# Calculus
# for
# Business
# and
# Social Sciences

Joseph C Foster
University of South Carolina
Spring 18

Last updated: March 29, 2018

# Contents

# Introduction

One day I will write an introduction, and it will go here.

# Chapter 1

# The Basics

**Learning Objectives**

After studying this chapter you should:

- be familiar with the different ways to represent a function
- understand the terms domain and range for a function
- be familiar with basic types of functions
- be able to create and use basic functions for modelling the real world
- understand what a composite function is
- know what the cost, revenue and profit functions are and what they represent
- know what the supply and demand curves are and what they represent
- understand and be able to calculate the average rate of change and relative change of a function

## 1.1 Functions

**Definition 1.1.1**

A **function** is a rule that takes certain values as inputs and assigns to each a definite output number. The set of all input numbers is called the **domain** of the function and the set of resulting output numbers is called the **range** of the function.

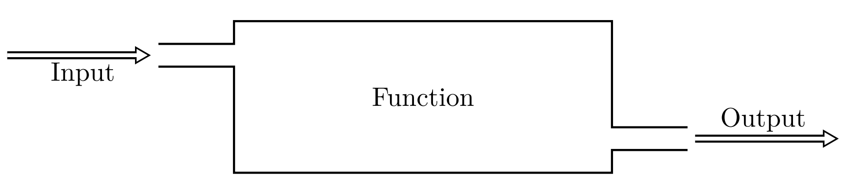
**Example 1.1.2**

The function $f$ which squares the number input and then adds 3 to the result can be represented by

$$f(x) = x^2 + 3$$

so that $f(1) = 1^2 + 3 = 1 + 3 = 4$ and $f(-4) = (-4)^2 + 3 = 16 + 3 = 19$.

Simplify put, a function is like a machine into which numbers are fed, and for each value input, the machine determines the *single* output value.




The input is called the *independent variable* and the output is called the *dependent variable*. For example, if you think of a cyclist, then time is the independent variable and distance travelled is the dependent variable.

We expressed the function in Example 1.1.2 as a *formula*. But we can just of easily expressed it differently. Almost any way of communicating the rule that the function follows is a way of representing a function. The most common ways we will see are; words, formulas, tables and graphs.

**Example 1.1.3**

The function $f$ from Example 1.1.2 can be represented in any of the following ways;

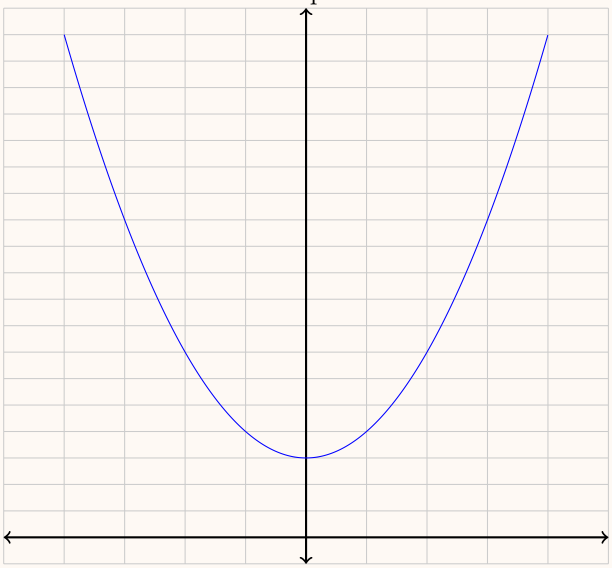
Words:
"$x$ squared plus 3"

Formula:
$f(x) = x^2 + 3$

Table:

| $x$ | $f(x)$ |
|---|---|
| $-4$ | 19 |
| $-3$ | 12 |
| $-2$ | 7 |
| $-1$ | 4 |
| 0 | 3 |
| 1 | 4 |
| 2 | 7 |
| 3 | 12 |
| 4 | 19 |

Graph:



Later on the course we will be focussing on using formulas to represent functions, but it important to be able to be able to read functions, more specifically data, that do not have formulas. For example there are no nice formulas for the following two examples.

**Example 1.1.4**

Below is a graph depicting the exchange rate between pounds (GBP) and dollars (USD). Clearly there is no neat formula relating the two, but the graph can be read much the same.


Taken from Yahoo Finance 3/1/2017.

**Example 1.1.5**

Below is a table depicting the average monthly precipitation from a town in Leicestershire, England. Again, no formula exists but the data is clearly communicated.

| Month | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Average Precipitation | 61.65 | 48.91 | 51.86 | 43.86 | 50.83 | 63.07 | 46.08 | 59.25 | 61.50 | 60.58 | 60.34 | 68.82 |

Taken from Wikipedia 3/1/2017

**Remark**

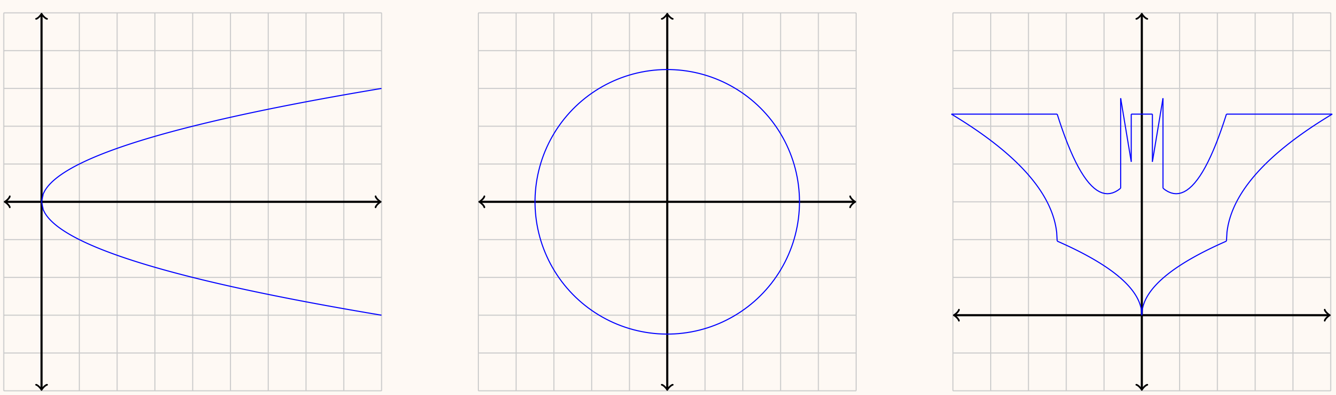
Remember, functions are like machines. Much like a machine, a function will only ever do what it is programmed to do, so if you input 3 into a function, its corresponding output, $f(3)$, will be its *only* output. Otherwise we would say the function is not *well-defined*.
Recall from your precalculus class this is the **vertical line test**. If you draw a vertical line on a graph and it intersects the graph in more than one place, the graph does not represent a function.

**Example 1.1.6**

You can verify the following using the vertical line test.

Functions:

Not Functions:



**Definition 1.1.7**

A function $f$ is **increasing** if the values of $f(x)$ increase as $x$ increases.
A function $f$ is **decreasing** if the values of $f(x)$ decrease as $x$ increases.

The graph of an *increasing* function *climbs* as we move from left to right.
The graph of a *decreasing* function *descends* as we move from left to right.

Increasing and decreasing are two important characteristics of a function. In the real world, businessmen are interested in when profits are increasing and decreasing so that they can make good decisions. It is very easy to see, from a graph, when functions are increasing and decreasing, that's why pictures like in Example 1.1.4 appear all the time in your typical image of someone working in business. But this is still a maths class so let's look at other ways to communicate the increasing and decreasing behaviour of functions.

**Definition 1.1.8**

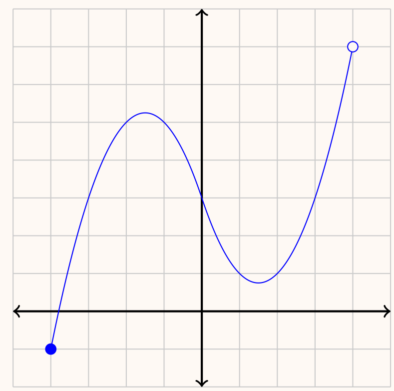
The set of numbers $t$ such that $a \le t \le b$ is called a **closed interval** and is written as $[a, b]$.
The set of numbers $t$ such that $a < t < b$ is called an **open interval** and is written as $(a, b)$.

In the previous definition we can also mix up the brackets and parenthesis. For example $(2, 5]$ would denote the interval $2 < t \le 5$. It might be easy to remember it as ( and ) leave out the endpoint, whereas [ and ] include the endpoint.

**Example 1.1.9**

Given the graph of a function $f$ below, we can see that;

- The domain of $f$ is $[-4, 4)$
- The range of $f$ is $[-1, 7)$
- $f$ increases on the interval $(-4, -1.5)$
- $f$ decreases on the interval $(-1.5, 1.5)$
- $f$ increases on the interval $(1.5, 4)$



Here we use the convention that a filled in circle on the graph *includes* that point and an empty circle *excludes* that point. The concept of increasing and decreasing functions is something we will explore in detail once we have studied the *derivative*.

**Remark**

Notice that in Example 1.1.9 our intervals on which the function increases and decreases are all open. I bet some of you are tempted to include the endpoints on these intervals, but we will see later on why we don't want to.

## 1.2 Linear Functions

The simplest functions one can consider are *linear* functions. You will have studied these a lot in your college algebra or precalculus class. They are a certain type of *polynomial* functions, which we will study a lot later in the course.

**Definition 1.2.1**

A **polynomial** (function) is an expression that involves only the operations of addition, subtraction, multiplication and non-negative integer exponents of a variable (most commonly, $x$).
The **degree** of a polynomial is the highest exponent that appears in the expression.

**Example 1.2.2**

Some examples of polynomials are;

| | | | | |
|---|---|---|---|---|
| $f(x) = 5$ | $f(x) = 2x - 3$ | $f(x) = x^2 + 7x - 5$ | $f(x) = 2x^3 - 11x + 1$ | $f(x) = x^{456}$ |
| degree$= 0$ | degree$= 1$ | degree$= 2$ | degree$= 3$ | degree$= 456$ |

It is important to remember that the domain of *every* polynomial function is the entire real line, denoted by $(-\infty, \infty)$ or $\mathbb{R}$ but the range of polynomial functions depend on, amongst other things, their degree.

**Definition 1.2.3**

A **linear function** is a polynomial function of degree one or zero.

The generic form in which we will denote a linear function is its *slope-intercept* form, $f(x) = mx + c$, where $m$ is the *slope* of the function and $c$ is the *y-intercept*.

**Example 1.2.4**

Linear:

$f(x) = 3x + 5$ $\quad$ $f(x) = -2x + 7$ $\quad$ $f(x) = 13$ $\quad$ $f(x) = \frac{3}{2}x - 1$ $\quad$ $f(x) = -3.2x + 4$

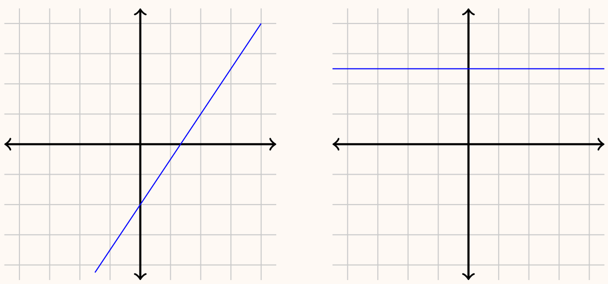
Not Linear:

$f(x) = x^2 + x + 1$ $\quad$ $f(x) = x^3 - x + 3$ $\quad$ $f(x) = -x^2 + 5x - 2$ $\quad$ $f(x) = 3^x$ $\quad$ $f(x) = \dfrac{x+1}{x^2+3}$

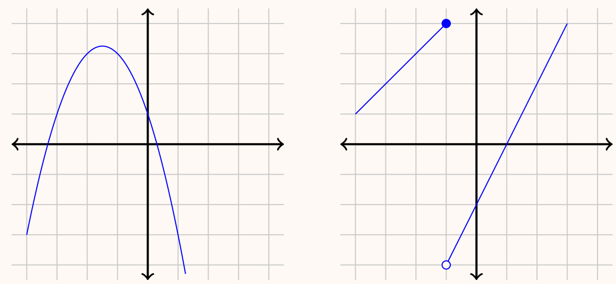
Graphically, a linear function consists of a single straight line.

**Example 1.2.5**

Linear:



Not Linear:



One important and useful property of linear functions is that they are completely determined by two points. That is, if you know the value of a linear function at two different input values, you can construct a formula for that function. Remember the graphs of linear functions are straight lines, and to draw a straight line you only need to connect two points.

Given two points $(x_1, f(x_1))$ and $(x_2, f(x_2))$, the slope of the line connecting them is given by

$$\text{slope } = \frac{\text{rise}}{\text{run}} = \frac{\text{difference in } y}{\text{difference in } x} = \frac{f(x_2) - f(x_1)}{x_2 - x_1} = m.$$

Once you have a value for $m$, you can calculate the $y$-intercept quite easily using the *point-slope* form of a line.

$$f(x) - f(x_1) = m\left(x - x_1\right).$$

So by rearranging we obtain

$$f(x) = m(x - x_1) + f(x_1) \implies f(x) = mx - mx_1 + f(x_1)$$

so in the slope-intercept from $c = -mx_1 + f(x_1)$. Don't worry if all this algebra looked scary with all the subscripts and what not. Let's look at an example.

**Example 1.2.6**

Below is data corresponding to the amount of money (in dollars) in a bank account after $t$ months. We wish to construct a linear model $A(t)$ to represent this amount.

| $t$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Amount | 200 | 220 | 240 | 260 | 280 | 300 | 320 | 340 | 360 | 380 | 400 | 420 |

We start by taking two points from this data, say $(3, 260)$ and $(5, 300)$. Then the slope, $m$, is given by

$$m = \frac{300 - 260}{5 - 3} = \frac{40}{2} = 20.$$

Putting this in the point-slope form of a line with the point $(5, 300)$, we obtain

$$A(t) - 300 = 20(t - 5) \implies A(t) = 20(t - 5) + 300 \implies A(t) = 20t - 100 + 300.$$

Thus the linear model representing this data is given by $A(t) = 20t + 200$.

Indeed we could have chosen *any* two points to deduce the equation of this linear function. To see this suppose we take $(0, 200)$ and $(8, 360)$ instead. Then

$$m = \frac{360 - 200}{8 - 0} = \frac{160}{8} = 20$$

and so

$$A(t) - 200 = 20(t - 0) \Longrightarrow A(t) = 20(t - 0) + 200 \Longrightarrow A(t) = 20t + 200.$$

**Remark**

Be careful when calculating the slope $m = \frac{f(x_2)-f(x_1)}{x_2-x_1}$. A common mistake is to switch the $x_1$ and $x_2$ in, say, the denominator but leaving the numerator the same, i.e. calculating $\frac{f(x_2)-f(x_1)}{x_1-x_2}$. This is *not* correct. Be careful!

Looking at the equation of a linear function, $f(x) = mx + c$, if $x$ increases by 1, $f(x)$ increases by $m$. If $x$ increases by 2, $f(x)$ increases by $2m$. If $x$ increases… you get the idea. The thing to notice is that the differences in $f(x)$ values are constant for equal differences of $x$. Indeed in Example 1.2.6 for every month that goes by, the amount in the account increases by 20. The next example is one of a table whose entries do *not* correspond to a linear function.

**Example 1.2.7**

The data in the table below does not represent a linear function.

| $x$ | 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 960 | 900 | 840 | 780 | 750 | 720 | 700 | 680 | 650 | 600 | 500 |

As $x$ changes from 2 to 4, $y$ changes from 900 to 840, corresponding to a slope of $m = 30$. But when $x$ changes from 6 to 8, $y$ changes from 780 to 715, corresponding to a slope of 15. Since a linear function must have a *constant* slope, this data cannot come from a linear function.

## 1.3 Modelling with Linear Functions

Let's now add some of the "business" part to this class.

**Definition 1.3.1**

The **cost function**, $C(q)$, gives the total cost of producing a quantity $q$ of some goods.

Intuitively, if you increase quantity you would expect the cost to increase. So $C(q)$ must be an *increasing* function. Costs of production fall under two categories: the *fixed costs*, which are the costs incurred no matter how much is purchased, and the *variable costs*, which depend on how much is purchased.

**Example 1.3.2**

Suppose a company manufactures t-shirts. The fixed costs for this company would be things like factory costs and machinery costs. On top of this would be the variable costs, such as labour or materials, that depend on the number of t-shirts made. Suppose the fixed costs for this company amount to \$1,200 and the variable costs are \$4 per t-shirt. The cost function for this company would then be

$$\text{Total costs for the company} = \text{Fixed costs} + \text{Variable costs} = 1,200 + 4 \cdot \text{Number of t-shirts}.$$

If $q$ represents the quantity of t-shirts made, then

$$C(q) = 1,200 + 4q.$$

You may recall from you college algebra or precalculus class that the $y$-intercept was sometimes called the *initial value*, or something similar. This initial value corresponds to the *fixed costs* of production. The variable cost for one additional unit is called the *marginal cost*. In the case of a linear cost function, the marginal cost is simply the slope, $m$.

**Definition 1.3.3**

The **revenue function**, $R(q)$, gives the total revenue received by a firm from selling a quantity, $q$, of some good.

$$\text{Revenue} = \text{Price} \cdot \text{Quantity}.$$

If the good sells for a price of $p$ per unit, and the quantity sold is $q$, then $R(q) = pq$. If $p$ is independent of the quantity sold, then $p$ is constant with respect to $q$. In this case $R(q)$ is a linear function passing through the origin.

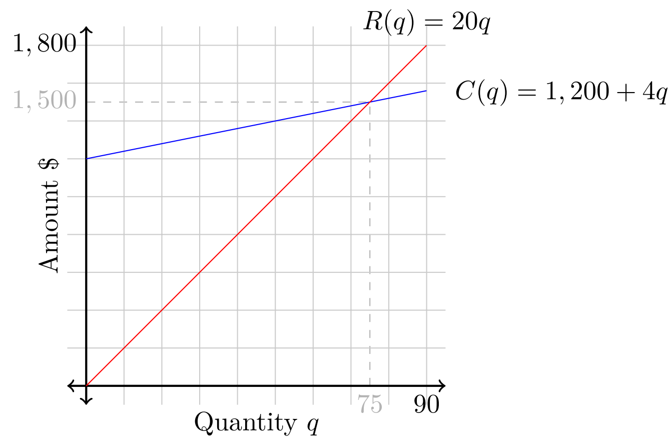
**Example 1.3.4**

Suppose the company from Example 1.3.2 sells each shirt for \$20. Then the revenue function is given by

$$R(q) = 20q.$$

The graph on the right shows the cost (blue) and revenue (red) functions relating to Example 1.3.2. How can we interpret this graph? Well, if the cost function lies above the revenue function the cost of manufacturing the goods is higher than the money gained from selling the goods. Thus the company is making a loss.

So, from a manufacturers standpoint, we are interested in the point at which the money gained from sales will be greater than the cost of manufacturing. In this case, we can see that as long as the company makes and sells more than 75 t-shirts, they will be making a *profit*. Any less and they will be losing money.




**Definition 1.3.5**

The **profit function**, $\pi(q)$, gives the total profit received by a firm from selling a quantity, $q$, of some good. It is the difference between revenue and cost, that is $\pi(q) = R(q) - C(q)$.
The **break-even point** is the value of $q$ so that $\pi(q) = 0$.

**Example 1.3.6**

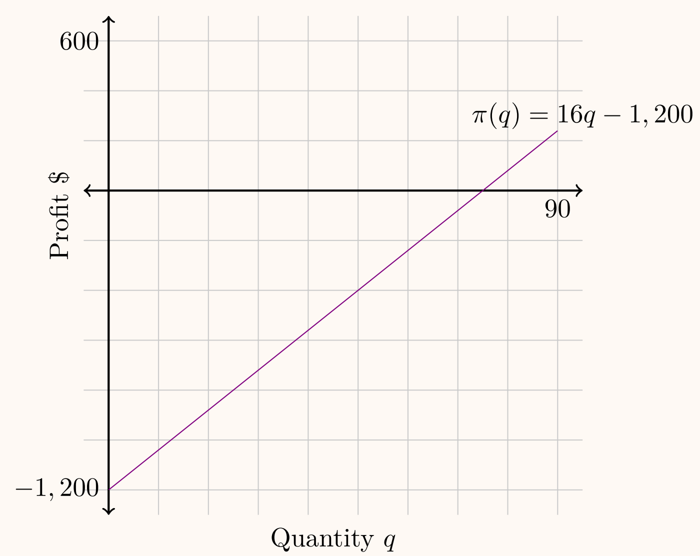
Continuing with Example 1.3.2, the profit function $\pi(q)$ is given by

$$\begin{aligned}\pi(q) = \text{Profit} &= \text{Revenue} - \text{Cost}\\ &= R(q) - C(q)\\ &= 20q - (1,200 + 4q)\\ &= 20q - 1,200 - 4q\\ &= 16q - 1,200.\end{aligned}$$

So $\pi(q) = 16q - 1,200$. Since the break-even point occurs when $\pi(q) = 0$, then

$$16q - 1,200 = 0 \Longrightarrow 16q = 1,200 \Longrightarrow q = 75,$$

so the company makes a profit provided they sell more than 75 t-shirts.

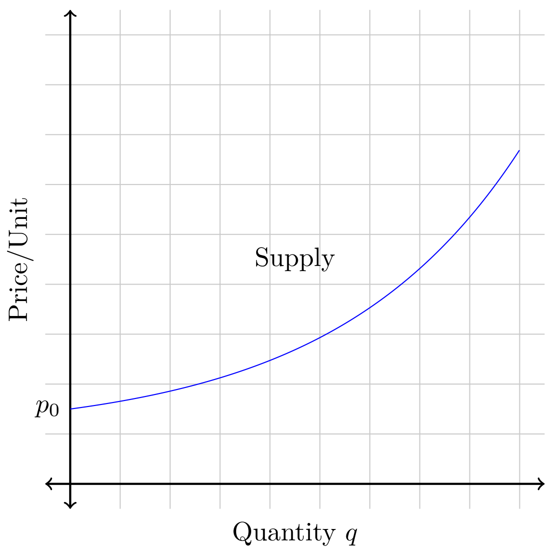


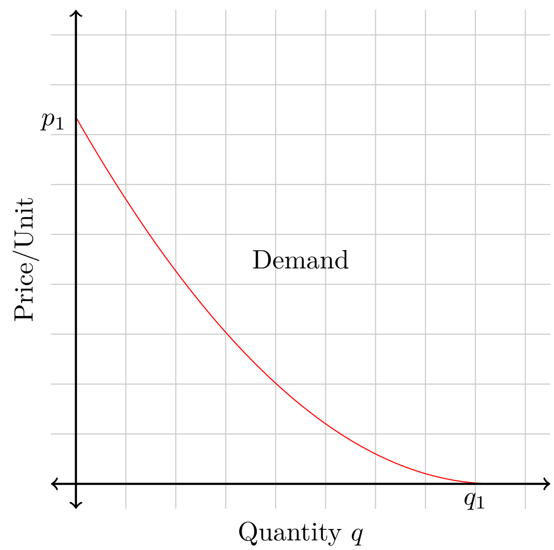
The quantity, $q$, of an item that is produced and sold depends on its price, $p$. The more expensive the item is, suppliers are more likely to want to supply more, whereas the quantity demanded by the consumers is likely to fall.

**Definition 1.3.7**

The **supply curve**, for a given item, relates the quantity, $q$, of an item that manufacturers are willing to make per unit time to the price, $p$, for which the item can be sold.
The **demand curve** relates the quantity, $q$, of an item demanded by consumers per unit time to the price, $p$, of the item.

While supply and demand are given as functions of price, for silly historical reasons economists put price on the vertical axis instead of the horizontal axis. Thus you end up with something that looks like the following;



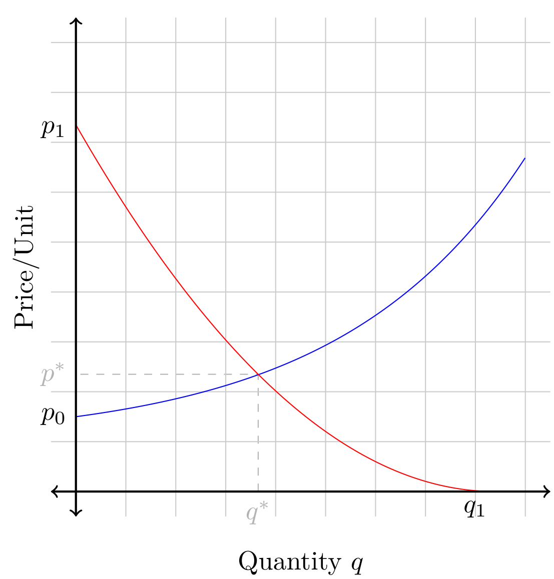

You will see that three points are labelled in the above graphs; $p_0, p_1$ and $q_1$. What do these points mean?

The vertical axis corresponds to a quantity value of zero. Since $p_0$ is the vertical intercept on the supply curve, this is the price per unit when the quantity supplied is zero. Plainly, any price lower than this and the supplier will not manufacture this item. Since $p_1$ is the vertical intercept on the demand curve, this is the price per unit when the quantity demanded is zero. That is, if the price is higher than this value then the consumer will not purchase any of the item.

The horizontal axis corresponds to a price of zero, so the quantity $q_1$ is the demand for the product if it were to be given away for free.

In summary, we can think of $p_0$ and $p_1$ as the minimum and maximum prices, respectively, at which to sell the item, and $q_1$ is the maximum quantity of items that could be sold. So, in the real world the supplier would be looking to choose a price $p$ such that $p_0 \leq p \leq p_1$, where the quantity sold is still high enough for them to make profit. The optimal price to sell the item at can be found by looking at where supply meets demand. Or, graphically, when the two graphs above intersect when plotted together.

This point is denoted by $(q^*, p^*)$ in the graph on the right. It is assumed that the market naturally settles to this equilibrium point.



**Definition 1.3.8**

The point $(q^*, p^*)$ on the above graph is called the **equilibrium point**. The value $q^*$ is the **equilibrium quantity** and the value $p^*$ is the **equilibrium price**.

**Example 1.3.9**

Suppose that the quantity of t-shirts supplied is given by $S(p) = 4p - 8$ and the quantity of t-shirts demanded is given by $D(p) = 212 - 6p$. To calculate the equilibrium point we simply set $S(p)$ and $D(p)$ equal to each other and solve for $p$;

$$S(p) = D(p) \implies 4p - 8 = 212 - 6p \implies 10p = 220 \implies p = 22.$$

So $p^* = 22$. Once you have this value you can deduce $q^*$ by plugging it in to either $S(p)$ or $D(p)$.

$$q^* = S(p^*) = 4(22) - 8 = 88 - 8 = 80.$$

Thus the equilibrium point is $(q^*, p^*) = (80, 22)$. We can read this as the quantity of t-shirts that should be manufactured is 80, and they should be sold for \$22 each.

The equilibrium point is the supply and demand analogue of the break-even point when we look at cost and revenue functions. They work hand in hand with each other, the cost and revenue functions tell you how much *should* be sold to make the current model profitable. The supply and demand functions tell you how much *could* be sold. If you need to sell 100 of some good to make a profit, but the demand isn't high enough for it, you probably should rethink your model.

**Remark**

None of the functions in this section *need* to be linear. We will be using these types of functions in the cases where they are both linear and non-linear later, so it is important that you understand what they represent.

## 1.4 Exponential Functions

The second type of function you should be familiar with are *exponential functions*.

**Definition 1.4.1**

If $a$ and $k$ are any numbers such that $a > 0$ and $a \neq 1$, then an **exponential function** is one of the form

$$f(x) = P_0 a^x.$$

We call $P_0$ the **initial value** or **initial population** and $a$ the **base** or **growth factor**.

Notice that unlike linear functions (and polynomials) the $x$ is now in the exponent and the base is a fixed number. Like polynomials however, the domain of *every* exponential function is all real numbers, $\mathbb{R}$.

**Remark**

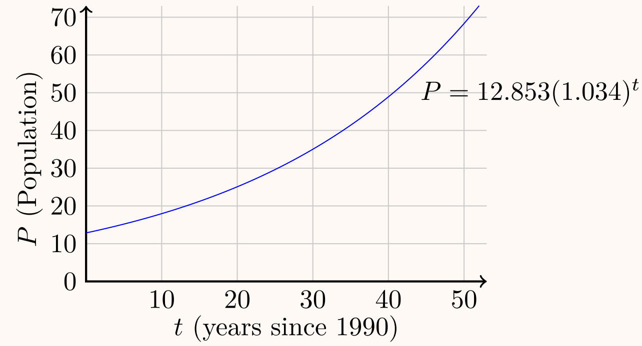
- Note that if we allow $a$ to be 0 or 1 we end up with constant functions rather than exponential functions, this is why we ignore these values for $a$.
- Note too that if $a < 0$ we cannot evaluate $f(x)$ for every value of $x$. Recall that $a^{1/2} = \sqrt{a}$. So if $a < 0$, then $f\left(\frac{1}{2}\right)$ is not defined.

Exponential functions are used to represent many phenomena in the natural and social sciences. One such thing they are used for is population.

**Example 1.4.2**

The (estimated) population of a certain country from 1990 to 1996 is given in the table below. In the third column the change in population is given. Notice that since the change in population is not constant, the population cannot be growing linearly. Another way to see this is to plot the points in the table and see that no straight line will intersect all of them at the same time.

| Year | Population (millions) | Change (millions) |
|---|---|---|
| 1990 | 12.853 | |
| 1991 | 13.290 | 0.437 |
| 1992 | 13.747 | 0.457 |
| 1993 | 14.225 | 0.478 |
| 1994 | 14.721 | 0.496 |
| 1995 | 15.234 | 0.513 |
| 1996 | 15.757 | 0.523 |




We can find a relationship in these changes if we look at the quotient of each years population with the previous years population.

$$\frac{\text{Population in 1991}}{\text{Population in 1990}} = \frac{13.290 \text{ million}}{12.853 \text{ million}} = 1.034 \quad \frac{\text{Population in 1994}}{\text{Population in 1993}} = \frac{14.721 \text{ million}}{14.225 \text{ million}} = 1.034$$

$$\frac{\text{Population in 1992}}{\text{Population in 1991}} = \frac{13.747 \text{ million}}{13.290 \text{ million}} = 1.034 \quad \frac{\text{Population in 1995}}{\text{Population in 1994}} = \frac{15.234 \text{ million}}{14.721 \text{ million}} = 1.034$$

$$\frac{\text{Population in 1993}}{\text{Population in 1992}} = \frac{14.225 \text{ million}}{13.747 \text{ million}} = 1.034 \quad \frac{\text{Population in 1996}}{\text{Population in 1995}} = \frac{15.757 \text{ million}}{15.234 \text{ million}} = 1.034$$

Since each quotient is 1.034, we see that the population grew by 3.4% each year. Whenever the percentage increase (or decrease) we have *exponential growth* (or *decay*). If $t$ is the number of years since 1990 and the population is in millions,

$$\begin{aligned}
&\text{When } t = 0, \text{ population} = 12.853 = 12.853(1.034)^0. \\
&\text{When } t = 1, \text{ population} = 13.290 = 12.853(1.034)^1. \\
&\text{When } t = 2, \text{ population} = 13.747 = 13.290(1.034) = 12.853(1.034)^2. \\
&\text{When } t = 3, \text{ population} = 14.225 = 13.747(1.034) = 12.853(1.034)^3 \ldots
\end{aligned}$$

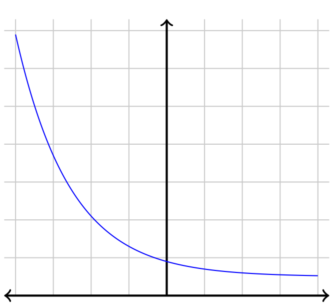
So the population in millions $t$ years after 1990 is given by $P = 12.853(1.034)^t$. Since the variable $t$ is in the exponent, this is an *exponential function*. The growth factor is 1.034 and the initial population (i.e. the population in 1990) is 12.853.

The above example illustrates that linear and exponential functions are different. But in some ways they are similar. A linear function has a constant rate of change, whereas an exponential function has a constant percentage rate of change, (or relative rate of change). The difference in these statements, plainly, means that a linear function changes at a constant rate regardless of the current value or population, and an exponential function changes at a rate dependent on current value or population.

**Definition 1.4.3**

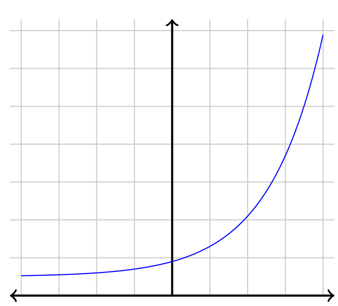
The **growth rate**, $r$, of an exponential function $f(x) = P_0 \cdot a^x$ is given by $r = a - 1$.

Since we allow for any positive number for $a$, it could very well be that the growth rate, $r$, is negative. For example if $a = \frac{1}{2}$. This is perfectly okay.



$\underline{r < 0}$ :

- *Decreasing* function.
- When modelling, one would expect a negative value of $r$ for any situation when something is decreasing. For example half-life of an isotope or expulsion of drugs.



$\underline{r > 0}$ :

- *Increasing* function.
- When modelling, one would expect a positive value of $r$ for any situation when something is increasing. For example population of bacteria or interest in an account.

The case when $r = 0$ is uninteresting, as this is when $a = 1$ and so the exponential function is actually just a constant function. Just like polynomials, the domain of all exponential functions is always $(-\infty, \infty)$. The range however is always $(0, \infty)$. These two properties are independent of any of the values of $k, a$ or $r$.

Exponential functions appear all of the place with lots of different bases. There is however one base that is preferred in calculus. This is the so called *natural base*.

**Definition 1.4.4**

The **natural exponential function** is $f(x) = e^x$, where the base $e$ is **Eulers number**;

$$e = 2.71828182845904523536028747135 27\ldots$$

This function is going to be crucial later on in the course, so it will be important to pay close attention to its properties as we proceed (don't worry you don't have to memorise its value). As far as what we have talked about so far however, the natural exponential functions behaves in the exact same way.

## 1.5 Logarithmic Functions

Logarithmic functions go hand in hand with exponential functions, they are *inverse functions* of each other. That is, logarithmic functions "undo" the rule that exponential functions preform and exponential functions "undo" the rule logarithmic functions perform. You will have used these in your college algebra or pre-calculus classes to solve exponential equations.

**Definition 1.5.1**

If $a$ is any number such that $a > 0$, then the **logarithm to base** $a$ is the function

$$f(x) = \log_a(x).$$

The statement $y = \log_a(x)$ is equivalent to $a^y = x$.

Again we refer to $a$ as being the *base* of the logarithm. In English, the logarithm to base $a$ takes the input $x$ and outputs the power to which $a$ must be raised to, to equal $x$. For example; $\log_2(8) = 3$ since $2^3 = 8$. This is what we mean when we say that the logarithm "undoes" the rule of the exponential function.

The domain and range of all logarithm functions are closely related to the domain and range of the exponential functions, this is a consequence of them being *inverses* of one another. Recall that the domain and range of all exponential functions are $(-\infty, \infty)$ and $(0, \infty)$, respectively. Well, the domain and range of all logarithm functions are $(0, \infty)$ and $(-\infty, \infty)$, respectively - they are just the opposite way round of the exponential functions!

**Example 1.5.2**

Using logarithm functions, we can solve exponential equations;

1. Solve $5 \cdot 3^x = 100$.
$$\begin{aligned} 5 \cdot 3^x &= 100 \\ \Rightarrow 3^x &= 20 \\ \Rightarrow x &= \log_3(20) \\ &\approx 2.7268 \end{aligned}$$

2. Solve $12 \cdot 4^x = 60$.
$$\begin{aligned} 12 \cdot 4^x &= 60 \\ \Rightarrow 4^x &= 5 \\ \Rightarrow x &= \log_4(5) \\ &\approx 1.1610 \end{aligned}$$

3. Solve $11 \cdot 6^x = 121$.
$$\begin{aligned} 11 \cdot 6^x &= 121 \\ \Rightarrow 6^x &= 11 \\ \Rightarrow x &= \log_6(11) \\ &\approx 1.3383 \end{aligned}$$

When working with logarithms, particularly with algebraic equations, we take advantage of certain properties that are unique to the logarithm functions. Since the logarithm functions are "undoing" the exponential functions, we can use all the laws of exponents to simplify things. Recall that for positive numbers $a, n$ and $m$, the following exponential laws hold;

| | | | | |
|---|---|---|---|---|
| $a^m a^n = a^{m+n}$ | $\dfrac{a^m}{a^n} = a^{m-n}$ | $(a^m)^n = a^{mn}$ | $1 = a^0$ | $a = a^1$ |

If you pay close attention to the relationship between the operation of the terms on the left hand sides and the operation of the exponents on the right hand sides, then you may see where the following laws of logarithms come from. If $a, A, B$ and $n$ are all positive numbers;

| | | |
|---|---|---|
| $\log_a(AB) = \log_a(A) + \log_a(B)$ | $\log_a\left(\frac{A}{B}\right) = \log_a(A) - \log_a(B)$ | $\log_a A^n = n\log_a(A)$ |

| | |
|---|---|
| $\log_a(1) = 0$ | $\log_a(a) = 1$ |

**Example 1.5.3**

Let's see how the above rules can be used to simplify algebraic expressions involving exponents.

1. Simplify $\log_4\left(x^3y^5\right)$.

$$\begin{aligned}\log_4\left(x^3y^5\right) &= \log_4\left(x^3\right) + \log_4\left(y^5\right)\\ &= 3\log_4(x) + 5\log_4(y)\end{aligned}$$

2. Simplify $\log_6\left(\frac{a^4}{b^7}\right)$.

$$\begin{aligned}\log_6\left(\frac{a^4}{b^7}\right) &= \log_6\left(a^4\right) - \log_6\left(b^7\right)\\ &= 4\log_6(a) - 7\log_6(b)\end{aligned}$$

3. Simplify $\log_7\left(\frac{x^9\sqrt{y}}{z^3}\right)$.

$$\begin{aligned}\log_7\left(\frac{x^9\sqrt{y}}{z^3}\right) &= \log_7\left(x^9y^{1/2}\right) - \log_7\left(z^3\right)\\ &= \log_7\left(x^9\right) + \log_7\left(y^{1/2}\right) - 3\log_z(z)\\ &= 9\log_7(x) + \frac{1}{2}\log_7(y) - 3\log_7(z)\end{aligned}$$

**Remark**

A common mistake when using logarithms is to bring down the power of an expression that is not attached to everything inside the log. That is, suppose you are trying to simplify $\log\left(5a^{2t}\right)$. This is not equal to $2t\log(5a)$ because the $2t$ is only the exponent of $a$ and not 5. If you were to simplify this using the multiplication property first, you could *then* bring down the power of just the $a$ part. That is $\log\left(5a^{2t}\right) = \log(5) + \log\left(a^{2t}\right) = \log(5) + 2t\log(a)$.

We saw that when it comes to exponential functions, we like to use the natural exponential, $e^x$. It makes sense then that we also like to deal with its corresponding logarithm, $\log_e(x)$. Since we use this logarithm a lot, we change its notation a little.

**Definition 1.5.4**

The **natural logarithm function** is $f(x) = \ln(x)$, where $\ln(x) = \log_e(x)$.

The natural logarithm behaves exactly like any other logarithm function. so all the properties above carry over. Don't let the notation confuse you - it is just a logarithm function, it just so happens to be the "favoured" one.

## 1.6 Modelling with Exponential Functions

### Compound Interest

Money deposited in a savings account increases exponentially, because the interest on the account is calculated by multiplying the current balance by a fixed factor - the interest rate. *Compound interest* means that the interest earned in one time period is split into smaller increments.

**Definition 1.6.1**

If an amount $P_0$ is invested at an annual interest rate $r$ compounded $n$ times each year, then the amount $A(t)$ of the investment after $t$ years is given by the formula

$$A(t) = P_0\left(1 + \frac{r}{n}\right)^{nt}.$$

Compounded $n$ times each year means how many times (per year) the interest on the account is calculated. For example, if the interest is compounded monthly then $n = 12$. If the interest is compounded quarterly, then $n = 4$. We end up with the $\frac{r}{n}$ since we essentially split the interest paid per year into equal instalments, so that the total interest paid is $r$ $\left(n \cdot \frac{r}{n} = r\right)$.

**Example 1.6.2**

Suppose that \$1,000 is deposited into a savings account paying 6% interest annually, compounded monthly. We want to find a model, $A(t)$, that represents the amount in the account after $t$ years. To summarise, we have:

Initial balance: $\$1,000 \Longrightarrow P_0 = 1,000$ Interest rate: $6\% \Longrightarrow r = 0.06$ Compounded: Monthly $\Longrightarrow n = 12$.

Putting these 3 pieces of information together, we obtain

$$A(t) = P_0\left(1+\frac{r}{n}\right)^{nt} = 1,000\left(1+\frac{0.06}{12}\right)^{12t} = 1,000\,(1.005)^{12t}.$$

In general, the more often interest is compounded, the more money will be earned. So,

$$P\left(1+\frac{r}{1}\right)^{1\cdot t} < P\left(1+\frac{r}{2}\right)^{2\cdot t} < P\left(1+\frac{r}{3}\right)^{3\cdot t} < P\left(1+\frac{r}{4}\right)^{4\cdot t} < \ldots P\left(1+\frac{r}{n}\right)^{n\cdot t} < \ldots.$$

So what happens if we increase $n$ without bound? Well, in calculus this is what is known as *limit*. Limits aren't a part of this course, so we will not discuss them, but essentially a limit tells you what happens when you let a variable tend towards something, in this case infinity. The result is quite remarkable, you can try it yourself. Let $r = 1$, and on your calculator calculate $\left(1+\frac{1}{n}\right)^n$ for larger and larger values of $n$. As large as your calculator can handle. What do you notice? Well, as $n$ increases, the quantity $\left(1+\frac{1}{n}\right)^n$ approaches $2.71828\ldots$. Do you remember this number? If you don't, turn back a couple of pages and you will see that this is precisely the number $e$! I told you this number was cool. This is precisely the way in which the number was discovered, letting the number of times interest is compounded tend towards infinity. This is made more interesting by the fact that if you let $n$ tend towards infinity in the quantity $\left(1+\frac{r}{n}\right)^n$, you obtain $e^r$.

Clearly we cannot actually compounded interest an infinite number of times a year. But after a certain value of $n$ we simply just use $e$ to calculate our interest. This is known as *continuously compounded interest*.

**Definition 1.6.3**

If an amount $P$ is invested at an annual interest rate $r$ **compounded continuously**, then the amount $A(t)$ of the investment after $t$ years is given by the formula

$$A(t) = P_0 e^{rt}.$$

**Example 1.6.4**

If \$4,000 is deposited into an account paying interest at a rate of %8 per year, compounded continuously, how much will be in the account after five years? How long will it take until the account contains \$10,000, assuming no other deposits? To start, we need to find an expression for $A(t)$. Breaking down the the information in the paragraph we obtain $A(t) = 4,000e^{0.08t}$. After five years then, the account will contain

$$A(5) = 4,000e^{0.08\cdot 5} \approx \$5967.30.$$

To find out the number of years required to reach \$10,000, we must solve $A(t) = 10,000$.

$$\begin{aligned} A(t) = 4,000e^{0.08\cdot t} &= 10,000 \\ \Rightarrow e^{0.08\cdot t} &= \frac{5}{2} \\ \Rightarrow 0.08\cdot t &= \ln\left(\frac{5}{2}\right) \\ \Rightarrow t &= \frac{\ln\left(\frac{5}{2}\right)}{0.08} \\ &\approx 11.4536 \text{ years} \end{aligned}$$

So in roughly 11.5 years the account will contain \$10,000.

## Exponential Growth and Decay

Taking a quick step away from the business applications, let's look at some other phenomena that can be modelled by exponential functions.

**Example 1.6.5**

Food poisoning is often cause by E. coli bacteria. To test for the presence of E. coli in a pot of beef stew, a biologist performs a bacteria count on a small sample of the beef stew kept at 25°C. She determines the count is 5 colony forming units per millilitre and that the number will double every 40 minutes. Find an exponential model $f(t) = k \cdot a^t$ for the bacteria count in the beef stew. Using this model, determine how long will it take for the amount of bacteria to reach 25 units.

Reading the scenario we can immediately that the initial population of bacteria is $k = 5$. Now we have to translate "doubling every 40 minutes" into mathematical language. The 40 minutes is simply our time period, so $t = 1$ corresponds to 40 minutes, $t = 2$ is 80 minutes, and so on.

We want an exponential model $f(t) = 5a^t$. Since the population doubles every 40 minutes, we know that $f(t+1) = 2 * f(t)$ (increasing $t$ by 1 corresponds to doubling the population at $f(t)$). So then $5a^{t+1} = 10a^t$. Dividing both sides by $5a^t$ yields $a = 2$. Thus our model is $f(t) = 5 \cdot 2^t$.

Now we want to know when $f(t) = 25$. So we simply solve for $t$ and interpret how long that is.

$$\begin{aligned} f(t) = 5 \cdot 2^t &= 25 \\ \Rightarrow 2^t &= 5 \\ \Rightarrow t &= \log_2(5) \\ &\approx 2.3219 \end{aligned}$$

So when $t$ is approximately 2.3219, the population is 25. In terms of time then, the population reaches 25 after $2.3219 \times 40 = 92.88$ minutes.

The above example is one that uses an exponential equation of the form $f(t) = P_0 \cdot a^t$. But we introduced the natural exponential for a reason, so let's bring that back. Much like what we did in continuously compounded interest, we can model quantities in nature with an exponential function of the form $P(t) = P_0 e^k t$. These models are what we refer to as *exponential growth/decay models*.

**Definition 1.6.6**

An **exponential growth/decay model** is one of the form

$$P(t) = P_0 e^{kt},$$

where $P_0$ is the *initial quantity* (or *initial population*) and $k$ is the *continuous growth factor*.
If $k > 0$ then the model represents exponential **growth**, and if $k < 0$ the model represents exponential **decay**.

It shouldn't be too alarming that we can represent exponential models in this way. We really aren't really changing anything from $P_0 \cdot a^t$. Indeed if $a$ is the growth factor of the model, then $a = e^{\ln(a)}$ and so $P_0 \cdot a^t = P_0 \cdot e^{\ln(a)t}$. We are simply relabelling our generic exponential function as one involving $e$.

It is important to remember that $k > 0$ is growth and $k < 0$ is decay. When reading a problem that describes a quantity increasing or decreasing, knowing what sign the $k$ value you are going to calculate should have will let you check your work easier. If you're looking at an increasing population and you calculate a negative $k$ value, you should instantly know something went wrong and that you should go back and check your work.

**Example 1.6.7**

We initially have 100 grams of a radioactive element and in $1,250$ years there will be 80 grams left.

(a) Determine the exponential decay equation for this element.

(b) How long will it take for half of the element to decay?

(c) How long will it take until there is only 1 gram of the element left?

(a) We can read from the paragraph that $P_0 = 100$, so $P(t) = 100e^{kt}$. We are also told that in $1,250$ years there will be 80 grams of the element left, so

$$\begin{aligned} P(1,250) &= 80 \\ \Rightarrow 100e^{k\cdot 1,250} &= 80 \\ \Rightarrow e^{k\cdot 1,250} &= \frac{4}{5} \\ \Rightarrow k\cdot 1,250 &= \ln\left(\frac{4}{5}\right) \\ \Rightarrow k &= \frac{1}{1,250}\ln\left(\frac{4}{5}\right) \\ &\approx -0.0001785 \end{aligned}$$

So $P(t) = 100e^{\frac{\ln(4/5)}{1,250}t} \approx 100e^{-0.0001785t}$. Note this negative $k$ value makes sense because the amount of element is *decreasing* (i.e. *decaying*).

(b) Now we are interested in when there will only be half of the amount of element remaining, that is when there is only 50 grams left. So we must solve $P(t) = 50$.

$$\begin{aligned} P(t) &= 50 \\ \Rightarrow 100e^{\frac{\ln(4/5)}{1,250}t} &= 50 \\ \Rightarrow e^{\frac{\ln(4/5)}{1,250}t} &= \frac{1}{2} \\ \Rightarrow \frac{\ln\left(\frac{4}{5}\right)}{1,250}t &= \ln\left(\frac{1}{2}\right) \\ \Rightarrow t &= \frac{1,250\ln\left(\frac{1}{2}\right)}{\ln\left(\frac{4}{5}\right)} \\ &\approx 3,882.8546 \end{aligned}$$

So the the amount of element will half in about $3,882.8546$ years.

(c) This time we do the exact same as in (b), but instead of solving for 50 we solve for 1. Thus we obtain $t \approx 25,797.1279$ years.

**Example 1.6.8**

A population of bacteria initially has 250 present and in 5 days there will be $1,600$ bacteria.

(a) Determine the exponential growth equation for this population.

(b) How long will it take for the population to grow from its initial population of 250 to a population of $2,000$?

(a) We solve this problem in the same way as the previous example. The initial population is $P_0 = 250$. Then,

$$\begin{aligned} P(5) &= 1,600 \\ \Rightarrow 250e^{k\cdot 5} &= 1,600 \\ \Rightarrow e^{k\cdot 5} &= \frac{32}{5} \\ \Rightarrow k\cdot 5 &= \ln\left(\frac{32}{5}\right) \\ \Rightarrow k &= \frac{1}{5}\ln\left(\frac{32}{5}\right) \\ &\approx 0.3713 \end{aligned}$$

(b) Solving for $P(t) = 250e^{\frac{\ln(32/5)}{5}} = 2,000$, we obtain $t = 5.6010$ days. The steps are the same as the previous example.

We will now do an example where the parameters we use in our model are not so clear.

**Example 1.6.9**

A certain sample collected in 2018 degrades at rate of one third every 367 years. That is, after 367 years $33.\bar{3}\%$ of the sample will have degraded. Upon collection it is believed that there is currently 96.1% of what there was when it was formed. How old is this sample?

What we want to do is to come up with a model $p(t) = P_0 e^{kt}$ to represent the percentage of the sample remaining at age $t$ years. Once we have this we can solve for $t$ to see how old the sample must be so that only 96.1% of it has survived until 2018.

We know two things:

$$p(0) = 100\% \text{ of the sample remaining} \Longrightarrow P_0 = 1$$
$$\text{and } p(367) = 66.\bar{6}\% \text{ of the sample remaining} \Longrightarrow e^{k\times 367} = \frac{2}{3}.$$

Using $p(367) = \frac{2}{3}$, we can solve for $k$ using natural logarithms,

$$\begin{aligned} e^{k\times 367} = \frac{2}{3} &\Longrightarrow \ln\left(e^{k\times 367}\right) = \ln\left(\frac{2}{3}\right) \\ &\Longrightarrow k \times 367 \ln(e) = \ln\left(\frac{2}{3}\right) \\ &\Longrightarrow k \times 367 = \ln\left(\frac{2}{3}\right) \\ &\Longrightarrow k = \frac{1}{367}\cdot\ln\left(\frac{2}{3}\right) \approx -1.105\times 10^{-3}. \end{aligned}$$

Now we have our $k$ value, we simply need to solve $p(t) = 0.961$ for $t$.

$$e^{kt} = 0.961 \Longrightarrow kt = \ln(0.961) \Longrightarrow t = \frac{1}{k}\cdot\ln(0.961) = \frac{367}{\ln(2/3)}\cdot\ln(0.961) \approx 36.007 \text{ years.}$$

So the sample is approximately 36 years old.

## 1.7 Polynomials

The last types of functions we will look at are *polynomial functions*. Recall these functions were defined in Definition 1.2.1. We will see in later chapters how “nice” polynomials of any degree are. The properties they exhibit will make you wish that every question you are given was based on a polynomial function rather than any other function. We have already looked at polynomial functions of degree 1 or 0, i.e. *linear functions*. So what about ones of higher degree? We will look closely at quadratic polynomials, i.e. degree 2 polynomials, and cubic polynomials, i.e. degree 3 polynomials. For polynomials of higher degree we will just comment on their basic properties.

### Quadratics

**Definition 1.7.1**

A **quadratic function** (or **quadratic polynomial**) is a polynomial function of degree 2.

The generic form in which we will denote a quadratic polynomial is $f(x) = ax^2 + bx + c$. Chances are you will have seen quadratic polynomials in your life. In fact the first thing you may have thought of when seeing the word “quadratic” is the *quadratic formula*.

**Proposition 1.7.2: Quadratic Formula**

Given a quadratic polynomial, $f(x) = ax^2 + bx + c$. We have the following;

1. If $b^2 - 4ac < 0$, the equation $f(x) = 0$ has no real solutions.
2. If $b^2 - 4ac = 0$, the equation $f(x) = 0$ has one unique real solution, given by $x = \frac{-b}{2a}$.
3. If $b^2 - 4ac > 0$, then the equation $f(x) = 0$ has two real solutions, given by the **quadratic formula**,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Each of the above cases rely on the value of the quantity $b^2 - 4ac$. This quantity is known as the *discriminant*. It is a useful tool that we can use to almost immediately see if a quadratic has roots (solutions) or not. In the case that this quantity is non-negative we find the expressions for the roots in each of the last two cases by *completing the square*. A technique of factorisation you may have been introduced to in a previous maths class.

*Proof.* Let $f(x) = ax^2 + bx + c$ be a quadratic equation with $a, b, c$ any real numbers and $a \neq 0$. Suppose $b^2 - 4ac \geq 0$. Then,

$$ax^2 + bx + c = a\left(x^2 + \frac{b}{a}x\right) + c = a\left(\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a^2}\right) + c = a\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a}.$$

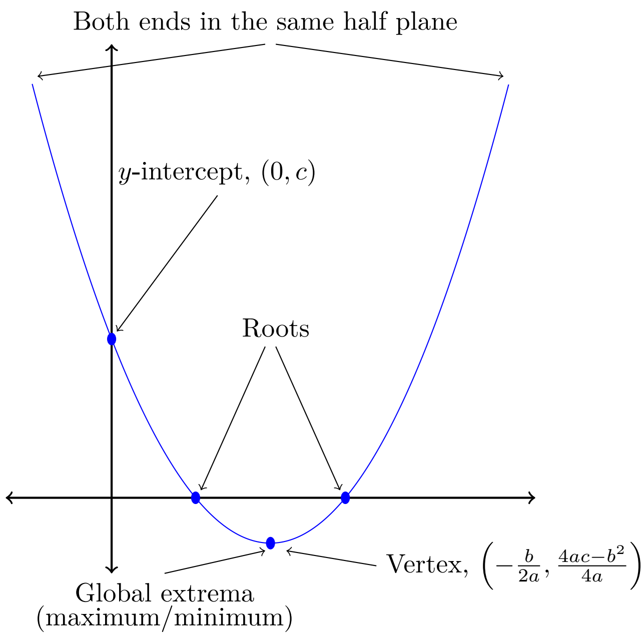
So if $f(x) = 0$, we have

$$a\left(x + \frac{b}{2a}\right)^2 + \frac{4ac - b^2}{4a} = 0 \Longrightarrow a\left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a} \Longrightarrow \left(x + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2} \Longrightarrow x + \frac{b}{2a} = \pm\frac{\sqrt{b^2 - 4ac}}{2a}$$
$$\Longrightarrow x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

□

So other than the nice formula above, what else is true about quadratics?

An example of $f(x) = ax^2 + bx + c$ with $a > 0$ and $b^2 - 4ac > 0$.




Depending on the sign of $a$, the graph of any quadratic will look like one of two things. In the case when $a > 0$, the graph will look like the picture to the left, like a cup. If $a < 0$ then take the image to the left and flip it upside down so that it looks like an arch - or a frown. *Every* single quadratic will look like this.

There will always be at most two real roots too. This comes down to the fact that a general polynomial of degree $n$ will have at most $n$ real roots. We have already discussed the quadratic formula that finds these roots should they exist.

The $y$-intercept will always be at the point $(0, c)$. This is easy to see algebraically since plugging in 0 for $x$ eliminates all terms with $x$'s attached to them, leaving only $c$.

The final thing of note is the *vertex* of a quadratic function. This is the point where the graph of the quadratic will change behaviour. In the case when $a > 0$, when $x < -\frac{b}{2a}$, the function will be *decreasing*. When $x > -\frac{b}{2a}$ the function will be *increasing*. In the case when $a < 0$ these cases will be flipped.

While the domain of every quadratic will be $(-\infty, \infty)$, its range depends on the vertex. If $a > 0$ the range is $\left[\frac{4ac-b^2}{4a}, \infty\right)$ and if $a < 0$ the range is $\left(-\infty, \frac{4ac-b^2}{4a}\right]$. This $y$-value of the vertex is known as the *global minimum* of the quadratic polynomial, if $a > 0$ (*global maximum* if $a < 0$). We will study global extrema in more detail later on in the course.

## Cubics

**Definition 1.7.3**

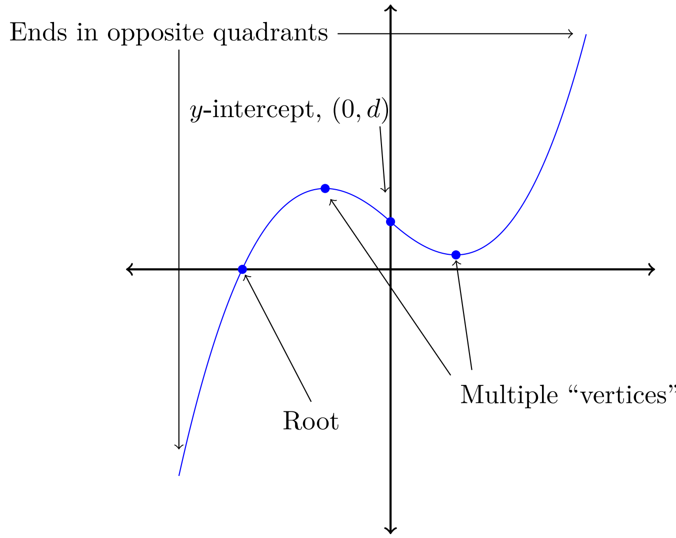
A **cubic function** (or **cubic polynomial**) is a polynomial function of degree 3.

Cubics behave a bit differently to quadratics. As in the case for quadratics, we write a general cubic as $f(x) = ax^3+bx^2+cx+d$.

The first you will notice is that the graph of a cubic function is that the graph looks no longer like a cup (or a frown). It is a wiggly line (like an s shape) from one quadrant of the plane to the opposite quadrant. The case when $a > 0$ is shown on the right. There is no longer a single vertex at which the graph changes direction, in some cases there are two such points and in some cases there are no such points, for example $f(x) = x^3$.

Another difference is that the range of cubic polynomials is always $(-\infty, \infty)$. This is because the leading power $x^3$ doesn't kill negative signs. In quadratics, for large values of $x$ the $x^2$ term contributes the most to the value of the function. Since squaring kills negative signs the function eventually would take on positive values. Since $x^3$ is the dominating term in cubics, and $x^3$ preserves sign, the function can take values of both signs.

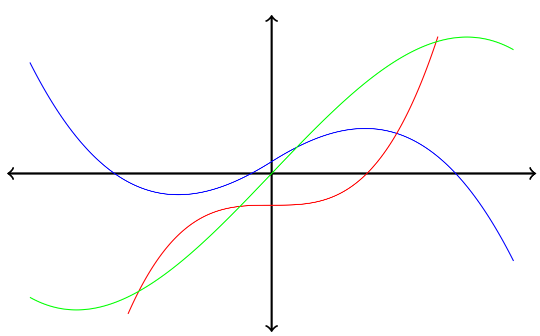
Since the range has changed in this way, this also means we are always guaranteed at least one real root to a cubic polynomial. In fact there is even a formula like the quadratic formula to calculate the roots of cubics. But we won't talk about that in this class, so don't worry about learning it - it's quite messy.




An example of $f(x) = ax^3 + bx^2 + cx + d$ with $a > 0$.

## Higher Degree Polynomials

Higher degree polynomials start to get more and more complicated, at least in terms of their algebraic properties. While there is a *quartic formula* for finding the roots of degree four polynomials, for degree five and higher no such formulae exist. But there are some properties that can be observed, these rely mainly on the parity of the degree of the polynomial. That is, is the degree odd or even?

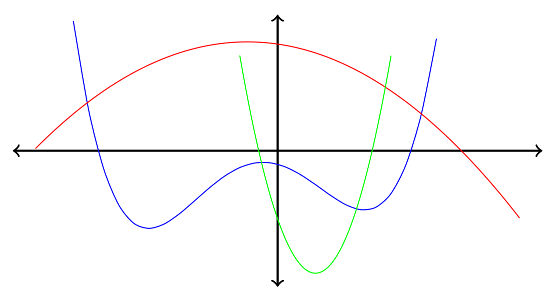


Odd degree polynomials: If the degree of a polynomial is odd, without looking at its expression we immediately know that the range of the polynomial is $(-\infty, \infty)$ and that the graph of the polynomial will be a wiggly line from one quadrant to its opposite quadrant. That is, if $a > 0$ (again $a$ is the coefficient of the highest power of $x$) the graph of the polynomial will range from the bottom left of the plane to the top right of the plane. If $a < 0$ it will range from the top left to bottom right. A consequence of this is that *every* odd degree polynomial has at least one real root.

**Proposition 1.7.4**

Every polynomial of odd degree has at least one real root.

Even degree polynomials: If the degree is even then the existence of at least one real root is no longer guaranteed. This is because there is a value $M$ so that the range of the polynomial is either $[M, \infty)$, if $a > 0$, or $(-\infty, M]$, if $a < 0$. This $M$ is analogous to the vertex of quadratic polynomials. Visually we see this since the "ends" of the graph both live in the top half ($a > 0$) or bottom half ($a < 0$) of the plane.



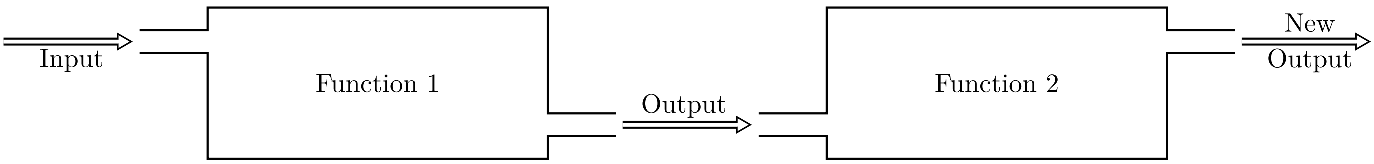
## 1.8 Composition of Functions

You should by now be happy with functions. You take an input $x$ and you apply the rule $f(x)$. But we need not just stick with the input of a variable, why not input an entire function into a function?

**Definition 1.8.1**

Let $f(x)$ and $g(x)$ be functions. Then the **composition** of $g$ with $f$, denoted by $g(f(x))$, is the function obtained by applying $g$ to $f(x)$.

Recall that functions are rules. So if we have two rules $f(x)$ and $g(x)$, then we mean by $g(f(x))$, "take $x$ and apply $f$, then take that output and apply $g$." So $g(f(x))$ is a new function that you can almost think of having 2 rules.




**Example 1.8.2**

Let $f(x) = 3x^2 - 5$ and $g(x) = 7 - 6x$. The composition of $g$ with $f$ is given by

$$g(f(x)) = 7 - 6f(x) = 7 - 6\left(3x^2 - 5\right) = 7 - 18x^2 + 30 = 37 - 18x^2.$$

The composition of $f$ with $g$ is given by

$$f(g(x)) = 3g(x)^2 - 5 = 3(7-6x)^2 - 5 = 3\left(49 - 84x + 36x^2\right) - 5 = 147 - 152x + 108x^2 - 5 = 142 - 152x + 108x^2$$

Notice that when we look at $g(f(x))$ we read from right to left. We take $x$, apply $f$ and then apply $g$.

You won't always want to calculate and simplify the expression when dealing with composition of functions. To avoid getting bogged down in algebra, just apply the rules $f$ and $g$ one at a time to get the desired output.

**Example 1.8.3**

Let $f(x) = 2x^2 - 9$ and $g(x) = x^5 - 3x^4 + 11x^2 - 12$. Calculate $g(f(2))$. You do not want to expand and simplify $g(f(x))$. Trust me. Instead lets do it one step at a time.

$$f(2) = 2(2)^2 - 9 = 2(4) - 9 = 8 - 9 = -1.$$

That was easy. Now,

$$g(f(2)) = g(-1) = (-1)^5 - 3(-1)^4 + 11(-1)^2 - 12 = -1 - 3(1) + 11(1) - 12 = -1 - 3 + 11 - 12 = -5.$$

So we calculated $g(f(2))$ without knowing what the expression for $g(f(x))$ looked like.

**Example 1.8.4**

Consider the 3 functions below.

| $x$ | 2 | 5 | 11 | 13 |
|---|---|---|---|---|
| $f(x)$ | 7 | 13 | 9 | 7 |

| $x$ | $-3$ | $-1$ | 0 | 2 |
|---|---|---|---|---|
| $g(x)$ | 5 | 1 | 7 | 8 |

| $x$ | $-2$ | 1 | 7 | 9 |
|---|---|---|---|---|
| $h(x)$ | 4 | 11 | 0 | 2 |

Verify yourself, by applying the rules from right to left, that the following are correct.

- $f(g(-3)) = 13$
- $g(h(7)) = 0$
- $h(g(0)) = 0$
- $h(f(2)) = 0$
- $g(h(f(13))) = 7$
- $f(h(g(-1))) = 9$
- $f(h(f(11))) = 7$
- $f(f(g(-3))) = 7$

## 1.9 Change

Now we start learning calculus and not algebra.

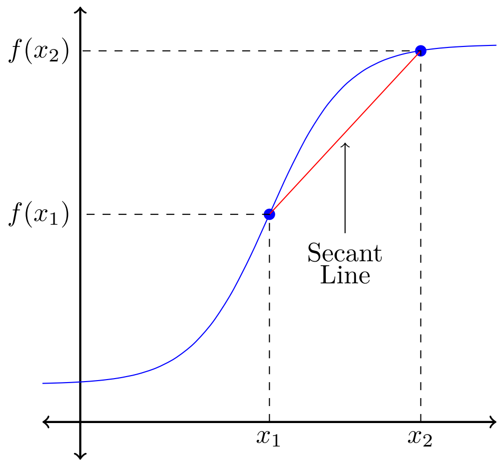
### Average Rate of Change

**Definition 1.9.1**

Let $y = f(x)$ be any function of $x$ and $x_1 \neq x_2$ be any real numbers in the domain of $f$. Then the **average rate of change** of $f(x)$ from $x = x_1$ to $x = x_2$ is given by the quotient

$$\frac{\Delta y}{\Delta x} = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

The above formula should look familiar. Indeed the slope of a linear function $f(x) = mx + c$ is given by the quotient $m = \frac{y_2 - y_1}{x_2 - x_1}$ for any two distinct points $(x_1, y_1)$, $(x_2, y_2)$ of $f$. This exactly the AROC (average rate of change) formula above where $y_1 = f(x_1)$ and $y_2 = f(x_2)$. This slope that we calculate when looking at the AROC is actually the slope of the secant line joining the two points $(x_1, f(x_1))$ and $(x_2, f(x_2))$.




So what does average rate of change mean? Recall that one way of thinking about the slope is

$$\text{slope} = \frac{\text{difference in } y}{\text{difference in } x}.$$

What this quotient represents is much does the value of $y$ change with respect to a change in $x$. For example, suppose you travel 100 miles in 2 hours. Then your average speed must have been 50mph. But you may not have necessarily travelled at 50mph for the entire journey. Maybe for the first hour you averaged 70mph and for the second hour 30mph. When we dealt with linear functions this quotient was the same regardless of what two points you picked, but for any given function it may vary depending on the interval you look at.

**Example 1.9.2**

A boat leaves the dock and travels down a river. The distance, in miles, travelled by the boat, $D$, can be modelled by the equation $D(t) = \dfrac{30t(t+1)}{t^2 + 2t + 2}$, where $t$ is the number of hours after departure. Find the average speed of the boat;

(a) Between $t = 0$ and $t = 2$,

(b) Between $t = 2$ and $t = 5$,

(c) Over the first 5 hours of its journey.

(d) Between the 5th and 10th hours of its journey.

You probably have heard or seen at some point in your life that speed $= \frac{\text{distance}}{\text{time}}$. So it shouldn't be too hard to wrap your head around distance $= D(t_2) - D(t_1)$ and time $= t_2 - t_1$ and so the average speed we are after is *exactly* the average rate of change of $D(t)$!

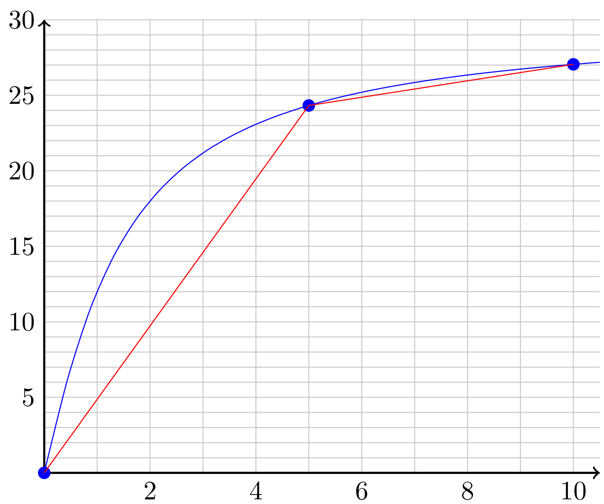
(a) Since $D(0) = 0$ and $D(2) = 18$, so speed $= \dfrac{D(2) - D(0)}{2 - 0} = \dfrac{18 - 0}{2 - 0} = \dfrac{18}{2} = 9$ miles/hour.

(b) As in (a), $D(2) = 18$ and $D(5) = \dfrac{900}{37} \approx 24.324$, so speed $= \dfrac{D(5) - D(2)}{5 - 2} \approx \dfrac{24.324 - 18}{5 - 2} = \dfrac{6.324}{3} =$ 2.108 miles/hour.

(c) We now look at the interval $t = 0$ to $t = 5$, so speed $= \dfrac{D(5) - D(0)}{5 - 0} \approx \dfrac{24.324 - 0}{5 - 0} = \dfrac{24.324}{5} = 4.8648$ miles/hour.

(d) As in (c), we look at the interval $t = 5$ to $t = 10$, since $D(10) \approx 27.049$, we have
$$\text{speed} = \frac{D(10) - D(5)}{10 - 5} \approx \frac{27.049 - 24.324}{10 - 5} = \frac{2.725}{5} = 0.545 \text{ miles/hour.}$$

Plot of $D(t)$




We see in Example 1.9.2 that the average rate of change was different over different intervals, something that did not occur with the slope of a linear function. This makes sense since the expression for $D(t)$ is clearly *not* linear. The graph to the left shows $D(t)$ over the intervals considered, and we can clearly see that the slopes of the secant lines are decreasing as the graph of the function, while still increasing, is "slowing down" in some sense.

This is going to be one of the focal points for the rest of this course. The main question you want to have in the back of your heads is "how is this function changing?" No matter what function we are dealing with, we can study this and we can interpret this.

In the case of Example 1.9.2, the speed (average rate of change), of the boat was decreasing as $t$ increased. The obvious interpretation here then is that the boat was slowing down.

## Distance, Velocity and Speed

In Example 1.9.2, we looked at the distance that the boat had travelled. Let's now modify that slightly.

**Example 1.9.3**

A boat leaves the dock and travels along a river. The distance, in miles, between the dock and the boat can be modelled by the equation $d(t) = \dfrac{9t(14-t)}{t^2+32}$, where $t$ is the number of hours after departure.

(a) Find the average rate of change of $d$ between $t = 2$ and $t = 5$,

(b) Find the average rate of change of $d$ between $t = 2$ and $t = 10$,

(c) Does speed as in Example 1.9.2 mean the same thing as it does here? That is, does the average rate of change for this scenario represent the average speed of the boat?

The calculations we need are exactly as before, $\text{AROC} = \frac{d(t_2)-d(t_1)}{t_2-t_1}$. We have $d(2) = 6$, $d(5) \approx 7.105$ and $d(10) \approx 2.727$. So,

(a) $\text{AROC} = \dfrac{d(5) - d(2)}{5 - 2} \approx \dfrac{7.105 - 6}{5 - 2} = \dfrac{1.105}{3} \approx 0.368$ miles/hour.

(b) $\text{AROC} = \dfrac{d(10) - d(2)}{10 - 2} \approx \dfrac{2.272 - 6}{10 - 2} = \dfrac{-3.728}{8} \approx -0.466$ miles/hour.

(c) The answer to part (b) was negative, so if the average rate of change for this boat represented its average speed, then we are saying that the boat travelled at a negative speed. Does that make sense? Nope. So the average rate of change here must be representing something different.

The boat clearly cannot travel at a negative speed, so what's going on here? In the first example $D(t)$ represented the total distance travelled by the boat. In the second, $d(t)$ represented the distance between the boat and the dock. The function $D(t)$ is an increasing function, which makes sense - once you've travelled a mile you can't un-travel a mile. The function $d(t)$ however is not an increasing function (see its graph on the next page). Over some initial interval the function is increasing, so the boat must be moving *away* from the dock. But at some point the function starts to decrease. So if the distance between the dock and the boat is decreasing, the boat must be moving *towards* the dock. So this boat must've turned around at some point - maybe they forgot their lunch. It could well be that the first boat turned around, but the information encoded into $D(t)$ cannot tell us this with any certainty.

So what does the average rate of change in the second example represent? When direction is incorporated into distance, we do not simply talk about speed, we also talk about *velocity*.

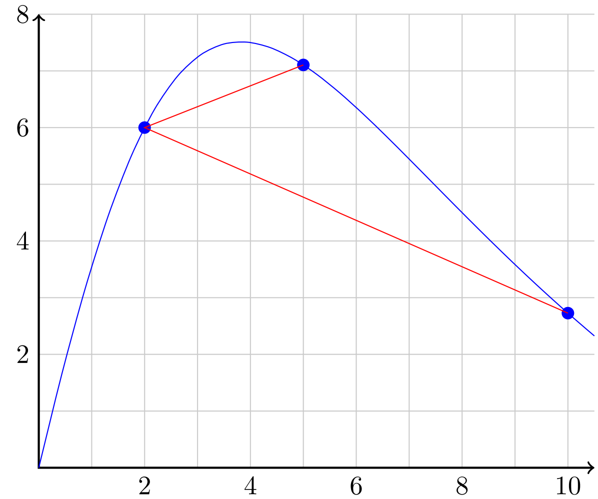
**Definition 1.9.4**

**Velocity** is the speed of something in a given direction.

$$\begin{array}{c}\text{Average}\\\text{Velocity}\end{array} = \frac{\text{Change in distance}}{\text{Change in time}} = \begin{array}{c}\text{Average rate of change}\\\text{with respect to time}\end{array}$$

If we wish to talk about the *speed* of the second boat we still can, the *speed* of an object is equal to the magnitude of the velocity, that is the absolute value of velocity. So in Example 1.9.3 (b), the boat travelled at a velocity of $-0.466$ miles/hour, but travelled at a speed of $0.466$ miles/hour.

Plot of $d(t)$




## Relative Change

How might we decide how significant a change is? If you lose $\$1,000$ playing poker, chances are that might hurt (if it wouldn't please make all cheques payable to Joseph C Foster). If Elon Musk lost $\$1,000$, he probably wouldn't care. Both you and Elon have lost $\$1,000$ dollars, so the change to your bank balance is the same, but the significance of this change to one of you is far less than the other. The significance depends on how much money you started with.

**Definition 1.9.5**

When a quantity $P$ changes from $P_0$ to $P_1$, we define the **relative change** to be

$$\text{Relative change in } P = \frac{\text{Change in } P}{P_0} = \frac{P_1 - P_0}{P_0}.$$

The relative change is a number without units and is often expressed as a percentage.

**Example 1.9.6**

A price decrease can be significant or inconsequential depending on the item.

(a) A gallon of gas currently costs $\$2.20$. If the price was cut by 50 cents, the relative change in price would be

$$\text{Relative change in } P = \frac{\text{Change in } P}{P_0} = \frac{0.5}{2.2} \approx 0.227.$$

So the price has been reduced by roughly 22.7%.

(b) A phone currently costs $\$595$. If the price was cut by 50 cents, the relative change in price would be

$$\text{Relative change in } P = \frac{\text{Change in } P}{P_0} = \frac{0.5}{595} \approx 0.00084.$$

So the price has been reduced by roughly 0.084%.

So if Shell cut their prices by 50 cents, you'd probably be |velocity|ing* over there right now. If Apple cut their price by 50 cents, you'd probably not move.

* If you understood this then you're a nerd. Also sorry not sorry.

We have talked about relative change before. We saw that exponential functions didn't have a constant average rate of change like linear function, but they did have a constant relative change - the growth rate $r$. So the rate of change of a linear function becomes less and less significant, but the rate of change of an exponential function holds the same significance regardless of the current value.

# Chapter 2

# Derivatives

## 2.1 Instantaneous Rate of Change

In Section 1.9 we introduced the average rate of change of a function over an interval. This lets us study the behaviour of a function over a given interval. But what if we want to study how a function is changing at a single point? For example; if you travel 100 miles over 5 hours, your average velocity would be 20mph. The average rate of change over the interval $t = 0$ to $t = 5$ would tell us this. But does this mean that at any given time during that interval you were travelling 20mph? Probably not.

**Example 2.1.1**

Suppose a ball is thrown straight up into the air. The table below gives its height after $t$ seconds. What is the velocity of the ball at exactly 1 second after it is thrown?

| $t$ (seconds) | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $y = h(t)$ (metres) | 1.8288 | 27.4320 | 43.2816 | 49.3776 | 45.7200 | 32.3088 | 9.1440 |

As in Definition 1.9.4,

$$\text{average velocity} = \frac{\text{change in distance}}{\text{change in time}}.$$

So the average velocity of the ball over the interval $[0, 1]$ would be 25.6032m/s, and over the interval $[1, 2]$ is 15.8496m/s. Thus the average velocity of the ball is in fact changing, depending on what interval we look at. But we are interested at the exact velocity of the ball at time $t = 1$. We expect then, that this velocity would be between these two values. But how to we calculate it? Let's look at the height of the ball over smaller intervals.

| $t$ (seconds) | 0.9 | 0.99 | 0.999 | 1 | 1.001 | 1.01 | 1.1 |
|---|---|---|---|---|---|---|---|
| $y = h(t)$ (metres) | 25.310592 | 27.224126 | 27.411273 | 27.432000 | 27.452726 | 27.638654 | 29.455872 |

So taking the average velocities over these smaller intervals, we obtain

$$\frac{h(1) - h(0.9)}{1 - 0.9} = \frac{2.121408}{0.1} = 21.21408, \quad \frac{h(1) - h(0.99)}{1 - 0.99} = \frac{0.207874}{0.01} = 20.7874, \quad \frac{h(1) - h(0.999)}{1 - 0.999} = \frac{0.020727}{0.001} = 20.727$$

$$\frac{h(1.1) - h(1)}{1.1 - 1} = \frac{2.02387}{0.1} = 20.23872, \quad \frac{h(1.01) - h(1)}{1.01 - 1} = \frac{0.20665}{0.01} = 20.6654, \quad \frac{h(1.001) - h(1)}{1.001 - 1} = \frac{0.020726}{0.001} = 20.726$$

So we see, by taking smaller and smaller intervals, the average velocity seems to be approaching a value between 20.726m/s and 20.727m/s. So the ball is approximately moving at a velocity of 20.7265m/s at the exact time $t = 1$.

By looking at smaller and smaller intervals we get a better idea of what the behaviour of a function is at a specific point. Intuitively, how much can you expect a function to change over some infinitesimally small interval? So by picking a super small interval, there's a good chance you will capture the behaviour of the function at the one point you're interested in. The value that arises has a name

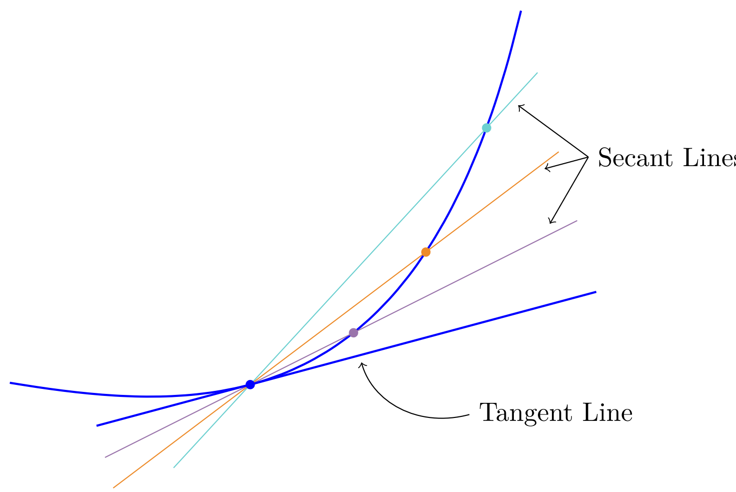
**Definition 2.1.2**

The **instantaneous rate of change** of a function $f(x)$ at $x = a$, or simply the **rate of change** of $f(x)$ at $x = a$, is defined to be the limit of the average rates of change of $f$ over shorter and shorter intervals.

The average rate of change of a function $f$ over the interval $[a, b]$ is equal to the slope of the *secant line* crossing the points $(a, f(a))$ and $(b, f(b))$. The instantaneous rate of change of the function at $x = a$ is the slope of the *tangent line* to $f$ at the point $(a, f(a))$.

The tangent line to $f$ at $a$ is the limit of the secant lines crossing $(a, f(a))$ and $(b, f(b))$ as $b$ approaches $a$. Intuitively, the tangent line is the secant line that would arise if we could use the same point twice to calculate the slope of the secant line. Of course we cannot *actually* do this since

$$\text{slope} = \frac{f(b) - f(a)}{b - a} \overset{\text{if } b=a}{\Longrightarrow} \frac{f(a) - f(a)}{a - a} = \frac{0}{0} =$$ ¯_(ツ)_/¯




**Example 2.1.3**

The quantity of a drug, in mg, in the blood at time $t$ minutes is modelled by the equation $Q(t) = 25(0.8)^t$. Let's approximate the instantaneous rate of change of the quantity of drug at time $t = 5$ minutes.

We will obtain a good estimate of the instantaneous rate of change by choosing a really small interval to calculate the average rate of change over. So lets choose the interval $5 \leq t \leq 5.001$. Then,

$$\text{Average rate of chage} = \frac{\text{Change in } Q}{\text{Change in } t} = \frac{25(0.8)^{5.001} - 25(0.8)^5}{5.001 - 5} = \frac{8.190172 - 8.192000}{5.001 - 5} = \frac{-0.001828}{0.001} = -1.827788\text{mg/m}.$$

So, we might assume then that the instantaneous rate of change of the quantity of the drug is about $-1.827788$mg/m.

What does that mean in this scenario? Well, the value is negative so the quantity of the drug must be *decreasing*, so then we can say the drug is leaving the blood at a rate of 1.827788mg/m at time $t = 5$ minutes.

## 2.2 The Derivative

The instantaneous rate of change of a function $f(x)$ is something we will study in great detail, and is one of the fundamental parts of calculus. So, instead of calling it the instantaneous rate of change of $f(x)$ all the time (which you've gotta admit is a bit of a mouthful) we call it the *derivative* of $f(x)$.

**Definition 2.2.1**

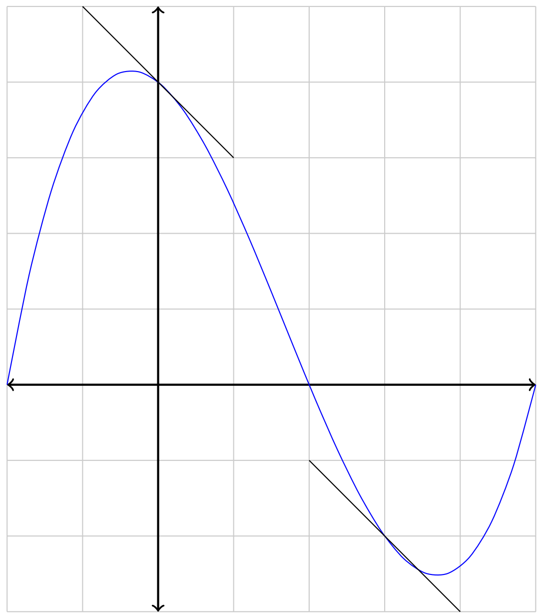
The **derivative of $f$ at $a$**, written $f'(a)$, is defined to be the instantaneous rate of change of $f$ at the point $a$.

In general we don't just want to study the derivative of $f(x)$ at just one point. We want to study the derivative of $f(x)$ at all of the points.

**Definition 2.2.2**

The **derivative of $f$** is the function $f'(x)$ with the rule

$$f'(x) = \text{the instantaneous rate of change of } f \text{ at } x$$



Remembering that the derivative of a function at $x = a$ is equal to the slope of the tangent line to the function at that point, we can approximate some of the values of the derivative just by looking at the graph.

Look at the slopes drawn on the function to the left. They may not exactly be tangent lines to the graph, but based on their position we could estimate the tangent lines to the points $x = 0$ and $x = 3$ to be roughly $-1$.

So we don't have an exact value for the derivatives at these two points. But we would be right in saying that whatever the values of these slopes, they do have to be at least negative. Is there significance to this? What about if we looked at a different point. If we try and draw a tangent line at the point $x = 4.5$ then that slope would be positive. What is the relationship?

Recall that when the slope $m$ of a linear function is positive, the graph showed an increasing function, and when $m$ was negative the graph showed a decreasing function.

Note too that here we have negative slopes on points in an interval where this function decreases and a positive slope occurring at points in an interval where the function is increasing.

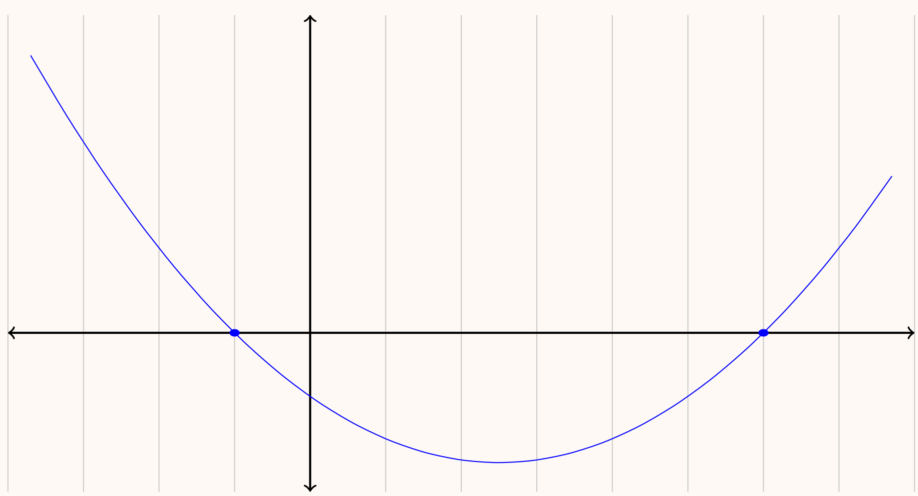
The sign of the value of the derivative at a point tells us the behaviour of the function at that point.

- If $f'(a) > 0$ then $f(x)$ is increasing at $x = a$
- If $f'(a) < 0$ then $f(x)$ is decreasing at $x = a$
- If $f'(a) = 0$ then $f(x)$ is stationary at $x = a$

When we say *stationary* at $x = a$ we mean that the tangent line to $f(x)$ at $x = a$ is perfectly horizontal, i.e. *constant*. We can learn a lot about what the graph of a function looks like, knowing its derivative.

**Example 2.2.3**

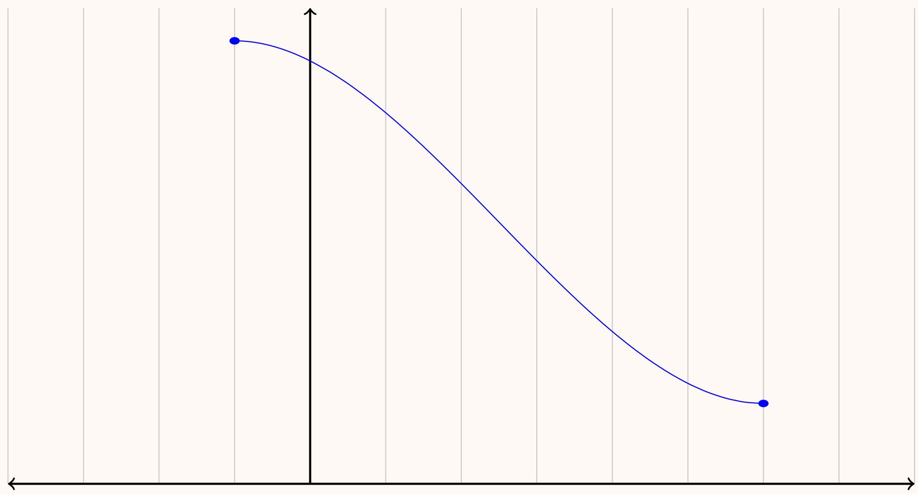
The graph of $f'(x)$ is given below.



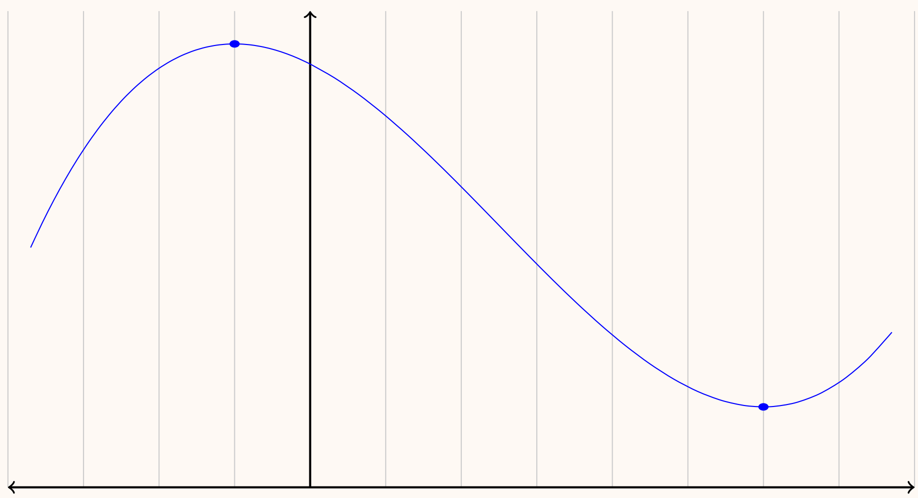
We see that the intervals in which $f'(x) > 0$ are $x < -1$ and $x > 6$. We don't know the values of $f'(x)$ here, but for now we only care that they are positive. We also have that $f'(x) < 0$ in the interval $-1 < x < 6$ and $f'(x) = 0$ at $x = -1$ and $x = 6$. So, using what we talked about above,

| $x < -1$ | $x = -1$ | $-1 < x < 6$ | $x = 6$ | $6 < x$ |
|---|---|---|---|---|
| $f'(x) > 0$ | $f'(x) = 0$ | $f'(x) < 0$ | $f'(x) = 0$ | $f'(x) > 0$ |
| $f(x)$ increasing | $f(x)$ stationary | $f(x)$ decreasing | $f(x)$ stationary | $f(x)$ increasing |

Let's put this together to try and guess how $f(x)$ might look. We know $f(x)$ is stationary at $x = -1$ and $x = 6$, and between those points, $f(x)$ is *decreasing*. So we have something like



On the other intervals, namely $x < -1$ and $x > 6$ we have that $f(x)$ is *increasing*, so then we have something like



Of course this is simply an approximation of what $f(x)$ might look like. The overall shape would look like what we have drawn, but since we do not know anything about the values of the function, so the scale may be off. It could be translated up or down or stretched in the $y$-axis.

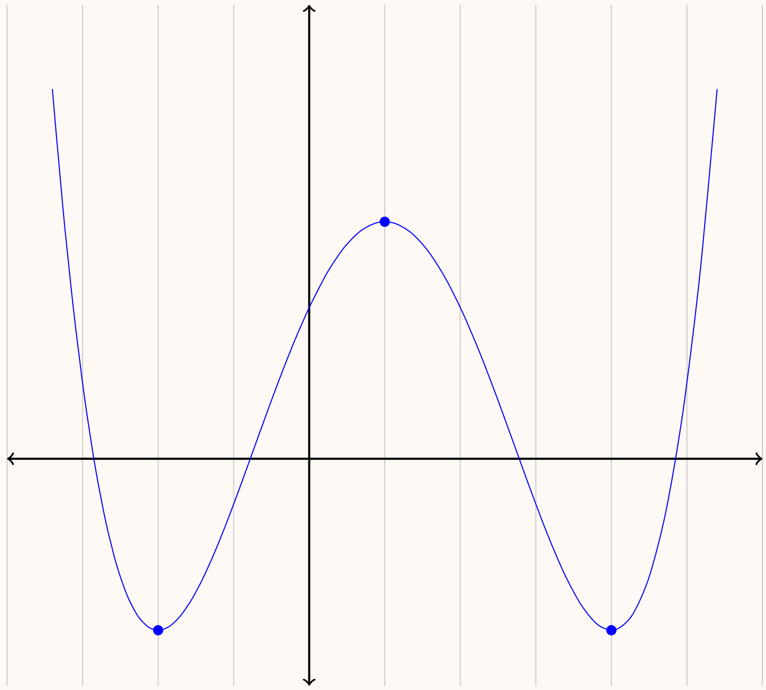
**Remark**

The important thing to remember when sketching a function from its derivative is that we only care about the *values* of the derivative and not whether or not the derivative itself is increasing/decreasing/stationary. In Example 2.2.3 in the interval $x < -1$, the graph of $f'(x)$ itself is clearly decreasing. *But* because the derivative is positive over that interval the function $f(x)$ is increasing. So it is possible that the behaviour of the derivative is different to the behaviour of the function at a given point.

We can also construct the derivative of a function given the graph of $f(x)$. Here we just read the relationships in the yellow box on the previous page from right to left.

**Example 2.2.4**

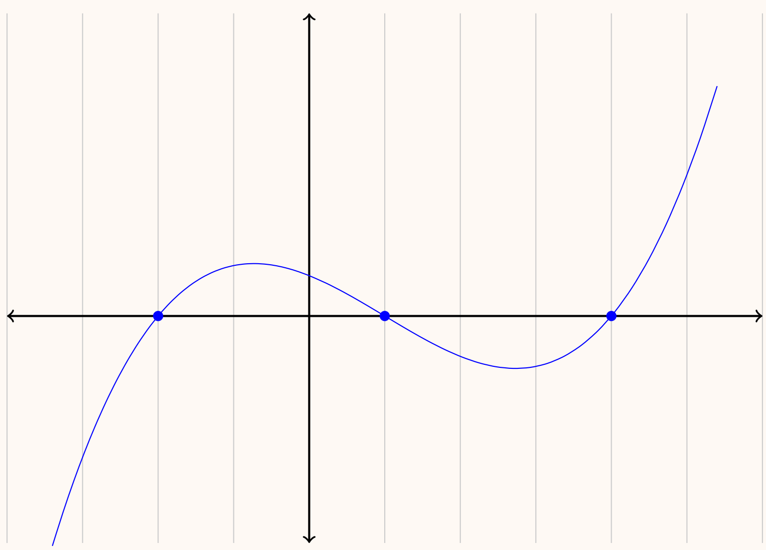
The graph of $f(x)$ is given below.



If we would like to construct the derivative from the function, we start not by looking at where $f(x) > 0$ etc., but here we look at where $f(x)$ is increasing/decreasing/stationary and then interpret that in terms of the derivative. So, we see for this function, $f(x)$ is increasing on the intervals $-2 < x < 1$ and $4 < x$, decreasing on the intervals $x < -2$ and $1 < x < 4$ and stationary at $x = -2, 1$ and 4. So we have,

| | | | | | | |
|---|---|---|---|---|---|---|
| $x < -2$ | $x = -2$ | $-2 < x < 1$ | $x = 1$ | $1 < x < 4$ | $x = 4$ | $4 < x$ |
| $f(x)$ decreasing | $f(x)$ stationary | $f(x)$ increasing | $f(x)$ stationary | $f(x)$ decreasing | $f(x)$ stationary | $f(x)$ increasing |
| $f'(x) < 0$ | $f'(x) = 0$ | $f'(x) > 0$ | $f'(x) = 0$ | $f'(x) < 0$ | $f'(x) = 0$ | $f'(x) > 0$ |

Drawing the derivative is arguably easier than the function. Simply plot the roots - i.e. the stationary points $x = -2, 1$ and 4 - and then connect them up by looping above the $x$-axis if $f(x)$ is increasing and below if $f(x)$ is decreasing.



Again this is just an approximation of what the derivative may look like.

## 2.3 Interpretations of the Derivative

### Alternative Notation

We have determined that $f'(x)$ gives the slope of the tangent to $f(x)$ at $x$. Recall that the slope (of a linear function) is given by

$$\text{slope} = \frac{y_2 - y_2}{x_2 - x_1} = \frac{\text{difference in } y}{\text{difference in } x}.$$

Gottfried Leibniz, a German mathematician who was one of the inventors of calculus, introduced different notation for the derivative. His notation is

$$f'(x) = \frac{dy}{dx}$$

where $y = f(x)$. We read this as "the derivative of $y$ with respect to $x$." This is similar to the slope formula above, where we can think of the $d$ as standing for *difference*.

Writing the derivative in this "fraction" way is useful for determining the units for the derivative. With $y$ on top and $x$ on the bottom, the units for $\frac{dy}{dx}$ are "units of $y$ per units of $x$." For example, if $y$ is distance, in miles say, a function of time $t$, in hours, then the derivative $\frac{dy}{dt}$ has the units "miles per hour." Knowing what the units for the derivative are helps with the interpretation of what the derivative represents.

**Example 2.3.1**

The cost $C$, in dollars, of building a fence of length $L$ metres is given by the function $C = f(L)$. What is the meaning of the statement $f'(11) = 15$?

Since $C$ is measured in dollars and $L$ is measured in metres, the units of $f'(L)$ are given by

$$\frac{dC}{dL} = \frac{\text{dollars}}{\text{metre}}.$$

We can think of $dC$ as the change in cost to build a fence $dL$ metres longer. So if the fence currently has a length of 11 metres, it would cost, approximately, an extra 15 dollars to build a fence of length 12 metres.

**Example 2.3.2**

The gas consumption, $C$, of a certain car is given as a function of the speed, $s$ of the car. That is, $C = f(s)$. If consumption is measured in gallons per gallons per hour and speed is measured in miles per hour, what is the meaning behind $f'(20) = -0.5$?

Since $C$ is measured in gallons per hour and $s$ is measured in miles per hour, the units of $f'(s)$ are given by

$$\frac{dC}{ds} = \frac{\text{gallons/hour}}{\text{miles/hour}} = \frac{\text{gallons}}{\text{hour}}\frac{\text{hour}}{\text{mile}} = \frac{\text{gallons}}{\text{mile}}.$$

We can think of $dC$ as the change in consumption if you were to go $ds$ more miles per hour. So if you are currently travelling at 20 miles per hour, the consumption to go 21 miles an hour would decrease by, approximately, 0.5 gallons per mile.

**Example 2.3.3**

The number of thousands of tons of zinc produced, $T$, is given as a function of the price in dollars per ton of zinc, $p$. That is, $T = f(p)$. What is the meaning of $f'(900) = 0.2$?

Since $T$ is measured in thousands of tons and $p$ is measured in dollars, the units of $f'(p)$ are given by

$$\frac{dT}{dp} = \frac{\text{thousand tons}}{\text{dollar}}.$$

We can think of $dT$ as the change in tons produced if the price was to increase by $dp$ dollars. So if the price per ton of zinc rises from \$900 to \$901 dollars, the production would increase by, approximately, 0.2 thousand tons (200 tons).

## Approximating a Function using its Derivative

The value of the derivative of a function at a certain point tells us something about how the function is changing at that point. In Example 2.3.1, when the fence is 11 metres long, the cost to increase the length to 12 metres was \$15. This is because $f'(11) = 15$ - the cost is changing at a rate of \$15 per metre. So here we used the current behaviour of a function to approximate a future value of the function. This approximation is called the *tangent line approximation*.

**Definition 2.3.4**

If $y = f(x)$ and $h$ is near 0, then $f(x+h) \approx f(x) + hf'(x)$. This is called the **tangent line approximation** of $f(x)$.

This "near 0" business is just saying that, using current behaviour to predict a future value of a function should be accurate over a short interval. For example, if a car is currently travelling at 20 metres per second, the amount of distance it will cover in the next second should be about 20 metres, since you can't really expect the car to accelerate to a much higher speed in just one second.

**Example 2.3.5**

Consider the following values of $f(x)$ and $f'(x)$.

| $x$ | 0 | 4 | 8 | 12 | 16 |
|---|---|---|---|---|---|
| $f(x)$ | 5 | 15 | 25 | 20 | 13 |
| $f'(x)$ | 2 | 3 | $-1$ | $-2$ | $-3$ |

- We can approximate $f(1)$ using the tangent line approximation and the value of $f(x)$ at $x = 0$.

$$f(1) \approx f(0) + 1 \cdot f'(0) = 5 + 1 \cdot 2 = 5 + 2 = 7.$$

- If we tried to use the same value of $f(x)$ to approximate $f(4)$, we obtain

$$f(4) \approx f(0) + 4 \cdot f'(0) = 5 + 4 \cdot 2 = 4 + 8 = 13.$$

  But note that this approximation really is just an approximation, since we know from the table that $f(4) = 15$.

- The same idea will work for other values,

$$f(9) \approx f(8) + 1 \cdot f'(8) = 25 + 1 \cdot (-1) = 25 - 1 = 24$$
$$f(17) \approx f(16) + 1 \cdot f'(16) = 13 + 1 \cdot (-3) = 13 - 3 = 10$$

- We can also work backwards,

$$f(3) \approx f(4) - 1 \cdot f'(4) = 15 - 1 \cdot 3 = 15 - 3 = 12$$
$$f(11) \approx f(12) - 1 \cdot f'(12) = 20 - 1 \cdot (-2) = 20 + 2 = 22$$

The idea of approximating a functions value based on known values will be revisited again in Section 2.5.

## Relative Rate of Change

In Chapter 1 we looked at average rate of change and relative change. Now we have taken the idea of average rate of change a step further and discovered the derivative, lets do the same for relative change.

**Definition 2.3.6**

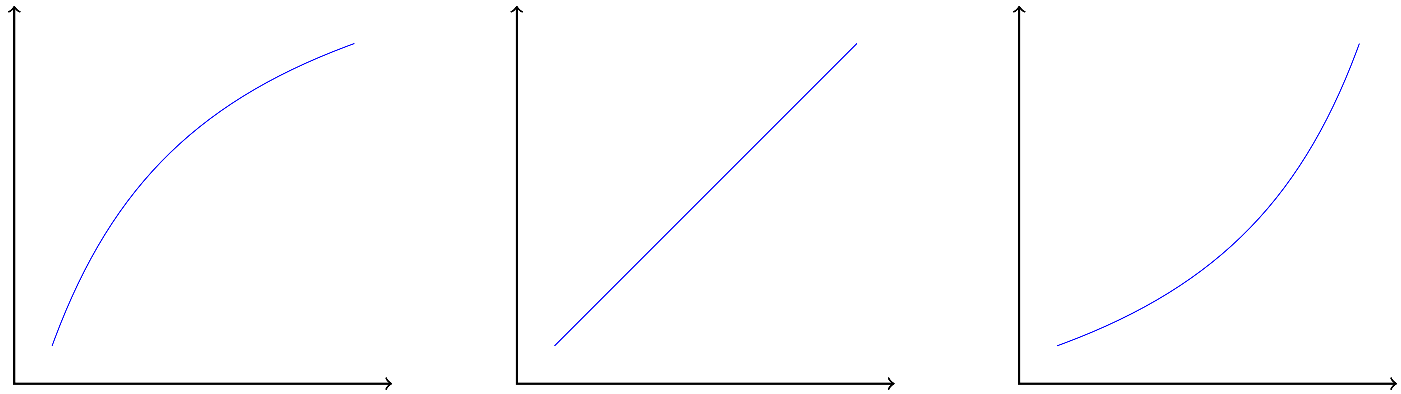
The **relative rate of change** of $y = f(x)$ at $x = a$ is defined to be

$$\frac{\text{relative change}}{\text{of } y \text{ at } a} = \frac{dy/dx}{y} = \frac{f'(a)}{f(a)}.$$

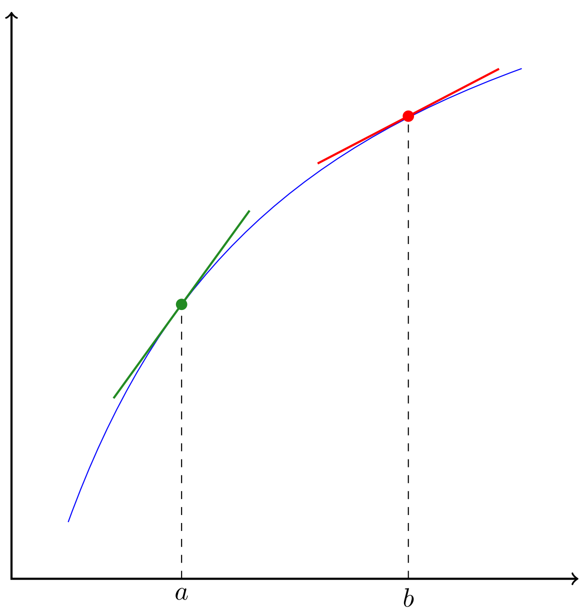
To continue the analogues; in Chapter 1 we observed that linear functions have constant average rates of change and exponential functions have constant relative change. In Chapter 2 so far we have seen that linear functions also have constant derivative. You may guess then that exponential functions have a constant relative rate of change.

## 2.4 The Second Derivative

Suppose I ask you to draw an increasing function. Suppose then I ask your friend to draw an increasing function. Suppose then that I ask your great aunt Margaret to draw an increasing function. What I get back could be these,



These would all be valid answers wouldn't they? Each function is increasing, and that was the only requirement. So what other information could I have given to ensure that all three of you drew the first function? Let's look closer,




Notice that if we pick two random points $a < b$, that the slope at $(a, f(a))$ is steeper than the slope at $(b, f(b))$. This is true for *any* two points on this arc. Saying that the slope at $(a, f(a))$ is steeper than at $(b, f(b))$ is equivalent to saying $f'(a) < f'(b)$. So, if we view $f'(x)$ as a function, we see that $f'(x)$ must be *decreasing*.

So if $f'(x)$ is decreasing, the *derivative of* $f'(x)$ must be negative. That is, derivative of the derivative is negative. This is called the *second derivative*.

**Definition 2.4.1**

The **second derivative** of a function $f(x)$, denoted $f''(x)$ is the derivative of the derivative. That is,

$$f''(x) = \frac{d}{dx} f'(x).$$

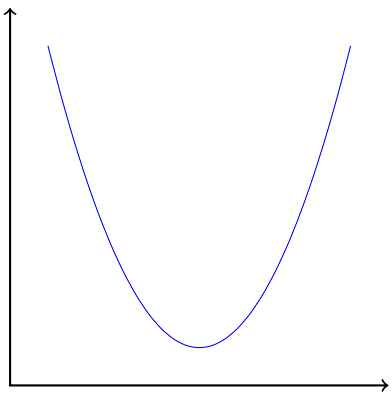
So the first curve represents an increasing function *with a decreasing derivative*. We can apply the same logic to see that the third curve represents an increasing function with an *increasing derivative*.

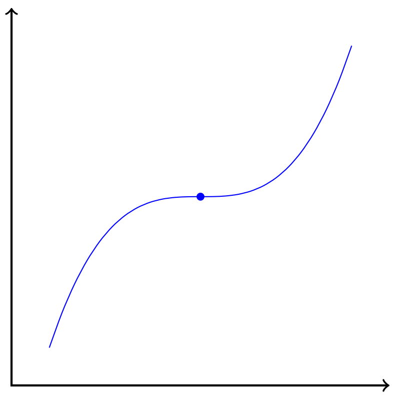
The second curve may be a little harder to interpret. But notice that this curve is a straight line, so this is a *linear* function. So the derivative of this function is constant, since the rate of change of a linear function is the same at every point, the derivative of this function is constant. So the second curve represents an increasing function with a *constant derivative*. The value that the second derivative takes is enough to describe the *concavity* of a function.

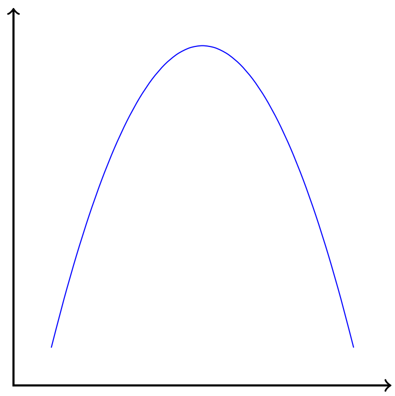
**Definition 2.4.2**

A function is said to be **concave up** at a point $x = a$ if $f''(a) > 0$.
A function is said to be **concave down** at a point $x = a$ if $f''(a) < 0$.
If $f''(a) = 0$ ,then $a$ is called a **point of inflection**. That is, a point at which the function changes **concavity**.

Concavity tells us information about the shape of the graph. Given information about the first derivative tells us the direction in which a graph is travelling, but the second derivative tells us the path the function takes. Its helpful to remember the rhyme "concave up like a cup, concave down like a frown." This is because the graph of a concave up function looks like a cup and the graph of a concave down function looks like a frown.

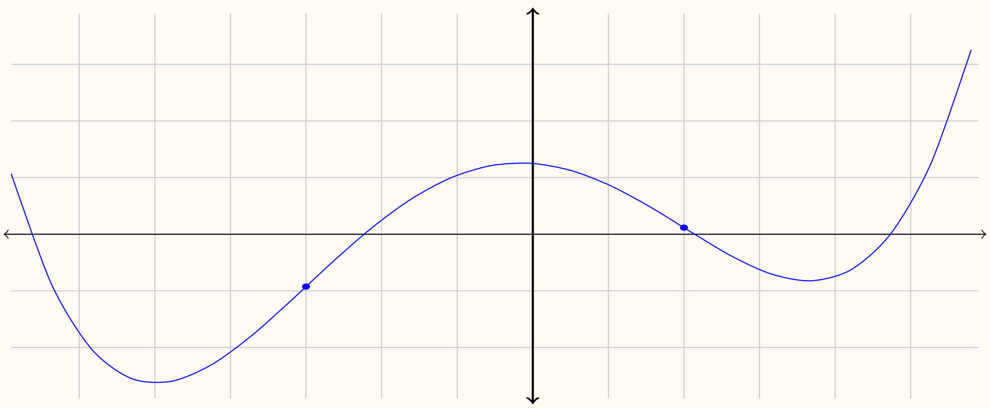

Concave up like a cup


A point of inflection


Concave down like a frown

**Example 2.4.3**

Consider the following function $f(x)$.

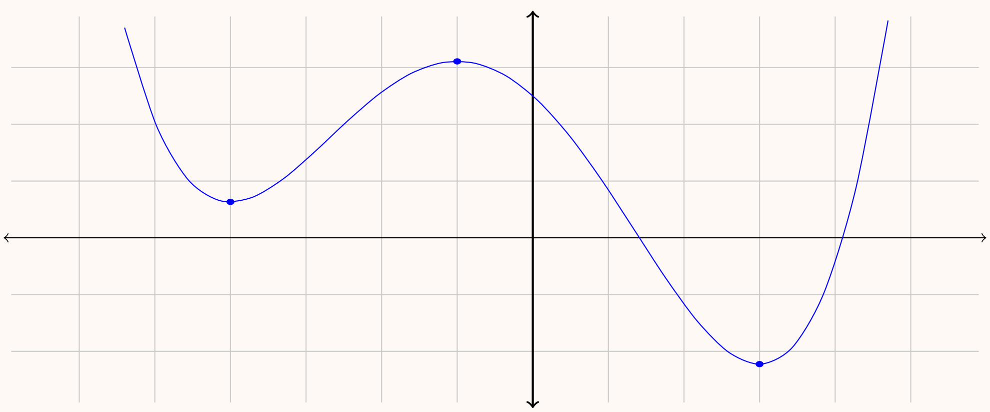


Then we have the following;

$$\begin{aligned}
f(x) \text{ is concave up over the interval } x < -3 &\iff f''(x) > 0 \text{ over the interval } x < -3 \\
f(x) \text{ changes concavity at } x = -3 &\iff f''(x) = 0 \text{ at } x = -3 \\
f(x) \text{ is concave down over the interval } -3 < x < 2 &\iff f''(x) < 0 \text{ over the interval } -3 < x < 2 \\
f(x) \text{ changes concavity at } x = 2 &\iff f''(x) = 0 \text{ at } x = 2 \\
f(x) \text{ is concave up over the interval } 2 < x &\iff f''(x) > 0 \text{ over the interval } 2 < x
\end{aligned}$$

**Example 2.4.4**

Consider the following function $f'(x)$.



Then we have the following;

$$
\begin{aligned}
f'(x) \text{ is decreasing over the interval } x < -4 &\iff f''(x) < 0 \text{ over the interval } x < -4 \\
f'(x) \text{ is stationary at } x = -4 &\iff f''(x) = 0 \text{ at } x = -4 \\
f'(x) \text{ is increasing over the interval } -4 < x < -1 &\iff f''(x) > 0 \text{ over the interval } -4 < x < -1 \\
f'(x) \text{ is stationary at } x = -1 &\iff f''(x) = 0 \text{ at } x = -1 \\
f'(x) \text{ is decreasing over the interval } -1 < x < 3 &\iff f''(x) < 0 \text{ over the interval } -1 < x < 3 \\
f'(x) \text{ is stationary at } x = 3 &\iff f''(x) = 0 \text{ at } x = 3 \\
f'(x) \text{ is increasing over the interval } 3 < x < &\iff f''(x) > 0 \text{ over the interval } 3 < x
\end{aligned}
$$

So given the original function $f(x)$, we determine values of $f''(x)$ by looking at the *shape* of $f(x)$.

Recall that if we want to know about $f(x)$ given $f'(x)$ we looked at the *values* of $f'(x)$. To deduce information about $f''(x)$ from $f'(x)$ we look at the *behaviour* of $f'(x)$ instead. The same way that we could have determined information about $f'(x)$ by looking at $f(x)$.

**Example 2.4.5**

Take another look at Example 2.2.3 and convince yourself that we did indeed draw the right graph of $f(x)$, using the properties of the second derivative.

## 2.5 Marginal Cost and Revenue

Recall in Section 1.3 we looked at cost, revenue and profit as applications of modelling with linear functions. In general cost, revenue and profit need not be linear.

Let's take cost for example. In Section 1.3 our cost functions were of the form $y = mx + c$. Here $c$ represented the *fixed costs*, such as the cost for machinery to start the production process, and $m$ represented the *variable costs*, i.e. how much each individual item cost to produce. It was called variable since the cost was directly proportional to the number of items made. Depending on the circumstances however, it may be appropriate to model the cost by a different function, such as a quadratic function or a cubic. How do we extract meaningful information from these functions.

**Definition 2.5.1**

Let $C(q)$ represent the total cost of producing $q$ units of a good or service. The **average cost** per unit is

$$AC(q) = \frac{C(q)}{q}.$$

The average cost of a product is useful in deciding what price an item should be sold at. If on average its costing \$30 to produce an item, you want to be selling them for more than \$30 dollars. In terms of variable and fixed costs, if $q$ is low (i.e. if you produce a small number of items) the average cost would be quite high, since the fixed costs are being distributed amongst a a small number of items. But increasing $q$ decreases the impact of the fixed costs on the average cost. If the average cost decreases as $q$ increases, then the company can sell their item at a lower price if they produce more.

If we wish to ignore the fixed costs and examine only how cost changes in terms of production then we go back to looking at the rate of change of a function, or the variable costs. In terms of linear cost functions we concluded that this was just $m$. But this is because linear functions are nice in that thy have a constant rate of change. If the function is not linear then this may not be a nice constant.

**Definition 2.5.2**

Let $C(q)$ represent the total cost of producing $q$ units of a good or service. The **marginal cost** is

$$MC(q) = C'(q) = \frac{dC}{dq}.$$

The marginal cost represents the immediate effect on cost if production is increased.

In a very similar manner then, it makes sense to talk about the average revenue and marginal revenue.

**Definition 2.5.3**

Let $R(q)$ represent the revenue earned by selling $q$ units of a good or service. The **average revenue** and **marginal revenue** are

$$AR(q) = \frac{R(q)}{q} \text{ and } = MR(q) = R'(q) = \frac{dR}{dq}.$$

What use are the marginal cost and revenue? Well, using the tangent line approximation as in Section 2.3 we can attempt to determine whether or not production of some item should be increased or not. Remembering that the goal is to make a profit, if we determine that an increase in production pushes the costs to be higher than the revenue, one would conclude that production should not be increased.

**Example 2.5.4**

It costs a company $C(q)$ thousand dollars to produce $q$ thousand widgets. These widgets then sell for $R(q)$ thousand dollars. If $C(17.5) = 10.1$, $MC(17.5) = 1.6$, $R(17.5) = 13.4$ and $MR(17.5) = 1.7$. Calculate the following;

1. The profit earned by producing 17.5 thousand widgets.
2. The approximate change in cost if the production increases from 17.5 to 18.1 thousand widgets.
3. The approximate change in revenue if the production increases from 17.5 to 18.1 thousand widgets.
4. The approximate change in profit if the production increases from 17.5 to 18.1 thousand widgets.

Recall the tangent line approximation if $f(x+h) \approx f(x) + hf'(x)$. So then from our definitions, $C'(q) = MC(q)$ and $R'(q) = MR(q)$. These questions are then simple applications of this.

1. Profit is what it has always been, $\pi(q) = R(q) - C(q)$. So

$$\pi(17.5) = R(17.5) - C(17.5) = 13.4 - 10.1 = \$3.3 \text{ thousand}.$$

2. Using the tangent line approximation, $C(q+h) \approx C(q) + hMC(q)$, so

$$C(18.1) \approx C(17.5) + 0.6C(17.5) = 10.1 + 0.6(1.6) = \$11.06 \text{ thousand}.$$

   So the *change* in cost is approximately $11.06 - 10.1 = 0.96$ thousand dollars.

3. Using the tangent line approximation, $R(q+h) \approx R(q) + hMR(q)$, so

$$R(18.1) \approx R(17.5) + 0.6R(17.5) = 13.4 + 0.6(1.7) = \$14.42 \text{ thousand}.$$

   So the *change* in revenue is approximately $14.42 - 13.4 = 1.02$ thousand dollars.

4. Profit is then as in a);

$$\pi(18.1) = R(18.1) - C(18.1) \approx 14.42 - 11.06 = \$4.36 \text{ thousand}.$$

   So the *change* in profit is approximately $4.36 - 3.3 = 1.06$ thousand dollars.

Companies always want to increase profit. So the thing they want to look at is the change in profit as production increases.

$$\text{Change in Profit} = \text{New Profit - Old Profit}$$

So if a company is considering increasing production from $q$ unites to $q+h$ units,

$$\pi(q+h) - \pi(q) = [R(q+h) - C(q+h)] - [R(q) - C(q)] \approx R(q) + hMR(q) - C(q) - hMC(q) - R(q) + C(q) = h\left(MR(q) - MC(q)\right).$$

So, provided $h\left(MR(q) - MC(q)\right) > 0$ the company would want to go ahead with the increase in production.

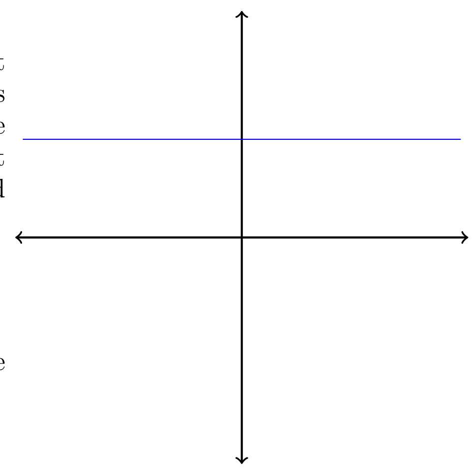
Remember that this is just an approximation however, and that the accuracy of approximation decreases as $h$ increases. So this approximation is only really used for small $h$, such as $h = 1$. So to check whether production should be increased from $q$ to $q+1$, we need only check that $MR(q) - MC(q) > 0$.

# Chapter 3

# Derivative Rules

In this Chapter we will finally learn techniques to calculate, explicitly, the derivatives of different types of functions. To do this we will learn some basic rules of the derivative and use this as building blocks to find derivatives of different kinds of functions.

## 3.1 The Constant, Power and Sum Rules



Consider a constant function $y = c$. The key thing about a constant function is that it *does not change*. As you can see in the graph to the right. Since the derivative measures change then, its easy to believe that the derivative of a constant function is simply 0. We can prove this more vigorously using some simple algebra. We saw in Section 2.1 that the derivative is simply the limit of the slopes of secant lines as the points get closer and closer. That is,

$$\text{Slope of Secant Line } = \frac{f(b) - f(a)}{b - a} \overset{b \to a}{\longrightarrow} \text{ Derivative.}$$

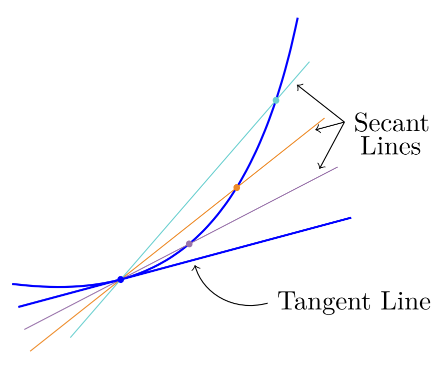
As in the tangent line approximation, we can let $h$ be a small number and consider the *difference quotient*,

$$\frac{f(x+h) - f(x)}{(x+h) - x} = \frac{f(x+h) - f(x)}{h}$$

So then if we let $h$ go to 0, we get exactly the slope at $f(x)$. So, if $f(x) = c$,

$$\frac{f(x+h) - f(x)}{h} = \frac{c - c}{h} = \frac{0}{h} = 0.$$

So now letting $h$ be zero on the right hand side we simply have 0, as expected.




**Proposition 3.1.1: Constant Rule**

For any real number $c$,

$$\frac{d}{dx}(c) = 0.$$

This difference quotient $\dfrac{f(x+h) - f(x)}{h}$ is a powerful tool in calculating some derivatives of common functions. Observe, $\underline{f(x) = x}$

$$\frac{f(x+h) - f(x)}{h} = \frac{x + h - x}{h} = \frac{h}{h} = 1 \overset{h \to 0}{\longrightarrow} 1$$

$\underline{f(x) = x^2}$

$$\frac{f(x+h)-f(x)}{h} = \frac{(x+h)^2 - x^2}{h} = \frac{x^2+2hx+h^2-x^2}{h} = \frac{2hx+h^2}{h} = 2x+h \overset{h\to 0}{\longrightarrow} 2x$$

$\underline{f(x) = x^3}$

$$\frac{f(x+h)-f(x)}{h} = \frac{(x+h)^3 - x^3}{h} = \frac{x^3+3hx^2++3h^2x+h^3-x^3}{h} = \frac{3hx^2+3h^2x+h^3}{h} = 3x^2+3hx+h^2 \overset{h\to 0}{\longrightarrow} 3x^2$$

All that we are doing here is expanding the expression $f(x+h)$, cancelling terms and once the $h$ is gone from the denominator we let it equal 0. This is precisely what we mean when we say the secant lines approach the tangent line. You may also notice a pattern,

$$\frac{d}{dx}(x) = 1 \qquad \frac{d}{dx}(x^2) = 2x \qquad \frac{d}{dx}(x^3) = 3x^2$$

In each case the resulting expression is one degree lower that what we started with. Further the 'old degree' is now the coefficient of the term that is left. If you did the algebra on $x^4, x^5, x^6$ and so on you would see this pattern continuing,

$$\frac{d}{dx}(x^4) = 4x^3 \qquad \frac{d}{dx}(x^5) = 5x^4 \qquad \frac{d}{dx}(x^6) = 6x^5.$$

Thus we obtain the *power rule*.

**Proposition 3.1.2: Power Rule**

For *any* real number $n$, we have

$$\frac{d}{dx}(x^n) = nx^{n-1}.$$

Note that in the statement of the above we let $n$ be *any* real number. We really do mean any number, not just positive integers. Using more advanced techniques than foiling the power rule holds for any decimal or negative number. If you recall your exponent laws you even get expressions for radicals.

**Example 3.1.3**

Find the derivative of each of the following:

1. $f(x) = x^{365}$
2. $g(t) = \frac{1}{t^2}$
3. $h(w) = \sqrt[3]{w}$

We simply apply the power rule to each of them.

1. $f(x) = x^{3.7}$

$$f'(x) = \frac{d}{dx}f(x) = \frac{d}{dx}(x^{3.7}) = 3.7x^{2.7}$$

2. $g(t) = \frac{1}{t^2} = t^{-2}$

$$g'(t) = \frac{d}{dt}g(t) = \frac{d}{dt}(t^{-2}) = -2t^{-3}$$

3. $h(w) = \sqrt[3]{w} = w^{\frac{1}{3}}$

$$h'(w) = \frac{d}{dw}h(w) = \frac{d}{dw}\left(w^{\frac{1}{3}}\right) = \tfrac{1}{3}w^{-\frac{2}{3}}$$

Given these rules we can start thinking about taking derivatives of polynomials, which are just sums of multiples of powers of a variable. Of course we need to develop rules that deal with multiples of powers and sums. It should not be hard to convince yourself of the next two rules if you think of a linear function $y = mx + c$. Since the slope of linear functions is constant, and the derivative at $x$ is equal to the slope $x$,

$$y' = \frac{d}{dx}y = \frac{d}{dx}(mx + c) = m.$$

Similarly then, the following hold:

**Proposition 3.1.4: Constant Multiple Rule**

Let $m$ be any real number. Then,

$$\frac{d}{dx}(mf(x)) = m\frac{d}{dx}f(x) = mf'(x).$$

**Proposition 3.1.5: Sum Rule**

$$\frac{d}{dx}\left(f(x) + g(x)\right) = \frac{d}{dx}f(x) + \frac{d}{dx}g(x) = f'(x) + g'(x).$$

**Example 3.1.6**

Find the derivative of $y = 6x^3 + 4x^2 - 2x + 7$.

$$\begin{aligned}
y' = \frac{d}{dx}y = \frac{d}{dx}&(6x^3 + 4x^2 - 2x + 7) \\
&= \frac{d}{dx}(6x^3) + \frac{d}{dx}(4x^2) + \frac{d}{dx}(-2x) + \frac{d}{dx}(7) && \text{Sum Rule} \\
&= 6\frac{d}{dx}(x^3) + 4\frac{d}{dx}(x^2) - 2\frac{d}{dx}(x) + \frac{d}{dx}(7) && \text{Constant Multiple Rule} \\
&= 6(3x^2) + 4(2x) - 2(1) + 0 && \text{Power and Constant Rules} \\
&= 18x^2 + 8x - 2 && \text{Combine}
\end{aligned}$$

## 3.2 The Derivative of Exponential and Logarithmic Functions

Obviously not all functions are polynomial-like. The two other functions we have focussed on have been Exponential and logarithmic functions. We will not go into the calculations involved to derive these rules as they use techniques not relevant to our interests. We simply state them.

**Proposition 3.2.1: Exponential Rule**

Let $a > 0$, $a \neq 1$. Then,

$$\frac{d}{dx}(a^x) = \ln(a)a^x.$$

**Proposition 3.2.2: Logarithmic Rule**

Let $a > 0$, $a \neq 1$. Then,

$$\frac{d}{dx}(\log_a(x)) = \frac{1}{\ln(a)x}.$$

When we study exponential and logarithmic functions, there are two particular ones we like the most. Namely, the *natural* exponential and logarithmic function. Recall that $e = 2.71828\ldots$ and $\ln(x) = \log_e(x)$. So then using our formulas above we obtain

$$\frac{d}{dx}(e^x) = e^x \text{ and } \frac{d}{dx}(\ln(x)) = \frac{1}{x}.$$

If you are going to remember two things from Section 3.2 it should be these two derivatives.

## 3.3 The Chain Rule

Recall that in Section 1.8 we looked at the composition of functions. That is, if $f(x)$ and $g(x)$ are functions, then the *composition of* $f(x)$ with $g(x)$ is $f(g(x))$. Supposing that we can find the derivative of both $f(x)$ and $g(x)$, how may we find the derivative of $f(g(x))$? We use what is called the *chain rule*.

**Proposition 3.3.1: Chain Rule**

Let $u(x), v(x)$ be differentiable functions. Then,

$$\frac{d}{dx}u(v(x)) = v'(x)u'(v(x)).$$

Implementing the chain rule is as easy as using the rules from Section 3.1 and 3.2 on the functions $u(x)$ and $v(x)$, and them simply combining them as in the formula above. The *hard* part of the chain rule can be identifying how to break up a function so that it takes the form $u(v(x))$. We will do some examples that will hopefully show common ways in which functions take on this form.

**Example 3.3.2**

Find the derivative of $y = 5(2t+1)^3$.

$$\begin{aligned} u(t) &= 5t^3 & v(t) &= 2t+1 & y &= u(v(t)) \\ u'(t) &= 15t^2 & v'(t) &= 2 & y' &= v'(t)u'(v(t)) \\ &&&& &= 2 \cdot 15v(t)^2 \\ &&&& &= 2 \cdot 15(2t+1)^2 \\ &&&& &= 30(2t+1)^2 \end{aligned}$$

Thus $y' = 30(2t+1)^2$.

**Example 3.3.3**

Find the derivative of $f(x) = e^{-6x}$.

$$\begin{aligned} u(x) &= e^x & v(x) &= -6x & f(x) &= u(v(x)) \\ u'(x) &= e^x & v'(x) &= -6 & f'(x) &= v'(x)u'(v(x)) \\ &&&& &= -6 \cdot e^{v(x)} \\ &&&& &= -6 \cdot e^{-6x} \\ &&&& &= -6e^{-6x} \end{aligned}$$

Thus $f'(x) = -6e^{-6x}$.

Example 3.3.3 highlights an important result, one that we will use frequently throughout,

$$\frac{d}{dx}\left(e^{v(x)}\right) = v'(x)e^{v(x)}.$$

In particular the case when $v(x) = c$, a constant function, should be noted

$$\frac{d}{dx}\left(e^{cx}\right) = ce^{cx}.$$

**Example 3.3.4**

Find the derivative of $g(w) = (4w^2 + w + 3)^7$.

$$\begin{aligned} u(w) &= w^7 & v(w) &= 4w^2 + w + 3 & g(w) &= u(v(w)) \\ u'(w) &= 7w^6 & v'(w) &= 8w + 1 & g'(w) &= v'(w)u'(v(w)) \\ &&&&&= (8w+1) \cdot 7v(w)^6 \\ &&&&&= (8w+1) \cdot 7(4w^2+w+3)^6 \\ &&&&&= (8w+1)(4w^2+w+3)^6 \end{aligned}$$

Thus $g'(w) = (8w+1)(4w^2+w+3)^6$.

## 3.4 The Product and Quotient Rules

We begin our discussion of products and quotients with two non-rules. These are two of the most common mistakes people make with the derivative and every time you do them a mathematician cries,

**Remark: Don't ever do this**

$$\frac{d}{dx} f(x)g(x) \neq f'(x)g'(x) \text{ and } \frac{d}{dx}\frac{f(x)}{g(x)} \neq \frac{f'(x)}{g'(x)}.$$

No matter how nice this would be, it is wrong. Now lets see what the actual rules are.

**Proposition 3.4.1: Product Rule**

Let $u(x)$ and $v(x)$ be differentiable functions. Then

$$\frac{d}{dx} u(x)v(x) = u'(x)v(x) + u(x)v'(x).$$

If you remember the difference quotient $\frac{f(x+h) - f(x)}{h}$ we discussed in Section 3.1, then we can supply a quick justification as to why this is in fact the correct rule.

$$\begin{aligned} \frac{u(x+h)v(x+h) - u(x)v(x)}{h} &= \frac{u(x+h)v(x+h) - u(x)v(x+h) + u(x)v(x+h) - u(x)v(x)}{h} \\ &= \frac{u(x+h) - u(x)}{h} v(x+h) + u(x)\frac{v(x+h) - v(x)}{h} \\ &\xrightarrow{h \to 0} u'(x)v(x) + u(x)v'(x). \end{aligned}$$

Let's see it in action.

**Example 3.4.2**

Differentiate $f(x) = x^2 e^{2x}$.

$$\begin{aligned} u(x) &= x^2 & v(x) &= e^{2x} & f(x) &= u(x)v(x) \\ u'(x) &= 2x & v'(x) &= 2e^{2x} & f'(x) &= u'(x)v(x) + u(x)v'(x) \\ &&&&&= 2x \cdot e^{2x} + x^2 \cdot 2e^{2x} \\ &&&&&= 2xe^{2x} + 2x^2e^{2x} \\ &&&&&= 2(1+x)xe^{2x} \end{aligned}$$

Thus $f'(x) = 2(1+x)xe^{2x}$.

**Example 3.4.3**

Differentiate $y = t^3 \ln(t+1)$.

$$u(t) = t^3 \qquad v(t) = \ln(t+1) \qquad y = u(t)v(t)$$

$$u'(t) = 3t^2 \qquad v'(t) = \frac{1}{t+1} \qquad \begin{aligned} y' &= u'(t)v(t) + u(t)v'(t) \\ &= 3t^2 \cdot \ln(t+1) + t^3 \cdot \frac{1}{t+1} \\ &= 3t^2 \ln(t+1) + \frac{t^3}{t+1} \end{aligned}$$

Thus $y' = 3t^2 \ln(t+1) + \frac{t^3}{t+1}$.

Now let us deal with quotients. If you believe the chain and product rules then we can derive the rule for quotients, simply by applying both of them. Indeed,

$$\frac{d}{dx}\frac{u(x)}{v(x)} = \frac{d}{dx} = u(x)[v(x)]^{-1} = u'(x)[v(x)]^{-1} + u(x) \cdot (-v'(x))\,[v(x)]^{-2} = \frac{u'(x)v(x) - u(x)v'(x)}{[v(x)]^2}.$$

This is precisely the *quotient rule*.

**Proposition 3.4.4: Quotient Rule**

Let $u(x)$ and $v(x)$ be differentiable functions. Then

$$\frac{d}{dx}\frac{u(x)}{v(x)} = \frac{u'(x)v(x) - u(x)v'(x)}{[v(x)]^2}.$$

**Example 3.4.5**

Differentiate $f(x) = \dfrac{3x + x^2}{5 + x}$.

$$u(x) = 3x + x^2 \qquad v(x) = 5 + x \qquad f(x) = \frac{u(x)}{v(x)}$$

$$u'(x) = 3 + 2x \qquad v'(x) = 1 \qquad \begin{aligned} f'(x) &= \frac{u'(x)v(x) - u(x)v'(x)}{[v(x)]^2} \\ &= \frac{(3+2x)\cdot(5+x) - (3x+x^2)\cdot 1}{[5+x]^2} \\ &= \frac{(3+2x)(5+x) - (3x+x^2)}{(5+x)^2} \\ &= \frac{15 + 13x + x^2 - 3x - x^2}{(5+x)^2} \\ &= \frac{15 + 10x}{(5+x)^2} \\ &= 5\frac{3+2x}{(5+x)^2} \end{aligned}$$

Thus $f'(x) = 5\dfrac{3+2x}{(5+x)^2}$.

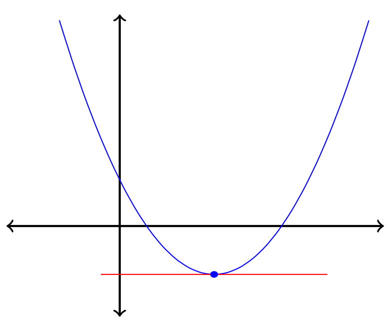
If you prefer not to use the quotient rule and instead use the product and the chain rules together, then you may and you will still obtain the same answer.

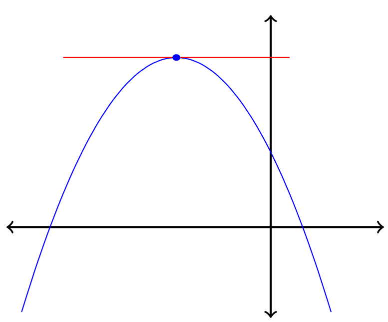
# Chapter 4

# Using the Derivative

## 4.1 Extrema

Recall from Section 2.2 that we said that a function $f(x)$ was *stationary* at $x = a$ if $f'(a) = 0$. So a stationary point occurs when the tangent line is a flat, horizontal line, as depicted in the two graphs below. These graphs are suggestive in that they show what appears to be maximum and minimum values of their respective functions. This is often the case of stationary points. In fact we now define a more general kind of point, to which stationary points fall under.





**Definition 4.1.1**

A **critical point** of $f(x)$ is a point $(a, f(a))$ such that $f'(a) = 0$ or does not exist.

Note that in our definition we write $(a, f(a))$. This implicitly says that we require $f(x)$ to be defined at the point $x = a$, even if its derivative does not. It is common that we simply refer to $x = a$ as being a critical point, rather than $(a, f(a))$, but you should remember that this is a point with both an $x$-value and a $y$-value - the $y$-value being what we will be interested in shortly.

**Example 4.1.2**

Find all critical points of $f(x) = x^3 - 6x^2 - 26x - 216$.

We must first calculate $f'(x)$ and then find the points such that $f'(x) = 0$ or is not defined.

We have that $f'(x) = 3x^2 - 12x - 26$, so solving by the quadratic formula (or factoring would work in this case) we obtain the values $x = \pm 2$ as critical points. Further, since $f'(x)$ is a polynomial, its domain is $(-\infty, \infty)$, so there are no points at which $f'(x)$ is undefined. Thus $f(x)$ has exactly two critical points, $x = 2$ and $x = -2$.

**Example 4.1.3**

Find all critical points of $f(x) = \sqrt[3]{x}$.

Note that $f(x) = \sqrt[3]{x} = x^{1/3}$ .Differentiating, we obtain

$$f'(x) = \frac{1}{3}x^{-2/3} = \frac{1}{3\sqrt[3]{x^2}}.$$

We see that $f'(x) = 0$ has no solutions. But $f'(x)$ is not defined at $x = 0$. We check that $f(x)$ is defined at $x = 0$, which it does, and conclude that the only critical point of $f(x)$ is $x = 0$.

We care about critical points as these are potential points for which a function attains its maximum and minimum values. Its easy to see why we would be interested in such values, in the contexts of business we want to maximise profit and minimise costs.

**Definition 4.1.4**

Consider a function $f(x)$ and a point $(a, f(a))$.

- The number $f(a)$ is called a **local maximum** of $f(x)$ if $f(a) \geq f(x)$ when $x$ is 'close to' $a$.
- The number $f(a)$ is called a **local minimum** of $f(x)$ if $f(a) \leq f(x)$ when $x$ is 'close to' $a$.

What do we mean when we say 'close to'? Well, we mean that there is some open interval, no matter how small, such that the claim holds. Graphically its easy to see where local maximums and minimums occur. But we can also test for them algebraically, using the *second derivative test*.

**Proposition 4.1.5: The Second Derivative Test**

Let $f(x)$ be a function such that both $f'(x)$ and $f''(x)$ exist. Then,

- If $f'(a) = 0$ and $f''(a) > 0$ then $f(x)$ has a local minimum at $x = a$.
- If $f'(a) = 0$ and $f''(a) < 0$ then $f(x)$ has a local minimum at $x = a$.

Note that in the case that $f''(a) = 0$, the test is inconclusive. Though we do know that $x = a$ would be a point of inflection of $f(x)$.

**Example 4.1.6**

Find and classify all critical points of $f(x) = 3x^5 - 5x^3$.

First we solve $f'(x) = 0$, so

$$f'(x) = 15x^4 - 15x^2 = 15x^2(x^2 - 1) = 15x^2(x + 1)(x - 1) = 0 \Longrightarrow x = 0, -1, 1.$$

So we have 3 critical points. To classify them we can simply apply the second derivative test. We have $f''(x) = 60x^3 - 30x = 30x(2x^2 - 1)$, so

$$\begin{aligned} f''(0) &= 30 \cdot 0(2 \cdot 0^2 - 1) = 0 \Longrightarrow \text{ inconclusive}, \\ f''(-1) &= 30 \cdot (-1)(2(-1)^2 - 1) = -30(2 - 1) = -30 < 0 \Longrightarrow \text{ local maximum}, \\ f''(1) &= 30 \cdot 1(2(1)^2 - 1) = 30(2 - 1) = 30 > 0 \Longrightarrow \text{ local minimum}. \end{aligned}$$

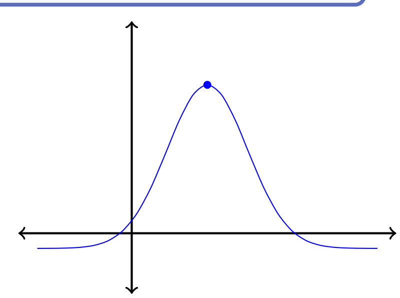
We have described *local* extrema (minima and maxima), and noted that this is defined as the smallest or largest point on some, potentially small, open interval. But what about the biggest point possible? What about a point $x = a$ such that $f(a) \geq f(x)$ for all $x$ where $f(x)$ is defined? This is precisely the notion of *global* extrema.

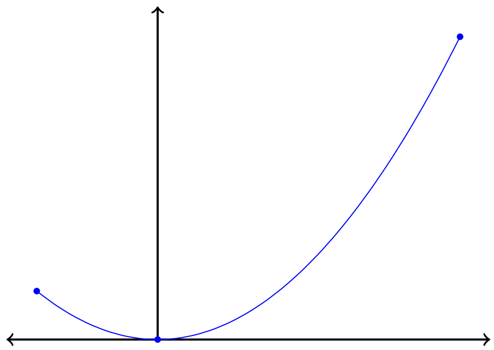
**Definition 4.1.7**

Let $a$ be a number in the domain of $f(x)$. Then,

- $f(a)$ is the **global minimum** of $f(x)$ on the domain if $f(a) \leq f(x)$ for all $x$ in the domain.
- $f(a)$ is the **global maximum** of $f(x)$ on the domain if $f(a) \geq f(x)$ for all $x$ in the domain.

Graphically, these points are very easy to pick out, if they exist. For example in the graph given to the right. They are simply the biggest and smallest points one would see. Note however that they need not exist. For example the function $f(x) = x^2$ grows without bound as $x$ increases. So there is no point on the domain bigger than every other point.





If we were to restrict the domain of $f(x) = x^2$ from all real numbers to just a closed interval, say $[-2, 5]$, then we would indeed have a global maximum. From the graph to the left its easy to see that $x = 5$ would be this global maximum. Since there are no more points to the right fro where $f(x)$ is now defined - we *restricted* its domain to the interval $[-2, 5]$. In fact given any continuous function on a closed interval, we can *always* find global extrema. We also don't have to work very hard to find them.

**Proposition 4.1.8: The Closed Interval Method for Global Extrema**

To find the global maximum and minimum of a continuous function $f(x)$ and a *closed interval* $[a, b]$, do the following;

1. Find all critical points of $f(x)$ in the interval $(a, b)$ and compute the values of $f(x)$ corresponding to these points.
2. Compute $f(x)$ at the endpoints, i.e. find $f(a)$ and $f(b)$.
3. The largest of all the points found in 1. and 2. is the global maximum of $f(x)$ on $[a, b]$ and the smallest is the global minimum of $f(x)$ on $[a, b]$.

**Example 4.1.9**

Find the global maximum and minimum values of $f(x) = x^3 - 3x^2 + 1$ on the interval $[-\frac{1}{2}, 4]$.

We will follow the closed interval method.

1. First we find the critical points of $f(x)$ that lie in the interval $(-\frac{1}{2}, 4)$.

$$f'(x) = 3x^2 - 6x = 3x(x - 2) \implies x = 0, 2.$$

   The values of $f(x)$ corresponding to these critical points are

$$f(0) = 1, \qquad f(2) = -3.$$

2. Next we compute the value of $f(x)$ at the endpoints of $[-\frac{1}{2}, 4]$.

$$f\left(-\tfrac{1}{2}\right) = \tfrac{1}{8}, \qquad f(4) = 17.$$

3. Finally we compare all of the values of $f(x)$ we have computed to see that the global minimum of $f(x)$ on $[-\frac{1}{2}, 4]$ is $f(2) = -3$ and the global maximum of $f(x)$ on $[-\frac{1}{2}, 4]$ is $f(4) = 17$.

## 4.2 Points of Inflection

We have seen points of inflection in Section 2.4. These are the points where the graph of a function *changes concavity*. Which algebraically translates to $f''(x) = 0$. Now we have the tools from Chapter 3 to find derivatives and second derivatives, we can now find points of inflections for many functions, without having to approximate them by picking a point on a graph where concavity changes.

**Example 4.2.1**

Find all inflection points of $f(x) = x^3 - 9x^2 - 48x + 52$.

We simply take the derivative twice and solve $f''(x) = 0$.

$$f'(x) = 3x^2 - 18x - 48, \qquad f''(x) = 6x - 18$$
$$f''(x) = 0 \Longrightarrow 6x - 18 = 0 \Longrightarrow 6x = 18 \Longrightarrow x = 3.$$

So $f(x)$ has a single point of inflection at $x = 3$.

**Example 4.2.2**

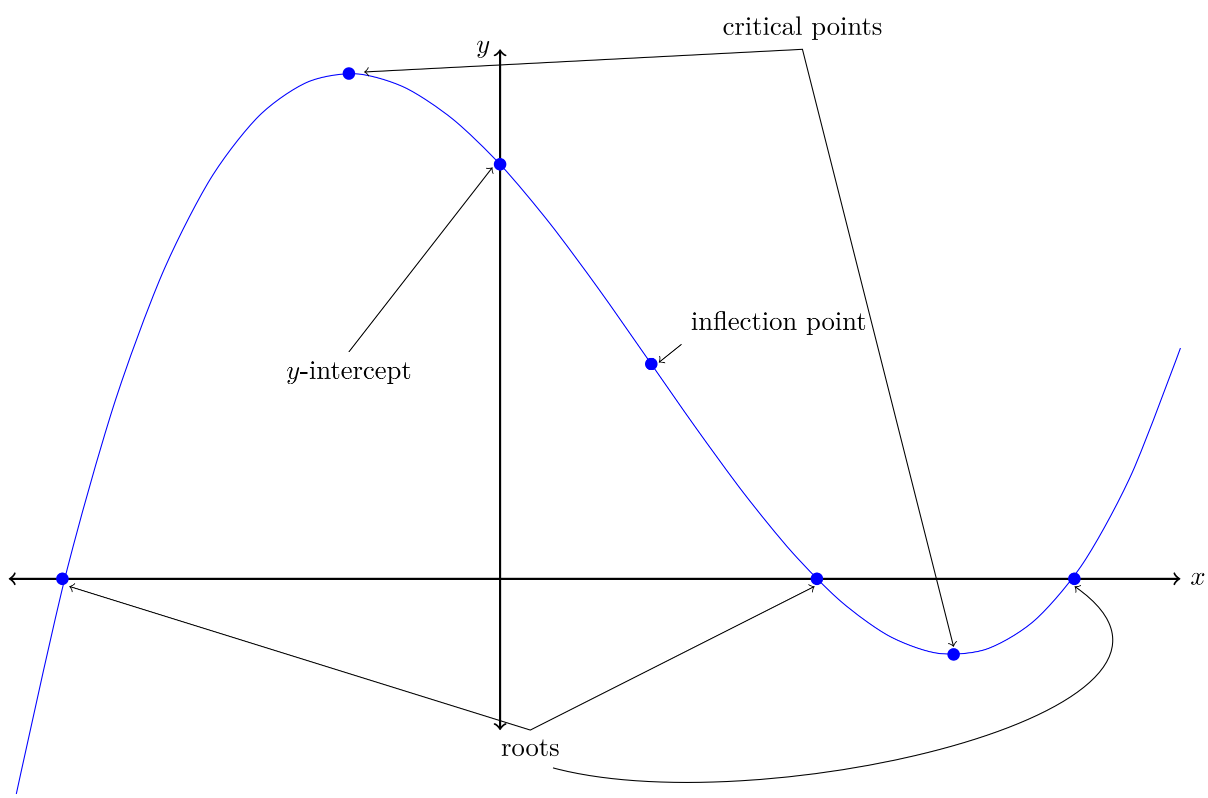
Find all inflection points of $g(t) = 3t^4 - 4t^3 + 6$.
As before, we take derivatives and solve $g''(t) = 0$.

$$g'(t) = 12t^3 - 12t^2, \qquad g''(t) = 36t^2 - 24t = 12t(3t - 2)$$
$$g''(t) = 0 \Longrightarrow 12t(3t - 2) = 0 \Longrightarrow 12t = 0 \text{ or } 3t - 2 = 0 \Longrightarrow t = 0 \text{ or } t = \frac{2}{3}.$$

So $g(t)$ has two inflection points at $t = 0$ and $t = \frac{2}{3}$.

Most of what we want to know about a function is found by solving something equal to zero. Namely; roots, critical points and inflection points.

- If $x = 0$ then $f(0)$ is the *y-intercept* of $f(x)$.
- If $f(a) = 0$ then $x = a$ is a *root* of $f(x)$.
- If $f'(a) = 0$ then $x = a$ is a *critical point* of $f(x)$.
- If $f''(a) = 0$ then $x = a$ is an *inflection point* of $f(x)$.



## 4.3 Profit, Cost and Revenue

In this section we will look at applying everything we know about the derivative to the setting of profits, costs and revenues. This is a natural setting to look at applications of the derivative to the real world as producers and companies want to *maximise profit*.

Thinking back to marginal cost and revenue, we have seen already that it is advised that companies increase production so long as $MR > MC$. If we apply what we have learnt in Chapters 3 and 4, we see that $\pi(q)$ will be maximised at the point where $\pi'(q) = 0$. But since,

$$\pi'(q) = \frac{d}{dq}\pi(q) = \frac{d}{dq}\left(R(q) - c(q)\right) = \frac{d}{dq}R(q) - \frac{d}{dq}C(q) = R'(q) - C'(q) = MR(q) - MC(q),$$

we see that profit can be maximised when $MR(q) = MC(q)$. Note it can also be *minimised* here - which would be bad.

**Example 4.3.1**

Find the quantity $q$ which maximises profit if the total revenue and cost, in dollars, are given by

$$R(q) = 5q - 0.003q^2 \text{ and } C(q) = 300 + 1.1q,$$

where $0 \leq q \leq 1000$ units.

To this we apply the closed interval method described in Section 4.1.

1. First we mind $MR(q)$ and $MC(q)$,

$$MR(q) = 5 - 0.006q \text{ and } MC(q) = 1.1.$$

So then $\pi'(q) = MR(q) - MC(q) = 5 - 0.006q - 1.1 = 3.9 - 0.006q$. Finding the critical points then yields,

$$\pi'(q) = 0 \implies 3.9 - 0.006q = 0 \implies 3.9 = 0.006q \implies 650 = q.$$

The profit at $q = 650$ is $\pi(650) = \$967.50$.

2. We must also check the endpoints, $q = 0$ and $q = 1000$.

$$q = 0 \implies \pi(0) = R(0) - C(0) = -\$300 \text{ and } q = 1000 \implies \pi(1000) = R(1000) - C(1000) = \$600.$$

3. The maximum profit we have found then is \$967.50 when $q = 650$.

It is important here to note that we do need to check endpoints given some *constraint*. It is very plausible that if a company just produced more and more of a good, that will increase their profit without bound. But in terms of real life, manufacturing an infinite number of products is not possible. There is always some constraint at play. In the previous example it was that we could not produce more than 1000 of the item. While mathematically it is possible for $\pi(q)$ to increase without bound, when applying to real life we have to think about these things.

If a company has fixed costs, that is costs that don't depend on $q$, then maximising profit is really just maximising revenue, as we see in this example.

**Example 4.3.2**

A tourist company can attract 300 customers for a day trip to the beach when tickets are sold at \$80. Every \$1 increase in price attracts 6 less customers. What is the price the company should sell a ticket at to maximise its revenue?

We start by setting up an equation relating price to demand. This is simply a linear expression, we have the point $(p, q) = (80, 300)$ and we know that $q$ decreases at a rate of $-6$ people per dollar - i.e. the slope is $-6$. So,

$$q - 300 = -6(p - 80) \implies q = 300 - 6(p - 80) \implies q = 780 - 6p.$$

Since revenue is given by $R = pq$, we can think of revenue as a function of price $p$ instead of quantity $q$ as we so often do. So, given that $q = 780 - 6p$,

$$R(p) = pq = p(780 - 6p) = 780p - 6p^2.$$

Now we simply maximise $R(p)$ as in the previous example. Note that here we can price a ticket at any value between $0 \leq p \leq 130$. Why is \$130 the most we can charge? Well, if we lose 6 customers per dollar and we can only lose 300 customers, we can only increase the price by \$50.

1. First we find the critical points of $R(p)$,

$$R'(p) = 780 - 12p = 0 \Longrightarrow 780 = 12p \Longrightarrow 65 = p.$$

   At $p = 65$ we have a revenue of $R(65) = \$25,350$.

2. The end points $p = 0$ and $p = 130$ give $R(0) = R(130) = \$0$. (This makes sense seeing as we either have free admission and no income or a high price with no customers, and so again no income).

3. So we see that revenue is maximised when tickets are priced at $p = \$65$.

# Chapter 5

# Antiderivatives and the Integral

## 5.1 Antiderivatives and the Indefinite Integral

Hopefully we now have a good idea of what the derivative is and we can calculate it. The main thing to take away is that the derivative tells us how a function *changes*. Now lets put the real life interpretations of the derivative to one side for a moment. We learnt in Chapter 3 how we can find expressions for the derivative. So actually calculating the derivative of a generic function $f(x)$ is really just some mathematical operation. An operation that we can reverse. For example, suppose we know

$$F'(x) = x^8.$$

How can we determine what $f(x)$ is? We need a way of *reversing* differentiation (taking the derivative). This is what an *anti derivative* is.

**Definition 5.1.1**

A function $F(x)$ is called an **antiderivative** of $f(x)$ if $F'(x) = f(x)$.

So what do we know? Well, we have the *power rule* from Chapter 3: $\frac{d}{dx}x^n = nx^{n-1}$. This means that the derivative of a power function is again a power function. So if $F'(x) = x^8$, it is reasonable to assume that $F(x)$ is a power function. When we apply the power rule, the power *decreases by 1*. So if $F'(x) = x^8$, the power of $F(x)$ must be $8 + 1 = 9$. So our first guess could be $F(x) = x^9$. Indeed,

$$\frac{d}{dx}(x^9) = 9x^8 \neq F'(x).$$

So $F(x)$ is not $x^9$. Recall that the power rule doesn't just decrease the power by 1, it brings down the power as a coefficient. So we need to get rid of that coefficient by dividing by, in this case, 9. So, $F(x) = \frac{1}{9}x^9$ is our next guess. This one works!

$$\frac{d}{dx}\left(\frac{1}{9}x^9\right) = 9\frac{1}{9}x^8 = x^8 = F'(x).$$

So $F(x) = \frac{1}{9}x^9$ is our answer. Great. But hang on.

$$\frac{d}{dx}\left(\frac{1}{9}x^9 + 17\right) = 9\frac{1}{9}x^8 + 0 = x^8 = F'(x).$$

So $F(x) = x^9 + 17$ too? How can $F(x)$ be two different functions at the same time? This is why in Definition 5.1.1 we say "*an* antiderivative" - there are more than 1!. In fact, it is easy to see what functions could be antiderivatives given that you know one of them. This is because they all differ by a constant.

**Definition 5.1.2**

The set of all antiderivatives of the function $f(x)$ is called the **indefinite integral** of $f(x)$ and is denoted by $\int f(x)\,dx$. It is given by,

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

where $F(x)$ is any antiderivative of $f(x)$ and $C$ is an arbitrary constant called the **constant of integration**.

**Remark**

Note that whenever we write the integral sign $\int$ we *always* write $dx$ after the expression we want to integrate. This functions the same way as the $dx$ in $\frac{d}{dx}$.

$$\frac{d}{dx}f(x) = \text{ differentiate } f(x) \text{ with respect to } x,$$
$$\int f(x)\,dx = \text{ integrate } f(x) \text{ with respect to } x.$$

We will learn how we can interpret the integral as we do the derivative in Chapter 6. For now however, we will focus on learning the techniques of integration thinking it purely as a mathematical operation.

We developed our tools for taking the derivative by learning some basic rules and piecing them together. We do the same here by merely *reversing* what we had for our derivative rules.

**Proposition 5.1.3: Rules for Integration**

1. Constant Rule for Integrals
$$\int 0\,dx = C \text{ for any real number } C$$
2. Power Rule for Integrals
$$\int x^n\,dx = \frac{1}{n+1}x^{n+1} + C \text{ for any real numbers } C \text{ and } n \text{ with } n \neq -1$$
3. Exponential Rule for Integrals
$$\int a^x\,dx = \frac{1}{\ln(a)}a^x + C \text{ for any number } a > 0 \text{ with } a \neq 1$$
4. Logarithmic Rule for Integrals
$$\int \frac{1}{x}\,dx = \ln(x) + C$$
5. Constant Multiple Rule for Integrals
$$\int mf(x)\,dx = m\int f(x)\,dx \text{ for any real number } m$$
6. Sum Rule for Integrals
$$\int f(x) + g(x)\,dx = \int f(x)\,dx + \int g(x)\,dx$$

We have not said much here, we have just listed the reversal of the rules we learnt in Chapter 3. The important thing to take away from this list is to understand that integration and differentiation behave like inverse operations of each other, just like addition and subtraction do. The only difference is this $+C$ that we throw in. Since taking the derivative means we lose all information about constants - since they become 0 - there is no definitive way of recovering that via integration. Let's see some examples putting these rules to use.

**Example 5.1.4**

Integrate $f(x) = 4x^3 + 6x^2 - 7x + 3$ with respect to $x$.

Our goal here is to find $\int 4x^3 + 6x^2 - 7x + 3\,dx$. We start by applying the sum and constant multiple rules to break this up into smaller parts,

$$\int 4x^3 + 6x^2 - 7x + 3\,dx = 4\int x^3\,dx + 6\int x^2\,dx - 7\int x\,dx + 3\int x^0\,dx.$$

Now looking at each of these integrals, we see that we just have to apply the power rule for integrals. Thus we obtain,

$$4\int x^3\,dx + 6\int x^2\,dx - 7\int x\,dx + 3\int x^0\,dx = 4\frac{1}{4}x^4 + 6\frac{1}{3}x^3 - 7\frac{1}{2}x^2 + 3x + C = x^4 + 2x^3 - \frac{7}{2}x^2 + 3x + C.$$

So our final answer is $\int 4x^3 + 6x^2 - 7x + 3\,dx = x^4 + 2x^3 - \frac{7}{2}x^2 + 3x + C$.

Note that we broke this up into smaller pieces just to illustrate that, as with the derivative, we can integrate term by term. In the upcoming examples we will not break it up.

**Example 5.1.5**

Integrate $g(x) = 6\sqrt{x} - 5\sqrt[3]{x} + 2\sqrt[3]{x^2} - 3$ with respect to $x$.

Again, our goal is to find $\int 6\sqrt{x} - 5\sqrt[3]{x} + 2\sqrt[3]{x^2} - 3\,dx$. At first this one looks scary. But if we simply use our knowledge of exponents to convert $\sqrt{\ }$ and $\sqrt[3]{\ }$ into powers, we are back to just applying the power rule.

$$\begin{aligned}\int 6\sqrt{x} - 5\sqrt[3]{x} + 2\sqrt[3]{x^2} - 3\,dx &= \int 6x^{1/2} - 5x^{1/3} + 2x^{2/3} - 3\,dx \\ &= 6\frac{1}{3/2}x^{3/2} - 5\frac{1}{4/3}x^{4/3} + 2\frac{1}{5/3}x^{5/3} - 3x + C \\ &= 6\frac{2}{3}x^{3/2} - 5\frac{3}{4}x^{4/3} + 2\frac{3}{5}x^{5/3} - 3x + C \\ &= 4x^{3/2} - \frac{15}{4}x^{4/3} + \frac{6}{5}x^{5/3} - 3x + C.\end{aligned}$$

Aside from the horrible fractions in that last one, you are hopefully beginning to see how to reverse our derivative rules. Also you are hopefully noting that $+C$ is begin religiously inserted into everything. Now lets try one that is not just the power rule.

**Example 5.1.6**

Integrate $h(x) = e^x - \frac{5}{x} + e^{-2x} + 7$ with respect to $x$.

Here we apply the exponential and logarithmic rules for integrals.

$$\begin{aligned}\int e^x - \frac{5}{x} + e^{-2x} + 7\,dx &= e^x - 5\ln(x) + \frac{1}{-2}e^{-2x} + 7x + C \\ &= e^x - 5\ln(x) - \frac{1}{2}e^{-2x} + 7x + C.\end{aligned}$$

**Remark**

From this point forward you will want to be completely comfortable with using the basic derivative rules so that you will be able to calculate these integrals with relative ease. If you still find yourself struggling, go back over the rules and practice some of the problems in the exercises section of Chapter 3.

## 5.2 Integration by Substitution

## 5.3 The Definite Integral

# Chapter 6

# Applications of the Integral

## 6.1 Distance and Accumulated Change

## 6.2 The Definite Integral as Area

## 6.3 Approximations

## 6.4 Total Change and the Fundamental Theorem of Calculus

## 6.5 Other Interpretations