right) coordinate systems used in three dimensions. The y and z axes are in the plane of the page, whereas the x axis comes out of the page.

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 a cylinder.

The cylindrical coordinates are related to the Cartesian coordinates by:

$$\rho = \sqrt{x^2 + y^2}$$
$$\tan(\phi) = \frac{y}{x}$$
$$z = z$$

In spherical coordinates, a point P is described by the radius, r, the distance from the point to the origin, the *polar angle* θ between the position vector $\vec{r}$ and the z-axis, and the *azimuthal angle*, ϕ, defined in the same way as in cylindrical 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:

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

4.2.2   Describing motion in three dimensions

Using Cartesian coordinates, it is straightforward to extend our description of motion from two dimensions to three. 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 $\vec{r}(t)$:

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

Note that sometimes, people use $\hat{i}$, $\hat{j}$, and $\hat{k}$ to denote the unit vectors in the x, y, and z directions. The unit vectors, in three dimensions, are given by:

$$\hat{x} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\hat{y} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$\hat{z} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

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:

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

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 4.11 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, as $x^2(t)(t) + y^2 = R^2$ for motion on the circle. Instead, we could think of an axis that is bent around the circle (as shown by the curved arrow, the d axis). The d axis is such that $d = 0$ where the circle intersects the x axis, and the value of d increases as we move counter-clockwise along the circle. Instead of using d, we can also use the angle, θ, between the position of the object on the circle and the x axis.
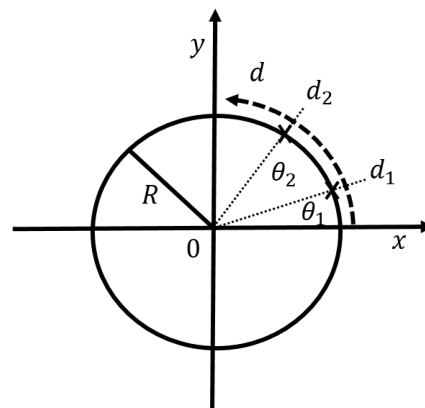

Figure 4.11:   Describing the motion of an object around a circle of radius R.

If we express the angle θ in radians, then it easy to convert between d 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:

$$\boxed{\theta = \frac{d}{R}} \tag{4.7}$$

In particular, if the object has gone around the whole circle, then $d = 2\pi R$ (the circumference of a circle), and the corresponding angle is, $\theta = \frac{2\pi R}{r} = 2\pi$, namely 360°.
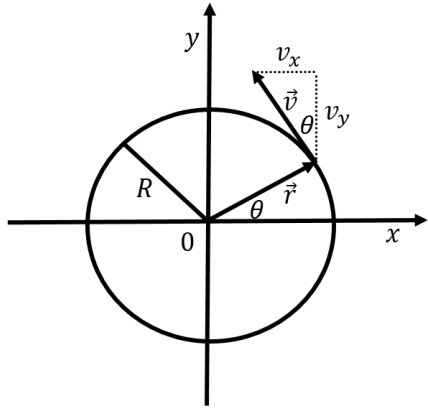
By using the angle, θ, instead of x and y, we are effectively using polar coordinates, with a fixed radius. As we already saw, the x and y positions are related to θ by:

$$x(t) = R\cos(\theta(t))$$
$$y(t) = R\sin(\theta(t))$$

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

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

and the velocity vector is thus given by:

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

$$= R \begin{pmatrix} \frac{d}{dt}\cos(\theta(t)) \\ \frac{d}{dt}\sin(\theta(t)) \end{pmatrix}$$

$$= R \begin{pmatrix} -\sin(\theta(t))\frac{d\theta}{dt} \\ \cos(\theta(t))\frac{d\theta}{dt} \end{pmatrix}$$

Figure 4.12: The position vector, $\vec{r}(t)$ is always perpendicular to the velocity vector, $\vec{v}(t)$, for motion on a circle.

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:

$$||\vec{v}|| = \sqrt{v_x^2 + v_y^2}$$

$$= \sqrt{\left(-R\sin(\theta(t))\frac{d\theta}{dt}\right)^2 + \left(R\cos(\theta(t))\frac{d\theta}{dt}\right)^2}$$

$$= \sqrt{R^2 \left(\frac{d\theta}{dt}\right)^2 [\sin^2(\theta(t)) + \cos^2(\theta(t))]}$$

$$= R\left|\frac{d\theta}{dt}\right|$$

The position and velocity vectors are illustrated in Figure 4.12 for an angle θ in the first quadrant ($\theta < \frac{\pi}{2}$). 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 4.12, showing that **the velocity vector is tangent to the circle** and perpendicular to the position vector.

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

along the d axis is:

$$v_d(t) = \frac{d}{dt}d(t)$$
$$= \frac{d}{dt}R\theta(t)$$
$$= R\frac{d\theta}{dt}$$

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

If the object has a **constant velocity** along the circle and started at a position along the circumference $d = d_0$, then its position can be described as:

$$d(t) = d_0 + v_d t$$

or, in terms of θ:

$$\theta(t) = \frac{d(t)}{R} = fracd_0 R + \frac{v_d}{R}t$$
$$= \theta_0 + \frac{d\theta}{dt}t$$
$$= \theta_0 + \omega t$$
$$\boxed{\therefore \omega = \frac{d\theta}{dt}}$$

where we introduced θ_0 as the angle corresponding to d_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. The relation between the "linear" velocity v_d and ω is:

$$\boxed{v_d = R\omega}$$

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

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

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

$$a_d(t) = \frac{d}{dt}v_d$$
$$= \frac{d}{dt}\omega R = R\frac{d\omega}{dt}$$
$$\boxed{a_d(t) = R\alpha}$$

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

$$d = R\theta \tag{4.8}$$
$$v_d = R\omega \tag{4.9}$$
$$a_d = R\alpha \tag{4.10}$$

If the object started at $t = 0$ with a position $d = d_0$ ($\theta = \theta_0$), and an initial linear velocity v_{0d} (angular velocity ω_0), and has a **constant linear acceleration** around the circle, a_d (angular acceleration, α), then the position of the object can be described as:

$$d(t) = d_0 + v_{d0}t + \frac{1}{2}a_d t^2$$
$$\theta(t) = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2$$

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

As you recall from section 4.1.4, we can compute the acceleration **vector** and identify components that are parallel and perpendicular to the velocity vector:

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

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 d direction (the linear acceleration). This component of the acceleration is responsible for increasing (or decreasing) the speed of the object.

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 $\hat{v} = -\sin(\theta)\hat{x} + \cos(\theta)\hat{y}$:

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

$$= (\omega R)\frac{d}{dt}\left[-\sin(\theta)\hat{x} + \cos(\theta)\hat{y}\right]$$

$$= \omega R\left[-\frac{d}{dt}\sin(\theta)\hat{x} + \frac{d}{dt}\cos(\theta)\hat{y}\right]$$

$$= \omega R\left[-\cos(\theta)\frac{d\theta}{dt}\hat{x} - \sin(\theta)\frac{d\theta}{dt}\hat{y}\right]$$

$$= \omega R[-\cos(\theta)\omega\hat{x} - \sin(\theta)\omega\hat{y}]$$

$$\boxed{\vec{a}_\perp(t) = \omega^2 R[-\cos(\theta)\hat{x} - \sin(\theta)\hat{y}]} \qquad (4.11)$$

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:

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

where in the last equal sign, we wrote the centripetal acceleration in terms of the speed around the circle.

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 4-7: 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, then we can define the **period and frequency** of the motion.

The period, T, is defined to be the time that it takes to complete one revolution around the circle. If the object has constant angular speed ω, we can find the time, T, that it takes to

complete one full revolution, from $\theta = 0$ to $\theta = 2\pi$:

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

$$\boxed{\therefore T = \frac{2\pi}{\omega}} \tag{4.13}$$

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:

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

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

Think of the period as the time that it takes to complete one revolution.

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 $f = 1\,\text{Hz}$, then the object goes around the circle once per second.

Another common unit for frequency is the rpm, or "rotations per minute". 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 4.13, where two objects can be travelling around two circles of radius R_1 and R_2 with the same angular velocity ω. If they have the same angular velocity, then it will take them the same amount of time to complete a revolution. However, the outer object has to cover a much larger distance (the circumference is larger), and thus has to move with a larger linear speed.
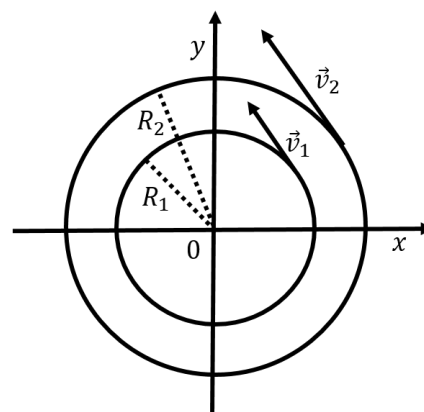

Figure 4.13: For a given angular velocity, the linear velocity will be larger on a larger circle ($v = \omega R$).

Checkpoint 4-8: A motor is rotating at $3000\,\text{rpm}$, what is the corresponding frequency in Hz?

   A) $5\,\text{Hz}$
   B) $50\,\text{Hz}$
   C) $500\,\text{Hz}$

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:

$$f(x) = x^2$$
$$\therefore f(2) = 4$$

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:

$$g(x, y) = x^2 + 2y^2$$
$$\therefore g(2, 3) = 22$$

We can easily visualize a function of 1 variable, for example by plotting it in python:

```python
#import pacakges for creating arrays of values and for plotting
import numpy as np #arrays
import pylab as pl #plotting

#define the function:
def f(x):
    return x*x

#create 100 values of x between -5 and +5
xvals = np.linspace(-5,5,100)

#Plot the function evaluated at the values of x against the values of x:
pl.plot(xvals,f(xvals))
pl.xlabel('x')
pl.ylabel('f(x)')
pl.title('f(x)=$x^2$')
pl.grid()
pl.show()
```
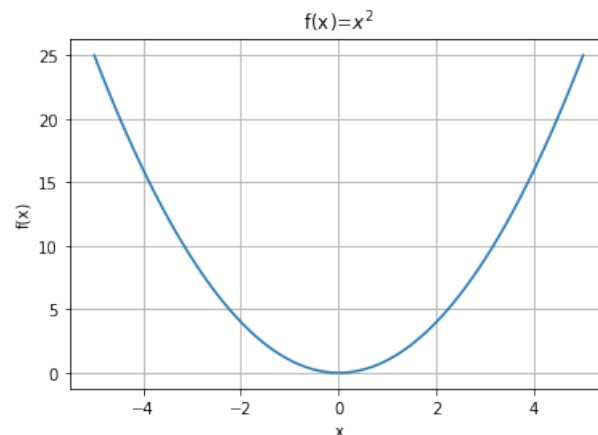
Output A.1:

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

Plotting a function of 2 variables is a little trickier, since we need to do it in three dimensions (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

```python
#import pacakges 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
```

```python
#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()
```
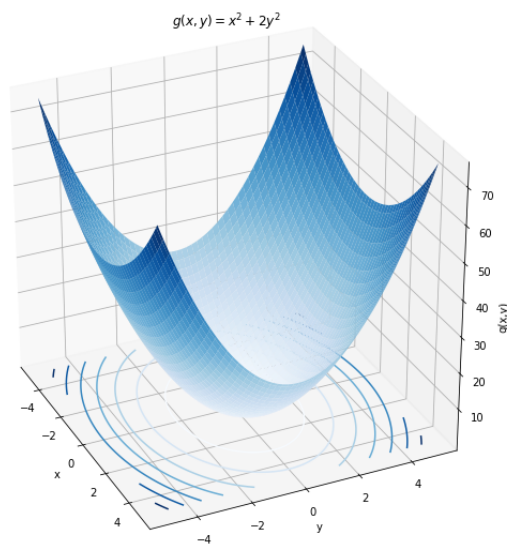
Output A.2:

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 A.1. For any value of x, we can define the slope of the function as the "steepness of the curve". For values of $x > 0$ the function increases as x increases, so we say that the slope is positive. For values of $x < 0$, the function decreases as x increases, so we say that the slope is negative. A synonym for the word slope is "derivative", which is the word that we prefer to use in calculus. The derivative of a function $f(x)$ is given the symbol $\frac{df}{dx}$ to indicate that we are referring to the slope of $f(x)$ when plotted as a function of x.

We need to specify which variable we are taking the derivative with respect to when the function has more than one variable but only one of them should be considered *independent*. For example, the function $f(x) = ax^2 + c$ 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 $\Delta x \to 0$, we get the derivative. In fact, this is the formal definition of the derivative:

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

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

As an example, let us determine the function $f'(x)$ that is the derivative of $f(x) = x^2$. We

start by calculating Δf:

$$\begin{aligned}
\Delta f &= f(x + \Delta x) - f(x) \\
&= (x + \Delta x)^2 - x^2 \\
&= x^2 + 2x\Delta x + \Delta x^2 - x^2 \\
&= 2x\Delta x + \Delta x^2
\end{aligned}$$

We now calculate $\frac{\Delta f}{\Delta x}$:

$$\begin{aligned}
\frac{\Delta f}{\Delta x} &= \frac{2x\Delta x + \Delta x^2}{\Delta x} \\
&= 2x + \Delta x
\end{aligned}$$

and take the limit $\Delta x \to 0$:

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

We have thus found that the function, $f'(x) = 2x$, is the derivative of the function $f(x) = x^2$. This is illustrated in Figure 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.
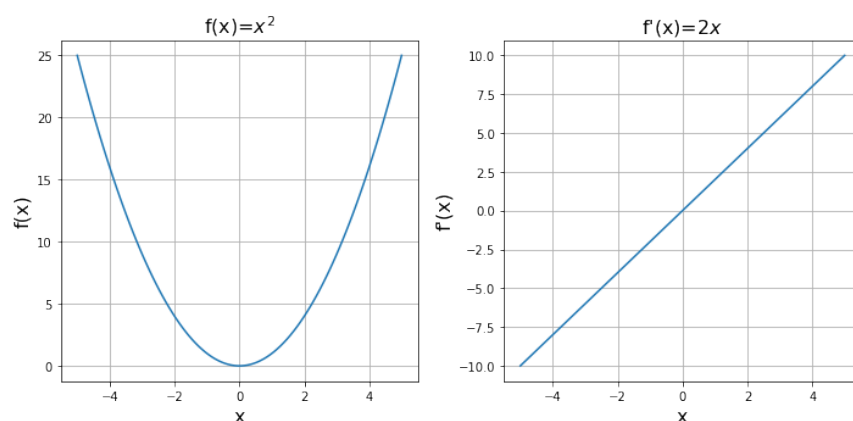

Figure A.3: $f(x) = x^2$ and its derivative, $f'(x) = 2x$ plotted for x between -5 and +5.

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:*

$$h(x) = \frac{f(x)}{g(x)}$$
$$f(x) = \sin(x)$$
$$g(x) = \cos(x)$$

Using the fourth row in Table A.2, and the common derivatives from Table A.1, we have:

$$f'(x) = \cos(x)$$
$$g'(x) = -\sin(x)$$
$$g^2(x) = \cos^2(x)$$
$$h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$$
$$= \frac{\cos(x)\cos(x) - \sin(x)(-\sin(x))}{\cos^2}$$
$$= \frac{\cos^2(x) + \sin^2(x)}{\cos^2}$$
$$= \frac{1}{\cos^2(x)}$$

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:

$$h(x) = \sin^2(x) = f(g(x))$$
$$f(x) = x^2$$
$$g(x) = \sin(x)$$

Using the common derivatives from Table A.1, we have:

$$f'(x) = 2x$$
$$g'(x) = \cos(x)$$

Applying the Chain Rule, we have:

$$h'(x) = f'(g(x))g'(x)$$
$$= 2\sin(x)g'(x)$$
$$= 2\sin(x)\cos(x)$$

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 processed 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.
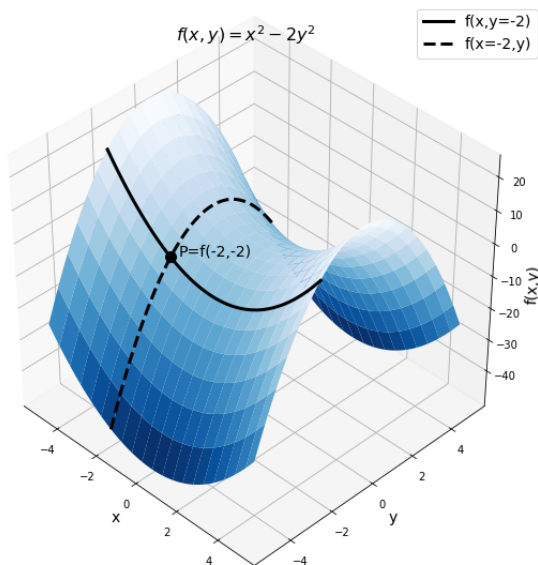

Figure A.4: $f(x, y) = x^2 - 2y^2$ plotted for x between -5 and +5 and for y between -5 and +5. The point P labelled on the figure shows the value of the function at $f(-2, -2)$. The two lines show the function evaluated when one of x or y is held constant.

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:

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

Calculating the partial derivative is very easy, as we just treat all variables as constants except for the variable with respect to which we are differentiating[1]. For the function $f(x, y) = x^2 - 2y^2$, we have:

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

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 are b constants and can be thought of as numbers that we do not know.

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

Since the partial derivatives tell us how the function changes in a particular direction, we can use them to find the direction in which the function changes *the most rapidly*. For example, suppose that the surface from Figure 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

---

[1]To 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 $y = -2$:

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

Evaluating the gradient at $x = -2$ and $y = -2$:

$$\nabla f(x,y) = 2x\hat{x} - 4y\hat{y}$$
$$= -4\hat{x} + 8\hat{y}$$
$$= 4(-\hat{x} + 2\hat{y})$$

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 function 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 how the function $f(x)$ varies as a function of x. In physics, we often know how a function varies, but we do not know the actual function. In other words, we often have the opposite problem: we are given the derivative of a function, and wish to determine the actual function. For this case, we will limit our discussion to functions of a single independent variable.

Suppose that we are given a function $f(x)$ and we know that this is the derivative of some other function, $F(x)$, which we do not know. We call $F(x)$ the **anti-derivative** of $f(x)$. The anti-derivative of a function $f(x)$, written $F(x)$, thus satisfies the property:

$$\frac{dF}{dx} = f(x)$$

Since we have a symbol for indicating that we take the derivative with respect to x ($\frac{d}{dx}$), we also have a symbol, $\int dx$, for indicating that we take the anti-derivative with respect to x:

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

Earlier, we justified the symbol for the derivative by pointing out that it is like $\frac{\Delta f}{\Delta x}$ but for the case when $\Delta x \to 0$. Similarly, we will justify the anti-derivative sign, $\int f(x)dx$, by showing that it is related to a sum with $f(x)\Delta x$, in the limit $\Delta x \to 0$. The $\int$ sign looks like an "S" for sum.

While it is possible to exactly determine the derivative of a function $f(x)$, the anti-derivative can only be determined up to a constant. Consider for example a different function, $\tilde{F}(x) = F(x) + C$, where C is a constant. The derivative of $\tilde{F}(x)$ is given by:

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

Hence, the function $\tilde{F}(x) = F(x) + C$ is also an anti-derivative of $f(x)$. The constant C can often be determined using additional information (sometimes called "initial conditions"). Recall the function, $f(x) = x^2$, shown in Figure 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 $f(x) = \frac{dF}{dx}$
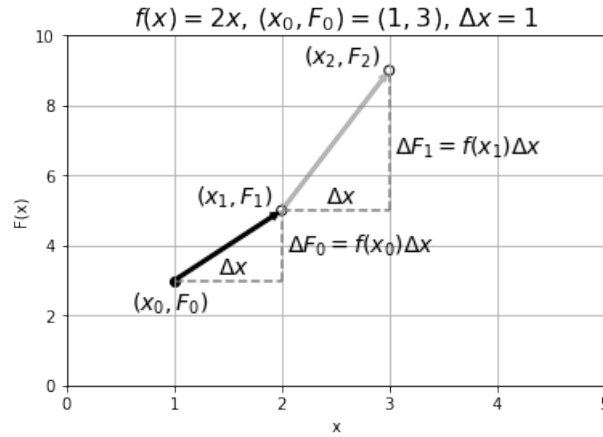2. the condition that $F(x)$ must pass through a specific point, $F(x_0) = F_0$.

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:

$$\frac{\Delta F_0}{\Delta x} \approx f(x_0) \quad (\text{in the limit} \Delta x \to 0)$$
$$\therefore \Delta F_0 = f(x_0)\Delta x$$

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

$$F_1 = F(x_1)$$
$$= F(x + \Delta x)$$
$$\approx F_0 + \Delta F_0$$
$$\approx F_0 + f(x_0)\Delta x$$

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 $\Delta x \to 0$:

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

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).
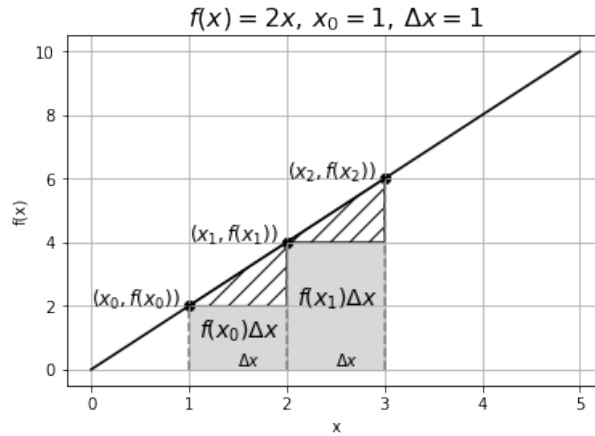

Figure A.6:   The function $f(x) = 2x$ and illustration of the terms $f(x_0)\Delta x$ and $f(x_1)\Delta x$ as the area between the curve $f(x)$ and the x axis when $\Delta x \to 0$.

As you can see, each term in the sum corresponds to the area of a rectangle between the function $f(x)$ and the x axis (with a piece missing). In the limit where $\Delta x \to 0$, the missing pieces (shown by the hashed areas in Figure 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:

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

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

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

We see that if we determine the area between $f(x)$ and the x axis from x_0 to x_N, we can obtain the difference between the anti-derivative at two points, $F(x_N) - F(x_0)$

The difference between the anti-derivative, $F(x)$, evaluated at two different values of x is called the **integral** of $f(x)$ and has the following notation:

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

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:

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

Recall that we wrote the anti-derivative with the same $\int$ symbol earlier:

$$\int f(x)dx = F(x)$$

The symbol $\int f(x)dx$ without the limits is called the **indefinite integral**. You can also see that when you take the (definite) integral (i.e. the difference between $F(x)$ evaluated at two points), any constant that is added to $F(x)$ will cancel. Physical quantities are always based on definite integrals, so when we write the constant C it is primarily for completeness and to emphasize that we have an indefinite integral.

As an example, let us determine the integral of $f(x) = 2x$ between $x = 1$ and $x = 4$, as well as the indefinite integral of $f(x)$, which is the case that we illustrated in Figures A.5 and A.6. Using equation A.3, we have:

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

$$= \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} 2x_{i-1}\Delta x$$

where we have:

$$x_0 = 1$$
$$x_N = 4$$
$$\Delta x = \frac{x_N - x_0}{N}$$

Note that N is the number of times we have Δx in the interval between x_0 and x_N. Thus, taking the limit of $\Delta x \to 0$ is the same as taking the limit $N \to \infty$. Let us illustrate the sum for the case where $N = 3$, and thus when $\Delta x = 1$, corresponding to the illustration in Figure A.6:

$$\sum_{i=1}^{i=N=3} 2x_{i-1}\Delta x = 2x_0\Delta x + 2x_1\Delta x + 2x_2\Delta x$$

$$= 2\Delta x(x_0 + x_1 + x_2)$$

$$= 2\frac{x_3 - x_0}{N}(x_0 + x_1 + x_2)$$

$$= 2\frac{(4) - (1)}{(3)}(1 + 2 + 3)$$

$$= 12$$

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:

$$\sum_{i=1}^{i=6} 2x_{i-1}\Delta x = 2\frac{x_6 - x_0}{N}(x_0 + x_1 + x_2 + x_3 + x_4 + x_5)$$

$$= 2\frac{4 - 1}{6}(1 + 1.5 + 2 + 2.5 + 3 + 3.5)$$

$$= 13.5$$

Writing this out again for the general case so that we can take the limit $N \to \infty$, and factoring out the $2\Delta x$:

$$\sum_{i=1}^{i=N} 2x_{i-1}\Delta x = 2\Delta x \sum_{i=1}^{i=N} x_{i-1}$$

$$= 2\frac{x_N - x_0}{N} \sum_{i=1}^{i=N} x_{i-1}$$

Now, consider the combination:

$$\frac{1}{N} \sum_{i=1}^{i=N} x_{i-1}$$

that appears above. This corresponds to the arithmetic average of the values from x_0 to x_{N-1} (sum the values and divide by the number of values). In the limit where $N \to \infty$,

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

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

Putting everything together:

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

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:

$$F(x) = x^2 + C$$
$$\int 2x\,dx = x^2 + C$$

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 show 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 A.7. If the rod is uniform in density, then if we cut it into, say, two equal pieces, those two pieces will

weigh the same. We can define a "linear mass density", μ, for the rod, as:

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

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 $10\,\text{cm}$ 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.
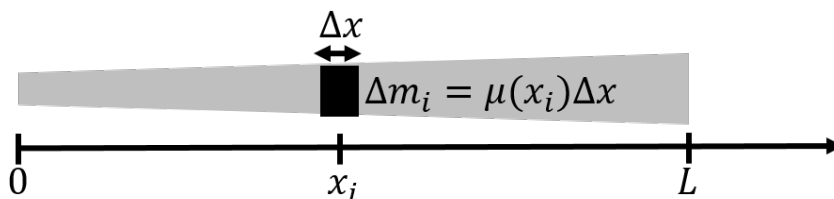

Figure A.7: A rod with a varying linear density. To calculate the mass of the rod, we consider a small mass element Δm_i of length Δx at position x_i. The total mass of the rod is found by summing the mass elements.

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 $\Delta x \to 0$:

$$M = \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} \Delta m_i$$

$$= \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} \mu(x_i)\Delta x$$

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

$$M = \lim_{\Delta x \to 0} \sum_{i=1}^{i=N} \mu(x_i)\Delta x$$

$$= \int_0^L \mu(x)dx$$

$$= G(L) - G(0)$$

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:

$$M = \int_0^L \mu(x)dx$$

$$= \int_0^L ax^3 dx$$

$$= G(L) - G(0)$$

$$= \left[a\frac{1}{4}L^4\right] - \left[a\frac{1}{4}0^4\right]$$

$$= 5\,\text{kg/m}^4 \frac{1}{4}(0.5\,\text{m})^4$$

$$= 78\,\text{g}$$

With a little practice, you can solve this type of problem without writing out the sum explicitly. Picture an *infinitesimal* piece of the rod of length dx at position x. It will have an *infinitesimal* mass, dm, given by:

$$dm = \mu(x)dx$$

The total mass of the rod is the then the sum (i.e. the integral) of the mass *elements*

$$M = \int dm$$

and we really can think of the $\int$ sign as a sum, when the things being summed are *infinitesimally* small. In the above equation, we still have not specified the range in x over which we want to take the sum; that is, we need some sort of index for the mass elements to make this a meaningful definite integral. Since we already know how to express dm in terms of dx, we can substitute our expression for dm using one with dx:

$$M = \int dm = \int_0^L \mu(x)dx$$

where we have made the integral definite by specifying the range over which to sum, since we can use x to "label" the mass elements.

One should note that coming up with the above integral is physics. Solving it is math. We will worry much more about writing out the integral than evaluating its value. Evaluating the integral can always be done by a mathematician friend or a computer, but determining which integral to write down is the physicist's job!