Inspiring Excellence

Differential Calculus and Co-ordinate Geometry

MAT 110

Lecture Notes

Preface and Acknowledgements

This lecture note compilation has been prepared for aiding students who are taking the course MAT110 (Differential Calculus & Co-ordinate Geometry) that is offered by BRAC University. These notes are a compilation of parts taken from three other books (listed below) that has been shortened down and altered so that it is adequate for students taking this course. These notes were created under the strict supervision of eminent mathematician, Dr. Syed Hasibul Hasan Chowdhury. The main goal of this compilation is to help keep things organized for the students and ensure continuity of the course content among every section. Since this is the first version of this compilation, there may be some errors and typing mistakes. If any mistakes are found please report them to **ahmed.rakin@bracu.ac.bd**.

Ahmed Rakin Kamal

Project Coordinator:

Dr. Syed Hasibul Hasan Chowdhury
Ahmed Rakin Kamal
Sayeda Tashnuba Jahan

Typeset by:

Rahul Pal *(Editor & Template Designer)*
Ahmed Nafis Farhan
Ruhan Habib
Md. Sakib Hasan

Reference Books:

- Calculus of One Variable by Keith E. Hirst
- Calculus and Analytic Geometry by Thomas and Finney
- Calculus Early Transcendentals by James Stewart

Contents

Chapter 1

Limits and Continuity

1.1 Limits: motivation and cursory look

In this first set of lecture notes, the goal will be to develop the intuitive framework of limit relying on graphical tools. This will lead us to the concept of asymptotes, a key concept of elementary algebra and Calculus. We will, subsequently, state the rules for calculating limits algebraically. We will be a little formal along the way by systematically approaching such calculations based on various kinds of algebraic methods. A good number of examples will be introduced in this first set of notes emphasizing on developing the skill of various types of calculation rather than on a more formal treatment of the subject matter.

Informally speaking, a continuous function can be regarded as one whose curve can be drawn without taking one's pencil off the paper. Mathematically speaking, a function $f(x)$ will be continuous at $x = a$ if $\lim_{x\to a} f(x)$ exists and is equal to $f(a)$. The notation $\lim_{x\to a} f(x)$ represents the number to which the function $f(x)$ approaches as x approaches the real number a in a strict mathematical sense to be explained in detail later in the lecture.

Examples of functions with continuous curves include polynomials, exponential functions, sine and cosine, square root. Combinations of continuous functions, using operations such as addition, multiplication and composition, are also continuous.

Tangent function ($f(x) = \tan x$) and reciprocal function ($f(x) = \frac{1}{x}$) are examples of functions with discontinuities. Using the definition of the tangent function as $\tan x = \dfrac{\sin x}{\cos x}$, we would suspect that discontinuities are associated with all values of $x = (2n \pm 1)\frac{\pi}{2}$ for n being in the set of natural numbers. Indeed such points do not belong to the domain. The graphs of tan and cot in Figure 1.1 show such discontinuities clearly.

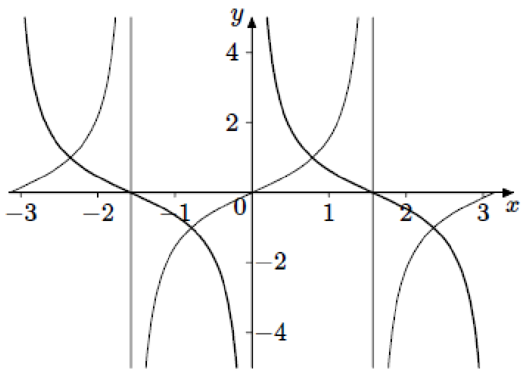

Figure 1.1: *Graphs of tangent (light) and cotangent (dark)*

Example 1.1.1.

Another example of discontinuity is illustrated in Figure 1.3. The rational expression $\dfrac{1}{x^2 - 1}$ blows up when the denominator is set to zero, i.e., when $x = \pm 1$, and this indicates a discontinuity at each of these points, as Figure 1.3 suggests. The numbers $x = \pm 1$ are not in the domain of the function. More examples of graphs exhibiting this kind of behaviour can be seen in Figures 1.2 and 1.15.

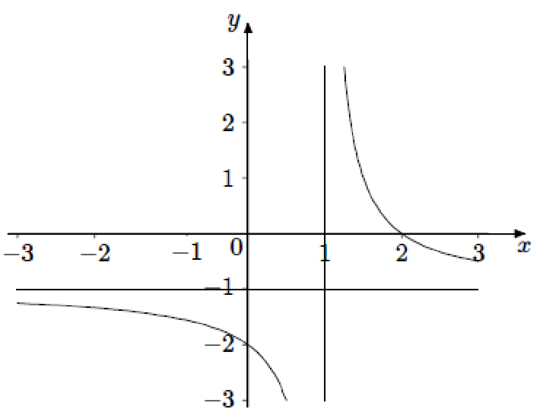

Figure 1.2: *Graph of* $f(x) = \frac{2-x}{x-1}$

As stated earlier, when we have a continuous function, as x approaches a, the value of $f(x)$ approaches the value of the function at $x = a$, namely $f(a)$. There are three possible ways in which a function $f(x)$ may become discontinuous at a given point $x = a$:

- $f(x)$ fails to approach a well-defined limiting value as x approaches a, i.e. $\lim_{x\to a} f(x)$ is not a finite real number,
- $f(a)$ is either undefined or infinite,
- both $\lim_{x\to a} f(x)$ and $f(a)$ are finite real numbers but unequal.

The last case has a special name called "removable discontinuity." As the name suggests, the discontinuity can be removed by redefining the function value by setting it equal to the limiting value of the function at the given point.

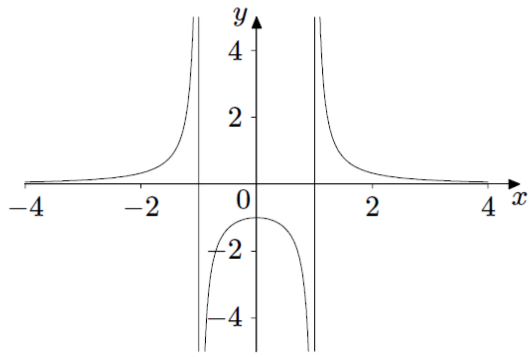

Figure 1.3: *Graph of* $\frac{1}{x^2-1}$

Remark 1. *It is not illegitimate to look for the behaviour of a function near a point which does not belong to the domain of the function, i.e., for which the function is undefined. For example, the function* $f(x) = \dfrac{\sin x}{x}$ *is not defined at* $x = 0$. *However if we plot its graph it appears to suggest that the value of* $f(x)$ *approaches 1 as* x *becomes increasingly closer to zero. The essence of this remark will be rejuvenated in Example 1.5.1. A word of caution, is however, in order here. The graph of* $f(x) = \dfrac{\sin x}{x}$ *as plotted by* ***MAPLE*** *in Figure 1.4 appears to be continuous at* $x = 0$, *although it is discontinuous there, the discontinuity being removable though.*

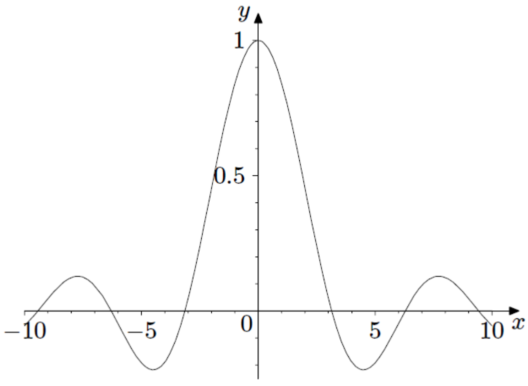

Figure 1.4: *Graph of* $f(x) = \frac{\sin x}{x}$

Upon removal of the discontinuity at $x = 0$, a continuous function $\tilde{f}$ can be constructed as follows

$$\tilde{f}(x) = \begin{cases} \frac{\sin x}{x} & \text{for } x \neq 0 \\ 1 & \text{for } x = 0 \end{cases} \tag{1.1}$$

We say that the definition of $f(x)$ has been extended by continuity.

Example 1.1.2.

Here is another example that looks at the behaviour of the function defined by $f(x) = \sin\left(\frac{1}{x}\right)$ where the point $x = 0$ does not belong to the domain. If one plots the graph, as in Figure 1.5, one finds that the function starts behaving erratically near the point $x = 0$. This is a good example exhibiting the shortcomings of a plotting device. The program calculates the value of the function at a finite number of points and joins them up. Although, for many functions the result appears to be a smooth curve, it doesn't work as expected in this example. In fact the graph oscillates between ± 1 infinitely many times as x approaches zero. The peaks of all the oscillations should have been located on the line $y = 1$ rather than the somewhat variable positions indicated on the **MAPLE** plot in Figure 1.5. The calculational details are as follows.

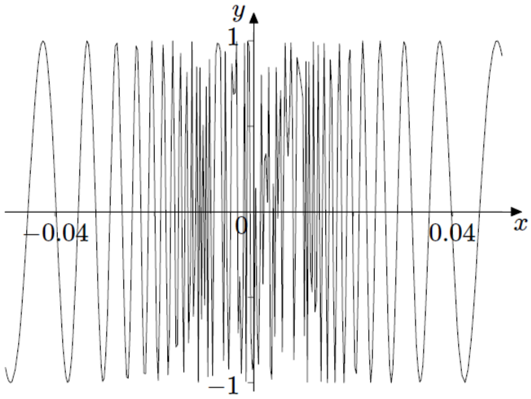

Figure 1.5: ***MAPLE*** *Plot of* $\sin\left(\frac{1}{x}\right)$

Firstly, because $-1 \le \sin t \le 1$, we can be certain that $-1 \le \sin\left(\dfrac{1}{x}\right) \le 1$.

Next we shall see where the graph crosses the x-axis. Now $\sin t = 0$ when $t = n\pi$, for all integers n. Therefore $\sin\left(\dfrac{1}{x}\right) = 0$ when $\dfrac{1}{x} = n\pi$, i.e., $x = \dfrac{1}{n\pi}$ (provided $n \neq 0$). This is a sequence of numbers which tends towards zero as n increases. More importantly, the students should intuitively understand that infinitely many terms of this sequence (due to the ones occurring after some finite value n) are all located in any arbitrarily small interval containing 0.

Now $\sin t = 1$ when $t = \dfrac{(4n+1)\pi}{2}$, for all integers n. Therefore

$$\sin\left(\frac{1}{x}\right) = 1 \tag{1.2}$$

when

$$\frac{1}{x} = \frac{(4n+1)\pi}{2};\ x = \frac{2}{(4n+1)\pi} \tag{1.3}$$

This is also a sequence of numbers which tends towards zero as n increases and, as in the previous case, infinitely many terms of this sequence will be located in any infinitesimally small interval around 0.

Finally, $\sin t = -1$ when $t = \dfrac{(4n-1)\pi}{2}$, for all integers n. Hence we have

$$\sin\left(\frac{1}{x}\right) = -1 \tag{1.4}$$

when

$$\frac{1}{x} = \frac{(4n-1)\pi}{2}; \; x = \frac{2}{(4n-1)\pi} \tag{1.5}$$

Again this is a sequence of numbers which tends towards zero as n increases and infinitely many terms of it are all densely spaced, as before, around any arbitrarily small interval about 0.

The crux of the matter is, therefore, the fact that however small an interval is chosen containing zero, there are always values of x inside that interval where $f(x) = 0$, $f(x) = 1$ and $f(x) = -1$. So, $f(x)$ does not tend to a limiting value as x tends to zero. The function turns out to be discontinuous at $x = 0$. Such discontinuity is termed as "oscillating discontinuity" since the function starts oscillating extremely rapidly as the point of discontinuity is approached from both sides.

Definition 1.1.1. We summarise the discussion and examples above in this informal definition of a limit, and introduce the two notations most commonly used for limits.

The function $f(x)$ is said to have the limit l as x approaches (or tends to) a if the values of $f(x)$ can be made as close as we like to l by taking x sufficiently close to a. We use the two notations

$\lim\limits_{x\to a} f(x) = l$ reads as "the limit of $f(x)$ as x tends to a is equal to l",
$f(x) \to l$ as $x \to a$, reads as "$f(x)$ tends to l as x tends to a".

It is assumed here that the domain of $f(x)$ includes an interval containing a, but not necessarily a itself.

The function $f(x)$ is said to be continuous at $x = a$ if $\lim\limits_{x\to a} f(x) = f(a)$, or, in the other notation, if $f(x) \to f(a)$ as $x \to a$.

The much more precise definitions employed in Real Analysis, analyse in terms of sets of real numbers the phrases "as close as we like to" and "taking x sufficiently close to" used in the informal definition above.

1.2 Left and right hand limits

At times, it may happen that the behaviour of a function near a differs, depending on whether x approaches a from below or above. Mathematically, these ideas are explained in the following way.

- As x tends to a from below (or from the left), the limit of $f(x)$ is equal to l : $\lim_{x \to a^-} f(x) = l$.
- As x tends to a from above (or from the right), the limit of $f(x)$ is equal to m : $\lim_{x \to a^+} f(x) = m$.

Example 1.2.1.

A simple example called the floor function is illustrated in Figure 1.6. Its graph consists of an array of floating stair like horizontal line segments of unit length. The value of the function at x is the greatest integer less than or equal to x. As far as its behaviour near $x = 2$ is concerned, **MAPLE** indicates, with a solid dot on the left-hand end of the topmost horizontal line segment, that the value of the function at 2 itself is 2. Just to the right of $x = 2$, the greatest integer less than or equal to 2.01, for example, is 2. In addition, just to the left of 2, the greatest integer less than or equal to 1.99 is 1. An open dot is used at the right-hand end of the second horizontal line to indicate that the value attained by the floor function at $x = 2$ is not 1.

So using the notation introduced above one can write

$$\lim_{x \to 2^-} \text{floor}(x) = 1, \lim_{x \to 2^+} \text{floor}(x) = 2. \tag{1.6}$$

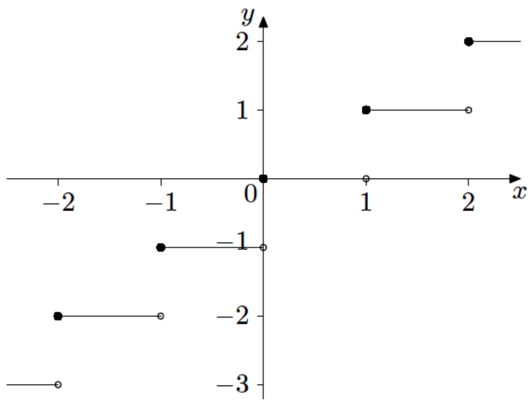

Figure 1.6: *Graph of the floor function*

The floor function has a discontinuity at each integer value of x. Such discontinuity is termed as "jump discontinuity" in contrast to the "removable discontiuity"discussed following Example 1.1.1. The value of the function jumps between two values as x moves across the point of discontinuity. This also contrasts with the oscillating discontinuity of $f(x) = \sin\left(\frac{1}{x}\right)$, discussed in Remark 1 earlier. In that case neither the left-hand limit nor the right-hand limit exists.

1.3 On asymptotes of various kinds

One may often get interested in understanding the behaviour of the graph of a function $y = f(x)$ when x or y increases or decreases without any bound. The state of x increasing without an upper bound is mathematically phrased as "x tends to infinity", symbolised by $x \to \infty$. When x decreases without any bound, we use the phrase "x tends to minus infinity", denoted by $x \to -\infty$. It is to be emphasised that the symbol ∞ does not represent a real number.

The following examples are provided to familiarize the readers with the use of this language and notation.

Example 1.3.1.

In Figure 1.7 we observe that the graph of the function x^2-1 appears to increase without bound as x tends to infinity (and to minus infinity).

Using the notation introduced above for limits, we write

$$x^2 - 1 \to \infty \text{ as } x \to \infty, x^2 - 1 \to \infty \text{ as } x \to -\infty. \tag{1.7}$$

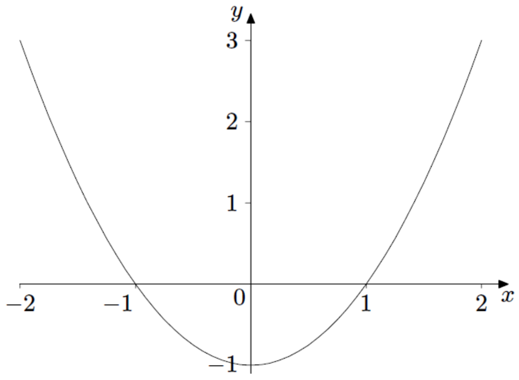

Figure 1.7: *Graph of* $x^2 - 1$

If we consider the graph of $y = x^3 + 1$ shown in Figure 1.8 we describe the limiting behaviour using the notation

$$x^3 + 1 \to \infty \text{ as } x \to \infty, x^3 + 1 \to -\infty \text{ as } x \to -\infty. \tag{1.8}$$

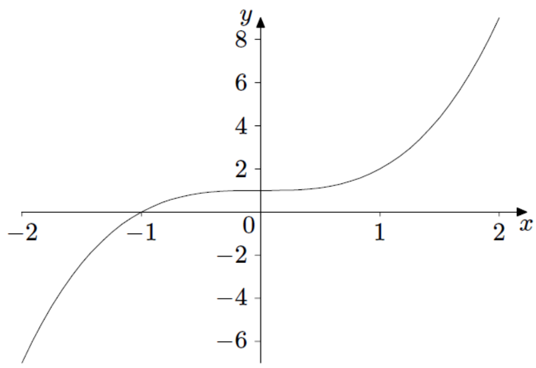

Figure 1.8: *Graph of* $x^3 + 1$

Example 1.3.2.

In the example in Figure 1.9 the function has finite limits as x tends to infinity and as x tends to minus infinity. The two notations in this case are

$$\frac{1+3e^x}{1+e^x} \to \infty \text{ as } x \to \infty, \frac{1+3e^x}{1+e^x} \to -\infty \text{ as } x \to -\infty. \tag{1.9}$$

$$\lim_{x\to\infty} \frac{1+3e^x}{1+e^x} = 3, \lim_{x\to-\infty} \frac{1+3e^x}{1+e^x} = 1. \tag{1.10}$$

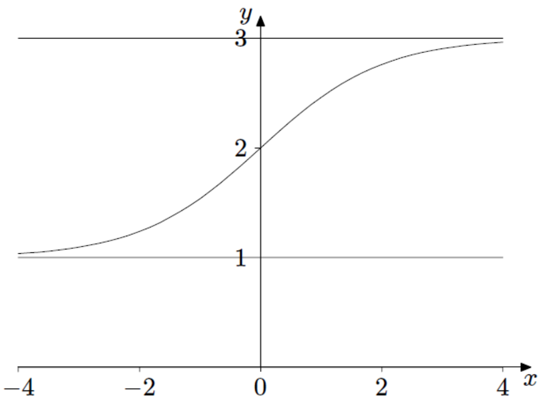

Figure 1.9: *Graph of* $\dfrac{1+3e^x}{1+e^x}$

Finally, we consider examples where the function is unbounded in the neighbourhood of some point on the x-axis.

Example 1.3.3.

Consider the graph shown in Figure 1.10

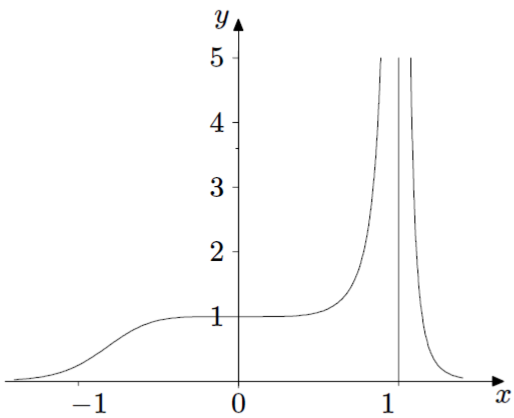

Figure 1.10: *Graph of* $\dfrac{1}{(x^5-1)^2}$

We use the notation

$$\frac{1}{(x^5-1)^2} \to \infty \text{ as } x \to 1.$$

In this case the limiting behaviour is the same to the left and to the right of $x = 1$.

If we look at the example shown in Figure 1.3 we see that the behaviour is not the same on each side of $x = 1$, or of $x = -1$. We have different one-sided limiting behaviour, symbolised by

$$\frac{1}{x^2-1} \to \infty \text{ as } x \to 1^+, \; \frac{1}{x^2-1} \to -\infty \text{ as } x \to 1^-$$
$$\frac{1}{x^2-1} \to \infty \text{ as } x \to -1^+, \; \frac{1}{x^2-1} \to -\infty \text{ as } x \to -1^-.$$

We used the idea of an asymptote graphically in a number of examples in Chapter 1. We can now give a description in terms of limits. In Figure 1.9 the graph approaches the horizontal line $y = 3$ as x tends to infinity, and it approaches $y = 1$ as x tends to minus infinity. In Figure 1.1 the graph approaches the vertical line $x = 1$ as x tends to 1 from above and below, with similar behaviour near $x = -1$. This motivates us to come up with the following definition.

Definition 1.3.1. If $f(x) \to m$ as $x \to \infty$ (or as $x \to -\infty$), then the line $y = m$ is called a **horizontal asymptote**. (This includes the possibility $m = 0$, in which case the x-axis is a horizontal asymptote.)

If $f(x) \to \pm\infty$ as $x \to a^+$ or as $x \to a^-$ then the line $x = a$ is called a **vertical asymptote**. (This includes the possibility $a = 0$, in which case the y-axis is a vertical asymptote.)

We can see further examples of horizontal asymptotes in Figures 1.2, 1.15, 1.3, 1.10. In Figure 1.4, the x-axis is a horizontal asymptote. This is clearer if the graph is plotted for a larger domain, and if the graphs of $y = \pm\frac{1}{x}$ are added. The **MAPLE** command to plot this is

```
plot([sin(x)/x,1/x,-1/x],x=-50..50,y=-0.25..1,color=black);
```

Vertical asymptotes are shown in Figures 1.2, 1.15, 1.13, 1.3.

1.4 Various Rules for calculating Limits

Example 1.4.1.

Find,

$$\lim_{x\to 1}\left(x^2+\frac{x^3+1}{x^2+1}+2x\sin(\pi\sqrt{3x^2+1})\right). \tag{1.11}$$

Since the denominator x^2+1 of the above expression of the function is never zero, its graph will be continuous everywhere so that the limit can just be found by substituting $x=1$ yielding

$$1^2+\frac{1^3+1}{1^1+1}+2\sin(\pi\sqrt{s+1})=1+1+2\sin 2\pi=2$$

Implicitly, the following rules are used to evaluate the limit above,

If

$$\begin{aligned}
&f(x)\to l \text{ as } x\to a \text{ and } g(x)\to m \text{ as } x\to a \text{ then}\\
&f(x)+g(x)\to l+m \text{ as } x\to a \text{ (addition rule);} && (1.12)\\
&f(x)-g(x)\to l-m \text{ as } x\to a \text{ (subtraction rule);} && (1.13)\\
&f(x)g(x)\to lm \text{ as } x\to a \text{ (multiplication rule);} && (1.14)\\
&\frac{f(x)}{g(x)}\to\frac{l}{m} \text{ as } x\to a \text{ (provided } m\neq 0) \text{ (division rule).} && (1.15)
\end{aligned}$$

if

$$\begin{aligned}
&f(t)\to l \text{ as } t\to a \text{ and } g(x)\to a \text{ as } x\to b\\
&\text{then } f(g(x))\to l \text{ as } x\to b \text{ (composition rule).}
\end{aligned} \tag{1.16}$$

Each of the above rules has an analogue for one-sided limits, for example

If

$$\begin{aligned}
&f(x)\to l \text{ as } x\to a^+ \text{ and } g(x)\to m \text{ as } x\to a^+\\
&\text{then } f(x)+g(x)\to l+m \text{ as } x\to a^+.
\end{aligned} \tag{1.17}$$

There is a similar set of rules for limits at infinity (and at minus infinity).

If

$$
\begin{aligned}
&f(x) \to l \text{ as } x \to \infty \text{ and } g(x) \to m \text{ as } x \to \infty \text{ then} \\
&f(x) + g(x) \to l + m \text{ as } x \to \infty \text{ (addition rule);} && (1.18)\\
&f(x) - g(x) \to l - m \text{ as } x \to \infty \text{ (subtraction rule);} && (1.19)\\
&f(x)g(x) \to lm \text{ as } x \to \infty \text{ (multiplication rule);} && (1.20)\\
&\frac{f(x)}{g(x)} \to \frac{l}{m} \text{ as } x \to \infty \text{ (provided } m \neq 0\text{) (division rule).} && (1.21)
\end{aligned}
$$

if

$$
\begin{aligned}
f(t) \to l \text{ as } t \to \infty &\text{ and } g(x) \to \infty \text{ as } x \to b \\
\text{then } f(g(x)) \to l &\text{ as } x \to b \text{ (composition rule).} \qquad (1.22)
\end{aligned}
$$

There are similar rules involving infinite limits for addition and multiplication, namely

if

$$f(x) \to \infty \text{ as } x \to a \text{ and } g(x) \to \infty \text{ as } x \to a$$

then

$$
\begin{aligned}
&f(x) + g(x) \to \infty \text{ as } x \to a \text{ (addition rule);} && (1.23)\\
&f(x)g(x) \to \infty \text{ as } x \to a \text{ (multiplication rule).} && (1.24)
\end{aligned}
$$

It is noteworthy, however, that there are no analogous rules for infinite limits involving subtraction or division. A variety of outcomes is possible, depending on the functions involved and this is where considerable care has to be taken. Example 1.4.2 demonstrates this.

Example 1.4.2.

Let $f(x) = \dfrac{1}{x^2}$ and $g(x) = \dfrac{1}{x^4}$. We can see, by plotting graphs for example, that $f(x) \to \infty$ as $x \to 0$ and $g(x) \to \infty$ as $x \to 0$. Then, $\dfrac{f(x)}{g(x)} = x^2 \to 0$ as $x \to 0$. However, $\dfrac{g(x)}{f(x)} = \dfrac{1}{x^2} \to \infty$ as $x \to 0$.

In both of the above cases of rational expressions, the numerator and denominator tend to infinity individually. However, in the first case the quotient tends to zero while in the second case the quotient tends to infinity.

1.5 Techniques that come in handy

Example 1.4.2 involves expressions for which we cannot find limits by a simple application of algebraic rules. Examples of such types of expression are:

$$\frac{f(x)}{g(x)} \text{ where } f(x) \to 0 \text{ and } g(x) \to 0 \quad \text{“}\frac{0}{0}\text{”}$$
$$\frac{f(x)}{g(x)} \text{ where } f(x) \to \infty \text{ and } g(x) \to \infty \quad \text{“}\frac{\infty}{\infty}\text{”}$$
$$f(x) \times g(x) \text{ where } f(x) \to 0 \text{ and } g(x) \to \infty \quad \text{“}0 \times \infty\text{”}$$
$$f(x) - g(x) \text{ where } f(x) \to \infty \text{ and } g(x) \to \infty \quad \text{“}\infty - \infty\text{”}$$
$$f(x)^{g(x)} \text{ where } f(x) \to 1 \text{ and } g(x) \to \infty \quad \text{“}1^{\infty}\text{”}$$

There are plenty of examples of functions f and g for which such limits are zero, infinity, negative infinity, finite and non-zero, or even non-existent.

We shall see in Chapter 2 that the derivative of a function is defined as

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

Both the numerator and denominator of the above quotient tend to zero so that the limit is of “$\frac{0}{0}$” type. Study of the limiting values of such indeterminate forms is therefore central to the development of calculus.

We move on to investigate some of these types supported by useful examples rather than outlining the general theory.

We shall discuss four techniques in the following subsections: squeezing, algebraic manipulation, change of variable, and l'Hôpital's Rule.

1.5.1 Squeezing

Example 1.5.1.

This is actually a motivating example of an important theorem in Calculus called the squeeze theorem. It will also serve the purpose of demonstrating a geometrical proof of the standard result $\lim_{x\to 0}\dfrac{x}{\sin x} = 1$. The domain of the expression $\frac{x}{\sin x}$ does not include $x = 0$. The limit is of "$\frac{0}{0}$" type. Now $\dfrac{x}{\sin x}$ is an even function, so its behaviour as x tends to zero from above will be the same as that from below. Since we are considering the limit as $x \to 0$ we can assume that $0 < x < \frac{\pi}{2}$ as in Figure 1.11. Another elementary fact from trigonometry that is required in this context is that the area of a circular sector is $\frac{1}{2}r^2x$, where r is the radius and x is the angle in radians subtended by the arc in question.

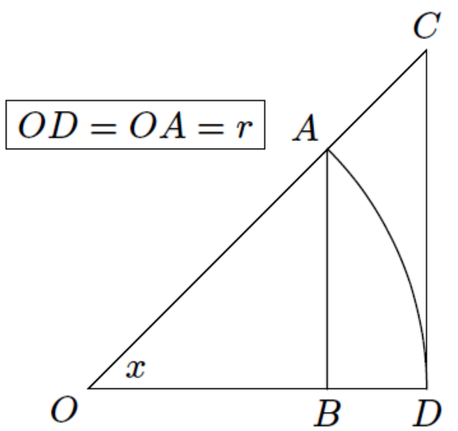

Figure 1.11: *Diagram for* $\lim_{x\to 0}\dfrac{x}{\sin x}$

In Figure 1.11 AD is a circular arc, and comparing areas we have

$$\text{Area of } \triangle OAB < \text{Area of sector } OAD < \text{Area of } \triangle OCD;$$
$$\frac{1}{2}(OB)(OA\sin x) < \frac{1}{2}OA^2x < \frac{1}{2}(OD)(OC\sin x);$$
$$\frac{1}{2}r^2\cos x\sin x < \frac{1}{2}r^2x < \frac{1}{2}r\frac{r}{\cos x}\sin x.$$

Dividing by $\frac{1}{2}r^2$ and by $\sin x$, which we are assuming to be positive, gives

$$\cos x < \frac{x}{\sin x} < \frac{1}{\cos x}$$

Now $\cos x \to 1$ as $x \to 0$, so $1 \le \lim_{x\to 0} \frac{x}{\sin x} \le 1$, i.e., $\lim_{x\to 0} \frac{x}{\sin x} = 1$.

This is an example of the "squeezing" technique derived from the more general squeeze theorem of Calculus as stated below.

Theorem 1.5.1. *Given the following inequality of functions*

$$g(x) \le f(x) \le h(x),$$

in an interval around $x = c$, one has

$$\lim_{x\to c} g(x) \le \lim_{x\to c} f(x) \le \lim_{x\to c} h(x)$$

provided those limits exist.

Figure 1.12 shows that the graph of $\frac{x}{\sin x}$ is indeed "squeezed" between the graphs of $\cos x$ and $\frac{1}{\cos x}$ for $-1 \le x \le 1$.

Thus all three functions have the same limit as $x \to 0$.

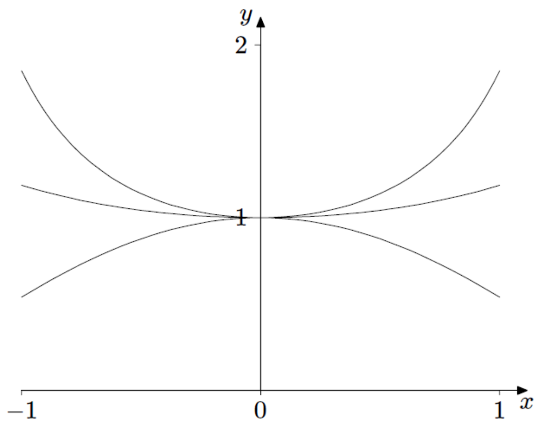

Figure 1.12: *Squeezing graph for* $\lim_{x\to 0} \frac{x}{\sin x}$

Example 1.5.2. *Find* $\lim_{x\to 0} x^2 \sin\left(\frac{1}{x}\right)$

We have demonstrated in Example 1.1.2 that $\sin\left(\frac{1}{x}\right)$ does not have a limiting value as tends to zero. However, we can use the squeezing technique to show that the above expression when multiplied by x^2 has a limiting value as x approaches zero. Using the fact that $\sin\frac{1}{x}$ always lies between 1 and -1, one deduces that $-x^2 \le x^2 \sin\left(\frac{1}{x}\right) \le x^2$ for all $x \neq 0$. Both the outer expressions in the above inequality tend to zero as x tends to zero so that the squeezing technique tells us that $\lim_{x\to 0} x^2 \sin\left(\frac{1}{x}\right) = 0$. From the Figure 1.13, one also finds that the graph of the function $x^2 \sin\left(\frac{1}{x}\right)$ is indeed "squeezed" between the graphs of $y = \pm x^2$ as $x \to 0$.

In this Example, it is important to note that the limiting point ($x = 0$ in this case) does not belong to the domain of the function $x^2 \sin\left(\frac{1}{x}\right)$ as we cannot substitute $x = 0$ in the expression $\sin\left(\frac{1}{x}\right)$. However, the fact that $\lim_{x\to 0} x^2 \sin\left(\frac{1}{x}\right)$ is zero means that we can extend the definition of the function by continuity, as we discussed in Remark 1.

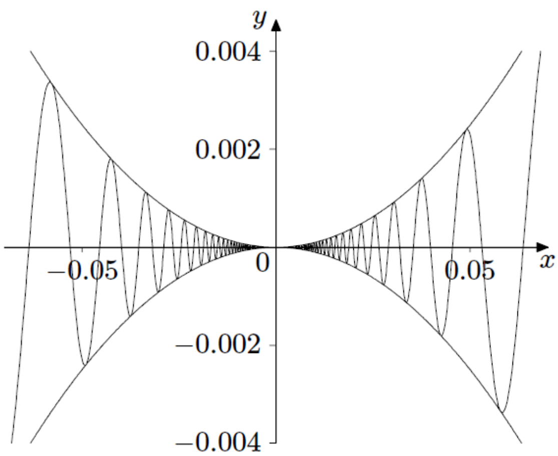

Figure 1.13: *Graph of* $x^2 \sin\left(\frac{1}{x}\right)$

1.5.2 Algebraic gymnastics

Example 1.5.3. *Find* $\lim\limits_{x\to 1} \dfrac{x^3-1}{x-1}$

We observe that we can cancel the factor $(x-1)$ from the numerator and denominator of the rational expression upon factorising the expression of the numerator. In fact, we are allowed to do so as $(x-1)$ is never equal to zero being infinitesimally close to zero. We then use algebraic rules of limits calculating separately the limit of each term in the resulting expression. Combining all these individual limits yields the final limiting value. We therefore have,

$$\lim_{x\to 1} \frac{x^3-1}{x-1} = \lim_{x\to 1} \frac{(x-1)(x^2+x+1)}{x-1}$$
$$= \lim_{x\to 1}(x^2+x+1) = 1+1+1 = 3.$$

Example 1.5.4. *Describe the behaviour of* $\dfrac{x^3-2x+3}{x^2+4}$ *as* $x\to\infty$ *and as* $x\to-\infty$.

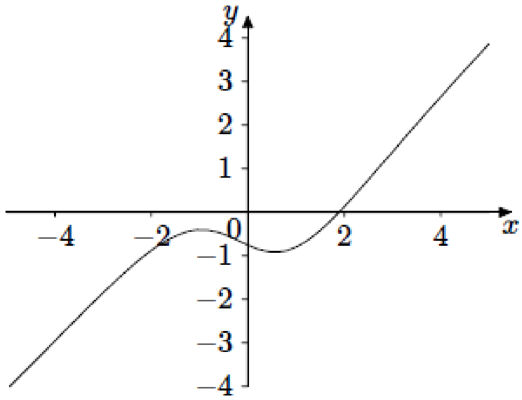

Figure 1.14: *Graph of* $\dfrac{x^3-2x+3}{x^2+4}$

The graph of this rational function is shown in Figure 1.14. The limit is of "$\frac{\infty}{\infty}$" type. For both the polynomials in the numerator and denominator of the given rational expression, the term with the greatest power of x dominates the rest of the terms. For the example in hand, x^3 and x^2 dominate in the numerator and in the denominator, respectively. Therefore, we expect the behaviour of the graph of $\dfrac{x^3-2x+3}{x^2+4}$ to be like that of $y=x$ as x grows without any bound in both the directions, i.e. as $x\to\pm\infty$. We demonstrate this algebraically by dividing the numerator and denominator by x^2, the largest power of x in the denominator. This yields,

$$\frac{x^3-2x+3}{x^2+4}=\frac{x-\frac{2}{x}+\frac{3}{x^2}}{1+\frac{4}{x^2}}.$$

The second and third terms in the numerator and the second term in the denominator tend to zero as $x\to\pm\infty$. The denominator therefore tends to 1. The numerator tends to ∞ as $x\to\infty$, and to $-\infty$ as $x\to-\infty$.We therefore conclude that

$$\frac{x^3-2x+3}{x^2+4}\to\infty \text{ as } x\to\infty \text{ and } \frac{x^3-2x+3}{x^2+4}\to-\infty \text{ as } x\to-\infty.$$

Example 1.5.5. *Describe the behaviour of* $f(x)=\dfrac{x^2-2}{x^3-2x^2-x+2}$ *as* $x\to\infty$

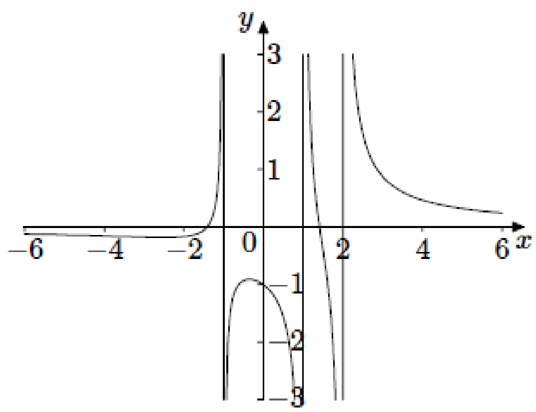

Figure 1.15: *Graph of* $\dfrac{x^2-2}{x^3-2x^2-x+2}$

The graph of this rational function is shown in Figure 1.15. The limit is again of "$\frac{\infty}{\infty}$" type. Following example 1.5.4, we divide the numerator and the denominator by the greatest power of x in the denominator, i.e. x^3 in this case, yielding

$$\frac{x^2-2}{x^3-2x^2-x+2}=\frac{\frac{1}{x}-\frac{2}{x^3}}{1-\frac{2}{x}-\frac{1}{x^2}+\frac{2}{x^3}} \tag{1.25}$$

The denominator tends to 1 as $x \to \pm\infty$ because all the terms except for the first one tend to zero. Both terms in the numerator tend to zero as $x \to \pm\infty$. Therefore, we have the numerator tending to zero and the denominator tending to 1. Hence, the quotient tends to zero. We have, therefore, been able to show algebraically that

$$\lim_{x\to\pm\infty}\frac{x^2-2}{x^3-2x^2-x+2}=0. \tag{1.26}$$

With the help of the Definition 1.3.1, equation 1.26 accounts the fact that the x-axis is a horizontal asymptote of the rational expression $\frac{x^2-2}{x^3-2x^2-x+2}$. This fact is graphically illustrated in Figure 1.15.

Example 1.5.6. *Determine the limiting behaviour of* $f(x)=\dfrac{(x+\sqrt{2})(x-\sqrt{2})}{(x+1)(x-1)(x-2)}$ *at the values of x where the denominator is zero.*

This is the function considered in the previous example with the numerator and denominator expressed in their factorised forms. It is presented graphically in Figure 1.15 . We, therefore, have to study the behaviour of the given function as x approaches each of -1, 1 and 2 from both sides. For these values, the numerator tends to -1, -1 and 2, respectively. The denominator tends to zero as x approaches -1, 1 and 2 from the right or from the left. It is a common mistake to naively conclude that the rational expression tends to infinity based on the fact that the denominator tends to zero while the numerator tends to a non-zero finite number. This error is sometimes written as "$\frac{1}{0}=\infty$". It is clear from Figure 1.15 that at each of $x=-1,1,2$, we have different one-sided limiting behaviours. The mistake is remedied by a careful bookkeeping of the signs of the various parts of the expression. We can do this either by plotting the graphs of the numerator and the denominator or by means of a rather convenient tabular approach as described below.

	$(-\infty,-\sqrt{2})$	$(-\sqrt{2},-1)$	$(-1,1)$	$(1,\sqrt{2})$	$(\sqrt{2},2)$	$(2,\infty)$
$x+\sqrt{2}$	$-$	$+$	$+$	$+$	$+$	$+$
$x-\sqrt{2}$	$-$	$-$	$-$	$-$	$+$	$+$
$x+1$	$-$	$-$	$+$	$+$	$+$	$+$
$x-1$	$-$	$-$	$-$	$+$	$+$	$+$
$x-2$	$-$	$-$	$-$	$-$	$-$	$+$
$f(x)$	$-$	$+$	$-$	$+$	$-$	$+$

At the top of the table we have put down the real numbers where either the numerator or the denominator is zero. In the leftmost column of the table, we list the factors appearing in the numerator and the denominator of the underlying rational expression given in example 1.5.6. The table indicates the sign of each factor in the intervals between these numbers. At the bottommost row of the table, we have listed the sign of $f(x)$ itself obtained from the set of signs above according to whether there is an even or an odd number of negative signs. Since the numbers $\pm\sqrt{2}$ are zeros of the numerator, the graph crosses the x-axis at those points making $f(x)$ change its sign there as confirmed by its graph in Figure 1.15.

Now let us study the behaviour of the function near $x = -1$. We can see from the table that if $x < -1$, then $f(x) > 0$, whereas if $x > -1$ then $f(x) < 0$. In both cases, the numerator tends to -1 and the denominator tends to zero. We therefore conclude that

$$f(x) \to \infty \text{ as } x \to -1^-, \text{ and } f(x) \to -\infty \text{ as } x \to -1^+.$$

Using similar reasoning near $x = 1$ and $x = 2$ we conclude that

$$\begin{gathered} f(x) \to -\infty \text{ as } x \to 1^-, \text{ and } f(x) \to \infty \text{ as } x \to 1^+ \\ f(x) \to -\infty \text{ as } x \to 2^-, \text{ and } f(x) \to \infty \text{ as } x \to 2^+. \end{gathered}$$

Therefore, the lines $x = -1, 1, 2$ are vertical asymptotes.

Example 1.5.7. *Discuss the limiting behaviour of* $\dfrac{1}{(x^5-1)^2}$.

The graph of this function is illustrated in Figure 1.10. Firstly, we observe that the denominator tends to infinity as $x \to \pm\infty$ and hence the function itself tends to zero, i.e., the x-axis is a horizontal asymptote. Secondly, the denominator has its only real zero at $x = 1$. Because the denominator is a square, it is positive for all $x \neq 1$, and so the expression tends to infinity as x tends either from left or from the right. Therefore, we can write

$$\frac{1}{(x^5-1)^2} \to \infty \text{ as } x \to 1.$$

Example 1.5.8. *Show that*

$$\lim_{x\to-\infty} \frac{1+3e^x}{1+e^x} = 1, \quad \lim_{x\to\infty} \frac{1+3e^x}{1+e^x} = 3.$$

The graph of this function is illustrated in Figure 1.9. We first note that $e^x \to 0$ as $x \to -\infty$ Therefore using the rules for limits tells us that

$$\lim_{x\to-\infty} \frac{1+3e^x}{1+e^x} \to \frac{1+0}{1+0} = 1$$

To deal with the other limit, we use a similar division procedure as in the rational function examples above. This time we divide numerator and denominator by e^x. We use the fact that $e^{-x} \to 0$ as $x \to \infty$ and the rules of limits yield

$$\lim_{x\to\infty} \frac{1+3e^x}{1+e^x} = \lim_{x\to\infty} \frac{e^{-x}+3}{e^{-x}+1} = \frac{0+3}{0+1} = 3$$

In the example above, we used our knowledge of the exponential function while in the previous examples, we needed to know the behaviour of powers of x, i.e. power functions as $x \to \infty$. Most of the examples in this section illustrate the principle of finding limits using knowledge of the limiting behaviour of basic functions.

1.5.3 Change of Variable

The technique of change of variable may turn useful in many circumstances, for example, in the case of integration and differential equations. In the present setting, the technique is employed in evaluating limits of various trigonometric and algebraic expressions. The general procedure, here, is to use the method of substitution to reduce the given expression to one whose limit is known beforehand.

Example 1.5.9. *Find* $\lim_{x\to\frac{\pi}{2}} \dfrac{\cos x}{\sin(\cos x)}$

The example above is of "$\frac{0}{0}$" type reminding us of the function $\dfrac{x}{\sin x}$.This suggests that we substitute $t = \cos x$. Then $t \to 0$ as $x \to \dfrac{\pi}{2}$.

Therefore $\lim_{x\to\frac{\pi}{2}} \dfrac{\cos x}{\sin(\cos x)} = \lim_{t\to 0} \dfrac{t}{\sin(t)} = 1$ using the result of Example 1.5.1. Note that while changing the variable, one has to pay attention to the limiting point. In the present example, $x = \frac{\pi}{2}$ corresponds to $t = 0$. The limiting value of the given trigonometric expression can then be easily computed using the composition rule for limits given in Section 1.4 which is as follows:

if $f(t) \to l$ as $t \to a$ and $g(x) \to a$ as $x \to b$ then $f(g(x)) \to l$ as $x \to b$.

In this case, we have applied the composition rule stated above with

$$f(t) = \frac{t}{\sin t}, l = 1, a = 0, g(x) = \cos x, b = \frac{\pi}{2}.$$

Example 1.5.10. *Find* $\lim_{x\to 0} \dfrac{\tan^{-1} x}{x}$.

This again is of "$\frac{0}{0}$" type and the composition rule can be used. A possible approach to simplifying the expression above is to let $t = \tan^{-1} x$. Hence, $x = \tan t$ with $t \to 0$ as $x \to 0$. Therefore,

$$\lim_{x\to 0} \frac{\tan^{-1} x}{x} = \lim_{t\to 0} \frac{t}{\tan t} = \lim_{t\to 0} \frac{t}{\sin t}. \cos t = \lim_{t\to 0} \frac{t}{\sin t}. \lim_{t\to 0} \cos t = 1,$$

by using the result of Example 1.5.1 and the product rule for limits given in Section 1.4.

1.5.4 L'Hôpital's Rule

This rule is designed specifically to deal with "$\frac{0}{0}$" and "$\frac{\infty}{\infty}$" types.

The rule involves differentiation which is discussed at length in Chapter 2. Readers are expected to know basic differentiation rules from high school mathematics.

Statement of the Rule

Suppose that the functions f and g are both differentiable in an interval containing $x = a$. Suppose also that $f(x) \to 0$ as $x \to a$ and that $g(x) \to 0$ as $x \to a$. (These are the conditions for the rule to be applicable and must always be checked)

Under the conditions stated above, if $\dfrac{f'(x)}{g'(x)} \to l$ as $x \to a$ then $\dfrac{f(x)}{g(x)} \to l$ as $x \to a$ (where l can be finite or infinite).

The rule applies with appropriate modifications for one-sided limits, for limits at infinity, and also when $f(x) \to \infty$ and $g(x) \to \infty$ as $x \to \infty$.

Example 1.5.11. *Find* $\displaystyle\lim_{x\to 1} \frac{\ln x}{x-1}$.

We let $f(x) = \ln x$, $g(x) = x - 1$. We must first check that the conditions for the application of l'Hôpital's Rule are satisfied. Both f and g are differentiable near $x = 1$, and both tend to zero as x tends to 1. Now $f'(x) = \dfrac{1}{x}$, $g'(x) = 1$, so $\dfrac{f'(x)}{g'(x)} = \dfrac{1}{x} \to 1$ as $x \to 1$. Therefore by l'Hôpital's Rule, $\dfrac{f(x)}{g(x)} \to 1$ as $x \to 1$.

We can make an interesting and important deduction from this example using a change of variable. If we let $x = 1 + \frac{a}{t}$ then $x \to 1$ as $t \to \infty$. Therefore,

$$\begin{aligned} 1 &= \lim_{x\to 1} \frac{\ln x}{x-1} = \lim_{t\to\infty} \frac{\ln(1+\frac{a}{t})}{\frac{a}{t}} \\ &= \lim_{t\to\infty} \frac{t\ln(1+\frac{a}{t})}{a} = \lim_{t\to\infty} \frac{\ln(1+\frac{a}{t})^t}{a} \end{aligned}$$

Multiplying by a gives $\lim_{t\to\infty} \ln\left(1+\frac{a}{t}\right)^t = a$, so $\lim_{t\to\infty}\left(1+\frac{a}{t}\right)^t = e^a$.

The limiting value of the function $f(x) = \left(1+\frac{x}{t}\right)^t$ as t increases without any bound is one of the many ways of defining the exponential function e^x.

Example 1.5.12. *Find* $\lim_{x\to 0^+} x\ln x$.

This example is of "$0 \times \infty$" type and hence it is not immediately clear if we can at all apply l'Hôpital's rule to compute the desired limit. However, if we re-write the expression of $x\ln x$ as

$$x\ln x = \frac{\ln x}{1/x},$$

we immediately see that it is of "$\frac{\infty}{\infty}$" type for which l'Hôpital's rule applies. Therefore, in this case, we let $f(x) = \ln x$, $g(x) = \frac{1}{x}$. One then deduces that $f(x) \to -\infty$ and $g(x) \to \infty$ as $x \to 0^+$. Calculating the derivatives yields,

$$f'(x) = \frac{1}{x},\ g'(x) = -\frac{1}{x^2},\ \text{so}\ \frac{f'(x)}{g'(x)} = \frac{1/x}{-1/x^2} = -x \to 0 \text{ as } x \to 0^+$$

Therefore, l'Hôpital's rule asserts that $\lim_{x\to 0^+} x\ln x = 0$.

Chapter 2

Differentiation

The differential calculus has two major areas of use and origin. One is geometry, and the problem of finding tangents to curves. The other is motion (speed, velocity, acceleration) and other rates of change. Both of these lead to the definition of the derivative in terms of a limit.

2.1 Definition using limit

We shall explore the definition of derivative of a function $f(x)$ at a given point by considering the problem of finding the gradient of the curve $y = f(x)$ at the given point by evaluating the slope of the tangent to $y = f(x)$ at the point in question.

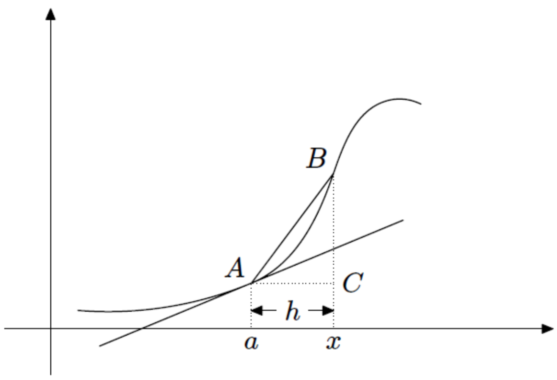

Figure 2.1: *Chord Slope Diagram*

Figure 2.1 represents the graph of a function $y = f(x)$. The tangent line at the point A can be considered as the limiting position of the chord BA as B tends towards A. This is achieved by letting x tend to a or equivalently by letting h tend to 0, since $x = a+h$. It is important to note that this is not just a one-sided limit and hence a diagram, where B is to the left of A with h being negative, is equally valid. The gradient of the tangent will, therefore, be the limit of the gradient of the chord as x tends to a.

The coordinates of the relevant points labelled in Figure 2.1 are as follows.

$$A(a, f(a)), B(x, f(x)), C(x, f(a)). \tag{2.1}$$

Therefore, the gradient (slope) of the chord is given by,

$$\frac{BC}{AC} = \frac{f(x) - f(a)}{x - a} = \frac{f(a+h) - f(a)}{h}. \tag{2.2}$$

Taking limit of the expression in Equation 2.2 as $h \to 0$, therefore, yields the gradient of the tangent. The gradient of the tangent to the curve $y = f(x)$ at a given point is precisely the derivative of $f(x)$ at the point in question as proposed in the following definition.

Definition 2.1.1. *The function $f(x)$ whose domain includes some interval containing the point a is said to be differentiable at a if the following limit exists.*

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}. \tag{2.3}$$

The value of this limit is called the derivative of f at a, denoted by $f'(a)$. We therefore have,

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}. \tag{2.4}$$

In many cases, a function will be differentiable for all (or most) of the values of x in the domain. In such cases, we think of a as a variable and use the term derivative for the function whose value at $x = a$ is $f'(a)$.

A variety of synonymous terminology for "derivative" of a function is encountered in the literature. Of them, differential coefficient, derived function, differential and first derivative are most notable. We shall use the term derivative to refer not only to the function resulting from the process of differentiation but also to the value of this derived function at some point of its domain.

Two types of notation are commonly used in the literature: the dash notation $f'(x)$, $f''(x)$, etc., and the Leibniz notation $\frac{dy}{dx}$, $\frac{d^2y}{dx^2}$, etc. Readers must have encountered these notational facts in high school calculus.

2.2 Applications of the Limit Definition

In this section, we will study some examples where we can find the derivative directly from the definition using limit together with an example where the derivative does not exist. We will also prove a basic property of derivatives that turns out to be very useful in graph sketching. In practice, we do not rely much on the limit definition though. Instead, we use algebraic rules for differentiation and apply them to functions whose derivatives are well-known. The first two examples show that the basic derivatives can be found using the limit definition. This is sometimes referred to as "finding the derivative from first principles", the first principles in question being the limit definition.

Example 2.2.1. *Use the limit definition to find the derivative of $f(x) = x^2$. Applying the limit definition gives,*

$$f'(a) = \lim_{x\to a}\frac{f(x)-f(a)}{x-a} = \lim_{x\to a}\frac{x^2-a^2}{x-a} = \lim_{x\to a}(x+a) = 2a.$$

Example 2.2.2. *Use the limit definition to find the derivative of $f(x) = sinx$.*

$$\begin{aligned} f'(a) &= \lim_{h\to 0}\frac{f(a+h)-f(a)}{h} = \lim_{h\to 0}\frac{sin(a+h)-sin(a)}{h} \\ &= \lim_{h\to 0}\frac{2cos(a+\frac{h}{2})sin(\frac{h}{2})}{h} = \lim_{h\to 0} cos\left(a+\frac{h}{2}\right)\frac{sin(\frac{h}{2})}{\frac{h}{2}} \\ &= cos(a).1 = cos(a). \end{aligned}$$

Here, we have used the limit obtained in Example 1.5.1 with $x = \frac{h}{2}$.

Example 2.2.3. *Show that* $f(x) = |x|$ *is not differentiable at* 0.

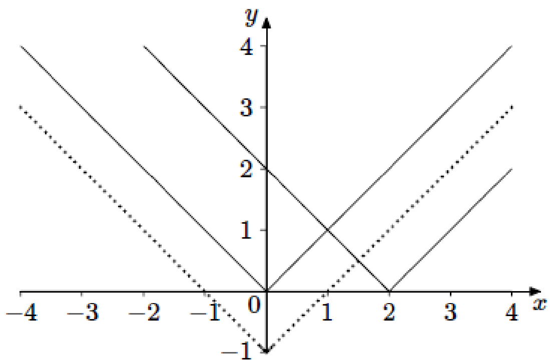

Figure 2.2: *Graphs of modulus functions.png*

We recall the graph of $y = |x|$ as shown in Figure 2.2 and notice that it has a sharp corner at $x = 0$. The gradient to the right is 1 and the gradient to the left is -1, indicating that the gradient at 0 cannot be well-defined. The limit definition confirms this, as follows.

$$\frac{f(0+h) - f(0)}{h} = \begin{cases} \frac{h}{h} = 1 & \text{for } h > 0 \\ \frac{-h}{h} = -1 & \text{for } h < 0. \end{cases}$$

This fact can also be rewritten as

$$\lim_{h \to 0^+} \frac{f(0+h) - f(0)}{h} = 1$$
$$\lim_{h \to 0^-} \frac{f(0+h) - f(0)}{h} = -1,$$

so that the left- and right-sided limits are different proving that the limit $\lim_{h \to 0} \dfrac{f(0+h) - f(0)}{h}$ doesn't exist. Hence, the modulus function is not differentiable at 0.

Example 2.2.4.

Consider the following piecewise defined function.

$$k(x) = \begin{cases} x^2 & \text{if } x \geq 0, \\ x^3 & \text{if } x < 0. \end{cases}$$

If $x > 0$, then $k(x) = x^2$ so that $k'(x) = 2x$. On the other hand, if $x < 0$, $k(x) = x^3$ yielding $k'(x) = 3x^2$. But to investigate differentiability at $x = 0$, we need to use the limit definition as follows.

$$\frac{k(0+h) - k(0)}{h} = \begin{cases} \frac{h^2-0}{h} = h & \text{if } x > 0, \\ \frac{h^3-0}{h} = h^2 & \text{if } x < 0. \end{cases}$$

We conclude that

$$\lim_{h\to 0^+} \frac{k(0+h) - k(0)}{h} = \lim_{h\to 0^+} h = 0,$$
$$\lim_{h\to 0^-} \frac{k(0+h) - k(0)}{h} = \lim_{h\to 0^-} h^2 = 0.$$

The left-hand and right-hand limits are equal and therefore we obtain

$$\lim_{h\to 0} \frac{k(0+h) - k(0)}{h} = 0.$$

Hence, $k(x)$ is differentiable at $x = 0$ and $k'(0) = 0$.

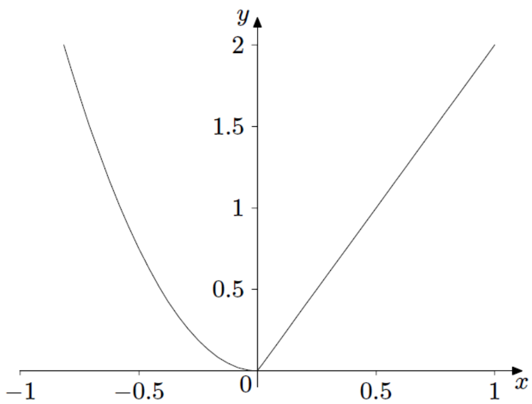

Figure 2.3: *Graph of the derivative function $k'(x)$*

Figure 2.3 shows the graph of $k'(x)$. We can see that there appears to be a sharp corner at $x = 0$, as there is for $|x|$. This suggests that $k'(x)$ is also not differentiable at $x = 0$. In what follows, we prove this using the limit definition.

$$\frac{k'(0+h) - k'(0)}{h} = \begin{cases} \frac{2h-0}{h} = 2 & \text{if } x > 0, \\ \frac{3h^2-0}{h} = 3h & \text{if } x < 0, \end{cases}$$

leading to the fact that

$$\lim_{h\to 0^+} \frac{k'(0+h) - k'(0)}{h} = \lim_{h\to 0^+} 2 = 2,$$
$$\lim_{h\to 0^-} \frac{k'(0+h) - k'(0)}{h} = \lim_{h\to 0^-} 3h = 0.$$

The left-and right-hand limits are not equal and hence

$$\lim_{h\to 0} \frac{k'(0+h) - k'(0)}{h} \quad \text{does not exist.}$$

This demonstrates the fact that $k'(x)$ is not differentiable at $x = 0$.

Example 2.2.5. *Prove that if a differentiable function is increasing then its derivative is non-negative.*

Suppose that for all a, b in the domain of f satisfying $a \leq b$, we have $f(a) \leq f(b)$. Let x denote an arbitrary number in the domain of f. Then if $h > 0$, we have $\dfrac{f(x+h) - f(x)}{h} \geq 0$, because the numerator is always non-negative (positive or zero) and the denominator is always positive (not zero). If $h < 0$, we also have $\dfrac{f(x+h) - f(x)}{h} \geq 0$ because in this case, the numerator is a non-positive (negative or zero) real number while the denominator is a negative (not zero) real number. Therefore we must have,

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \geq 0,$$

the existence of the above limit is guaranteed by the differentiability of the function f.

We can prove similarly that if a differentiable function is decreasing, then its derivative is non-positive.

2.3 Rules of Differentiation

The basic algebraic rules of differentiation enable us to differentiate expressions involving sums, products and quotients of functions whose derivatives are already known. The readers are expected to be familiar with the derivatives of some elementary functions presented in the table below from High school mathematics.

$f(x)$	$f'(x)$
x^n	nx^{n-1}
$sin(x)$	$cos(x)$
$cos(x)$	$-sin(x)$
$tan(x)$	$sec^2(x)$
e^x	e^x
lnx	$\frac{1}{x}$

The basic rules of differentiation are summarised as follows.

Suppose that f and g are differentiable functions. Then for any constants A and B the following hold

Sum rule:

$$\frac{d}{dx}(Af(x) + Bg(x)) = Af'(x) + Bg'(x). \tag{2.5}$$

Product rule:

$$\frac{d}{dx}(f(x)g(x)) = f(x)g'(x) + g(x)f'(x). \tag{2.6}$$

Quotient rule:

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}, \ (g(x) \neq 0). \tag{2.7}$$

2.4 The indispensable Chain Rule

The Chain Rule (or function of a function rule to refer to what it is good for) tells us how to differentiate composite functions. It is described as follows.

Suppose that the function g is differentiable at x and the function f is differentiable at $g(x)$. Then the derivative of the composite function $f \circ g$ with $(f \circ g)(x) = f(g(x))$ is given by $(f \circ g)'(x) = f'(g(x))g'(x)$. Using Leibniz notation, the rule can be stated as follows.

If $y = f(u)$ and $u = g(x)$ then $\dfrac{dy}{dx} = \dfrac{dy}{du}\dfrac{du}{dx}$.

We will not derive the chain rule rigorously. Rather, we will try to provide a somewhat intuitive justification based on the limit definition. A rigorous derivation can be found in any standard text on real analysis.

$$\frac{f(g(x+h)) - f(g(x))}{h} = \left[\frac{f(g(x+h)) - f(g(x))}{g(x+h) - g(x)}\right]\left[\frac{g(x+h) - g(x)}{h}\right]. \tag{2.8}$$

Here, we have introduced the term $g(x+h) - g(x)$ in the numerator and in the denominator. This helps separate the behaviour of f and that of g. We, then, let $g(x) = u$ and $g(x+h) = u + k$ so that $k \to 0$ as $h \to 0$. We, therefore, obtain,

$$\frac{f(g(x+h)) - f(g(x))}{h} = \left[\frac{f(u+k) - f(u)}{k}\right]\left[\frac{g(x+h) - g(x)}{h}\right]. \tag{2.9}$$

Setting the right sides of Equations 2.8 and 2.9 equal and taking the limit $h \to 0$, one obtains

$$f'(u)g'(x) = f'(g(x))g'(x) \text{ as } h \to 0$$

The argument presented above is not flawless since we cannot be sure of $g(x+h) - g(x)$ being always nonzero for some values of h arbitrarily close to 0 and dividing by zero is not a mathematically valid operation. The argument does, however, provide an intuitive justification relating the chain rule to the limit definition.

Example 2.4.1. *Find the derivative of* $\ln \cos x$.

First, let $y = \ln u$, $u = \cos x$. Using chain rule, one then obtains

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = \frac{1}{u}(-\sin x) = \frac{-\sin x}{\cos x} = -\tan x.$$

Example 2.4.2. *Differentiate* a^x *with respect to* x.

It is a common mistake to write the derivative as xa^{x-1}. This is **WRONG** since the variable x appears in the exponent of the expression a^x with $a > 0$. To do the calculation correctly, one has to invoke the relationship between an arbitrary and natural exponential function given by $a^x = e^{x \ln a}$. We, then, use the chain rule following the substitutions of $y = e^u$ and $u = x \ln a$ to obtain,

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = e^u \ln a = e^{x \ln a} \ln a = a^x \ln a.$$

Example 2.4.3. *Differentiate* $f(x) = x^2 \cos \frac{1}{x}$.

In examples as the one above, more than one of the differentiation rules are to be employed. Firstly, we need the product rule since the function is x^2 multiplied by a cos term. Secondly, the cos term itself is composite calling for the chain rule. Applying both rules yields,

$$f'(x) = 2x \cos \frac{1}{x} + x^2 \left(-\sin \frac{1}{x}\right)\left(-\frac{1}{x^2}\right) = 2x \cos \frac{1}{x} + \sin \frac{1}{x}. \quad (2.10)$$

Figure 2.4 shows a MAPLE plot of this formula for $f'(x)$.

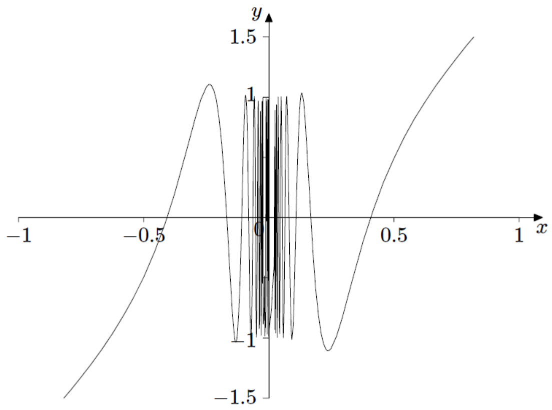

Figure 2.4: *A Discontinuous Derivative*

Now this calculation is not valid when $x = 0$, and indeed $f(x)$ is not defined there. However, the squeezing argument used in Example 1.5.2 shows that $f(x)$ has the limit zero as x tends to zero and that if we extend the definition of $f(x)$ by letting $f(0) = 0$, the resulting function is continuous at 0. What about the differentiability of the continuous function we have just constructed by extending the original one? We can't substitute $x = 0$ in Equation 2.10, the expression for the derivative of $f(x)$ as it is only valid for nonzero x. Therefore, we have to go back to the limit definition and investigate whether the appropriate limit exists.

$$\frac{f(0+h) - f(0)}{h} = \frac{h^2 \cos \frac{1}{h}}{h} = h \cos \frac{1}{h} \to 0 \text{ as } h \to 0,$$

by the squeezing argument. Therefore, f is differentiable at 0 and $f'(0) = 0$ provided we extend the definition of the original function by continuity. In spite of f being differentiable at 0, $f'(x)$ does not have a limit as $x \to 0$, because the sin term in Equation 2.10 oscillates infinitely often in any interval containing $x = 0$ as we established in Example 1.1.2. So, f (after extending the definition by continuity) is differentiable everywhere but its derivative is discontinuous at $x = 0$ which is, of course, not of "removable" type.

2.5 Derivatives of higher orders

Given a function f specified by a formula $y = f(x)$, one can differentiate it to obtain the formula for its first derivative $f'(x)$ which, in turn, can be differentiated to yield $f''(x)$, the formula for the second derivative of f. One can keep repeating this procedure to obtain a sequence of derivatives, denoted by,

$$f'(x),\ f''(x),\ f'''(x),\ f^4(x),\ \ldots,\ f^n(x),\ \ldots,$$

or, using the Leibniz notation for derivatives, by

$$\frac{dy}{dx}, \frac{d^2y}{dx^2}, \frac{d^3y}{dx^3}, \ldots, \frac{d^ny}{dx^n}, \ldots \quad .$$

Example 2.5.1. *Find the n-th derivative of* $f(x) = \ln(2x+3)$.

Calculating the first few derivatives, using the chain rule, is relatively straightforward, giving,

$$f(x) = \ln(2x+3);$$

$$f'(x) = \frac{2}{2x+3};$$

$$f''(x) = \frac{-4}{(2x+3)^2};$$

$$f'''(x) = 2.\frac{8}{(2x+3)^3};$$

$$f^4(x) = 3.2.\frac{-16}{(2x+3)^4};$$

$$f^5(x) = 4.3.2.\frac{32}{(2x+3)^5}$$

The trick, here, is to find the general pattern present in the expressions of the higher order derivatives of f. This enables us to conjecture a formula for the nth order derivative, namely,

$$f^n(x) = (n-1)!\frac{(-1)^{(n+1)}2^n}{(2x+3)^n}. \tag{2.11}$$

We leave it to the reader to use the method of mathematical induction to prove the result 2.11.

Chapter 3

Differentiation techniques

In this chapter, we will try to master some techniques of differentiation which deal with functions specified in various forms. We will also consider functions defined implicitly, functions defined parametrically, functions involving powers and inverse functions. The chapter also includes discussions on Leibniz Theorem, a result which enables us to calculate higher order derivatives of products of functions.

3.1 Technique of implicit differentiation

Often it happens to be the case that y is not given as a function of x explicitly. Instead, we have an "entangled" type of equation in x and y using which neither x nor y can be expressed in terms of the other variable explicitly. We may still want to find $\dfrac{dy}{dx}$ and as the result suggests, the final expression involves both variables.

The following example illustrates the discussion conducted above.

Example 3.1.1. *Find the gradient* $\dfrac{dy}{dx}$ *at the point* $(1, 2)$ *on the curve whose equation is given by*

$$x^3 - 5xy^2 + y^3 + 11 = 0$$

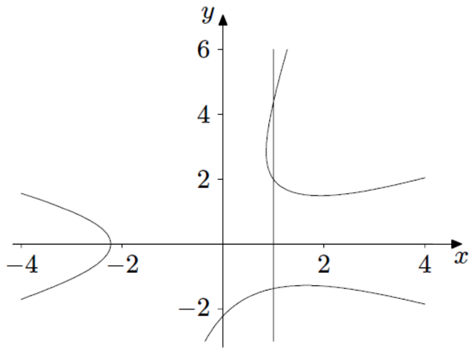

Figure 3.1: *Graph of* $x^3 - 5xy^2 + y^3 + 11 = 0$

Figure 3.1 suggests that the curve is not the graph of y as a function of x. Indeed when $x = 1$ there are three possible values of y on the part of the graph shown. This is demonstrated by the inclusion of the line $x = 1$ in Figure 3.1. One of the three points of intersection of this line with the graph is $(1, 2)$ as understood from the figure. A small part of the curve in the vicinity of $(1, 2)$ represents the graph of a function $y = y(x)$ which solves the equation of the curve as well and encapsulates the local behaviour of the graph near $(1, 2)$. We cannot find $y(x)$ globally in terms of x depicting the full curve $x^3 - 5xy^2 + y^3 + 11 = 0$ though.

The function $y(x)$ satisfies the equation of the curve, namely

$$x^3 - 5x(y(x))^2 + (y(x))^3 + 11 = 0.$$

We, therefore, have to use the chain rule to differentiate the y^2 and y^3 terms and the product rule for the second term involving x and y. Using the chain rule for the terms involving powers of $y(x)$ yields

$$\frac{d}{dx}\left(y(x)^3\right) = 3(y(x))^2\frac{dy}{dx},$$
$$\frac{d}{dx}\left(y(x)^2\right) = 2y(x)\frac{dy}{dx}.$$

Differentiating the equation of the curve with respect to x, therefore, gives

$$3x^2 - 5\left[x \cdot 2y(x)\frac{dy}{dx} + (y(x))^2\right] + 3(y(x))^2\frac{dy}{dx} = 0.$$

Rearranging the terms, one obtains

$$[3(y(x))^2 - 10xy(x)]\frac{dy}{dx} = 5(y(x))^2 - 3x^2,$$

leading to

$$\frac{dy}{dx} = \frac{5(y(x))^2 - 3x^2}{3(y(x))^2 - 10xy(x)} = \frac{5y^2 - 3x^2}{3y^2 - 10xy}.$$

The gradient at the point $(1, 2)$ is, then, found by substituting these values for x and y in the expression above yielding $-\dfrac{17}{8}$. This value is consistent with the Figure 3.1 where the tangent line at $(1, 2)$ does indeed appear to have a fairly steep negative gradient.

Note also that, in the last expression, we have introduced the short-hand notation y for $y(x)$ in order to maintain notational simplicity. We will stick to this practice for the rest of the chapter.

Example 3.1.2. *Given* $\cos(xy) = \exp(x + y)$*, find* $\dfrac{dy}{dx}$ *in terms of* x *and* y*.*

This is purely an algebraic problem. We first apply the chain rule to both sides and obtain

$$-\sin(xy)\frac{d}{dx}(xy) = \exp(x + y)\frac{d}{dx}(x + y).$$

One then requires the product rule to be applied on the left hand side yielding

$$-\sin(xy)\left(y + x\frac{dy}{dx}\right) = \exp(x + y)\left(1 + \frac{dy}{dx}\right).$$

We now collect all the terms involving the derivative and then divide to isolate the derivative as we did in Example 3.1.1. One then obtains

$$\frac{dy}{dx} = -\frac{\exp(x+y) + y\sin(xy)}{x\sin(xy) + \exp(x+y)},$$

provided that the denominator is not zero.

Note that assigning arbitrary value to the ordered pair (x, y) in the expression above is meaningless. The point (x, y) requires to satisfy the original equation in order for $\frac{dy}{dx}$ to be interpreted as the gradient of the curve at the given point.

Example 3.1.3. *Given $xy + e^y = 0$, find $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$ in terms of x and y.*

Differentiating the equation with respect to x, one obtains

$$y + x\frac{dy}{dx} + e^y\frac{dy}{dx} = 0. \tag{3.1}$$

We differentiate Equation (3.1) once more to obtain

$$\frac{dy}{dx} + \frac{dy}{dx} + x\frac{d^2y}{dx^2} + e^y\frac{dy}{dx}\cdot\frac{dy}{dx} + e^y\frac{d^2y}{dx^2} = 0.$$

We now rearrange both the equations above to obtain

$$\begin{aligned}
\frac{dy}{dx} &= -\frac{y}{x+e^y};\\
\frac{d^2y}{dx^2} &= -\frac{2\frac{dy}{dx} + e^y\left(\frac{dy}{dx}\right)^2}{x+e^y}\\
&= -\frac{-\frac{2y}{x+e^y} + e^y\left(\frac{-y}{x+e^y}\right)^2}{x+e^y}\\
&= -\frac{-2y(x+e^y) + y^2e^y}{(x+e^y)^3}.
\end{aligned}$$

3.2 Logarithmic Differentiation

This topic is an offshoot of implicit differentiation technique encountered in the preceding section. This is especially useful when one encounters expressions involving the variable in an exponent. It can also be applied to products of functions expressed in a complicated way.

Example 3.2.1. *Differentiate* $y = x^{\sin x}$.

We take logarithm on both sides of the equation to obtain

$$\ln y = \ln\left(x^{\sin x}\right) = \sin x \ln x.$$

Differentiation on both sides with respect to x, then, yields

$$\frac{1}{y}\frac{dy}{dx} = \cos x \cdot \ln x + \sin x \cdot \frac{1}{x},$$

by successive application of implicit differentiation technique and product rule.

Rearranging the terms, one finally obtains

$$\frac{dy}{dx} = y\left(\cos x \cdot \ln x + \sin x \cdot \ln x\right) = x^{\sin x}\left(\cos x \cdot \ln x + \sin x \cdot \ln x\right).$$

Example 3.2.2. *Differentiate* $y = x^2 e^x \sin x \cosh x$.

We can use the product rule to attack this problem. But taking logarithm on both sides of the equation will help us convert the expression of the right hand side into a sum that can be handled in a much simpler way. To this end, one obtains

$$\begin{aligned}
\ln y &= \ln(x^2) + \ln(e^x) + \ln(\sin x) + \ln(\cosh x) \\
&= 2\ln x + x + \ln(\sin x) + \ln(\cosh x)
\end{aligned}$$

Differentiation gives, $\dfrac{1}{y}\dfrac{dy}{dx} = \dfrac{2}{x} + 1 + \cot x + \tanh x.$

$$\begin{aligned}
\text{Therefore, } \frac{dy}{dx} &= y\left(\frac{2}{x} + 1 + \cot x + \tanh x\right) \\
&= (x^2 e^x \sin x \cosh x)\left(\frac{2}{x} + 1 + \cot x + \tanh x\right) \\
&= 2xe^x \sin x \cosh x + x^2 e^x \cos x \cosh x \\
&\quad + x^2 e^x \sin x \sinh x + x^2 e^x \sin x \cosh x.
\end{aligned}$$

3.3 Parametric Differentiation

Equations of curves are often given parametrically, for example, the equation of an ellipse specified by

$$x = a\cos t,\ y = b\sin t,\ 0 \le t \le 2\pi. \tag{3.2}$$

We want to find the gradient $\dfrac{dy}{dx}$. The stumbling block here is the fact that the given parametric equations can only be differentiated with respect to t.

We can look for the solution in two different ways. Firstly, we can use the chain rule to obtain

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}, \quad \text{leading to} \quad \frac{dy}{dx} = \frac{dy}{dt}\bigg/\frac{dx}{dt}, \quad \text{provided } \frac{dx}{dt} \neq 0. \tag{3.3}$$

Secondly, one can trace back to the definition of derivative using limit and express the gradient $\dfrac{dy}{dx}$ at a point specified by $t = k$ by calculating the chord slope limit as $t \to k$. The reasoning goes as follows.

$$\begin{aligned}\frac{dy}{dx} &= \lim_{t\to k}\frac{y(t)-y(k)}{x(t)-x(k)} = \lim_{t\to k}\left[\frac{y(t)-y(k)}{x(t)-x(k)}\right]\left[\frac{t-k}{t-k}\right] \\ &= \lim_{t\to k}\left[\left\{\frac{y(t)-y(k)}{t-k}\right\}\left\{\frac{t-k}{x(t)-x(k)}\right\}\right] \\ &= \lim_{t\to k}\left[\frac{y(t)-y(k)}{t-k}\right]\Bigg/\lim_{t\to k}\left[\frac{x(t)-x(k)}{t-k}\right] = \frac{dy}{dt}\Bigg/\frac{dx}{dt}, \\ &\text{provided } \frac{dx}{dt}\neq 0.\end{aligned}$$

For the ellipse given by Equation (3.2), one obtains

$$\frac{dy}{dx} = -\frac{b}{a}\cot t = -\frac{b^2x}{a^2y}, \tag{3.4}$$

where we used $\cot t = \frac{bx}{by}$ by manipulating the parametric equation (3.2). Also, note that the expression (3.4) for the gradient is meaningful only if $\sin t \neq 0$, which excludes the points given by $t = 0, \pm\pi, \pm 2\pi, ...,$. The tangents to the ellipse at these points are all parallel to the y-axis.

Example 3.3.1. *Given $x = a\cos t$, $y = b\sin t$, $0 \leq t \leq 2\pi$, find $\frac{d^2y}{dx^2}$.*

We recall that $\dfrac{d^2y}{dx^2} = \dfrac{dY}{dx}$ with $Y = \dfrac{dy}{dx}$. Applying the parametric differentiation formula (3.3) to Y, one then obtains

$$\frac{dY}{dx} = \frac{dY}{dt}\Bigg/\frac{dx}{dt}. \tag{3.5}$$

We have computed Y in Equation (3.4). Using this in Equation (3.5) then yields

$$\frac{d^2y}{dx^2} = \frac{dY}{dx} = \frac{dY}{dt}\Bigg/\frac{dx}{dt} = \frac{(b/a)\operatorname{cosec}^2 t}{-a\sin t} = -\frac{b}{a^2\sin^3 t} = -\frac{b^4}{a^2y^3}, \tag{3.6}$$

provided $\sin t \neq 0$, i.e. everywhere except for the points where the tangents to the ellipse are parallel to the y-axis.

Remark 2. *Here, one could have started with the gradient expressed in terms of x and y only as in the later part of Equation (3.4) and differentiate it again with respect to x using implicit differentiation technique. After some algebraic manipulations, one is supposed to obtain* $\dfrac{d^2y}{dx^2} = -\dfrac{b^4}{a^2y^3}$ *in agreement with Equation (3.6).*

3.4 Differentiation of Inverse Functions

In what follows, we assume that the readers are familiar with the notion of inverse function. In case if they are not, they are advised to consult any standard Precalculus book (for example, the one by Sullivan). In this section, differentiation of various inverse functions will concern us. Although, it is possible to find a general formula, as we shall demonstrate, in most cases it is more helpful to use an implicit function approach that will be explained through some examples in this section.

Suppose that we have a differentiable function f with its inverse g. So, $y = f(x)$ and $x = g(y)$ are equivalent so that $f(g(y)) = y$ and $g(f(x)) = x$ hold for all y in the domain of g (along with $g(y)$ in the domain of f) and for all x in the domain of f (along with $f(x)$ in the domain of g), respectively. We shall first establish differentiability of g using the limit definition.

$$\frac{dg}{dy} = \lim_{k\to 0} \frac{g(y+k) - g(y)}{k}$$

Now $y + k = f(x + h)$ for some h and since f is continuous it follows that $k \to 0$ as $h \to 0$. Also, since f has an inverse, it is 1-1. Therefore, for $h \neq 0$, we have $f(x + h) \neq f(x)$ so that $k \neq 0$. Hence,

$$\frac{g(y+k) - g(y)}{k} = \frac{g(y+k) - g(y)}{y+k-y} = \frac{x+h-x}{f(x+h) - f(x)} = \frac{h}{f(x+h) - f(x)}.$$

From which we deduce that

$$\frac{dg}{dy} = \lim_{k\to 0} \frac{g(y+k) - g(y)}{k} = \lim_{h\to 0} \frac{h}{f(x+h) - f(x)} = 1 \Big/ \frac{df}{dx}. \quad (3.7)$$

There is an alternative way of reaching Equation (3.7) by taking the differentiability of the inverse of g into account and using the inverse function relationship $g(f(x)) = x$. Differentiating this equation and using the chain rule, one obtains $g'(f(x))f'(x) = 1$. Also, since $y = f(x)$, we have

$$g'(y) = \frac{1}{f'(x)}. \tag{3.8}$$

Example 3.4.1. *Verify Equation (3.8) using the logarithmic function.*

Suppose $y = f(x) = \ln x$ so that $x = g(y) = e^y$ is the formula for the inverse function. Then

$$g'(y) = e^y = e^{\ln x} = x = \frac{1}{\frac{1}{x}} = \frac{1}{f'(x)}.$$

Example 3.4.2. *In this example, we consider the problem of differentiating the inverse sine function. By restricting the domain of the sine function to a suitable interval, say, $[-\frac{\pi}{2}, \frac{\pi}{2}]$, one obtains a 1-1 function. It is therefore legitimate to call for the inverse of this restricted sine function.*

Suppose $y = \sin^{-1} x$, which is equivalent to $x = \sin y$. Differentiation of the latter equation implicitly with respect to x yields

$$1 = \cos(y)\frac{dy}{dx} \quad \text{so that} \quad \frac{dy}{dx} = \frac{1}{\cos y}.$$

In order to express the result in terms of x, we use the identity $\cos^2 y + \sin^2 y = 1$ yielding $\cos y = \pm\sqrt{1 - \sin^2 y} = \pm\sqrt{1 - x^2}$.

As discussed at the onset that the restricted sine function reads,

$$x = \sin y; \quad \text{with} \quad -\frac{\pi}{2} \leq y \leq \frac{\pi}{2},$$

so that

$$x \in [-1, 1] \quad \text{as} \quad y = \sin^{-1} x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$$

Consequently,

$$\cos y \geq 0 \quad \text{as} \quad y \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$$

The argument above presents the rationale behind picking up the positive square root from the expression of $\cos y = \pm\sqrt{1-x^2}$. Therefore, sticking to the positive square root only, the equation

$$\frac{dy}{dx} = \frac{1}{\cos y} = \pm\frac{1}{\sqrt{1-x^2}},$$

as deduced earlier, translates to

$$\frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1-x^2}}.$$

On the other hand, also note that

$$x \in [-1, 1] \quad \text{as} \quad y = \sin^{-1} x \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right],$$

so that one obtains, in this case,

$$\cos y \leq 0 \quad \text{as} \quad y \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right].$$

In this situation, one picks up the negative square root from the expression of $\cos y = \pm\sqrt{1-x^2}$ so that one ends up with

$$\frac{d}{dx}(\sin^{-1} x) = -\frac{1}{\sqrt{1-x^2}}.$$

3.5 Leibniz Theorem

Finding the n-th derivative of a product can be fairly complicated. In this section, we will derive a general formula to deal with this problem. If we start applying the product rule three times to the general expression of the form $h(x) = f(x)g(x)$ and collect like terms together at each stage, we soon explore a pattern emerging. We, therefore, come up with

$$\begin{aligned}
h'(x) &= f'(x)g(x) + f(x)g'(x);\\
h''(x) &= [f''(x)g(x) + f'(x)g'(x)] + [f'(x)g'(x) + f(x)g''(x)]\\
&= f''(x)g(x) + 2f'(x)g'(x) + f(x)g''(x);\\
h'''(x) &= [f'''(x)g(x) + f''(x)g'(x)] + 2[f''(x)g'(x) + f'(x)g''(x)]\\
&\quad + [f'(x)g''(x) + f(x)g'''(x)]\\
&= f'''(x)g(x) + 3f''(x)g'(x) + 3f'(x)g''(x) + f(x)g'''(x);
\end{aligned}$$

where at each stage the square brackets indicate a pair of terms arising from the application of the product rule to a single term at the previous stage.

The pattern of numerical coefficients when the terms are collected together is that of the binomial coefficients from Pascal's triangle and this enables us to formulate the general result.

Theorem 3.5.1. *If the functions f and g are both differentiable n times, then their product is also differentiable n times and*

$$\begin{aligned}
\frac{d^n}{dx^n}(fg) &= \frac{d^n f}{dx^n}g + \binom{n}{1}\frac{d^{n-1}f}{dx^{n-1}}\frac{dg}{dx} + \binom{n}{2}\frac{dx^{n-2}f}{dx^{n-2}}\frac{d^2g}{dx^2} + \ldots\\
&\quad + \binom{n}{k-1}\frac{d^{n-(k-1)}f}{dx^{n-(k-1)}}\frac{d^{k-1}g}{dx^{k-1}} + \binom{n}{k}\frac{d^{n-k}f}{dx^{n-k}}\frac{d^kg}{dx^k} + \ldots + f\frac{d^ng}{dx^n}\\
&= \sum_{k=0}^{n}\binom{n}{k}\frac{d^{n-k}f}{dx^{n-k}}\frac{d^kg}{dx^k}. \qquad (3.9)
\end{aligned}$$

The proof of Leibniz Theorem uses the method of mathematical induction. It is omitted in this set of notes. The interested readers can consult the textbook by Hirst (page 104) for the proof.

Example 3.5.1. *Find a formula for the n-th derivative of $x^2 \ln(2x+3)$.*

Let $f(x) = \ln(2x+3)$; $g(x) = x^2$. Notice that for g, the third and subsequent derivatives are all zero so that only the first three terms in Leibniz formula are non-zero. We use the formula (2.11) for n-th derivative of $f(x)$, which is

$$f^{(n)}(x) = (n-1)!\frac{(-1)^{(n+1)}2^n}{(2x+3)^n}.$$

Leibniz Theorem, therefore, gives

$$\begin{aligned}(fg)^{(n)}(x) &= f^{(n)}(x)\cdot x^2 + \binom{n}{1} f^{(n-1)}(x)\cdot 2x + \binom{n}{2} f^{(n-2)}(x)\cdot 2 \\ &= (n-1)!\frac{(-1)^{(n+1)}2^n}{(2x+3)^n}\cdot x^2 \\ &+ n(n-2)!\frac{(-1)^n 2^{n-1}}{(2x+3)^{n-1}}\cdot 2x \\ &+ \frac{n(n-1)}{2!}\cdot (n-3)!\frac{(-1)^{(n-1)}2^{n-2}}{(2x+3)^{n-2}}\cdot 2\end{aligned}$$

If we extract the factor of $\dfrac{(n-3)!(-1)^{(n+1)}2^{n-2}}{(2x+3)^n}$ from each of the three terms above, we are left with

$$(n-1)(n-2)\cdot 2^2\cdot x^2 - n(n-2)\cdot 2\cdot 2x(2x+3) + n(n-1)(2x+3)^2,$$

which, after some straightforward algebraic manipulations, simplifies to $8x^2 + 12nx + 9n^2 - 9n$. We, therefore, have shown that

$$\frac{d^n}{dx^n}[x^2\ln(2x+3)] = \frac{(n-3)!(-1)^{(n+1)}2^n}{(2x+3)^n}(8x^2+12nx+9n^2-9n).$$

Chapter 4

Applications of the derivative

4.1 Gradient of tangent to a curve

We observed that $f'(a)$, the value of the derivative $f'(x)$ at $x = a$, corresponds to the gradient of the tangent to the graph of $y = f(x)$ at the point $(a, f(a))$. This enables us to write down the equation for the tangent line at $(a, f(a))$ as

$$y - f(a) = f'(a)(x - a).$$

This is used in many problems in coordinate geometry as explained in the following examples.

Example 4.1.1. *Let $A = (0, -a)$ be a point on the negative y-axis ($a > 0$). Find the equations of the tangents to the parabola $y = x^2$ which pass through the point A. Find the distance between the points where these tangents cross the x-axis.*

The equation of the tangent at a point (p, p^2) of the parabola is

$$y - p^2 = 2p(x - p),$$

which can be rearranged as,

$$y = 2px - p^2. \tag{4.1}$$

These lines meet the y-axis at $y = -p^2$. As we anticipate that this point be A, we require $p^2 = a$, i.e., $p = \pm\sqrt{a}$. Therefore, the coordinates of the points, on the parabola to which the required tangents are drawn, are given by $(\pm\sqrt{a}, a)$.

These tangent lines cross the x-axis where $y = 0$. Therefore, substituting $y = 0$ in Equation (4.1) leads to

$$x = \frac{p}{2} = \pm\frac{\sqrt{a}}{2},$$

delineating the fact that the two tangent lines cross the x-axis at $(-\frac{\sqrt{a}}{2}, 0)$ and at $(\frac{\sqrt{a}}{2}, 0)$. And therefore we conclude that the distance between the points where the two tangents cross the x-axis is $\frac{\sqrt{a}}{2} + \frac{\sqrt{a}}{2} = \sqrt{a}$.

Figure 4.1 (which corresponds to the case $a = 4$) exhibits the symmetry of the underlying problem.

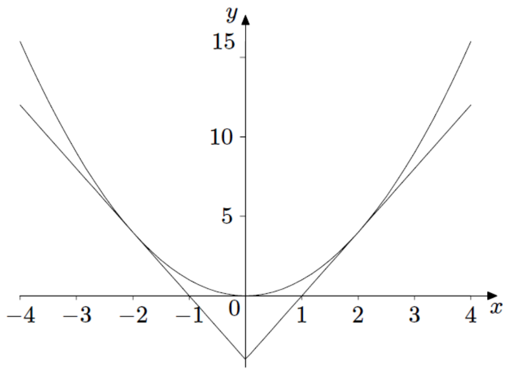

Figure 4.1: *Parabola and tangents*

Example 4.1.2. *Let A be a point on the ellipse given parametrically by $x = a\cos t$, $y = b\sin t$. Point B is the reflection of A about the y-axis. The tangents at A and B cross at right-angles. Find the admissible values of t.*

Since points A and B are reflections of each other about the y-axis, they must have coordinates $(\pm a\cos t,\ b\sin t)$, which, in turn, can be written in the form

$$(a\cos t, b\sin t),\ (a\cos(\pi - t), b\sin(\pi - t)). \tag{4.2}$$

To find the gradients of the respective tangents to the points of the ellipse with coordinates given by (4.2), one needs to employ the technique of parametric differentiation. Using Equation (3.4), the gradients of the respective tangents can be written down as

$$-\frac{b}{a}\cot t \text{ and } -\frac{b}{a}\cot(\pi - t),$$

respectively. For the tangents to cross at right-angles, the product of the gradients must be -1 as we know from elementary coordinate geometry. Therefore,

$$\left(\frac{b}{a}\cot t\right)\left(\frac{b}{a}\cot(\pi - t)\right) = -1.$$

Using $\cot(\pi - t) = -\cot t$, one arrives at

$$\frac{b^2}{a^2}\cot^2 t = 1, \text{ i.e., } \tan t = \pm\frac{b}{a},$$

so that the admissible values of t are the solutions of the equation above. In particular, if the ellipse is a circle with $a = b$, then $\tan t = \pm 1$ resulting in $t = \pm\dfrac{\pi}{4}$.

4.2 Maxima and Minima

Finding how large or small a given quantity is, can be of utmost importance in many problems of science, engineering, economics, etc. And when this quantity is given by an algebraic formula, we can use differentiation to determine these desired maximum or minimum values. The following definition classifies various kinds of maxima and minima.

Definition 4.2.1. *Suppose that f is a function with domain D and that a is a number in D.*

*(i) f is said to have a **global maximum** at a if $f(x) \leq f(a)$ for all x in D.*

*(ii) f is said to have a **global minimum** at a if $f(x) \geq f(a)$ for all x in D.*

*(iii) f is said to have a **local maximum** at a if there is an interval $(a-\delta, a+\delta)$ such that $f(x) \leq f(a)$ for all x in D belonging to this interval.*

*(iv) f is said to have a **local minimum** at a if there is an interval $(a-\delta, a+\delta)$ such that $f(x) \geq f(a)$ for all x in D belonging to this interval.*

Remark 3. *Global maximum/minimum is also a local maximum/minimum if it is attained at some interior point of the domain D and not at its boundary. The local maximum/minimum may also fail to be the global one.*

In some texts "absolute" and "relative" are used for "global" and "local", respectively. There is a single term called extrema (local or global) to refer to both maximum and minimum.

Example 4.2.1. *Locate the maxima and minima of the function defined by $f(x) = x^3 - 3x^2 - 9x$, for $-4 \leq x \leq 6$, by plotting its graph.*

In this example, we have set the domain D to be the interval specified by $-4 \leq x \leq 6$. We can see from the graph that f has a global minimum at $x = -4$. It is the point where the value of $f(x)$ is the smallest compared to all other function values attained in the domain D. Likewise, there is a global maximum at $x = 6$ because $f(6)$ is larger than all other values of $f(x)$ with x belonging to the domain D.

When $x = -1$, we have a local maximum. We can see from the graph that $f(x) \leq f(-1)$ for all x satisfying $-2 \leq x \leq 0$. This is not a global maximum because, for example, $f(6) > f(-1)$.

When $x = 3$ we have a local minimum. We can see from the graph that $f(x) \geq f(3)$ for all x satisfying $0 \leq x \leq 4$. This is not a global minimum because, for example, $f(-4) < f(3)$.

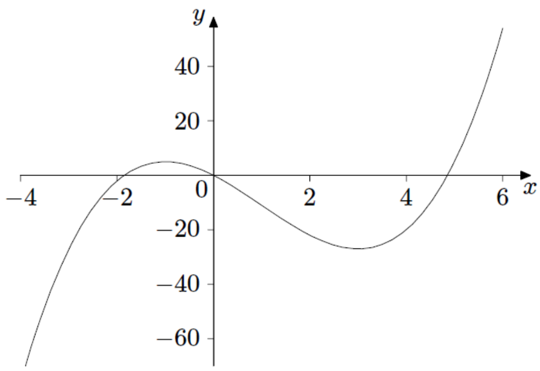

Figure 4.2: *Maximum and minimum*

To complete the description of the maxima and minima we have to calculate the value of the function at each of the corresponding values of x.

So, to summarise:

There is a local maximum at the point $(-1, f(-1)) = (-1, 5)$.
There is a local minimum at the point $(3, f(3)) = (3, -27)$.
There is a global maximum at the point $(6, f(6)) = (6, 54)$.
There is a global minimum at the point $(-4, f(-4)) = (-4, -76)$.

Note that the function, given in Example 4.2.1, attains a global maximum at $6 \in [-4, 6]$. Since 6 belongs to the boundary $\{-4, 6\}$ of the domain $[-4, 6]$, the attained global maximum is not a local maximum validating the remark 3. We shall now see how to find such points using differentiation.

Theorem 4.2.1. *Suppose that f is differentiable at $x = a$ and that $x = a$ is a local maximum or a local minimum. Then $f'(a) = 0$.*

Proof

We use the limit definition of the derivative (Definition 2.1.1), namely

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

This is a two-sided limit so that both the one-sided limits exist and are equal. Therefore, one obtains

$$f'(a) = \lim_{x\to a^+} \frac{f(x)-f(a)}{x-a} = \lim_{x\to a^-} \frac{f(x)-f(a)}{x-a}.$$

Suppose that the function has a local maximum at $x = a$. Then $f(x) \le f(a)$ for all x in some interval I containing a. Therefore, for all x in this interval, we have $f(x) - f(a) \le 0$. For the right hand limit, we have $x - a > 0$ so that for x in I, we deduce that $\dfrac{f(x)-f(a)}{x-a} \le 0$. Since we are considering the limit $x \to a$, x can always be restricted to the interval I. Therefore,

$$f'(a) = \lim_{x\to a^+} \frac{f(x)-f(a)}{x-a} \le 0. \tag{4.3}$$

For the left hand limit, we have $x - a < 0$ so that for all $x \in I$, we obtain, $\dfrac{f(x)-f(a)}{x-a} \ge 0$ leading to

$$f'(a) = \lim_{x\to a^-} \frac{f(x)-f(a)}{x-a} \ge 0. \tag{4.4}$$

Combining Equation (4.3) with Equation (4.4), one concludes that $f'(a) = 0$.

The case of a local minimum is handled exactly in the same way.

Remark 4. *The converse of the theorem 4.2.1 is not true in general, i.e., there are situations when the derivative is zero but we have neither a local maximum nor a local minimum. One of the simplest examples is given by $f(x) = x^3$. Differentiation yields $f'(x) = 3x^2$ so that one has $f'(0) = 0$. This point is neither a local maximum nor a local minimum because x^3 is negative for all $x < 0$ and positive for all $x > 0$. Nevertheless, points with such properties are also of considerable importance in interpreting graphs of various functions.*

In order to proceed further in an organized way, we need the following definitions.

Definition 4.2.2. *Any point a for which either $f'(a) = 0$ or $f'(a)$ does not exist i.e., f is not differentiable at the point a is called a critical point of f. There are two types of critical points. A point a for which $f'(a)$ does not exist or is undefined is called a singular point of the function f. If, on the other hand, $f'(a) = 0$, then a is called a stationary point of f. A turning point on the graph of a function is where the first derivative changes its sign.*

It is also important to note, in this case, that if $f'(x) \geq 0$ ($f'(x) > 0$) at each point in an interval I lying inside the domain, then f is said to be increasing (strictly increasing) in I. On the other hand, if $f'(x) \leq 0$ ($f'(x) < 0$) everywhere in an interval I lying in its domain, then it is said to be decreasing (strictly decreasing) in I. A few remarks related to Definition 4.2.2 are in order.

Remark 5. *If the function is differentiable, then the turning points are also stationary points. But there may be stationary points of the function that are not turning points. To illustrate this fact, consider the function f given by $f(x) = x^3$. Here, $x = 0$, being a stationary point of the function $f(x) = x^3$, is not a turning point because the function is increasing on both sides of the point $x = 0$, i.e., $f'(x) > 0$ for both $x > 0$ and $x < 0$.*

There are a few other important terms related to the second derivative of the given function that need to be defined separately. We do it at this stage.

Definition 4.2.3. *Given a twice differentiable function f in an interval I lying in its domain ($f''(x)$ exists everywhere in I), it is called concave upward (downward) if $f''(x) > 0$ ($f''(x) < 0$) everywhere in I. An inflection point of f is a point on its graph where the concavity changes from concave upward to concave downward or vice versa. A stationary inflection point of a given function f is not only an inflection point but also a stationary point of f. A non-stationary inflection point of f is, on the other hand, an inflection point that is not a stationary point of f.*

In order to gather a clear understanding of the terminologies introduced in Definitions 4.2.2 and 4.2.3, we urge the students to refer to the following figure.

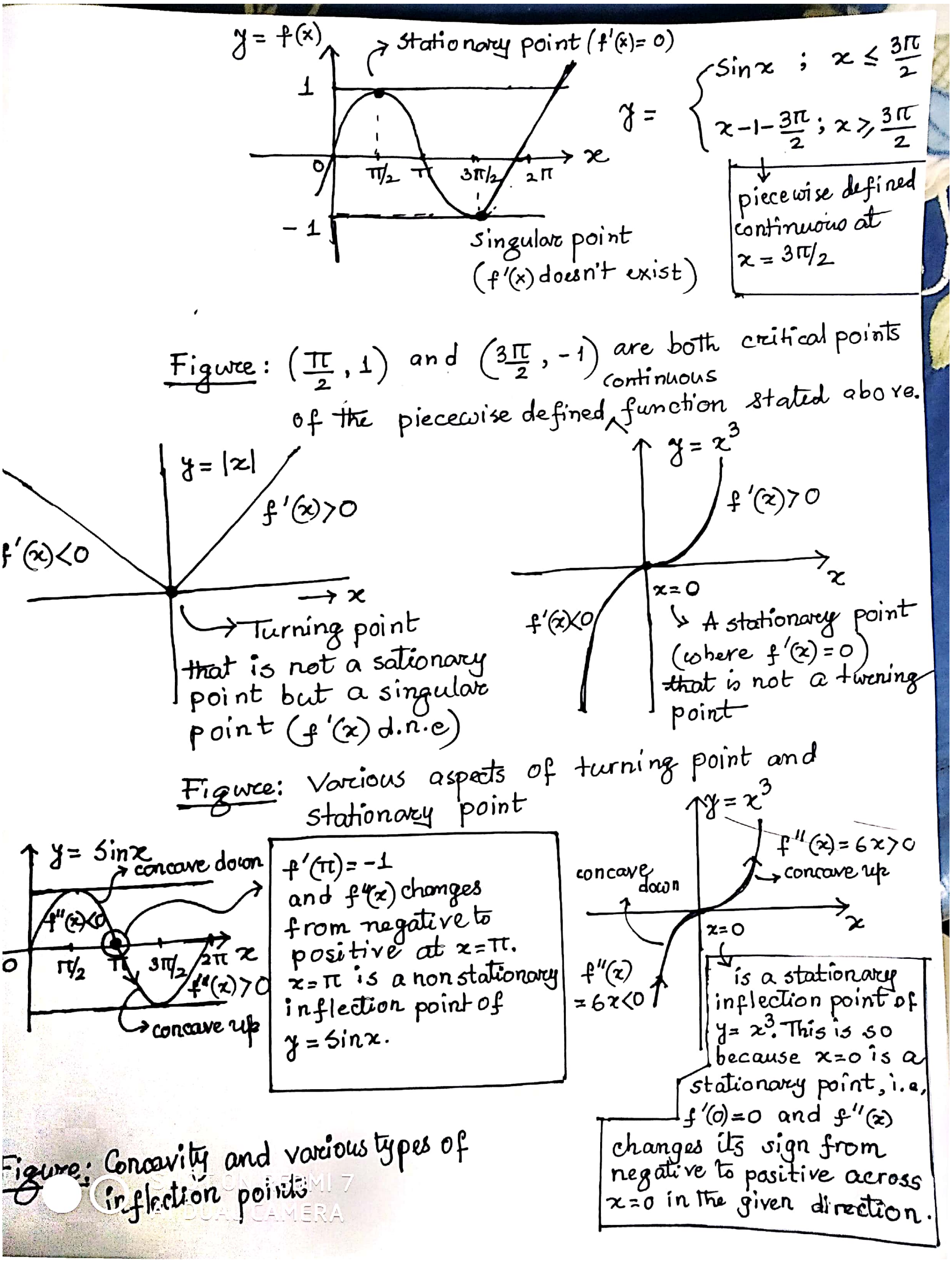

Figure 4.3: Various aspects of critical point, stationary point, turning point, concavity and inflection point of a given function

A few remarks related to Definition 4.2.3 are in order.

Remark 6. *In order to find all possible inflection points associated with the graph of a function, one first has to determine a suitable set of candidates to qualify as inflection point. These candidates are the ones where the second derivative f'' is either zero or undefined. In particular, one may have an inflection point on the graph of a function f where f'' is undefined. For example, $x = 0$ is an inflection point of the function f given by $f(x) = x^{\frac{1}{3}}$ because $f''(x) > 0$ (concave upward) when $x < 0$ and $f''(x) < 0$ (concave downward) when $x > 0$. But $f''(x) = -\frac{2}{9}x^{-\frac{5}{3}}$ is undefined at $x = 0$.*

Before embarking on a detailed discussion on how to sketch the graph of a function using derivatives, we state an important theorem.

Theorem 4.2.2. *(The second derivative test) Suppose that $f(x)$ has a stationary point at $x = a$, i.e., $f'(a) = 0$ and that f is twice differentiable at $x = a$. The following hold*

(i) If $f''(a) > 0$, then the stationary point $x = a$ is a local minimum.

(ii) If $f''(a) < 0$, then the stationary point $x = a$ is a local maximum.

(iii) $f''(a) = 0$ gives no information about the type of the stationary point $x = a$, i.e., it is neither a maximum nor a minimum.

Proof

We use the limit definition 2.1.1 of the derivative of a function to write the derivative of the first derivative of the given function f at $x = a$ as

$$(f')'(a) = f''(a) = \lim_{x \to a} \frac{f'(x) - f'(a)}{x - a}.$$

In this case, we have $f'(a) = 0$ and therefore we obtain from above

$$f''(a) = \lim_{x \to a} \frac{f'(x)}{x - a}.$$

(i) $f''(a) > 0$ tells us that for x near a, we must have $f'(x) < 0$ if $x < a$ and $f'(x) > 0$ if $x > a$. In other words, f is strictly decreasing to the left of a while f is strictly increasing to the right of a so that one ends up with a local minimum of f at $x = a$.

(ii) If $f''(a) < 0$, then for x being sufficiently close to a, $f'(x) > 0$ whenever $x < a$ and $f'(x) < 0$ when $x > a$. Therefore, f is strictly increasing to the left of a and it is strictly decreasing to the right of a so that f has a local maximum at $x = a$.

(iii) Here we will provide three examples satisfying $f'(a) = 0$ and $f''(a) = 0$. In the first example, $x = a$ is neither a maximum nor a minimum. In the second example, $x = a$ is a minimum while in the third example $x = a$ will be a maximum proving that $f''(a) = 0$ is an inconclusive test as claimed in the theorem. The first example is given by $f(x) = x^3$. Here, one has $f'(0) = 0$ and $f''(0) = 0$ but $x = 0$ is neither a maximum nor a minimum of f as studied earlier. The second example pertains to the function given by $f(x) = x^4$. Here, one also obtains $f''(x) = 12x^2$ resulting in $f''(0) = 0$. But contrary to the first example, the function, here, achieves a minimum at $x = 0$. The third example concerns the function given by $f(x) = -x^4$. The second derivative of this function is given by $f''(x) = -12x^2$ which also yields $f''(0) = 0$. This function, contrary to the ones provided in the first two examples, attains a maximum at $x = 0$.

Note that it is usually possible to investigate case (iii) using Taylor series as will be discussed in Chapter 5.

General guidelines for sketching graph of a function:

The following checklist is intended as a guide to sketching a curve $y = f(x)$ by hand. Not every item is relevant to every function. (For instance, a given curve might not have an asymptote or possess symmetry.) But the guidelines provide all the information you need to make a sketch that displays the most important aspects.

1. **Domain** It's often useful to start by determining the domain D of f, that is, the set of values of x for which $f(x)$ is defined.

2. **Intercepts** The y-intercept is $f(0)$ and this tells us where the curve intersects the y-axis. To find the x-intercepts, we set $y = 0$ and solve for x. (You can omit this step if the equation is difficult to solve).

3. **Symmetry**
(i) If $f(x) = f(-x)$ for all x in D , that is, the equation of the curve is unchanged when x is replaced by $-x$, then f is an **even function** and the curve is symmetric about the y-axis. This means that our work is cut in half. If we know what the curve looks like for $x \geq 0$, then we need only reflect about the x-axis to obtain the complete curve.

   (ii) If $f(-x) = -f(x)$ for all x in D, then f is an **odd function** and the curve is symmetric about the origin. Again we can obtain the complete curve if we know what it looks like for $x \geq 0$.[Rotate 180 about the origin].

   (iii) If $f(x + p) = f(x)$ for all x in D, where p is a positive constant, then is f called a **periodic function** and the smallest such number p is called the period. For instance, $y = \sin x$ has period 2π and $y = \tan x$ has period π . If we know what the graph looks like in an interval of length p, then we can use translation to sketch the entire graph.

4. **Asymptotes** (revisiting Definition 1.3.1)
(i) *Horizontal Asymptotes* If $\lim_{x\to\infty} f(x) = L$ or $\lim_{x\to-\infty} f(x) = L$, then the line $y = L$ is a horizontal asymptote of the curve $y = f(x)$. If it turns out that $\lim_{x\to\infty} f(x) = \infty$ (or $-\infty$), then we do not have an asymptote to the right, but that is still useful information for sketching the curve.

   (ii) *Vertical Asymptotes* The line $x = a$ is a vertical asymptote if at least one of the following statements is true:

$$\lim_{x\to a^+} f(x) = \infty \qquad \lim_{x\to a^-} f(x) = \infty$$

$$\lim_{x\to a^+} f(x) = -\infty \qquad \lim_{x\to a^-} f(x) = -\infty$$

(For rational functions you can locate the vertical asymptotes by equating the denominator to 0 after cancelling any common factors. But for other functions this method does not apply.) Furthermore, in sketching the curve it is very useful to know exactly which of the statements in above is true.

(iii) *Slant Asymptotes* Some curves have asymptotes that are oblique, that is, neither horizontal nor vertical. If

$$\lim_{x\to\infty} [f(x) - (mx + b)] = 0$$

then the line $y = mx + b$ is called a **slant asymptote** because the vertical distance between the curve $y = f(x)$ and the line $y = mx+b$ approaches 0. For rational functions, slant asymptotes occur when the degree of the numerator is one more than the degree of the denominator. In such a case the equation of the slant asymptote can be found by long division.

5. **Intervals of Increase or Decrease** Use the First Derivative Test. Compute $f'(x)$ and find the intervals on which $f'(x)$ is positive (f is increasing) and the intervals on which $f'(x)$ is negative (f is decreasing).

6. **Local Maximum and Minimum Values** Find the critical numbers of f [the numbers c where $f'(c) = 0$ or $f'(c)$ does not exist]. Then use the First Derivative Test. If f' changes from positive to negative at a critical number c, then $f(c)$ is a local maximum. If f' changes from negative to positive at c, then $f(c)$ is a local minimum. Although it is usually preferable to use the First Derivative Test, you can use the Second Derivative Test if $f'(c) = 0$ and $f''(c) \neq 0$. Then $f''(c) > 0$ implies that $f(c)$ is a local minimum, whereas $f''(c) < 0$ implies that $f(c)$ is a local maximum.

7. **Concavity and Points of Inflection** Compute $f''(x)$ and use the Concavity Test. The curve is concave upward where $f''(x) > 0$

and concave downward where $f''(x) < 0$. Inflection points occur where the direction of concavity changes.

8. **Sketch the Curve** Using the information in items 1 - 7, draw the graph. Sketch the asymptotes as dashed lines. Plot the intercepts, maximum and minimum points, and inflection points. Then make the curve pass through these points, rising and falling according to 5, with concavity according to 7, and approaching the asymptotes. If additional accuracy is desired near any point, you can compute the value of the derivative there. The tangent indicates the direction in which the curve proceeds.

Example 4.2.2. *Use the guidelines to sketch the curve* $y = \dfrac{2x^2}{x^2-1}$.

1. The domain is

$$\{x|x^2-1 \neq 0\} = \{x|x \neq \pm 1\} = (-\infty, -1) \cup (-1, 1) \cup (1, \infty)$$

2. The x- and y-intercepts are both 0.

3. Since $f(-x) = f(x)$, the function f is even. The curve is symmetric about the y-axis.

4.

$$\lim_{x\to\pm\infty} \frac{2x^2}{x^2-1} = \lim_{x\to\pm\infty} \frac{2}{1-1/x^2} = 2$$

Therefore, the line $y = 2$ is a horizontal asymptote.
Since the denominator is 0 when $x = \pm 1$, we compute the following limits:

$$\lim_{x\to 1^+} \frac{2x^2}{x^2-1} = \infty \qquad \lim x \to 1^- \frac{2x^2}{x^2-1} = -\infty$$

$$\lim_{x\to 1^+} \frac{2x^2}{x^2-1} = -\infty \qquad \lim x \to 1^- \frac{2x^2}{x^2-1} = \infty$$

Therefore, the lines $x = 1$ and $x = -1$ are vertical asymptotes. This information about limits and asymptotes enables us to draw a preliminary sketch shown below in Figure 4.4, showing the parts of the curve near the asymptotes.

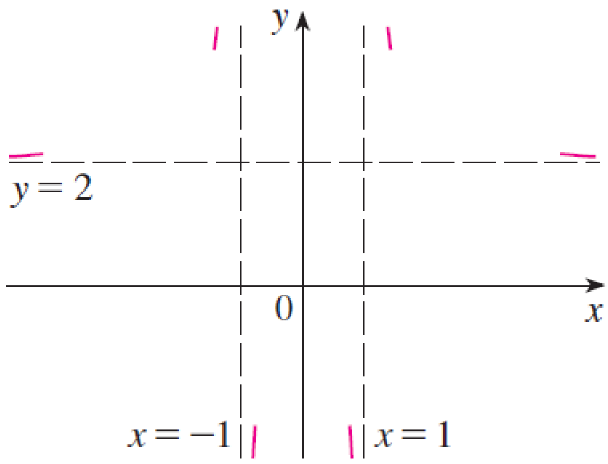

Figure 4.4: Preliminary Sketch

5.
$$f'(x) = \frac{4x(x^2-1) - 2x^2 \cdot 2x}{(x^2-1)^2} = \frac{-4x}{(x^2-1)^2}$$

Since $f'(x) \geq 0$ when $x < 0 (x \neq -1)$ and $f'(x) \leq 0$ when $x > 0 (x \neq 1)$, f is increasing on $(-\infty, -1)$ and $(-1, 0)$ and decreasing on $(0, 1)$ and $(1, \infty)$

6. The only critical number is $x = 0$. Since f' changes from positive to negative at 0, $f(0) = 0$ is a local maximum by the First Derivative Test.

7.
$$f''(x) = \frac{-4(x^2-1)^2 + 4x \cdot 2(x^2-1)2x}{(x^2-1)^4} = \frac{12x^2+4}{(x^2-1)^3}$$

Since $12x^2 + 4 > 0$ for all x, we have

$$f''(x) > 0 \iff x^2 - 1 > 0 \iff |x| > 1$$

and $f''(x) < 0 \iff |x| < 1$. Thus the curve is concave upward on the intervals $(-\infty, -1)$ and $(1, \infty)$ and concave downward on $(-1, 1)$. It has no point of inflection since 1 and -1 are not in the domain of f

8. Using the above information, we finish the sketch in Figure 4.5.

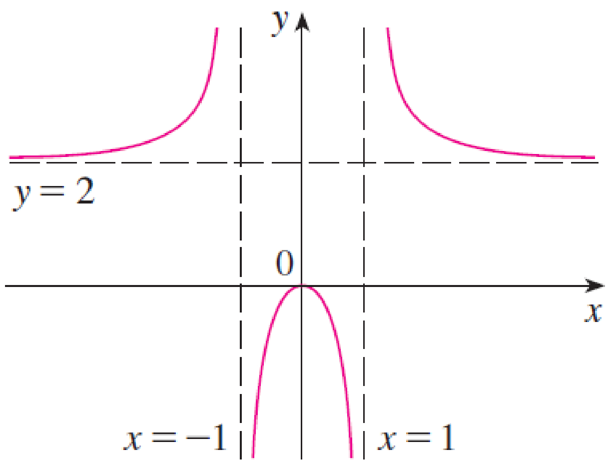

Figure 4.5: Finished sketch of $y = \dfrac{2x^2}{x^2 - 1}$

4.3 Optimisation Problems

In this section, we will be concerned with optimisation problems. This amounts to the fact that we will be trying to find a solution to a problem which optimises (finds the best value for) some quantity such as volume, cost etc. We shall consider problems where the optimum solution corresponds to a maximum or a minimum. Problems of this type are often stated verbally. There is a common strategy to attack such problems using the following steps.

1. Choose variables to represent the quantities described in the problem.
2. If possible draw a diagram and indicate these quantities on the diagram.
3. Use the diagram and the information provided to find expressions that establish relationship between the quantities in question.

4. Use these expressions to determine a formula for the quantity to be optimised as a function of one variable.
5. Determine the restrictions on the variables arising from the conditions stated in the problem (for example, an area should be positive).
6. Use the techniques of calculus to determine the stationary points of the function mentioned in step 4 taking the restrictions on the variables into full consideration.
7. Determine maxima or minima depending on the optimisation problem that one is presented with.

In this section, we restrict our attention to problems involving one variable only although, in practice, optimisation problems often involve many variables. Operational research involves study of such multivariable optimisation problems.

Example 4.3.1. *Find the point on the curve $y = x^3$ that is nearest to* $(4, 0)$.

The variables, here, will be the coordinates of an arbitrary point (x, y) on the curve and the distance D of that point from $(4, 0)$. The variables x and y are, of course, related by the equation of the curve given by $y = x^3$.

We know that the distance will involve a square root from Pythagoras' Theorem. It appears, therefore, to be easier algebraically to consider the square of the distance given by $D^2 = (x-4)^2+y^2$. Using the relationship $y = x^3$, one then obtains

$$D^2 = (x-4)^2 + x^6. \tag{4.5}$$

We have now obtained a formula for the quantity to be optimised as a

function of one variable, i.e., the x coordinate only. To find its minimum value, we look for the stationary points of D^2 expressed as a function of x. Differentiation yields

$$\frac{d(D^2)}{dx} = 2(x-4) + 6x^5 = 2(x-1)(3x^4 + 3x^3 + 3x^2 + 3x + 4). \tag{4.6}$$

Upon using chain rule in Equation (4.6), one obtains

$$\frac{d(D^2)}{dx} = 2D\frac{dD}{dx} = 2(x-1)p(x),$$

leading to

$$\begin{aligned}\frac{dD}{dx} &= \frac{(x-1)p(x)}{D} \\ &= \frac{(x-1)p(x)}{\sqrt{(x-4)^2+x^6}}.\end{aligned} \tag{4.7}$$

It is evident from the graph of $y = x^3$ that the nearest point to $(4, 0)$ on the curve will lie in the first quadrant where $x > 0$. The polynomial factor $p(x) = 3x^4+3x^3+3x^2+3x+4$ on the right hand side of Equation (4.7) is clearly positive for all positive values of x.

Therefore, the derivative of the distance function with respect to the x-coordinate can only be zero when $x = 1$. It reflects the fact that $(1, 1)$ is a stationary point of the distance function D. Now, based on the discussion conducted above in the light of Equation (4.7), notice that for $0 < x < 1$, $\frac{dD}{dx} < 0$ while for $x > 1$ one always has $\frac{dD}{dx} > 0$ so that the stationary point corresponding to $x = 1$ is indeed a minimum located in the first quadrant. By substituting $x = 1$ in Equation (4.5), one obtains the desired minimum value of the distance function D given by $D_{\text{min}} = \sqrt{(1-4)^2+1^6} = \sqrt{10}$.

Example 4.3.2. *A rectangular piece of plastic sheeting* $10m$ *long and* $2m$ *wide is folded in half lengthwise to form a tunnel in the shape of a triangular prism to protect plants in the garden from cold weather. What should be the height of the prism to maximise the volume of the prism?*

The variables involved in this problem are the volume V, the height h and half the base b as shown in Figure 4.6. Pythagoras' Theorem tells us that $h^2 + b^2 = 1$ and the volume is given by $V = 10hb$. It is evident from the preceding equations that optimisation of the square of the volume, in this case, will be a lot easier to handle. We, therefore, eliminate b to obtain

$$V^2 = 100h^2b^2 = 100h^2(1-h^2) = 100(h^2-h^4). \tag{4.8}$$

Differentiation with respect to h then yields,

$$\frac{d(V^2)}{dh} = 100(2h-4h^3) = 200h(1-2h^2) = 200h(1-h\sqrt{2})(1+h\sqrt{2}). \tag{4.9}$$

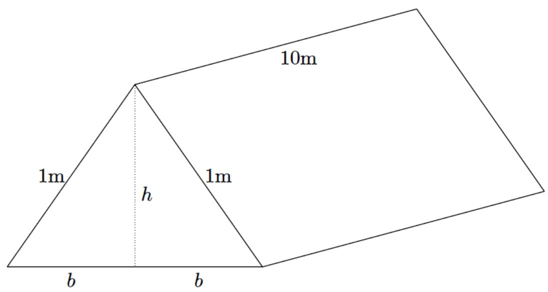

Figure 4.6: *Diagram for Example 4.3.2*

Now using chain rule, as in Example 4.3.1, one obtains the following from Equation 4.3.

$$\frac{d(V^2)}{dh} = 2V\frac{dV}{dh} = 200h(1-h\sqrt{2})(1+h\sqrt{2}),$$

leading to

$$\begin{aligned} \frac{dV}{dh} &= \frac{100h(1-h\sqrt{2})(1+h\sqrt{2})}{V} \\ &= \frac{10(1-h\sqrt{2})(1+h\sqrt{2})}{\sqrt{1-h^2}}. \end{aligned} \tag{4.10}$$

We want to set the derivative to zero in Equation (4.10) since we are looking for the stationary points of the volume V expressed as a function of the height h. Also, note the implicit constraint $0 < h < 1$ imposed on the parameter h that follows from Equation (4.8) and from the positivity of the height. The only value of $h \in (0,1)$ yielding a zero derivative in (4.10) is $h = \frac{1}{\sqrt{2}}$. Therefore, the stationary point of the volume V expressed as a function of height h corresponds to $h = \frac{1}{\sqrt{2}}$ only.

Let us now try to determine the type of this stationary point. Using Equation (4.10), one deduces that $\frac{dV}{dh} > 0$ when $0 < h < \frac{1}{\sqrt{2}}$ while $\frac{d(V^2)}{dh} < 0$ when $1 > h > \frac{1}{\sqrt{2}}$. Therefore, the stationary point corresponding to $h = \frac{1}{\sqrt{2}}$ is where the maximum of V, as a function of h, is achieved.

The maximum volume, therefore, follows from Equation (4.8).

$$V_{\max} = \sqrt{100\left(\frac{1}{2} - \frac{1}{4}\right)} = 5\mathrm{m}^3.$$

4.4 Problems involving growth and decay

We encounter images of colonies of bacteria in various medical programmes on television. The size of the colony seems to grow slowly at first but the rate of growth increases rapidly as time progresses. The experts tell us that the bacteria divide to produce new bacteria. So, for example, if there are 10^6 bacteria at the onset, then after one division there will be twice as many, namely, 2×10^6. Upon another division, there will again be twice as many, i.e., 4×10^6. Following this pattern, one finds that the total number of bacteria will rise to $2^n \times 10^6$ after n divisions (assuming that none die). We immediately find that the rate of growth depends on how many bacteria are present there at any given stage of the process. The number of bacteria in a colony is very large and division happens very quickly. Therefore, we model the process by taking the number of bacteria to be $X(t)$, a quantity that changes continuously with time t. With this assumption, $X(t)$ is not restricted to integer values only and we can use calculus to model the growth of the colony. It has been found that the model described above represents such growth processes in nature with high precision. The rate of growth of $X(t)$ will be modelled by the derivative and the assumption that this is proportional to the quantity present at time t leads to the equation

$$\frac{dX}{dt} = kX,$$

where k is the constant of proportionality. This is known as the exponential growth model. Solving such equations is the province of the study of differential equations, which is outside the scope of this book. However, in this case we are looking for a function $X(t)$ whose derivative is itself, multiplied by a constant k and our experience with differentiation suggests the following trial solution

$$X(t) = Ae^{kt}.$$

Since $X(0) = A$, the constant A represents the size of the colony of bacteria at initial time $t = 0$.

In such a growth process, we have $k > 0$ and $Ae^{kt} \to \infty$ as $t \to \infty$. It is evident, therefore, that the model only applies for a restricted interval of time, for otherwise the colony would eventually weigh more than the earth! What happens of course is that certain factors come into play which we did not build into the model. Bacteria die and growth slows down as food runs out.

What we need to know for a particular growth process is the value of the constants A and k. These can be determined using measurements as will be explained in the following examples.

Example 4.4.1. *A colony of bacteria weighs* $1\mu g$. *After 10 seconds, it weighs* $1.5\mu g$. *Determine the rate of growth assuming an exponential model and hence deduce the weight of the colony after 20 seconds.*

Let $X(t)$ denote the weight of the colony at time t with $X(t) = Ae^{kt}$. When $t = 0, X(0) = A = 1\mu\text{g}$. Also, when $t = 10$, we know that $X(10) = e^{10k} = 1.5$. Therefore, $k = \ln(1.5)/10$. After 20 seconds, we have

$$X(t) = e^{20\frac{\ln 1.5}{10}}\mu\text{g} = e^{\ln(1.5)^2}\mu\text{g} = 2.25\mu\text{g}.$$

Example 4.4.2. *A large circular water lily leaf floating on the surface of a pond grows and spreads out in such a way that the larger it becomes, the faster it grows. The gardener measures the area of the leaf and* *finds after 20 days that it has quadrupled in size. Assuming exponential growth, how long did it take from the gardener's initial measurement to double in size?*

Let $X(t)$ denote the area of the leaf at time t with $X(t) = Ae^{kt}$. We have $X(0) = A$ and $X(20) = 4A = Ae^{20k}$ so that $e^{20k} = 4$, yielding $k = \ln(4)/20$. We want to find the value of t for which $X(t) = 2A$, i.e., $2A = Ae^{kt}$. Therefore

$$t = \frac{\ln 2}{k} = \frac{20 \ln 2}{\ln 4} = \frac{20 \ln 2}{2 \ln 2} = 10.$$

In fact, common sense suggests that if it doubled in 10 days it would be quite likely to double again in another 10 days. This tells us that the exponential growth model agrees with our intuitive ideas about such growth processes.

The opposite of growth is decay and one of the best-known decay processes is the decomposition of radioactive material. We model this situation by assuming that the more atoms there are in a lump of material, the more decay or decomposition will take place suggesting that the rate of decay should be proportional to the number of atoms present in a lump of such material. Again, we can write $\frac{dX}{dt} = kX$ and since it is a decay process rather than a growth one, $X(t)$ should decrease as t increases yielding a negative derivative. And therefore a negative value for k will follow. For this reason, we often make the negative sign explicit and write the governing equation in the form

$$\frac{dX}{dt} = -kX, \quad \text{with the solution given as} \quad X(t) = Ae^{-kt}.$$

We call it an exponential decay. The rate of decay of a radioactive material is usually specified by providing its half-life. Half-life measures the time it takes for a given amount of the substance to become reduced by half as a consequence of decay. Some of the transuranic elements are rather evanescent, having a very short half-life. Others have a very long half-life and unfortunately some of the radioactive waste products from nuclear fission fall into this category.

Example 4.4.3. *A radioactive element has a half-life of H years. Every year a nuclear reactor produces a quantity of this element as waste.*

After how long will this batch of waste contain only 10% of the original amount of the radioactive material?

Let $X(t)$ denote the amount of material remaining after time t. Assuming exponential decay, we have $X(t) = Ae^{-kt}$, where the original amount of material is $X(0) = A$. When $t = H$, the half-life, $X(H) = \frac{1}{2}A$. Therefore,

$$\frac{A}{2} = Ae^{-kH}, \text{ giving } k = \frac{\ln 2}{H}$$

Writing $X(t) = X(0)/10$, one obtains

$$\frac{A}{10} = Ae^{-kt}, \quad \text{so that} \quad t = H\,\frac{\ln 10}{\ln 2} \approx 3.32H$$

One of the waste products from nuclear reactors is Plutonium 239. This has a half-life of 24,400 years. Therefore, the time taken for 90% of a batch of such material to decay will be approximately 81,000 years.

Example 4.4.4. *Newton's law of cooling states that a body cools at a rate which is proportional to the difference between its temperature and the temperature of its surroundings. On a warm day, when room temperature is 24°C, a cup of coffee is poured out at 80°C. After 5 minutes, it has cooled to 60°C. How long must I wait if I like my coffee to be at 45°C before drinking it?*

Let $T(t)$ denote the temperature of the coffee at time t. Newton's law of cooling tells us that $\frac{dT}{dt} = -k(T - 24)$. We first note that $\frac{dT}{dt} = \frac{d}{dt}(T - 24)$ so that we can write the equation as

$$\frac{d}{dt}(T - 24) = -k(T - 24) \text{ so that } T - 24 = Ae^{-kt}.$$

When $t = 0, T = 80$, one has $80 = 24 + A$, yielding $A = 56$. After 5 minutes, when $T = 60$, the equation tells us that

$$60 - 24 = 36 = 56e^{-5k}, \text{ from which } k = \frac{1}{5}\ln\left(\frac{56}{36}\right) = \frac{\ln 56 - \ln 36}{5}.$$

We want to find the value of t for which $T = 45$ and therefore, we require

$$45 - 24 = 21 = 56e^{-kt}, \text{ giving } t = 5\,\frac{\ln 56 - \ln 21}{\ln 56 - \ln 36} \approx 11\text{minutes}.$$

Chapter 5

Maclaurin and Taylor Expansions

How does your calculator work out values of the trigonometric functions, or any of the other functions on the various buttons? The circuitry can do addition, and therefore multiplication, which is in essence repeated addition. It can also do subtraction, and division, which can be thought of as repeated subtraction. It can therefore work out values of polynomials like

$$3 - x + 2x^2 + 4x^3 - 2.7x^4$$

for numerical values of x. We therefore need to be able to find polynomials which approximate to the standard functions, and these could then be programmed into the circuits in a calculator. Clearly such an exercise should include an analysis of the greatest possible errors arising with such approximations if the digits in the display are to be accurate. We start with a simple case.

5.1 Linear Approximation

In this section we concentrate on finding linear approximations to a given function over a specified interval.

The curve shown in Figure 5.1 has equation $y = 10 - x^3$, and we have drawn in three of many possible lines which might be considered as providing linear approximations for this function in the interval $0 \leq x \leq 2$. This interval specifies the two points $P(1, 9)$ and $Q(2, 2)$ indicated on the diagram.

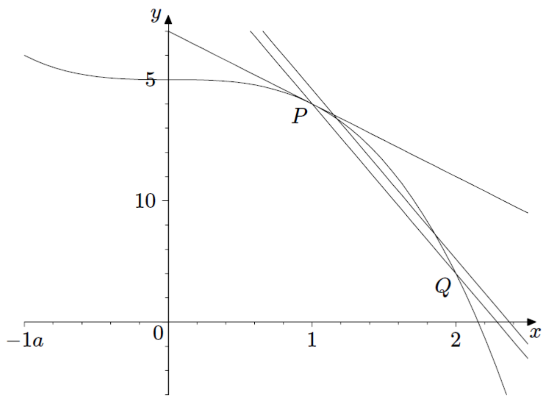

Figure 5.1: Linear approximations

Firstly we have drawn the tangent line at P . This approximates to the curve very well when x is near to 1, but as x increases towards 2 the line diverges from the curve by a considerable amount.

Secondly we have drawn the line passing through P and Q. This provides quite a good approximation across the interval, but near to P it is not so accurate as the tangent line.

Finally we have drawn a line parallel to the line PQ, for which the greatest difference between the y-value on the line and that on the curve, as x varies between 0 and 2, is smaller than it is for PQ. It is better in some places, but worse in others, especially near to $x = 1$ and $x = 2$.

Each of these lines is a possible candidate for a linear approximation. They have been chosen using different criteria, and this serves to emphasise that there is no "best" linear approximation.

Such considerations apply when we extend the discussion to approximating polynomials of higher degree than linear, and several different families of approximating polynomials have been investigated over the years.

In this chapter we shall base the choice of approximation on the notion of tangency, as with the first of the lines discussed above.

If we now generalise the situation so that the curve has equation

$y = f(x)$, and the point P is $(a, f(a))$, then we can work out the equation of the tangent line, and it can be rearranged in the form

$$y = f(a) + f'(a)(x - a) \tag{5.1}$$

The tangent line and the curve have the same gradient at P, and so y changes as x changes at approximately the same rate along the line and the curve in a small interval containing $x = a$. This suggests that the tangent line will give a reasonable approximation to the curve near to $x = a$.

Even simpler than a linear approximation is to use a constant, so that we could say that near to $x = a$, $f(x)$ is approximately equal to $f(a)$. In the next section we shall consider the error which this constant approximation involves.

5.2 The Mean Value Theorem

A proof of this theorem is beyond the scope of this course. It can be found in Real Analysis books. It is important however to have a clear statement of the theorem with all the requisite conditions, and to give a geometrical interpretation.

Theorem 5.2.1 (The Mean Value Theorem). *Suppose that $f(x)$ is a function which is continuous for all values of x satisfying $a \leq x \leq b$, and differentiable for all values of x satisfying $a < x < b$. Then there exists a number c satisfying $a < c < b$ for which*

$$\frac{f(b) - f(a)}{b - a} = f'(c) \tag{5.2}$$

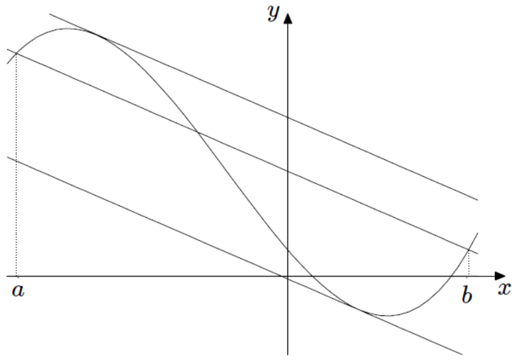

Figure 5.2: The Mean Value Theorem

Geometrically the left-hand side of the equation represents the slope of the chord joining the points $(a, f(a))$ and $(b, f(b))$. The right-hand side is the gradient of the tangent at the point $(c, f(c))$. So the conclusion of the theorem says that under the appropriate conditions of continuity and differentiability, given a chord joining two points on a graph, there is always a tangent at some intermediate point which is parallel to the chord. Figure 5.2 illustrates this, where we have shown a case when there is more than one possible value of c.

Example 5.2.1. *Use the Mean Value Theorem to show that if $f'(x) > 0$ for all x satisfying $a < x < b$, then f is strictly increasing. What can we conclude if the condition $f'(x) > 0$ is replaced by $f'(x) \geq 0$?*

Let x_1 and x_2 be any numbers satisfying $a \leq x1 < x2 \leq b$. Using the Mean Value Theorem gives

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} = f'(c) > 0$$

Since the denominator $x_2 - x_1 > 0$ we deduce that $f(x_2) - f(x_1) > 0$. So f is an increasing function.

If the condition $f'(x) > 0$ is replaced by $f'(x) \geq 0$ the same reasoning tells us that $f(x_2) - f(x_1) \geq 0$. So f is still an increasing function, although not necessarily strictly increasing.

We can re-write the equation for the Mean Value Theorem in the form

$$f(b) = f(a) + f'(c)(b - a) \tag{5.3}$$

To interpret this as providing an approximation it is convenient to regard a as a fixed number, and b as a variable, which in accordance with usual notation we shall replace by x. We can then write

$$f(x) = f(a) + f'(c)(x - a) \tag{5.4}$$

where c lies somewhere between x and a.

This equation tells us that if the function $f(x)$ is approximated using the constant $f(a)$, then the error has the form $f'(c)(x - a)$. Geometrically this makes sense, because we should expect the error to depend on how far x is from a and on how large the rate of change of f is. Now in most cases we cannot find the value of c exactly. Indeed if we could, this would tell us the exact value of $f(x)$ and there would be no need to consider approximations at all.

Sometimes we can find estimates for f' , usually in a form which tells us that for some real number M , and for all x satisfying $a \leq x \leq b$,

$$|f'(x)| \leq M$$

We can then deduce that the approximation error satisfies

$$|f'(c)(x-a)| \leq M|x-a|$$

The right hand side is called an **error bound** for the constant approximation.

In order to study the error involved in using a linear approximation we need a version of the Mean Value Theorem applied to two functions simultaneously. We first note that if we apply to Mean Value Theorem to two functions f and g we obtain

$$\frac{f(b)-f(a)}{b-a} = f'(c_1); \ \frac{g(b)-g(a)}{b-a} = g'(c_2)$$

If we divide these two equations the denominator $b-a$ cancels and we obtain

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c_1)}{g'(c_2)}$$

There is no reason why c_1 and c_2 should be the same. However the next theorem shows that we can find a common value of c for the latter equation.

Theorem 5.2.2 (Generalised Mean Value Theorem)**.** *If the two functions $f(x)$ and $g(x)$ are both continuous for $a \leq x \leq b$ and differentiable for $a < x < b$, and if $g'(x)$ is non-zero for all x satisfying $a < x < b$ then there is a number c between a and b such that*

$$\frac{f(b)-f(a)}{g(b)-g(a)} = \frac{f'(c)}{g'(c)} \tag{5.5}$$

Proof

We first comment that the denominator $g(b)-g(a)$ cannot be zero, for otherwise the Mean Value Theorem would tell us that $g'(c) = 0$ for some c between a and b, contrary to the conditions of the theorem.

We now define the function $k(x)$ by means of the equation

$$k(x) = (f(b) - f(a))(g(x) - g(a)) - (g(b) - g(a))(f(x) - f(a))$$

It is easy to see that $k(b) = k(a) = 0$, and that the conditions of the theorem ensure that $k(x)$ satisfies the conditions needed to apply the Mean Value Theorem. Therefore

$$0 = \frac{k(b) - k(a)}{b - a} = k'(c)$$

for some c between a and b. Differentiating k gives

$$k'(x) = (f(b) - f(a))g'(x) - (g(b) - g(a))f'(x)$$

and so we conclude that for some number c between a and b,

$$k'(x) = (f(b) - f(a))g'(c) - (g(b) - g(a))f'(c)$$

which can be rearranged in the form given in the statement of the theorem. □

Now recall that the linear approximation we obtained in Section 5.1 had the equation $y = f(a) + f'(a)(x - a)$. So the error will be given by

$$E(x) = f(x) - (f(a) + f'(a)(x - a)) \tag{5.6}$$

We now apply the Generalised Mean Value Theorem with $f(x)$ replaced by $E(x)$ and with $g(x)$ replaced by $(x-a)^2$. Noting that $E(a) = 0$ and $g(a) = 0$ we deduce that for some number c between a and b

$$\frac{E(b)}{(b-a)^2} = \frac{E'(c)}{2(c-a)} = \frac{f'(c) - f'(a)}{2(c-a)} = \frac{f''(d)}{2}$$

where d is number between a and c, obtained by applying the Mean Value Theorem to $f'(x)$ over the interval $a \leq x \leq c$.

We can summarise these results by saying that for some number d between a and x we have

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(d)}{2}(x-a)^2 \tag{5.7}$$

So the error term $\frac{f''(d)}{2}(x-a)^2$ depends on the second derivative, which tells us how fast the gradient of the graph of $f(x)$ is changing. If this is very large then the graph will quickly diverge from that of its linear approximation.

Sometimes we can find estimates for f'' which tell us that for some real number L, and for all x satisfying $a \leq x \leq b$,

$$|f''(x)| \leq L$$

We can then deduce that the approximation error satisfies

$$\left|\frac{f''(d)}{2}(x-a)^2\right| \leq \frac{L}{2}|x-a|^2$$

Now the error bound for the constant approximation involved $|x-a|$, whereas that for the linear involves $|x-a|^2$. So if for example $|x-a| \leq 10^{-3}$ then $|x-a|^2 \leq 10^{-6}$, and so it is possible that the linear approximation could have an accuracy of twice as many decimal places as the constant approximation (depending on the relative sizes of f' and f'').

Example 5.2.2. *Find the equation of the tangent line to the curve whose equation is $y = \sqrt[3]{x}$ at the point $(8, 2)$, and use it to find an approximate value for $\sqrt[3]{8.2}$. Find an error bound and hence discuss the accuracy of the approximation.*

To find the tangent and the error term we need the first two derivatives of $f(x) = \sqrt[3]{x}$. So we have

$$\begin{aligned} f(x) &= \sqrt[3]{x} = x^{\frac{1}{3}} \\ f'(x) &= \frac{1}{3}x^{-\frac{2}{3}} = \frac{1}{3(\sqrt[3]{x})^2} \\ f''(x) &= \frac{1}{3}\left(-\frac{2}{3}\right)x^{-\frac{5}{3}} = -\frac{2}{9(\sqrt[3]{x})^5} \end{aligned}$$

In this example we have $a = 8$, so the equation of the tangent line at $(8, 2)$ is

$$y = 2 + \frac{1}{12}(x - 8)$$

An approximation for $\sqrt[3]{8.2}$ is therefore

$$2 + \frac{1}{12}(0.2) = 2 + \frac{1}{60} = 2.01666...$$

The error term is $\frac{f''(d)}{2}(0.2)^2$, where $8 < d < 8.2$

Now f'' is a decreasing function of x for $x > 0$, so we have

$$|f''(d)| \leq |f''(8)| = \frac{2}{9(\sqrt[3]{8})^5} = \frac{2}{9 \times 32}$$

So an error bound will be

$$\frac{2}{9 \times 32} \times \frac{1}{2} \times (0.2)^2 = \frac{1}{9 \times 32 \times 25} = \frac{1}{7200} = 0.00013888...$$

Now f'' is negative for $x > 0$, so the error is negative. Therefore

$$2.01666... - 0.0001388.. \leq \sqrt[3]{8.2} \leq 2.01666...$$

which tells us that $2.01652 < \sqrt[3]{8.2} < 2.01667$. Therefore we can say that $\sqrt[3]{8.2} \approx 2.017$ with an error of less than 1 in the third decimal place.

Example 5.2.3. *Use the Generalised Mean Value Theorem to give a proof of l'Hôpital's Rule.*

Suppose that $f(x)$ and $g(x)$ satisfy the conditions of l'Hôpital's rule, and that $\dfrac{f'(x)}{g'(x)} \to l$ as $x \to a$. We apply the Generalised Mean Value Theorem, noting that $f(a) = g(a) = 0$. Therefore

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(x)} \to l \text{ as } x \to a$$

because $a < c < x$ and so $c \to a$ as $x \to a$. This completes the proof of l'Hôpital's Rule.

5.3 Taylor Polynomials

This family of polynomial approximations was investigated by the English mathematician Brook Taylor (1685–1731) and by the Scottish mathematicians Colin Maclaurin (1698–1756) and James Gregory (1638–1675).

We shall begin by investigating an expansion for a function in the form of an infinite series of powers which we shall then use to generate approximations near to $x = a$. Generalising the form of linear and quadratic approximations we assume a series of the form

$$f(x) = a_0 + a_1(x-a) + a_2(x-a)^2 + a_3(x-a)^3 +,,,$$

If we substitute $x = a$ then all the terms are zero except the first, giving $f(a) = a_0$, which tells us what the constant term must be in terms of f . This is the aim for all the coefficients in the series, to find them in terms of f. We shall assume that it is valid to differentiate this infinite series term-by-term, to obtain

$$f'(x) = a_1 + 2a_2(x-a) + 3a_3(x-a)^2 + 4a_4(x-a)^3 + ...$$

Substituting $x = a$ then gives $f'(a) = a_1$. Differentiating again gives

$$f''(x) = 2a_2 + 3 \cdot 2a_3(x-a) + 4 \cdot 3a_4(x-a)^2 + ...$$

and now substituting $x = a$ gives $f''(a) = 2a_2$

If we repeat this process we find out that $f'''(a) = 3 \cdot 2a_3$, $f^{(4)}(a) = 4 \cdot 3 \cdot 2a_4$, and so on. This generalises to suggest that $f^{(n)}(a) = n!a_n$. So we have formulae for the coefficients in terms of f and its derivatives.

Definition 5.3.1. *We define the* ***Taylor series expansion of*** $f(x)$ ***about*** $x = a$ *to be the series*

$$f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n + ... \tag{5.8}$$

When we truncate this series we obtain the **Taylor polynomial of $f(x)$ about $x = a$ of degree n**, denoted by

$$P_{n,a}(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{5.9}$$

The special case $a = 0$ is called the **Maclaurin expansion**, which is therefore a series involving just powers of x,

$$f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + ... + \frac{f^{(n)}(0)}{n!}x^n + ... \tag{5.10}$$

In finding Taylor series we shall assume some results from Real Analysis:

1. The Taylor series usually converges to the function $f(x)$ that we start with, but there are exceptions.
2. We can perform algebraic operations on series to get new series, such as addition, multiplication, substitution, differentiation and integration.
3. If a function is equal to some series expansion in powers of x (or $x - a$) then that must be the Maclaurin (Taylor) series, i.e., the series is unique.
4. There are precise conditions for the validity of these results, discussed in this lecture note, Chapters 1, 2 . The examples we shall consider have been chosen to satisfy those conditions.
5. Certain basic expansions for elementary functions form part of school math- ematics and should be memorised.

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ... + \frac{x^n}{n!} + ... \tag{5.11}$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + ... + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + ... \tag{5.12}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + ... + (-1)^n \frac{x^{2n}}{2n!} + ... \tag{5.13}$$

Example 5.3.1. *Find the Maclaurin expansion of* $\dfrac{1}{1-x}$.

We can recognise $\dfrac{1}{1-x}$ as the sum of an infinite Geometric Series, or we can expand $(1-x)^{-1}$ by the Binomial Theorem. In either case we obtain

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots + x^n + \ldots$$

valid for $|x| < 1$. Property 3 in the list above tells us that this must be the Maclaurin series.

Example 5.3.2. *Find the Taylor expansion for* $\cos x$ *about* $x = \dfrac{\pi}{2}$.

We shall show two methods. In the first we calculate the successive derivatives, and in the second we use a known expansion.

Method 1

The following table shows the successive derivatives, and in the right-hand column their values when $x = \dfrac{\pi}{2}$.

$f(x)$	$\cos x$	0
$f'(x)$	$-\sin x$	-1
$f''(x)$	$-\cos x$	0
$f'''(x)$	$\sin x$	1
$f^{(4)}(x)$	$\cos x$	0
$f^{(5)}(x)$	$-\sin x$	-1
$f^{(6)}(x)$	$-\cos x$	0

So we can write down the Taylor series as follows.

$$\begin{aligned} f\left(\frac{\pi}{2}\right) + f'\left(\frac{\pi}{2}\right)\left(x-\frac{\pi}{2}\right) + \frac{f''\left(\frac{\pi}{2}\right)}{2!}\left(x-\frac{\pi}{2}\right)^2 + \frac{f'''\left(\frac{\pi}{2}\right)}{3!}\left(x-\frac{\pi}{2}\right)^3 \\ + \frac{f^{(4)}\left(\frac{\pi}{2}\right)}{4!}\left(x-\frac{\pi}{2}\right)^4 + \frac{f^{(5)}\left(\frac{\pi}{2}\right)}{5!}\left(x-\frac{\pi}{2}\right)^5 + \ldots \\ = -\left(x-\frac{\pi}{2}\right) + \frac{1}{3!}\left(x-\frac{\pi}{2}\right)^3 - \frac{1}{5!}\left(x-\frac{\pi}{2}\right)^5 \end{aligned}$$

Method 2

The known expansions we have are Maclaurin expansions, about $x = 0$, so we need to perform a translation by putting $y = x - \frac{\pi}{2}$. In this way when $x = \frac{\pi}{2}$ we shall have $y = 0$.

Using this transformation $\cos x$ becomes $\cos\left(y + \frac{\pi}{2}\right) = -\sin y$, using a trigonometric identity. The expansion of $-\sin y$ about $y = 0$ is

$$-y + \frac{y^3}{3!} - \frac{y^5}{5!} + ...$$

Therefore the expansion for $\cos x$ about $x = \frac{\pi}{2}$ is

$$-\left(x - \frac{\pi}{2}\right) + \frac{1}{3!}\left(x - \frac{\pi}{2}\right)^3 - \frac{1}{5!}\left(x - \frac{\pi}{2}\right)^5 + ...$$

5.4 Taylor's Theorem

Theorem 5.4.1 (Taylor's Theorem). *If the* $(n+1)$*-th derivative of* f *exists throughout an interval containing* a *and* x*, and if* $P_{n,a}(x)$ *denotes the Taylor polynomial of* $f(x)$ *of degree* n *about the point* a*, then we have*

$$f(x) = P_{n,a}(x) + E_n(x) \tag{5.14}$$

where the error (or remainder) term $E_n(x)$ *is given by*

$$E_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \tag{5.15}$$

where c *is some number between* a *and* x*. This is known as Lagrange's form of the remainder, named after the French mathematician Joseph-Louis Lagrange (1736–1813).*

We shall not give a proof of Taylor's Theorem in this book. Proofs can be found in Real Analysis textbooks.

As we remarked in Section 5.4, we can use the Taylor expansion to obtain polynomial approximations for a function by truncating the series after a finite number of terms. As with the constant and linear approximations, we need to be able to find error bounds, and the error term in Taylor's Theorem enables us to do this.

Example 5.4.1. *Use Taylor's Theorem to calculate an approximate value for* $\ln 1.05$*, accurate to six places of decimals.*

We consider the expansion of $\ln(1+x)$ about $x = 0$,

$$\ln(1+x) = x - \frac{x^2}{2!} + \frac{x^3}{3!} + ... + (-1)^{n+1}\frac{x^n}{n!} + ...$$

However to investigate the error we need to find an expression for the $(n+1)$-th derivative. In the table below we have calculated the first few derivatives, to the extent that we can see a pattern and write down an expression for the $(n+1)$-th derivative. Strictly speaking we should give a proof of our formula, by mathematical induction, but we shall omit that here. Beginning with $f(x)$ and repeatedly differentiating gives

$$\begin{aligned}
f(x) &= \ln x \\
f'(x) &= (1+x)^{-1} \\
f''(x) &= -(1+x)^{-2} \\
f'''(x) &= 2(1+x)^{-3} \\
f^{(4)}(x) &= -3 \times 2(1+x)^{-4} \\
f^{(5)}(x) &= 4 \times 3 \times 2(1+x)^{-5} \\
&\vdots \\
f^{(n)}(x) &= (-1)^{n-1}(n-1)!(1+x)^{-n} \\
f^{(n+1)}(x) &= (-1)^{n}n!(1+x)^{-(n+1)}
\end{aligned}$$

The error term is given by

$$E_n(x) = \frac{(-1)^n n!}{(n+1)!}x^{n+1}(1+c)^{-(n+1)} = \frac{(-1)^n}{(n+1)}x^{n+1}(1+c)^{-(n+1)}$$

To find the approximation for $\ln 1.05$ we put $x = 0.05$. Therefore $0 < c < 0.05$ and so we know that $(1+c) > 1$, telling us that $\dfrac{1}{1+c} < 1$, and hence

$$|E_n(0.05)| \leq \frac{(0.05)^{n+1}}{n+1}$$

To achieve six decimal places of accuracy we need this error bound to be sufficiently small, and using a calculator gives error bounds for $n = 1, 2, 3, 4$ as follows.

n	Error bound
1	1.25×10^{-3}
2	4.167×10^{-5}
3	1.5625×10^{-6}
4	6.26×10^{-8}

So $n = 3$ will not give sufficient accuracy, but $n = 4$ will. This means that we should use the polynomial approximation

$$\ln(1+x) \approx x - \frac{x^2}{2!} + \frac{x^3}{3!} - \frac{x^4}{4!}$$
$$\ln(1.05) \approx 0.05 - \frac{0.05^2}{2!} + \frac{0.05^3}{3!} - \frac{0.05^4}{4!}$$

Working this out on a calculator gives a readout of 0.048790104 and the calculator manual suggests that no more than the last digit would be suspect.

Now with $n = 4$ the error term is positive, and so we can say that

$$0.04879010 < \ln(1.05) < 0.04879011 + 6.25 \times 10^{-8} < 0.04879018$$

So we can be certain that $\ln(1.05) \approx 0.048790$ to six decimal places of accuracy.

We can see in this example that not only do we have to take the error bound into account, but also any possible errors arising from the use of a calculator or a computer.

Chapter 6

Partial Derivative

6.1 Functions of two variables

The temperature T at a point on the surface of the earth at any given time depends on the longitude and latitude of the point. We can think of T as being a function of the two variables x and y or as a function of the pair (x, y). We indicate this functional dependence by writing $T = f(x, y)$.

The volume of a circular cylinder depends on its radius and its height. In fact, we know that $V = \pi r^2 h$. We say that is a function of r and h, and we write $V(r, h) = \pi r^2 h$.

Definition 6.1.1. *A function of two variables is a rule that assigns to each ordered pair of real numbers (x, y) in a set D a unique real number denoted by $f(x, y)$. The set D is the domain of f and its range is the set of values that f takes on i.e the value of $f(x, y)$.*

In other words, the set of values that we are allowed to plug in the function f is our domain and the output which is generated by the function is our range for the function f. Functions are like machines which generate output (Range) based on the allowed input(Domain) given. Lets see some examples of two variable functions and their domains and ranges.

Example 6.1.1. *Evaluate $f(3,2)$ and find the domain and range*

$$f(x,y)=\sqrt{x^2-y}$$

The constraint put by the square root is that the value inside cannot be negative. So, we require $x^2 - y \geq 0$ and thus $x^2 \geq y$ which is the domain of the function. The range of the function is all values which are greater than or equal to zero i.e $0 \leq range < \infty$.

6.2 The partial diffentiation

Definition 6.2.1. *Let $f(x, y)$ be a function of two variables. The partial derivative $\partial f/\partial x$ is the function obtained by differentiating f with respect to x, regarding y as a constant. Similarly, $\partial f/\partial y$ is obtained by differentiating f with respect to y, regarding x as a constant.*

We often use the alternative notation

$$f_x = \partial f/\partial x, f_y = \partial f/\partial y$$

If $f(x, y) = x^2 + xy + y^3 - 1$ then $f_x = 2x + y$, $f_y = x + 3y^2$

As a limit,

$$\partial f/\partial x = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h}$$

Also, the equation $z = f(x, y)$ defines a surface in 3-dimensional space with x, y, z-axes, and $\partial f/\partial x$ is the gradient of the tangent at a point in the x-direction.

Higher partial derivatives Can differentiate f_x, f_y partially with respect to x and y to get four second order derivatives:

$$\begin{aligned} f_{xx} &= \partial^2 f/\partial x^2, \\ f_{yy} &= \partial^2 f/\partial y^2, \\ f_{xy} &= \partial^2 f/\partial x\partial y, \\ f_{yx} &= \partial^2 f/\partial y\partial x, \end{aligned}$$

Example 6.2.1. *$f(x,y) = x^2 + xy + y^3 - 1$ as above: then*

$$f_{xx} = 2, f_{yy} = 6y, f_{xy} = f_{yx} = 1$$

Example 6.2.2. *$f(x,y) = \tan^{-1}(y/x)$ then we get*

$$\begin{aligned} &f_{xx} = 2xy/(x^2+y^2)^2, f_{yy} = -2xy/(x^2+y^2)^2, \\ &f_{xy} = f_{yx} = (y^2-x^2)/(x^2+y^2)^2. \end{aligned}$$

Notice that in both examples, $f_{xy} = f_{yx}$. This is a general fact:

Theorem 6.2.1. *If $f(x,y)$ is a function of two variables, and the second order partial derivatives $f_{xy} and f_{yx}$ both exist and are continuous, then $f_{xy} = f_{yx}$.*

Here, saying that a function $g(x,y)$ is continuous at a point (a,b) means that $\lim_{(x,y)\to(a,b)} g(x,y) = g(a,b)$ - this means that as (x,y) gets closer and closer to (a,b), $g(x,y)$ gets closer and closer to $g(a,b)$.

All the functions we meet in this chapter will satisfy the assumptions of the theorem, so from we on we always assume that $f_{xy} = f_{yx}$.

More than 2 variables For a function $f(x,y,z,...)$ of three or more variables, we define the partial derivative $f_x = \partial f/\partial x$ to be the function obtained by differentiating f with respect to x, regarding $y, z, ...$ as constants. Similarly for $\partial f/\partial y$ and $\partial f \partial z$ and so on.

Example 6.2.3. *If* $f(x,y,z) = (x^2+y^2+z^2)^{1/2}$ *then*

$$f_x = x(x^2+y^2+z^2)^{-1/2}, f_y = y(x^2+y^2+z^2)^{-1/2}, f_z = z(x^2+y^2+z^2)^{-1/2}.$$

6.3 Chain rule

Recall the chain rule for differentiating functions of one variable: if $f = f(u)$ and $u = u(t)$ then $df/dt = df/du.du/dt$. Now let $f = f(x,y)$ where x, y are both functions of one variable t - say

$$x = x(t), y = y(t).$$

If a small change δt in t gives corresponding changes $\delta x, \delta y$ in x and y, then $\delta x \approx dx/dt.\delta t, \delta y \approx dy/dt.\delta t$, so by Section 1.2,

$$\delta f \approx f_x.\delta x + f_y.\delta y \approx f_x dx/dt.\delta t + f_y dy/dt.\delta t.$$

Dividing through by δt and taking the limit as $\delta \to 0$, we get

Chain Rule I If $f = f(x,y)$ where $x = x(t), y = y(t)$, then

$$df/dt = f_x dx/dt + f_y dy/dt.$$

Similarly for functions $f(x,y,z)$ with $x = x(t), y = y(t), z = z(t)$ we get

$$df/dt = f_x dx/dt + f_y dy/dt + f_z dz/dt.$$

Example 6.3.1. *If* $f(x, y) = x^3y + \sin(x + y)$ *where* $x = t^2, y = \sin t$*, then*

$$df/dt = (3x^2y + \cos(x + y)).2t + (x^3 + \cos(x + y)). \cos t.$$

Could of course work this out by substituting for x, y to get f as a function of t, and then differentiating, but would be somewhat unpleasant.

Implicit functions An equation of the form

$$f(x, y) = 0$$

defines y as an implicit function of x. To find dy/dx, we differentiate the equation with respect to x using the Chain Rule I (noting that x and y are both functions of the single variable x). This gives $f_x + f_y dy/dx = 0$, hence

$$dy/dx = -fx/fy.$$

For example, if the equation is $x^2 - y \cos x + x^3y^2 = 0$, then differentiating as above gives $2x + y \sin x + 3x^2y^2 + (-\cos x + 2x^3y)dy/dx = 0$, hence

$$dy/dx = (2x + y \sin x + 3x^2y^2)/(\cos x - 2x^3y).$$

General chain rule Now suppose $f = f(x, y)$ where $x = x(s, t), y = y(s, t)$. Then f is also a function of s, t and we'd like a formula for $\delta f/\delta s$, $\delta f/\delta t$. Well, if we regard t as a constant then $x = x(s), y = y(s)$, so we can apply Chain Rule I to get

Chain Rule II : If $f = f(x, y)$ where $x = x(s, t), y = y(s, t)$, then

$$\partial f/\partial s = f_x \partial x/\partial s + f_y \partial y/\partial s,$$

and similarly

$$\partial f/\partial t = f_x \partial x/\partial t + f_y \partial y/\partial t,$$

For more than 2 variables the rule is entirely similar: if $f = f(x, y, z, ...)$ where $x = x(s, t, ...), y = y(s, t, ...), z = z(s, t, ...), ...$ then

$$\delta f/\delta s = f_x \delta x/\delta s + f_y \partial y/\partial s + f_z \partial z/\partial s + ...$$

and similarly for $\partial f/\partial t$ and so on.

Example 6.3.2. *Let $f = f(x, y)$ and let r, θ be polar coordinates, so that $x = r\cos\theta, y = r\sin\theta$. Express the Laplace equation*

$$\partial^2 f/\partial x^2 + \partial^2 f/\partial y^2 = 0$$

in polar coordinate form.

Step 1 First work out f_x, f_y in terms of r, θ using the Chain Rule. Well,

$$r = (x^2 + y^2)^{1/2}, \delta = \tan^{-1}(y/x),$$

so by the Chain Rule,

$$f_x = f_r \partial r/\partial x + f\theta\partial\theta/\partial x = f_r.x(x^2 + y^2)^{-1/2} - f_\theta.y(x^2 + y^2)^{-1},$$

and so we get

$$f_x = f_r \cos\theta - f_\theta(\sin\theta)/r.$$

Similarly

$$f_y = f_r \sin\theta - f_\theta(\cos\theta)/r.$$

Step 2 Now we work out f_{xx} in terms of r, θ. By Step 1,

$$\begin{aligned}
f_{xx} &= \cos\theta.\partial/\partial r(f_x) - ((\sin\theta)/r).\partial/\partial\theta(f_x) \\
&= \cos\theta.\partial f/\partial t/\partial r(f_r\cos\theta - f_\theta(\sin\theta)/r) - \\
&\quad ((\sin\theta)/r).\partial/\partial\theta(f_r\cos\theta - f_\theta(\ sin\theta)/r) \\
&= \cos\theta[\cos\theta f_{rr} + ((\sin\theta)/r^2)f\theta - ((\sin\theta)/r)f_{\theta r}] - \\
&\quad ((\sin\theta)/r)[\cos\theta f_{r\theta} - \sin\theta f_r - ((\cos\theta)/r)f_\theta - ((\sin\theta)/r)f_{\theta\theta}].
\end{aligned}$$

Hence we get

$$\begin{aligned}
f_{xx} = \cos^2\theta f_{rr} - &((2\sin\theta\cos\theta)/r)f_{\theta r} + ((\sin^2\theta)/r)f_r \\
&+ ((2\sin\theta\cos\theta)/r^2)f_\theta + ((\sin^2\theta)/r^2)f_{\theta\theta}.
\end{aligned}$$

Similarly we get

$$\begin{aligned}
f_{xx} = \sin^2\theta f_{rr} - &((2\sin\theta\cos\theta)/r)f_{\theta r} + ((\cos^2\theta)/r)f_r \\
&+ ((2\sin\theta\cos\theta)/r^2)f_\theta + ((\cos^2\theta)/r^2)f_{\theta\theta}.
\end{aligned}$$

Adding these two expressions, we obtain

$$f_{xx} + f_{yy} = f_{rr} + (1/r)f_r + (1/r^2)f_{\theta\theta}.$$

Therefore Laplace's equation in polar form is

$$f_{rr} + (1/r)f_r + (1/r^2)f_{\theta\theta}. = 0$$

6.4 Taylor expansion of Multivariable function

Recall in chapter 5 where we discussed Taylor/Maclaurin series. There you learned about how to approximate many different functions, where the approximations are polynomial functions. But, the prescription that we showed is only valid for single valued function. In this section, we shall see how such approximations look when we try it out for multivalued functions.

$$f(a+h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + ... + \frac{f_{(n)}(a)}{n!}h^n + R_n(h),$$

where $R_n(h)$ is the "error term".

Now we'll get a similar expansion for a function $f(x, y)$ of two variables at a point (a, b). We study f along the line joining (a, b) to $(a + h, b + k)$ for small h, k. A point on this line is $(a + th, b + tk)$, with t between 0 and 1. Define a function $F(t)$ of one variable by

$$F(t) = f(a + th, b + tk)(0 \leq t \leq 1).$$

So $F(0) = f(a, b)$ and $F(1) = f(a + th, b + tk)$. We are interested in the Maclaurin series for F, which is

$$F(t) = f(0) + f'(0)t + \frac{f''(0)}{2!}t^2 + ... \tag{6.1}$$

Let's work out $F'(0)$ and $F''(0)$ in terms of f and its partial derivatives. By Chain Rule I, $F'(t) = f_x\frac{dx}{dt} + f_y\frac{dy}{dt}$ where $x = a+th, y = b+tk$, and hence

$$F'(t) = hf_x + kf_y$$

(evaluated at $(a + th, b + tk)$). Applying the Chain Rule again,

$$\begin{aligned} F''(t) &= h\frac{d}{dt}(f_x(a + th, b + tk)) + k\frac{d}{dt}(fy(a + th, b + tk)) \\ &= h(hf_{xx} + kf_{xy}) + k(hf_{yx} + kf_{yy}) \\ &= h^2 f_{xx} + 2hkf_{xy} + k2f_{yy}. \end{aligned}$$

Hence

$$\begin{aligned} F(0) &= f(a,b), \\ F'(0) &= hf_x(a,b) + kf_y(a,b), \\ F''(0) &= h^2 f_{xx}(a,b) + 2hkf_{xy}(a,b) + k^2 f_{yy}(a,b). \end{aligned}$$

Substituting into (1), we get

Taylor expansion of $f(x,y)$ at (a,b): This is

$$\begin{aligned} f(a+h,b+k) = f(a,b) + hf_x(a,b) + kf_y(a,b) + \frac{1}{2}(h^2 f_{xx}(a,b) \\ + 2hkf_{xy}(a,b) + k^2 f_{yy}(a,b)) \\ + \text{terms of degree 3 or more in } h,k. \end{aligned}$$

For small h,k this gives an approximation to $f(a+h,b+k)$.

6.5 Maxima, Minima and Saddle point

Now we use the previous section to study maxima and minima of functions of 2 variables. For a function $f(x, y)$, we say f has a maximum at a point (a, b) if $f(a+h, b+k) < f(a, b)$ for all small values of h, k (not both 0). Similarly, f has a minimum at (a, b) if $f(a+h, b+k) < f(a, b)$ for all small h, k.

Finding maxima and minima Spose f has a maximum at (a, b). Then if we fix $y = b$ and vary x, we certainly have $f(a + h, b) < f(a, b)$ for all small h, which says that the function $f(x, b)$ of 1 variable (x) has a maximum at $x = a$. Hence $\partial f/\partial x = 0$ at (a, b), and likewise $\partial f/\partial y = 0$ at (a, b). This is also true at a minimum. Hence maxima and minima are stationary points of f, in the following sense:

Definition 6.5.1. *We say (a, b) is a stationary point of f if $f_x(a, b) = f_y(a, b) = 0$. A stationary point may or may not be a max/min. For example, if $f(x, y) = xy$ then $(0, 0)$ is a stationary point, but it is not a max or a min, as we can make $f(h, k)$ positive or negative for suitable choices of small h, k.*

We call a stationary point (a, b) a saddle point of f if it is not a max or a min.

Example 6.5.1. *Let $f(x, y) = x^2/2 - x + xy^2$.*

Then $f_x = x - 1 + y^2$, $f_y = 2xy$. At a stationary point, $2xy = 0$ so $x = 0$ or $y = 0$. If $x = 0$ then $-1 + y^2 = 0$ so $y = \pm 1$. If $y = 0$ then $x - 1 = 0$ so $x = 1$. So f has 3 stationary points:

$$(0, 1), (0, -1), (1, 0).$$

Suppose (a, b) is a stationary point of f, so $f_x(a, b) = f_y(a, b) = 0$. Write

$$A = f_{xx}(a, b), B = f_{xy}(a, b), C = f_{yy}(a, b).$$

Then the Taylor expansion of f at (a, b) is

$$f(a + h, b + k) = f(a, b) + \tfrac{1}{2}(Ah^2 + 2Bhk + Ck^2) + \text{ higher terms.}$$

Write

$$\Delta = Ah^2 + 2Bhk + Ck^2$$

Then we see that (a, b) is a max if $Delta < 0$ for all small h, k; (a, b) is a min if $\Delta > 0$ for all small h, k; and otherwise (a, b) is a saddle.

Suppose now that $A \neq 0$. Then

$$\Delta = Ah^2 + 2Bhk + Ck^2 = \frac{1}{A}[(Ah + Bk)^2 + (AC - B^2)k^2].$$

If $AC - B2 > 0$ and $A > 0$ then $\Delta > 0$ for all small h, k, and so (a, b) is a min. If $AC - B2 > 0$ and $A < 0$ then $\Delta < 0$ for all small h, k, and so (a, b) is a max. And if $AC - B2 < 0$ then Δ can be made positive or negative for suitably chosen small $h, k, so(a, b)$ is a saddle.

Now suppose $A = 0$ and $B \neq 0$. Then $\Delta = k(2Bh + Ck)$, which can be made positive or negative for small h, k, so again we have a saddle.

Summary Let (a, b) be a stationary point of $f(x, y)$, and let $A = f_{xx}(a, b), B = f_{xy}(a, b), C = f_y(a, b)$. The nature of the stationary point is as follows:

A	$AC - B^2$	nature
> 0	> 0	minimum
< 0	> 0	maximum
any	< 0	saddle

Chapter 7

Vector Calculus

7.1 Differentiation of Vectors

We are familiar with the term vector from high school physics. It is defined as a quantity having both magnitude and direction. Mathematically, In three dimensional Cartesian coordinate system, it is represented as,

$$\boldsymbol{A} = iA_x + jA_y + kA_z$$

with i, j and k representing unit vectors along x, y and z axis respectively. A_x, A_y and A_z are the components of $\boldsymbol{A}$. If $\boldsymbol{A}$ is a function of some variable, say t, the individual components would be a function of t as well. So, we can write like,

$$\boldsymbol{A}(t) = \mathbf{i}A_x(t) + \mathbf{j}A_y(t) + \mathbf{k}A_z(t)$$

Differentiation w.r.t t yields,

$$\frac{d\boldsymbol{A}}{dt} = \mathbf{i}\frac{dA_x}{dt} + \mathbf{j}\frac{dA_y}{dt} + \mathbf{k}\frac{dA_z}{dt}$$

The derivative $\frac{d\boldsymbol{A}}{dt}$ is also a vector whose components are the derivatives of the components of $\boldsymbol{A}$ itself.

Example 7.1.1.

Let (x, y, z) be the coordinates of a moving particle at time t. The displacement vector or position vector, $\boldsymbol{r}(t)$ is written as,

$$\boldsymbol{r}(t) = ir_x + jr_y + kr_z$$

The velocity vector $\boldsymbol{v}(t)$ is obtained by differentiating $\boldsymbol{r}(t)$.

$$\boldsymbol{v}(t) = \frac{d\boldsymbol{r}}{dt} = i\frac{dr_x}{dt} + j\frac{dr_y}{dt} + k\frac{dr_z}{dt}$$

Similarly, the acceleration vector is the derivative of $\boldsymbol{v}(t)$.

$$\boldsymbol{a}(t) = \frac{d\boldsymbol{v}}{dt} = i\frac{dv_x}{dt} + j\frac{dv_y}{dt} + k\frac{dv_z}{dt}$$

This technique can be applied for any combinations of vectors i.e product of a vector and a scalar or two vectors.

$$\begin{aligned}
\frac{d}{dt}(a\boldsymbol{A}) &= a\frac{d\boldsymbol{A}}{dt} + \frac{da}{dt}\boldsymbol{A} \\
\frac{d}{dt}(\boldsymbol{A}\cdot\boldsymbol{B}) &= \boldsymbol{A}\cdot\frac{d\boldsymbol{B}}{dt} + \frac{d\boldsymbol{A}}{dt}\cdot\boldsymbol{B} \qquad (7.1) \\
\frac{d}{dt}(\boldsymbol{A}\times\boldsymbol{B}) &= \boldsymbol{A}\times\frac{d\boldsymbol{B}}{dt} + \frac{d\boldsymbol{A}}{dt}\times\boldsymbol{B}
\end{aligned}$$

7.2 Fields and Directional Derivative

Many physical quantities have different values at different points in space. For example, the temperature in a room is different at different points, the electric field around a point charge is large near the charge and decreases as we go away from the charge. The term *field* refers to the physical quantity of interest and the region of space where it has a value. *Field* can be classified as *scalar field* or *vector field* depending on whether the physical quantity is a scalar or vector.

Let us consider temperature $T(x, y, z)$ of a room. If we start from a point in the room and move away from it, we will observe change in temperature. Also we will notice that it increases in some direction and decreases in other directions. As the rate of change of temperature depends on the direction we are moving, it is called *directional derivative.*

We will learn how to calculate *directional derivative* of a field. Let us now take a scalar field $\phi(x, y, z)$. We want to find $\dfrac{d\phi}{ds}$(The rate of change of ϕ with distance) at a point (x_0, y_0, z_0) in a given direction. Let $\mathbf{u} = \mathbf{i}a + \mathbf{j}b + \mathbf{k}c$ be a unit vector in given direction.

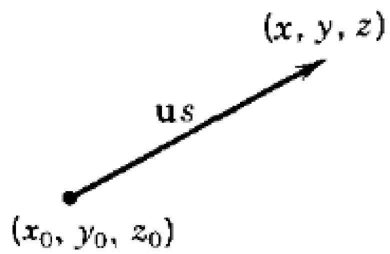

Figure 7.1: The directional derivative

We start at (x_0, y_0, z_0) and go a distance s to the point (x, y, z) in the direction of $\mathbf{u}$. The vector joining them is $\mathbf{u}s$ expressed as,

$$(x, y, z) - (x_0, y_0, z_0) = \mathbf{u}s = (a\mathbf{i} + b\mathbf{j} + c\mathbf{k})s$$

or,

$$
\begin{aligned}
x &= x_0 + as \\
y &= y_0 + bs \\
z &= z_0 + cs
\end{aligned}
\tag{7.2}
$$

Substituting for (x, y, z), $\phi(x, y, z)$ can be expressed as a function of s only. Applying chain rule of differentiation, we can find $\dfrac{d\phi}{ds}$,

$$
\begin{aligned}
\frac{d\phi}{ds} &= \frac{\partial\phi}{\partial x}\frac{dx}{ds} + \frac{\partial\phi}{\partial y}\frac{dy}{ds} + \frac{\partial\phi}{\partial z}\frac{dz}{ds} \\
&= \frac{\partial\phi}{\partial x}a + \frac{\partial\phi}{\partial y}b + \frac{\partial\phi}{\partial z}c
\end{aligned}
\tag{7.3}
$$

Differentiation of x, y, z w.r.t s is a, b, c respectively.

We can think the last line of equation (7.6) as a dot product of the vector $\mathbf{i}\left(\frac{\partial\phi}{\partial x}\right) + \mathbf{j}\left(\frac{\partial\phi}{\partial y}\right) + \mathbf{k}\left(\frac{\partial\phi}{\partial z}\right)$ and the unit vector $\mathbf{u}$.

$$
\frac{d\phi}{ds} = (\mathbf{i}\frac{\partial\phi}{\partial x} + \mathbf{i}\frac{\partial\phi}{\partial x} + \mathbf{i}\frac{\partial\phi}{\partial x}) \cdot \mathbf{u} \tag{7.4}
$$

The above mentioned vector is called *gradient* of ϕ and written as $\nabla\phi$. We will see more on this symbol ∇ in the next section.

With this operator, the directional derivative has more concise form,

$$
\frac{d\phi}{ds} = \nabla\phi \cdot \mathbf{u} \tag{7.5}
$$

Example 7.2.1. *Find the directional derivative of* $\phi = x^2y + xz$ *at (1,2,-1) in the direction of* $\mathbf{A} = 2\mathbf{i} - 2\mathbf{j} + \mathbf{k}$

Let us calculate $\mathbf{u}$ and $\nabla\phi$ and substitute it in equation (7.8).

$$
\begin{aligned}
\mathbf{u} &= \frac{\mathbf{A}}{|\mathbf{A}|} \\
&= \frac{1}{3}(2\mathbf{i} - 2\mathbf{j} + \mathbf{k})
\end{aligned}
$$

and

$$\begin{aligned}\nabla\phi &= \mathbf{i}\frac{\partial\phi}{\partial x} + \mathbf{j}\frac{\partial\phi}{\partial x} + \mathbf{k}\frac{\partial\phi}{\partial x}\\ &= (2xy+z)\mathbf{i} + x^2\mathbf{j} + x\mathbf{k}\end{aligned}$$

Evaluating $\nabla\phi$ at point $(1,2,-1)$,

$$\nabla\phi(1,2,-1) = 3\mathbf{i} + \mathbf{j} + \mathbf{k}$$

The directional derivative,

$$\frac{d\phi}{ds}(1,2,-1) = 2 - \frac{2}{3} + \frac{1}{3} = \frac{5}{3}$$

7.3 Expressions Involving ∇

We have already introduced the *vector operator* ∇ and $\nabla\phi$ in last section. We will see more expressions involving ∇ in this section.

When ∇ operates on a vector field $\mathbf{V}(x, y, z)$, we define it as *divergence* of $\mathbf{V}$ or div $\mathbf{V}$ or just $\nabla \cdot \mathbf{V}$.

$$\nabla \cdot \mathbf{V} = \frac{\partial V_x}{\partial x} + \frac{\partial V_x}{\partial x} + \frac{\partial V_x}{\partial x} \tag{7.6}$$

Another expression, *curl* of $\mathbf{V}$ is defined as,

$$\nabla \times \mathbf{V} = \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ V_x & V_y & V_z \end{vmatrix} \tag{7.7}$$

Laplacian is another very useful expresion of ∇.

$$\begin{aligned} \nabla^2\phi &= \nabla \cdot \nabla\phi \\ &= \frac{\partial}{\partial x}\frac{\partial\phi}{\partial x} + \frac{\partial}{\partial y}\frac{\partial\phi}{\partial y} + \frac{\partial}{\partial z}\frac{\partial\phi}{\partial z} \\ &= \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} \end{aligned} \tag{7.8}$$

Among these expressions, the *gradient* and *Laplacian* involve scalar fields and the *divergence* and *curl* involve vector fields.

7.4 Application of the Divergence Operator

We have defined the divergence of a vector function $V(x, y, z)$ as

$$div\, \mathbf{V}V = \nabla \cdot \mathbf{V} = \frac{\partial V_x}{\partial x} + \frac{\partial V_x}{\partial x} + \frac{\partial V_x}{\partial x} \tag{7.9}$$

Consider a region in which water is flowing. We can imagine drawing at every point a vector v equal to the velocity of the water at that point.

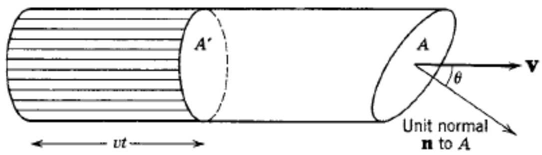

Figure 7.2: The direction of flow

For our example of water flow, let $V = v\rho$, where ρ is the density of the water. Then the amount of water crossing in time t an area A' which is perpendicular to the direction of flow, is (see Figure 7.2 the amount of water in a cylinder of cross-section A' and length vt. This amount of water is

$$(vt)(A')(\rho) \tag{7.10}$$

The same amount of water crosses area A (see Figure 7.2) whose normal is inclined at angle θ to v. Since $A' = A\cos\theta$

$$v\rho\cos\theta = V\cos\theta = V \cdot n \tag{7.11}$$

if n is a unit vector.

Now consider an element of volume $dxdydz$ in the region through which the water is flowing (Figure 7.3). Water is flowing either in or out of the volume $dxdydz$ through each of the six surfaces of the volume element; we shall calculate the net outward flow. In Figure 7.3, the rate at which water flows into $dxdydz$

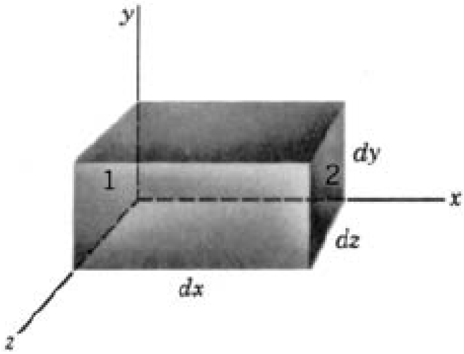

Figure 7.3: An element of $dxdydz$ in a region through which the water is flowing

through surface 1 is [by 7.3] $V \cdot i$ per unit area, or $(V \cdot i)\ dydz$ through the area $dydz$ of surface 1. Since $V \cdot i = V_x$, we find that the rate at which water flows across surface 1 is $V_x\ dydz$. A similar expression gives the rate at which water flows out through surface 2, except that V_x must be the x component of V at surface 2 instead of at surface 1. We want the difference of the two V_x values at two points, one on surface 1 and one on surface 2, directly opposite each other, that is, for the same y and z. These two values of V_x differ by ΔV_x which can be approximated by dV_x. For constant y and z, $dV_x = (\partial V_x/\partial x)dx$. Then the net outflow through these two surfaces is the outflow through surface 2 minus the inflow through surface 1, namely,

$$[(V_x \text{ at surface 2}) - (V_x \text{ at surface 1})]\ dydz = \left(\frac{\partial V_x}{\partial x}\partial x\ dx\right) dydz$$

We get similar expressions for the net outflow through the other two pairs of opposite surfaces:

$$\begin{array}{ll} \dfrac{\partial V_y}{\partial y}\ dxdydz & \quad \text{through top and bottom, and} \\ \dfrac{\partial V_z}{\partial z}\ dxdydz & \quad \text{through the other two sides.} \end{array}$$

Then the total net rate of loss of water from $dxdydz$ is

$$\left(\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z}\right) dxdydz = \nabla \cdot \mathbf{V}\ dxdydz \tag{7.12}$$

If we divide (7.12) by $dxdydz$, we have the rate of loss of water per unit volume. This is the physical meaning of a divergence: It is the net rate of outflow per unit volume evaluated at a point (let $dxdydz$ shrink to a point).

7.5 Combinations of *Gradient, Divergence* and *Curl*

We now consider the action of two vector operators in succession on a scalar or vector field. We can immediately discard four of the nine obvious combinations of *gradient, divergence* and *curl* since they clearly do not make sense. If ϕ is a scalar field and $\mathbf{a}$ is a vector field, these four combinations are $grad(grad\,\phi), div(div\,\mathbf{a}), curl(div\,\mathbf{a})$ and $grad(curl\,\mathbf{a})$.

In each case, the second(outer) vector operator is acting on the wrong type of field, i.e. scalar instead of vector or vice versa. In $grad(grad\,\phi)$, for example, $grad$ acts on $grad\,\phi$, which is a vector field, but we know that $grad$ only acts on scalar fields.

Of the five valid combinations, two are identically zero, namely

$$curl\;grad\,\phi = \nabla \times \nabla\phi = 0$$

$$div\;curl\,\mathbf{a} = \nabla \cdot (\nabla \times \mathbf{a}) = 0$$

The three remaining combinations of *gradient, divergence* and *curl* are

$$\begin{aligned}
div\;grad\,\phi &= \nabla \cdot \nabla\,\phi = \nabla^2\,\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} \\
grad\;div\,\mathbf{a} &= \nabla\,(\nabla \cdot \mathbf{a}) = \left(\frac{\partial^2 a_x}{\partial x^2} + \frac{\partial^2 a_y}{\partial x \partial y} + \frac{\partial^2 a_z}{\partial x \partial z}\right)\mathbf{i} + \\
&\qquad \left(\frac{\partial^2 a_x}{\partial y \partial x} + \frac{\partial^2 a_y}{\partial y^2} + \frac{\partial^2 a_z}{\partial y \partial z}\right)\mathbf{j} + \\
&\qquad \left(\frac{\partial^2 a_x}{\partial z \partial x}\frac{\partial^2 a_y}{\partial z \partial y} + \frac{\partial^2 a_z}{\partial^2 z}\right)\mathbf{k} \\
curl\;curl\,\mathbf{a} &= \nabla \times (\nabla \times \mathbf{a}) = \nabla(\nabla \cdot \mathbf{a}) - \nabla^2\,\mathbf{a}
\end{aligned}$$

Chapter 8

Conics and Co-ordinate Geometry

We begin this chapter with equations for conic sections. Interestingly, these are the paths traveled by planets, satellites, and other bodies (even electrons) whose motions are driven by inverse square forces. Conic sections are best described with the help of polar coordinates, so we also investigate curves in this new coordinate system. Finally we introduce cylindrical and spherical coordinates which are useful for describing some three dimensional objects.

8.1 Conic Sections and Quadratic Equations

This section shows how the conic sections from Greek geometry are described today as the graphs of quadratic equations in the coordinate plane. The Greeks of Plato's time described these curves as the curves formed by cutting a double cone with a plane (Fig. 8.1); hence the name conic section.

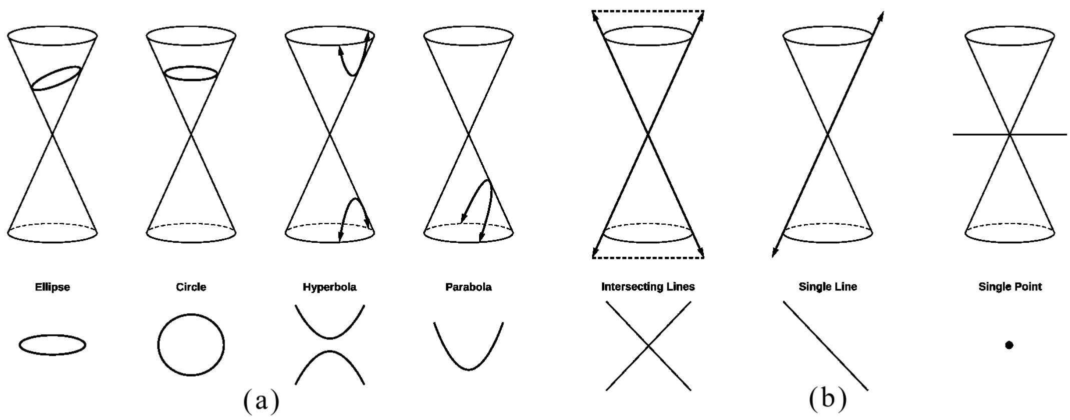

Figure 8.1: *(a) The standard conic sections, (b) The degenerate conic sections*

Circles

Definition 8.1.1. *A* ***circle*** *is a set of points in a plane whose distance from a given fixed point in the plane is constant. The fixed point is the* ***center*** *of the circle and the constant distance is the* ***radius****.*

The standard form equations for circles, derived from the distance formula, $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$, are these:

circle of radius a, centered at the origin:

$$x^2 + y^2 = a^2 \tag{8.1}$$

circle of radius a, centered at (h, k)

$$(x - h)^2 + (y - k)^2 = a^2 \tag{8.2}$$

Parabolas

Definition 8.1.2. *A set that consists of all the points in a plane, equidistant from a given fixed point and a given fixed line in the plane is a* ***parabola****. The fixed point is the* ***focus*** *and the fixed line is the* ***directrix****.*

If the focus F lies on the directrix L, the parabola is the line through F perpendicular to L. We consider this to be a degenerate case and assume henceforth that F does not lie on L.

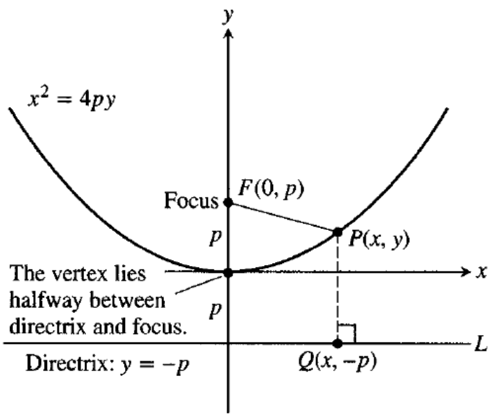

Figure 8.2: *The Parabola $x^2 = 4py$*

A parabola has its simplest equation when its focus and directrix straddle one of the coordinate axes. For example, suppose that the focus lies at the point $F(0, p)$ on the positive y-axis and that the directrix is the line $y = -p$ (Fig. 8.2). In the notation of the figure, a point $P(x, y)$ lies on the parabola if and only if $PF = PQ$. From the distance formula,

$$PF = \sqrt{(x-0)^2 + (y-p)^2} = \sqrt{x^2 + (y-p)^2}$$
$$PQ = \sqrt{(x-x)^2 + (y-(-p))^2} = \sqrt{(y+p)^2}$$

When we equate , square and simplify these expressions, we get,

$$y = \frac{x^2}{4p} \quad \text{or} \quad x^2 = 4py \qquad \text{Standard form} \tag{8.3}$$

These equations reveal the parabola's symmetry about the y-axis. We call the y-axis the **axis** of the parabola (short for "axis of symmetry").

The point where a parabola crosses its axis is the **vertex**. The vertex of the parabola $x^2 = 4py$ lies at the origin (Fig. 8.2). The positive number p is the parabola's **focal length**.

If the parabola opens downward, with its focus at $(0, -p)$ and its directrix the line $y = p$, then Eqs. (8.3) becomes

$$y = -\frac{x^2}{4p} \quad \text{and} \quad x^2 = -4py \tag{8.4}$$

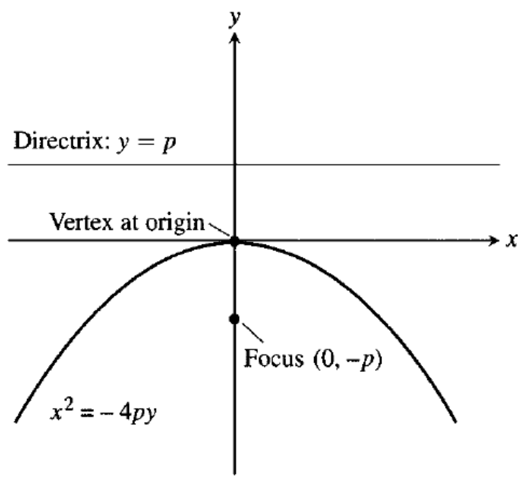

Figure 8.3: *The parabola* $x^2 = -4py$

(Fig 8.3). We obtain similar equations for parabolas opening to the right or to the left (Fig. 8.4, and Table 8.1).

Equation	Focus	Directrix	Axis	Opens
$x^2 = 4py$	$(0, p)$	$y = -p$	y-axis	Up
$x^2 = -4py$	$(0, -p)$	$y = p$	y-axis	Down
$y^2 = 4px$	$(p, 0)$	$x = -p$	x-axis	To the right
$y^2 = -4px$	$(-p, 0)$	$x = p$	x-axis	To the left

Table 8.1: *Standard-form equations for parabolas with vertices at the origin* $(p > 0)$

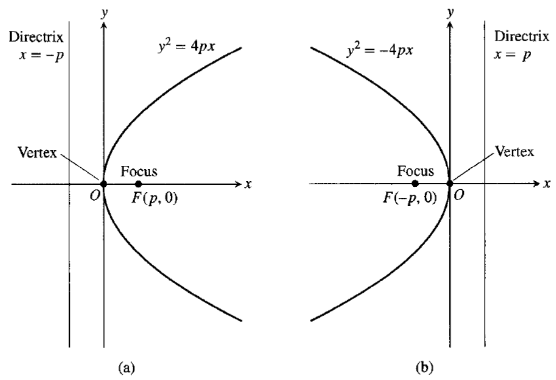

Figure 8.4: *(a) The parabola* $y^2 = 4px$*. (b) The parabola* $y^2 = -4px$

Example 8.1.1. *Find the focus and directrix of the parabola* $y^2 = 10x$*.*

Solution

We find the value of p in the standard equation $y^2 = 4px$:

$$4p = 10, \quad \text{so} \quad p = \frac{10}{4} = \frac{5}{2}$$

Then we find the focus and directrix for this value of p:

Focus: $\quad (p, 0) = \left(\frac{5}{2}, 0\right)$

Directrix: $\quad x = -p \quad \text{or} \quad x = -\frac{5}{2}$

Horizontal and vertical shift formulas can be applied to the equations in Table 8.1 to give equations for a variety of parabolas in other locations.

Ellipses

Definition 8.1.3. *An* ***ellipse*** *is the set of points in a plane whose distances from two fixed points in the plane have a constant sum. The two fixed points are the* ***foci*** *of the ellipse.*

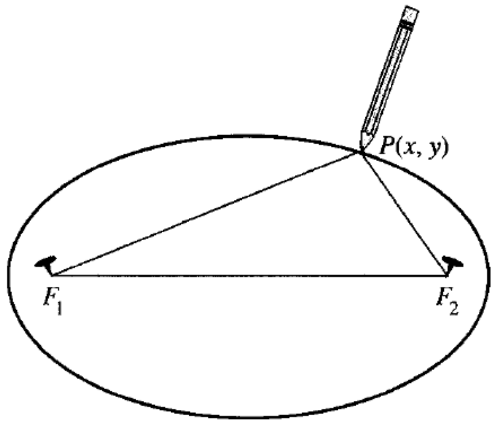

Figure 8.5: *How to draw an ellipse.*

The quickest way to construct an ellipse uses the definition. Put a loop of string around two tacks F_1 and F_2 , pull the string taut with a pencil point P, and move the pencil around to trace a closed curve (Fig. 8.5). The curve is an ellipse because the sum $PF_1 + PF_2$, being the length of the loop minus the distance between the tacks, remains constant. The ellipse's foci lie at F_1 and F_2 .

Definition 8.1.4. *The line through the foci of an ellipse is the ellipse's* ***focal axis****. The point on the axis halfway between the foci is the* ***center****. The points where the focal axis and the ellipse cross are the ellipse's* ***vertices*** *(Fig 8.6).*

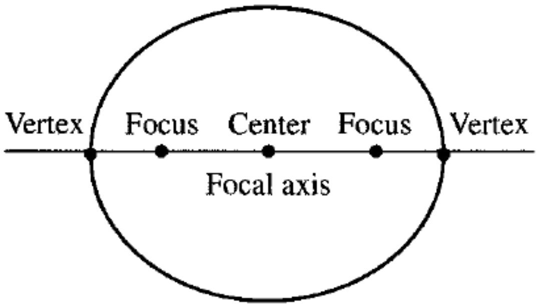

Figure 8.6: *Points on the focal axis of an ellipse.*

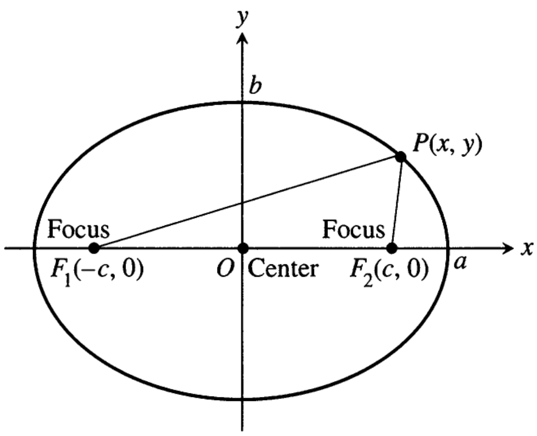

Figure 8.7: *The ellipse defined by the equation* $PF_1 + PF_2 = 2a$ *is the graph of the equation* $(x^2/a^2) + (y^2/b^2) = 1$.

If the foci are $F_1(-c, 0)$ and $F_2(c, 0)$ (Fig. 8.7), and $PF_1 + PF_2$ is denoted by $2a$, then the coordinates of a point P on the ellipse satisfy the equation

$$\sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a$$

To simplify this equation, we move the second radical to the right-hand side, square, isolate the remaining radical, and square again, obtaining

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1 \tag{8.5}$$

Since $PF_1 + PF_2$ is greater than the length F_1F_2 (triangle inequality for triangle PF_1F_2), the number $2a$ is greater than $2c$. Accordingly, $a > c$ and the number $a^2 - c^2$ in Eq. (8.5) is positive.

The algebraic steps leading to Eq. (8.5) can be reversed to show that every point P whose coordinates satisfy an equation of this form with $0 < c < a$ also satisfies the equation $PF_1 + PF_2 = 2a$. A point therefore lies on the ellipse if and only if its coordinates satisfy Eq. (8.5).

$$b = \sqrt{a^2 - c^2} \tag{8.6}$$

then $a^2 - c^2 = b^2$ and Eq (8.5) takes the form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{8.7}$$

Equation (8.7) reveals that this ellipse is symmetric with respect to the origin and both coordinate axes. It lies inside the rectangle bounded by the lines $x = \pm a$ and $y = \pm b$. It crosses the axes at the points $(\pm a, 0)$ and $(0, \pm b)$. The tangents at these points are perpendicular to the axes because

$$\frac{dy}{dx} = -\frac{b^2 x}{a^2 y} \qquad \text{Obtained from Eq (8.7) by implicit differentiation}$$

is zero if $x = 0$ and infinite if $y = 0$.

The Major and Minor Axes of an Ellipse

The **major axis** of the ellipse in Eq. (8.7) is the line segment of length $2a$ joining the points $(\pm a, 0)$. The **minor axis** is the line segment of length $2b$ joining the points $(0, \pm b)$. The number a itself is the **semimajor axis**, the number b the **semiminor axis**. The number c, found from Eq. (8.6) as

$$c = \sqrt{a^2 - b^2}$$

is the **center-to-focus distance** of the ellipse.

Example 8.1.2.

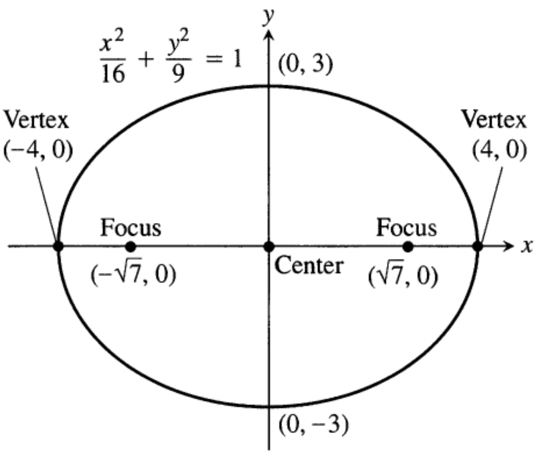

Figure 8.8: *Major axis horizontal (Example 7.2)*

The ellipse

$$\frac{x^2}{16} + \frac{y^2}{9} = 1$$

(8.8) has

Semimajor axis: $a = \sqrt{16} = 4$, Semiminor axis: $b = \sqrt{9} = 3$
Center-to-focus distance: $c = \sqrt{16 - 9} = \sqrt{7}$
Foci: $(\pm c, 0) = (\pm\sqrt{7}, 0)$
Vertices: $(\pm a, 0) = (\pm 4, 0)$
Center: $(0, 0)$

Standard-Form Equations for Ellipses Centered at the Origin

Foci on the x-axis: $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \ (a > b)$
Center-to-focus distance: $c = \sqrt{a^2 - b^2}$
Foci: $(\pm c, 0)$
Vertices: $(\pm a, 0)$
Foci on the y-axis: $\frac{x^2}{b^2} + \frac{y^2}{a^2} = 1 \ (a > b)$

Center-to-focus distance: $c = \sqrt{a^2 - b^2}$
Foci: $(0, \pm c)$
Vertices: $(0, \pm a)$

In each case, a is the semimajor axis and b is the semiminor axis.

Hyperbolas

Definition 8.1.5. *A* ***hyperbola*** *is the set of points in a plane whose distances from two fixed points in the plane have a constant difference. The two fixed points are the* ***foci*** *of the hyperbola.*

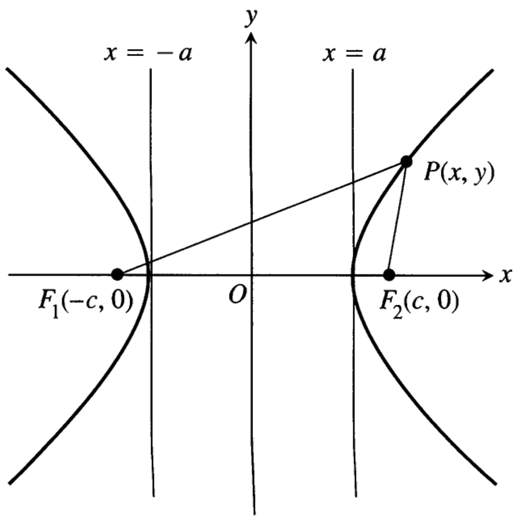

Figure 8.9: *Hyperbolas have two branches. For points on the right-hand branch of the hyperbola shown here, $PF_1 - PF_2 = 2a$. For points on the left-hand branch, $PF_2 - PF_1 = 2a$.*

If the foci are $F_1(-c, 0)$ and $F_2(c, 0)$ (Fig 8.9) and the common difference is $2a$, then a point (x, y) lies on the hyperbola if and only if

$$\sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} = \pm 2a \tag{8.8}$$

To simplify this equation, we move the second radical to the right-hand side, square, isolate the remaining radical, and square again, obtaining

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1 \tag{8.9}$$

So far, this looks just like the equation for an ellipse. But now $a^2 - c^2$ is negative because $2a$, being the difference of two sides of triangle PF_1F_2 , is less than $2c$, the third side.

The algebraic steps leading to Eq. (8.9) can be reversed to show that every point P whose coordinates satisfy an equation of this form with $0 < a < c$ also satisfies Eq. (8.8). A point therefore lies on the hyperbola if and only if its coordinates satisfy Eq. (8.9).

If we let b denote the positive square root of $c^2 - a^2$

$$b = \sqrt{c^2 - a^2} \tag{8.10}$$

then $a^2 - c^2 = -b^2$ and Eq (8.9) takes the more compact form

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{8.11}$$

The differences between Eq. (8.11) and the equation for an ellipse (Eq. 8.7) are the minus sign and the new relation

$$c^2 = a^2 + b^2 \qquad \text{From Eq. 8.10}$$

Like the ellipse, the hyperbola is symmetric with respect to the origin and coordinate axes. It crosses the x-axis at the points $(\pm a, 0)$. The tangents at these points are vertical because

$$\frac{dy}{dx} = \frac{b^2 x}{a^2 y} \qquad \text{Obtained from Eq. (8.11) by implicit differentiation}$$

is infinite when $y = O$. The hyperbola has no y-intercepts; in fact, no part of the curve lies between the lines $x = -a$ and $x = a$.

Definition 8.1.6. *The line through the foci of a hyperbola is the **focal axis**. The point on the axis halfway between the foci is the hyperbola's **center**. The points where the focal axis and the hyperbola cross are the **vertices** (Fig 8.10)*

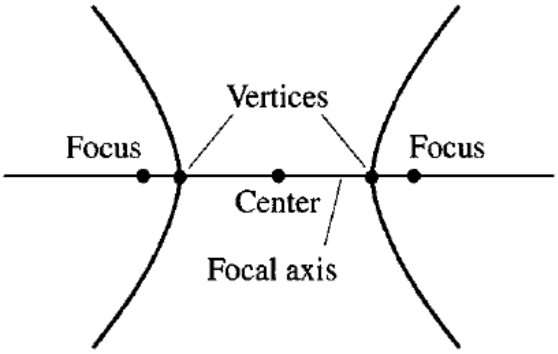

Figure 8.10: *Points on the focal axis of a hyperbola*

Asymptotes of Hyperbolas

The hyperbola

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

has two asymptotes, the lines

$$y = \pm\frac{b}{a}x$$

The asymptotes give us the guidance we need to graph hyperbolas quickly. The fastest way to find the equations of the asymptotes is to replace the 1 in the equation for the hyperbola by 0 and solve the new equation for y:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0 \Rightarrow y = \pm\frac{b}{a}x$$

Standard-Form Equations for Hyperbolas Centered at the Origin

Foci on the x-axis: $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$

Center-to-focus distance: $c = \sqrt{a^2 + b^2}$

Foci: $(\pm c, 0)$

Vertices: $(\pm a, 0$

Asymptotes: $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 0 \quad \text{or} \quad y = \pm \dfrac{b}{a} x$

Foci on the x-axis: $\dfrac{y^2}{a^2} - \dfrac{x^2}{b^2} = 1$

Center-to-focus distance: $c = \sqrt{a^2 + b^2}$

Foci: $(0, \pm c)$

Vertices: $(0, \pm a)$

Asymptotes: $\dfrac{y^2}{a^2} - \dfrac{x^2}{b^2} = 0 \quad \text{or} \quad y = \pm \dfrac{a}{b} x$

8.2 Classifying Conic Sections by Eccentricity

We now show how to associate with each conic section a number called the conic section's eccentricity. The eccentricity reveals the conic section's type (circle, ellipse, parabola, or hyperbola) and, in the case of ellipses and hyperbolas, describes the conic section's general proportions.

Eccentricity
Although the center-to-focus distance c does not appear in the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \quad (a > b)$$

for an ellipse, we can still determine c from the equation $c = \sqrt{a^2 - b^2}$. If we fix a and vary c over the interval $0 < c < a$, the resulting ellipses will vary in shape. They are circles if $c = 0$ (so that $a = b$) and flatten as c increases. If $c = a$, the foci and vertices overlap and the ellipse degenerates into a line segment.

We use the ratio of c to a to describe the various shapes the ellipse can take. We call this ratio the ellipse's eccentricity.

Definition 8.2.1. *The **eccentricity** of the ellipse* $(x^2/a^2) + (y^2/b^2) = 1$ $(a > b)$ *is*

$$e = \frac{c}{a} = \frac{\sqrt{a^2 - b^2}}{a} \tag{8.12}$$

The planets in the solar system revolve around the sun in elliptical orbits with the sun at one focus. Most of the orbits are nearly circular. Pluto has a fairly eccentric orbit, with $e = 0.25$, as does Mercury, with $e = 0.21$. Other members of the solar system have orbits that are even more eccentric. Icarus, an asteroid about 1 mile wide that revolves around the sun every 409 Earth days, has an orbital eccentricity of 0.83.

Example 8.2.1. *The orbit of Halley's comet is an ellipse 36.18 astronomical units long by 9.12 astronomical units wide. (One astronomical unit [AU] is 149,597,870 km, the semimajor axis of Earth's orbit.) Its eccentricity is*

$$e = \frac{\sqrt{a^2 - b^2}}{a} = \frac{\sqrt{(36.18/2)^2 - (9.12/2)^2}}{(1/2)(36.18)} = \frac{\sqrt{(18.09)^2 - (4.56)^2}}{18.09} \approx 0.97$$

Whereas a parabola has one focus and one directrix, each ellipse has two foci and two directrices. These are the lines perpendicular to the major axis at distances $\pm a/e$ from the center. The parabola has the property that

$$PF = 1 \cdot PD \tag{8.13}$$

for any point P on it, where F is the focus and D is the point nearest P on the directrix. For an ellipse, it can be shown that the equations that replace (7.13) are

$$PF_1 = e \cdot PD_1, \quad PF_2 = e \cdot PD_2 \tag{8.14}$$

In each equation in (7.14) the directrix and focus must correspond; that is, if we use the distance from $PtoF_1$, we must also use the distance from P to the directrix at the same end of the ellipse. The directrix $x = -a/e$ corresponds to $F_1(-c, 0)$, and the directrix $x = a/e$ corresponds to $F_2(c, 0)$.

The eccentricity of a hyperbola is also $e = c/a$, only in this case c equals $\sqrt{a^2 + b^2}$ instead of $\sqrt{a^2 - b^2}$. In contrast to the eccentricity of an ellipse, the eccentricity of a hyperbola is always greater than 1.

Definition 8.2.2. *The* ***eccentricity*** *of the hyperbola* $(x^2/a^2) - (y^2/b^2) = 1$ *is*

$$e = \frac{c}{a} = \frac{\sqrt{a^2 + b^2}}{a} \tag{8.15}$$

In both ellipse and hyperbola, the eccentricity is the ratio of the distance between the foci to the distance between the vertices (because $c/a = 2c/2a$).

$$\text{Eccentricity} = \frac{\text{distance between foci}}{\text{distance between vertices}} \tag{8.16}$$

In an ellipse, the foci are closer together than the vertices and the ratio is less than 1. In a hyperbola, the foci are farther apart than the vertices and the ratio is greater than 1.

Example 8.2.2. *Locate the vertices of an ellipse of eccentricity 0.8 whose foci lie at the points* $(0, \pm 7)$.

Solution Since $e = c/a$, the vertices are the points $(0, \pm a)$ where

$$a = \frac{c}{e} = \frac{7}{0.8} = 8.75$$

or $(0, \pm 8.75)$

Definition 8.2.3. *The* ***eccentricity*** *of a parabola is* $e = 1$.

The "focus-directrix" equation $PF = e \cdot PD$ unites the parabola, ellipse, and hyperbola in the following way. Suppose that the distance PF of a point P from a fixed point F(the focus) is a constant multiple of its distance from a fixed line (the directrix). That is, suppose

$$PF = e \cdot PD \tag{8.17}$$

where e is the constant of proportionality. Then the path traced by P is

a) a *parabola* if $e = 1$

b) an *ellipse* of eccentricity e if $e < 1$, and

c) a *hyperbola* of eccentricity e if $e > 1$

Equation (8.17) may not look like much to get excited about. There are no coordinates in it and when we try to translate it into coordinate form it translates in different ways, depending on the size of e. At least, that is what happens in Cartesian coordinates. However, in polar coordinates, the equation $PF = e \cdot PD$ translates into a single equation regardless of the value of e, an equation so simple that it has been the equation of choice of astronomers and space scientists for nearly 300 years.

Given the focus and corresponding directrix of a hyperbola centered at the origin and with foci on the x-axis, we can find e. Knowing e, we can derive a Cartesian equation for the hyperbola from the equation $PF = e \cdot PD$, as in the next example. We can find equations for ellipses centered at the origin and with foci on the x-axis in a similar way.

Example 8.2.3. *Find a Cartesian equation for the hyperbola centered at the origin that has a focus at $(3, 0)$ and the line $x = 1$ as the corresponding directrix.*

Solution We first find the hyperbola's eccentricity. The focus is

$$(c, 0) = (3, 0), \quad \text{so} \quad c = 3$$

The directrix is the line

$$x = \frac{a}{e} = 1, \quad \text{so} \quad a = e$$

When combined with the equation $e = c/a$ that defines eccentricity, these results give

$$e = \frac{c}{a} = \frac{3}{e}, \quad \text{so} \quad e^2 = 3 \text{ and } e = \sqrt{3}$$

Knowing e, we can now derive the equation we want from the equation $PF = e \cdot PD$. We have

$$\begin{aligned} PF &= e \cdot PD \\ \sqrt{(x-3)^2 + (y-0)^2} &= \sqrt{3}|x - 1| \\ x^2 - 6x + 9 + y^2 &= 3(x^2 - 2x + 1) \\ 2x^2 - y^2 &= 6 \\ \frac{x^2}{3} - \frac{y^2}{6} &= 1 \end{aligned}$$

8.3 Polar Co-ordinates

In this section, we study polar coordinates and their relation to Cartesian coordinates. While a point in the plane has just one pair of Cartesian coordinates, it has infinitely many pairs of polar coordinates. This has interesting consequences for graphing, as we will see in the next section.

Definition 8.3.1.

To define polar coordinates, we first fix an origin O (called the pole) and an initial ray from O (Fig. 8.11). Then each point P can be located by assigning to it a polar coordinate pair (r, θ) in which r gives the directed distance from O to P and θ gives the directed angle from the initial ray to ray OP.

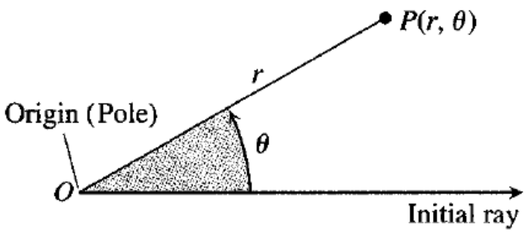

Figure 8.11: *To define polar coordinates for the plane, we start with an origin, called the pole, and an initial ray.*

Polar Coordinates

$$p(r, \theta) \tag{8.18}$$

As in trigonometry, θ is positive when measured counterclockwise and negative when measured clockwise. The angle associated with a given point is not unique. For instance, the point 2 units from the origin along the ray $\theta = \pi/6$ has polar coordinates $r = 2, \theta = \pi/6$. It also has coordinates $r = 2, \theta = -11\pi/6$ (Fig. 8.11).

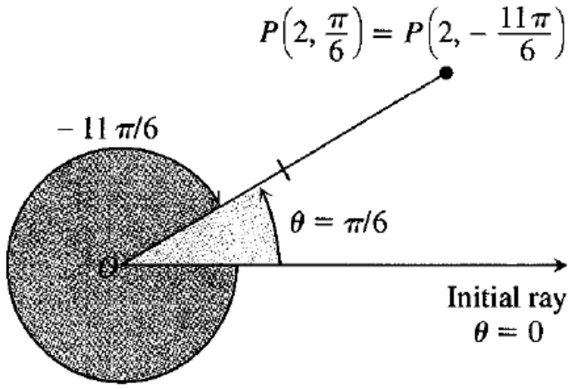

Figure 8.12: *Polar co-ordinates are not unique*

Negative Values of r

There are occasions when we wish to allow r to be negative. That is why we use directed distance in (Eq. 8.18). The point $P(2, 7\pi/6)$ can be reached by turning $7\pi/6$ rad counterclockwise from the initial ray and going forward 2 units (Fig. 8.13). It can also be reached by turning $\pi/6$ rad counterclockwise from the initial ray and going backward 2 units. So the point also has polar coordinates $r = -2, \theta = \pi/6$.

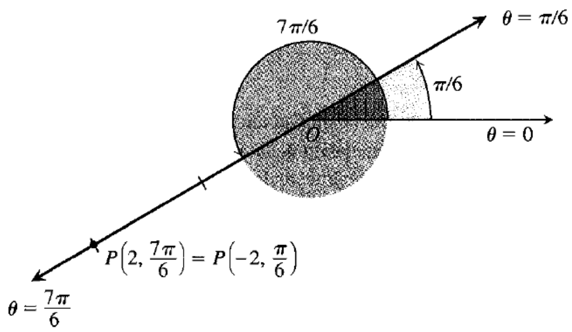

Figure 8.13: *Polar co-ordinates can have negative r values*

Example 8.3.1. *Find all the polar coordinates of the point $P(2, \pi/6)$.*

We sketch the initial ray of the coordinate system, draw the ray from the origin that makes an angle of $\pi/6$ rad with the initial ray, and mark the point $(2, \pi/6)$ (Fig. 8.14). We then find the angles for the other coordinate pairs of P in which $r = 2$ and $r = -2$.

For $r = 2$, the complete list of angles is

$$\frac{\pi}{6}, \frac{\pi}{6} \pm 2\pi, \frac{\pi}{6} \pm 4\pi, \frac{\pi}{6} \pm 6\pi...$$

For $r = -2$, the angles are

$$-\frac{5\pi}{6}, -\frac{5\pi}{6} \pm 2\pi, -\frac{5\pi}{6} \pm 4\pi, -\frac{5\pi}{6} \pm 6\pi, ...$$

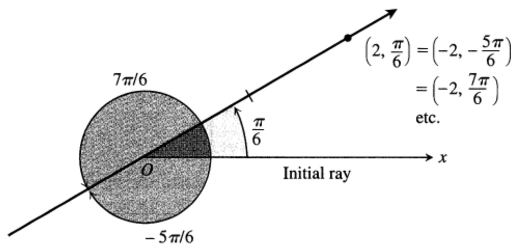

Figure 8.14: *The point $P(2, \pi/6)$ has infinitely many polar coordinate pairs.*

The corresponding coordinate pairs of P are

$$(2, \frac{\pi}{6} + 2n\pi), \quad n = 0, \pm1, \pm2...$$

$$(-2, \frac{5\pi}{6} + 2n\pi), \quad n = 0, \pm1, \pm2...$$

When $n = 0$, the formulas give $(2, \pi/6)$ and $(-2, -5\pi/6)$. When $n = 1$, they give $(2, 13\pi/6)$ and $(-2, 7\pi/6)$, and so on.

Cartesian Versus Polar Coordinates

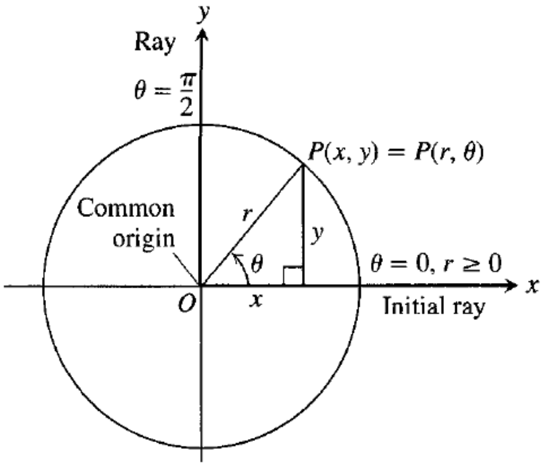

Figure 8.15: *The usual way to relate polar and cartesian co-ordinates.*

When we use both polar and Cartesian coordinates in a plane, we place the two origins together and take the initial polar ray as the positive x-axis. The ray $\theta = \pi/2, r > 0$, becomes the positive y-axis (Fig. 8.15). The two coordinate systems are then related by the following equations.

Equations Relating Polar and Cartesian

$$x = r\cos\theta, \quad y = r\sin\theta, \quad x^2 + y^2 = r^2, \quad \frac{y}{x} = \tan\theta \tag{8.19}$$

We use (Eq 8.19)to rewrite polar equations in Cartesian form and vice versa.

Example 8.3.2.

Polar equation	Cartesian equivalent
$r\cos\theta = 2$	$x = 2$
$r^2\cos\theta\sin\theta = 4$	$xy = 4$
$r^2\cos^2\theta - r^2\sin^2\theta = 1$	$x^2 - y^2 = 1$
$r = 1 + 2r\cos\theta$	$y^2 - 3x^2 - 4x - 1 = 0$
$r = 1 - \cos\theta$	$x^4 + y^4 + 2x^2y^2 + 2x^3 + 2xy^2 - y^2 = 0$

With some curves, we are better off with polar coordinates; with others, we aren't.

Example 8.3.3. *Find a polar equation for the circle* $x^2 + (y-3)^2 = 9$.

$$\begin{aligned} x^2 + (y-3)^2 &= 9 \\ x^2 + y^2 - 6y + 9 &= 9 \\ x^2 + y^2 - 6y &= 0 \\ r^2 - 6r\sin\theta &= 0 \\ r = 0 \text{ or } r - 6\sin\theta &= 0 \\ r &= 6\sin\theta \end{aligned}$$

We will say more about polar equations of conic sections in the next Section.

8.4 Polar Equations for Conic Sections

Polar coordinates are important in astronomy and astronautical engineering because the ellipses, parabolas, and hyperbolas along which satellites, moons, planets, and comets move can all be described with a single relatively simple coordinate equation. We develop that equation here.

Lines

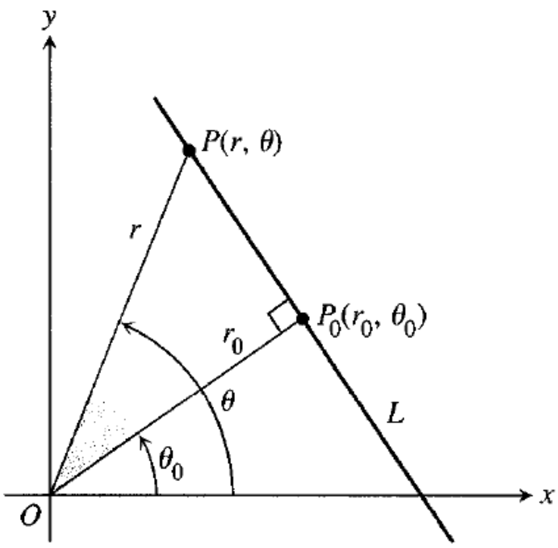

Figure 8.16: *We can obtain a polar equation for the line L by reading the relation $r_0/r = \cos(\theta - \theta_0)$*

Suppose the perpendicular from the origin to line L meets L at the point $P_0(r_0, \theta_0)$, with $r_0 > 0$ (Fig. 8.16). Then, if $P(r, \theta)$ is any other point on L, the points P, P_0, and O are the vertices of a right triangle, from which we can read the relation

$$\frac{r_0}{r} = \cos(\theta - \theta_0)$$

or

$$r\cos(\theta - \theta_0) = r_0$$

The Standard Polar Equation for Lines

If the point $P_0(r_0, \theta_0)$ is the foot of the perpendicular from the origin to the line L, and $r_0 > 0$, then an equation for L is

$$r\cos(\theta - \theta_0) = r_0 \tag{8.20}$$

Example 8.4.1. *Use the identity* $\cos(A-B) = \cos A\cos B+\sin A\sin B$ *to find a Cartesian equation for the line in Fig. 8.17.*

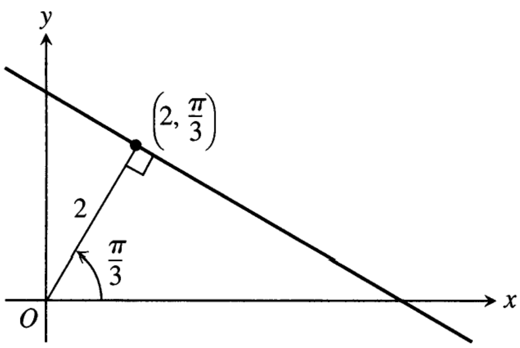

Figure 8.17: *The standard polar equation of this line is* $r\cos(\theta - \frac{\pi}{3}) = 2$

$$\begin{aligned} r\cos(\theta - \frac{\pi}{3}) &= 2 \\ r(\cos\theta\cos\frac{\pi}{3} + \sin\theta\sin\frac{\pi}{3}) &= 2 \\ \frac{1}{2}r\cos\theta + \frac{\sqrt{3}}{2}r\sin\theta &= 2 \\ \frac{1}{2}x + \frac{\sqrt{3}}{2}y &= 2 \\ x + \sqrt{3}y &= 4 \end{aligned}$$

Circles

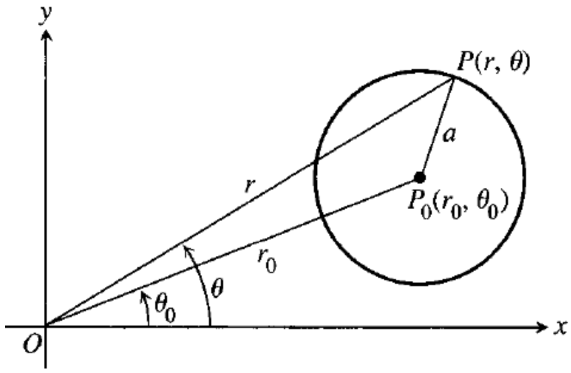

Figure 8.18: *We can obtain a polar equation for this circle by applying the Law of Cosines to triangle OP_0P*

To find a polar equation for the circle of radius a centered at $P_0(r_0, \theta_0)$, we let $P(r, \theta)$ be a point on the circle and apply the Law of Cosines to triangle OP_0P (Fig. 8.18). This gives

$$a^2 = r_0^2 + r^2 - 2r_0 r\cos(\theta - \theta_0). \tag{8.21}$$

If the circle passes through the origin, then $r_0 = a$ and Eq. 8.21 simplifies to

$$\begin{aligned} a^2 &= a^2 + r^2 - 2ar\cos(\theta - \theta_0) \\ r^2 &= 2ar\cos(\theta - \theta_0) \\ r &= 2a\cos(\theta - \theta_0) \qquad (8.22) \end{aligned}$$

If the circle's center lies on the positive x-axis, $\theta_0 = 0$ and Eq. 8.22 becomes

$$r = 2a\cos\theta \tag{8.23}$$

If the center lies on the positive y-axis, $\theta = \pi/2, \cos(\theta - J\pi/2) = \sin\theta$, and Eq. 8.22 becomes

$$r = 2a\sin\theta \tag{8.24}$$

Equations for circles through the origin centered on the negative x- and y-axes can be obtained from Eqs. 8.23 and 8.24 by replacing r with $-r$.

Example 8.4.2. *Circles through the origin*

Radius	Center (polar coordinates)	Equation
2	$(3, 0)$	$r = 6\cos\theta$
3	$(2, \pi/2)$	$r = 4\sin\theta$
$\frac{1}{2}$	$(-1/2, 0)$	$r = -\cos\theta$
1	$(-1, \pi/2)$	$r = 2\sin\theta$

Ellipses, Parabolas, and Hyperbolas Unified

To find polar equations for ellipses, parabolas, and hyperbolas, we place one focus at the origin and the corresponding directrix to the right of the origin along the vertical line $x = k$ (Fig. 8.19). This makes

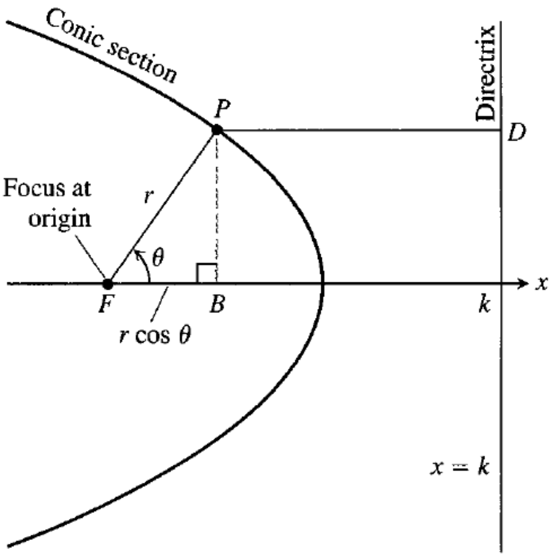

Figure 8.19: *If a conic section in put in this position, then $PF = r$ and $PD = k - r\cos\theta$*

$$PF = r$$

and

$$PD = k - FB = k - r\cos\theta.$$

The conic's focus-directrix equation $PF = e \cdot PD$ then becomes

$$r = e(k - r\cos\theta),$$

which can be solved for r to obtain

$$r = \frac{ke}{1 + c\cos\theta} \tag{8.25}$$

This equation represents an ellipse if $0 < e < 1$, a parabola if $e = 1$, and a hyperbola if $e > 1$. And there we have it- ellipses, parabolas, and hyperbolas all with the same basic equation.

Example 8.4.3. *Typical conics from Eq. 8.25*

$e = \frac{1}{2}$:	ellipse	$r = \frac{k}{2+\cos\theta}$
$e = 1$:	parabola	$r = \frac{k}{1+\cos\theta}$
$e = 2$:	hyperbola	$r = \frac{k}{1+2\cos\theta}$

You may see variations of Eq. 8.25 from time to time, depending on the location of the directrix. If the directrix is the line $x = -k$ to the left of the origin (the origin is still a focus), we replace Eq. 8.25 by

$$r = \frac{ke}{1 - e\cos\theta}$$

The denominator now has a $(-)$ instead of a $(+)$. If the directrix is either of the lines $y = k$ or $y = -k$, the equations we get have sines in them instead of cosines, as shown in Fig 8.20.

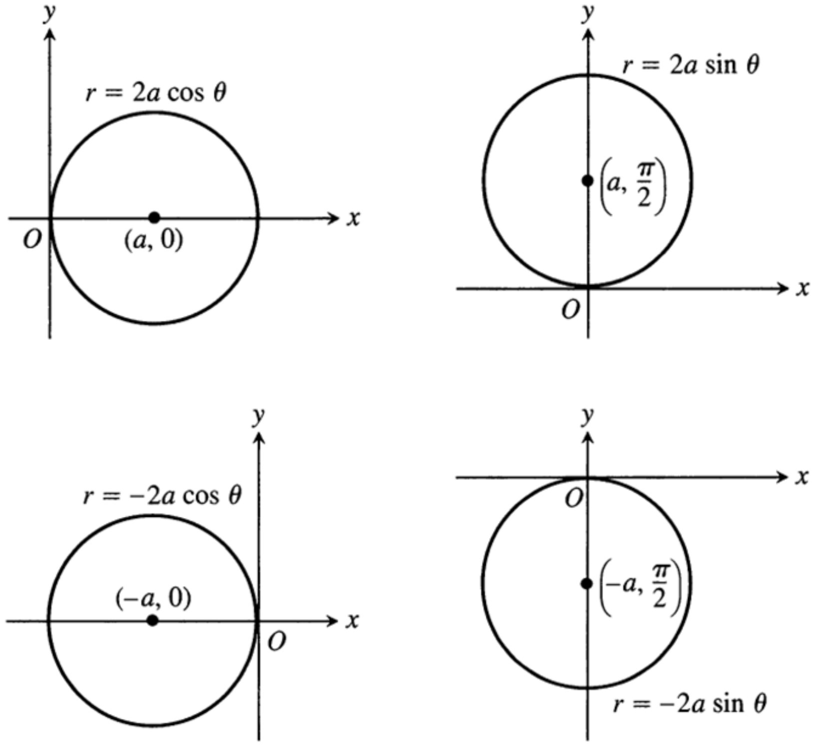

Figure 8.20: *Polar Equations for Circles Through the Origin Centered on the x- and y-axes, Radius a*

8.5 Cylindrical and Spherical Co-ordinates

This section introduces two new coordinate systems for space: the cylindrical coordinate system and the spherical coordinate system. Cylindrical coordinates simplify the equations of cylinders. Spherical coordinates simplify the equations of spheres and cones.

Cylindrical Coordinates

We obtain cylindrical coordinates for space by combining polar coordinates in the xy-plane with the usual z-axis. This assigns to every point in space one or more coordinate triples of the form (r, θ, z), as shown in Fig. 8.21.

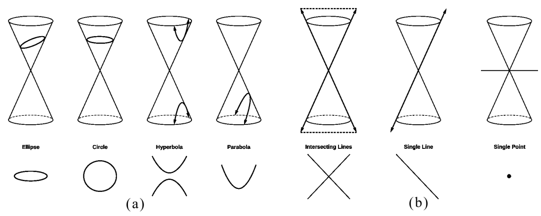

Figure 8.21: *The cylindrical co-ordinates of a point in space are r, θ, z*

Definition 8.5.1. ***Cylindrical coordinates*** *represent a point P in space by ordered triples (r, θ, z) in which*

1. *r and θ are polar coordinates for the vertical projection of P on the xy-plane,*
2. *z is the rectangular vertical coordinate.*

The values of x, y, r, and θ In rectangular and cylindrical coordinates are related by the usual equations.

Equations Relating Rectangular (x, y, z) and Cylindrical $(r; \theta, z)$ Coordinates

$$x = r\cos\theta \;\; y = r\sin\theta \;\; z = z \tag{8.26}$$
$$r^2 = x^2 + y^2 \;\; \tan\theta = \frac{y}{x} \tag{8.27}$$

In cylindrical coordinates, the equation $r = a$ describes not just a circle in the xy-plane but an entire cylinder about the z-axis(Fig 8.22). The z-axis is given by $r = O$. The equation $\theta = \theta_0$ describes the plane that contains the z-axis and makes an angle θ_0 with the positive x-axis. And, just as in rectangular coordinates, the equation $z = z_0$ describes a plane perpendicular to the z-axis.

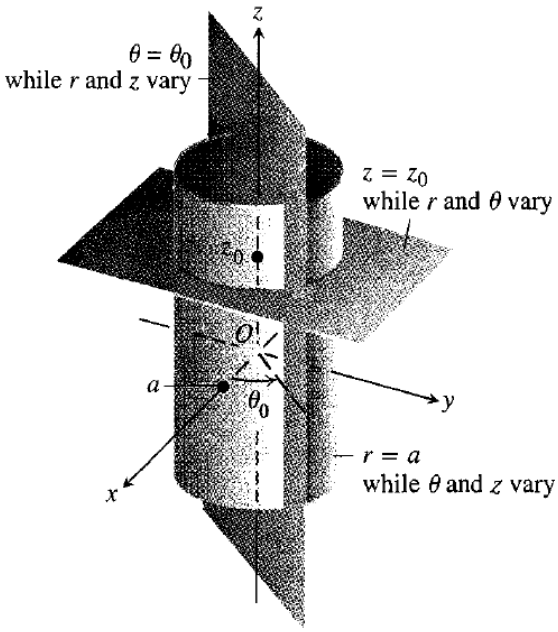

Figure 8.22: *Constant co-ordinate equations in cylindrical co-ordinates yeild cylinders and planes*

Example 8.5.1. *What points satisfy the equations*

$$r = 2, \theta = \frac{\pi}{4}$$

These points make up the line in which the cylinder $r = 2$ cuts the portion of the plane $\theta = \pi/4$ where r is positive. This is the line through the point $(2, \pi/4, 0)$ parallel to the z-axis.

Example 8.5.2. *Sketch the surface* $r = 1 + \cos\theta$*.*

The equation involves only r and θ; the coordinate variable z is missing. Therefore, the surface is a cylinder of lines that pass through the cardioid $r = 1 + \cos\theta$ in the $r\theta$-plane and lie parallel to the z-axis. The rules for sketching the cylinder are the same as always: sketch the x-, y-, and z-axes, draw a few perpendicular cross sections, connect the cross sections with parallel lines, and darken the exposed parts

Spherical Coordinates

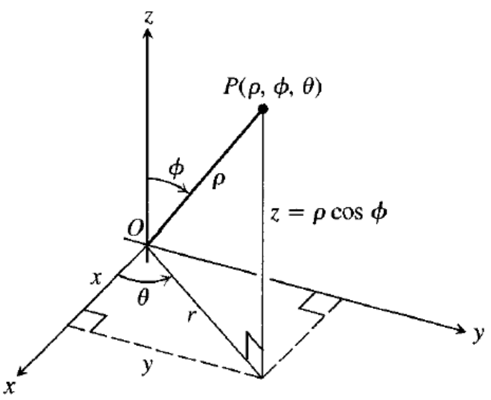

Figure 8.23: *The Spherical Co-ordinates ρ, ϕ, θ and their relation to x, y, z, and r*

Spherical coordinates locate points in space with angles and a distance, as shown in Spherical coordinates locate points in space with angles and a distance, as shown in Fig.8.23

The first coordinate, p $= |\vec{OP}|$, is the point's distance from the origin. Unlike r, the variable p is never negative. The second coordinate, ϕ, is the angle $\vec{OP}$ makes with the positive z-axis. It is required to lie in the interval $[0, \pi]$. The third coordinate is the angle θ as measured in cylindrical coordinates.

Definition 8.5.2. ***Spherical coordinates*** *represent a point P in space by ordered triples (ρ, ϕ, θ) in which*

1. *ρ is the distance from P to the origin*
2. *ϕ is the angle $\vec{OP}$ makes with the positive z-axis$(0 \leq \phi \leq \pi)$*
3. *θ is the angle from cylindrical coordinates*

The equation $p = a$ describes the sphere of radius a centered at the origin (Fig 8.24). The equation $\phi = \phi_0$ describes a single cone whose vertex lies at the origin and whose axis lies along the z-axis. (We broaden our interpretation to include the xy-plane as the cone $\phi = \pi/2$.) If ϕ_0 is greater than $\pi/2$, the cone $\phi = \phi_0$ opens downward.

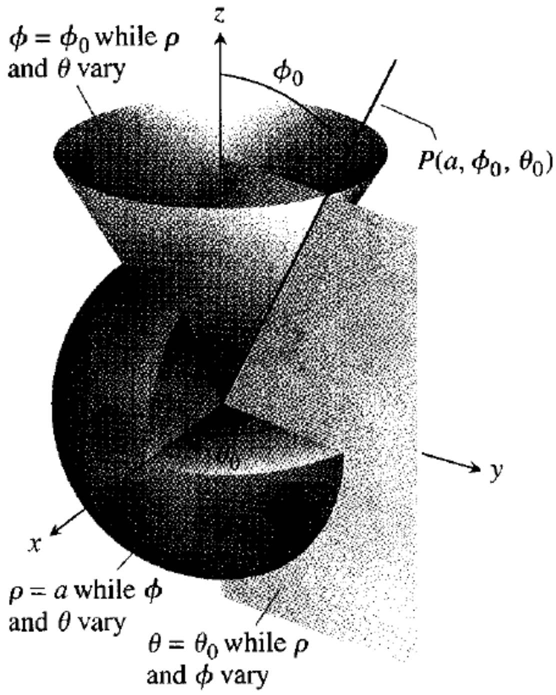

Figure 8.24: *The cylindrical co-ordinates of a point in space are r, θ, z*

Equations Relating Spherical Coordinates to Cartesian and Cylindrical Coordinates

$$
\begin{aligned}
r = \rho \sin\phi, \quad x &= r\cos\theta = \rho\sin\phi\cos\theta, \\
z = \rho\cos\phi, \quad y &= r\sin\theta = \rho\sin\phi\sin\theta, \\
\rho &= \sqrt{x^2+y^2+z^2} = \sqrt{r^2+z^2}
\end{aligned}
\tag{8.28}
$$

Example 8.5.3. *Find a spherical coordinate equation for the sphere*

$$x^2 + y^2 + (z-1)^2 = 1$$

We use Eqs. 8.28 to substitute for x, y, and z:

$$
\begin{aligned}
x^2 + y^2 + (z-1)^2 &= 1 \\
\rho^2 \sin^2\phi \cos^2\theta + \rho^2 \sin^2\phi \sin^2\theta + (\rho\cos\phi - 1)^2 &= 1 \\
\rho^2 \sin^2\phi(\cos^2\theta + \sin^2\theta) + \rho^2\cos^2\phi - 2\rho\cos\phi + 1 &= 1 \\
\rho^2(\sin^2\phi + \cos^2\phi) &= 2\rho\cos\phi \\
\rho^2 &= 2\rho\cos\phi \\
\rho &= 2\cos\phi
\end{aligned}
$$

Appendices

List of Figures