Chapter 2. Mathematical appendix

In this appendix to the Mathematica notes, we look into a few mathematical points that arose in the Mathematica examples but may not be familiar.

2.19 Remarks. By a function f from one set A of objects to another B, we mean any clearly defined <u>rule</u> that associates <u>one and only one</u> element of B to each element of A. If $a \in A$ (read "a is an element of A") then we denote by f(a) the element of B that f associates to a.

The set A is called the *domain* of the function.

We will mostly deal with the case where $A, B \subset \mathbb{R}$, that is they are sets of numbers, and very often the function will be specified by giving a formula for the value f(x) in terms of x. For example we could have $f(x) = x^2 + 1$, meaning that (say) when x = 5, $f(x) = f(5) = 5^2 + 1 = 26$. That is the function f associates the number 26 to x = 1.

A good way to picture a function is via its graph y = f(x). To be more exact, if $f: A \to B$ where A and B are sets of numbers, then when we plot the graph we plot all the points (x, y) in the plane where $x \in A$. Typically we end up with a curve, though in complicated examples we could end up with several disjointed curves.

Notice that a graph of a function is a 'curve' in the plane with the property that each vertical line crosses it at most once.



We can think of x on the horizontal axis (where we envisage the domain) and the corresponding value y = f(x) as lying on the vertical axis. Sometimes it is useful to think of an arrow from the domain x to y = f(x) as indicated. The function or rule is all the arrows together, telling us where each x ends up.

The (vertical) line $x = x_0$ crosses the graph (once) if $x_0 \in A$ = the domain of f. If the domain A is smaller than the whole real line, then there will be vertical lines that miss the graph.

2.20 Intervals. In many cases we will be dealing with functions $f: A \to B$ where the sets A and B are *intervals*. Here we will review the notation for intervals.

If $a, b \in \mathbb{R}$ and $a \leq b$, then the *closed interval* with end points a and b is the set of all real numbers between a and b inclusive of the end points:

$$[a,b] = \{x \in \mathbb{R} : a \le x \text{ and } x \le b\} = \{x \in \mathbb{R} : a \le x \le b\}$$

The *open interval* between a and b is

$$(a,b) = \{x \in \mathbb{R} : a < x < b\}$$

There are various infinite intervals (no endpoint on one side or the other) and we use the notations ∞ and $-\infty$ as convenient replacements for the missing endpoints to the right or left. We do not mean to imply that there are any numbers ∞ or $-\infty$.

Here are the semi-infinite open and closed intervals $(a, b \in \mathbb{R})$ and the notations we use

$$[a, \infty) = \{x \in \mathbb{R} : a \le x\}$$

$$(a, \infty) = \{x \in \mathbb{R} : a < x\}$$

$$(-\infty, b] = \{x \in \mathbb{R} : x \le b\}$$

$$(-\infty, b) = \{x \in \mathbb{R} : x < b\}$$

Note the convention that round brackets (or parentheses) are used for end points that are not included in the set. If all (finite) end points are included we refer to the interval as closed. If no finite intervals are included we call the interval open. So $[a, \infty)$ and $(-\infty, b]$ are closed, while (a, ∞) and $(-\infty, b)$ are open. There is one doubly infinite interval

$$(-\infty,\infty)=\mathbb{R}$$

and it counts as both open and closed.

There remain two other types of intervals we may encounter once or twice, the half-open and half-closed intervals (which are neither open nor closed)

$$\begin{array}{ll} [a,b) &=& \{x \in \mathbb{R} : a \leq x < b\} \\ (a,b] &=& \{x \in \mathbb{R} : a < x \leq b\} \end{array}$$

where we restrict to a < b.

Technically we could allow a = b in the case [a, b] but $[a, a] = \{a\}$ is just a one-point set and this will either be a very simple case or a case we will not want to consider in future theorems. The case (a, a) is the empty set and we will probably never want to consider that.

2.21 *Examples.* Consider the function $f \colon \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$.

We may sometimes wish to discuss the same function x^2 but concentrate on a range of values of x like $0 \le x \le 1$ and then we are dealing with a different function $g: [0, 1] \to \mathbb{R}$ given by the same formula $g(x) = x^2$.

Another type of example is a function given by a rule like f(x) = 1/x which clearly does not make sense for x = 0 (when we would be trying to divide by 0). So the natural thing is to just make that restriction and consider $f: \{x \in \mathbb{R} : x \neq 0\} \to \mathbb{R}$ given by f(x) = 1/x. Here we have strayed outside having the domain A being an interval. $\{x \in \mathbb{R} : x \neq 0\} = (-\infty, 0) \cup (0, \infty)$ is a union of two intervals. 2.22 *Remark.* In section 2.10 we saw how to use the Plot [] command in Mathematica to draw graphs. Notice that we have to specify the domain (the range of x) for the graph, and Mathematica will try to understand if there are values of x where the formula given does not make sense (say Plot $[1/x, \{x, -1, 1\}]$ for example).

However, there are curves that are not graphs (for example a circle where there are vertical lines that cross the graph more than once). We can use parametric equations to describe such graphs.

2.23 Parametric curves. One way to think of parametric curves is to imagine the process of drawing it. As you draw the curve (on a pieces of paper marked with a horizontal and a vertical axis) the tip of the pen will have a position at any given time.

We have to idealise a little to a 'very' sharp pen which covers just one point at a time and then the point will have position (x, y) described by two coordinates. However, the coordinates x and y of the pen will change as you move the pen and so we need to describe how they change with time. We can do that by imagining that we record the position at all times, and that we (say) count time in seconds from some starting time that we agree upon. Then the moving pen will have position given by x = x(t) and y = y(t) at time t (seconds). We have a rule that gives xand y in terms of t. So we have two function x(t) and y(t), where t runs over the time interval where we draw the curve.

Another way to imagine it, is to imagine a tiny particle (in fact a point particle) that moves in the plane according to the rule that at time t it is at position (x, y) with

$$\begin{cases} x = x(t) \\ y = y(t) \end{cases}$$

(and t in some range). The curve is the trajectory traced out by the moving particle.

2.24 Examples. (i) Consider the example

$$\begin{cases} x = \cos t \\ y = \sin t \end{cases} \quad (0 \le t \le 2\pi).$$

We can see that every point (x, y) described in this way, we have

$$x^2 + y^2 = \cos^2 t + \sin^2 t = 1$$

and this means that all the points lie on the unit circle (points of distance 1 away from the origin).



In fact this is one approach to explaining what $\cos t$ and $\sin t$ mean. Notice that here (and **always always** in calculus) we are taking the angle t in **radians**).

(ii) We can describe the circle of radius r > 0 and center at (x_0, y_0) using the distance formula and the requirement that points (x, y) on the circle are at distance r from the centre. We get

$$\sqrt{(x-x_0)^2 + (y-y_0)^2} = r,$$

or, squaring both sides, the more usual form

$$(x - x_0)^2 + (y - y_0)^2 = r^2.$$

We can parametrise this as

$$\begin{cases} x = x_0 + r\cos t \\ y = y_0 + r\sin t \end{cases} \quad (0 \le t \le 2\pi).$$

If we don't allow t to run all the way from 0 to 2π we will only get part of the circle. If, on the other hand we let t run longer (say $0 \le t \le 4\pi$) the same points will be traced out again. In this way, the trajectory or curve traced out is not the full story. In a parametric curve we also give the information about how the curve was traced out, in which direction, and whether some parts are gone over more than once.

When we multiply out the equation of a circle we get an equation

$$x^{2} + y^{2} - 2x_{0}x - 2y_{0}y = r^{2} - x_{0}^{2} - y_{0}^{2}$$

of the form

$$x^2 + y^2 + Ax + By = C$$

where x^2 and y^2 have the same coefficient and there is no xy term. These are quadratic equations in x and y of a very special form.

(iii) Quadratic equations of the type

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

describe *ellipses* centred at the origin.



It is called an ellipse in 'standard position' where the longest and shortest diameters are along the axes. The ellipse fits inside a box centered at the origin extending from x = -a to x = a horizontally and from y = -b to y = b. (This box is included in the diagram above as well as the actual curve, the ellipse.)

$$\begin{cases} x = a\cos t \\ y = b\sin t \end{cases} \quad (0 \le t \le 2\pi).$$

describes the curve in parametric terms and this can be used in a Mathematica command. In the case a = 5 and b = 2, we would type

ParametricPlot[{ 5 Cos[t], 2 Sin[t] }, {t, 0, 2 Pi}]

to plot the ellipse

$$\frac{x^2}{25} + \frac{y^2}{4} = 1$$

(iv) There is another rather different type of curve that arises from a quadratic equation in x and y. It is called a *hyperbola* and in standard position (centered on the origin) it has an equation

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1$$

One way to draw this curve is to draw the two diagonals of a rectangle centered at the origin extending from x = -a to x = a horizontally and from y = -b to y = b. Extend the diagonals infinitely far and the hyperbola has two sections, which follow the diagonals very closely in the far distance.



To parametrise this curve, we cannot use trigonometric functions but we can do it with something called *hyperbolic functions*.

2.25 Hyperbolic functions. We explain now what the *hyperbolic cosine* $\cosh x$ is and what the *hyperbolic sine* function $\sinh x$. Our explanation is in terms of the exponential function e^x which is still to come in course 1S1.

The definitions are

and

$$\cosh x = \frac{e^x + e^{-x}}{2}$$
$$\sinh x = \frac{e^x - e^{-x}}{2}.$$

These look completely different from the trigonometric functions $\cos \theta$ and $\sin \theta$ (which we have seen relate to a circle), but they have very similar properties. The big difference is that the trig functions are periodic with period 2π , that is

$$\cos(\theta + 2\pi) = \cos\theta, \qquad \sin(\theta + 2\pi) = \sin\theta$$

do *not correspond* to any properties of hyperbolic functions but just about every other formula for trig functions has a counterpart for hyperbolic functions. The difference is that the minus signs are not in the same places.

For example $\cosh^2 x - \sinh^2 x = 1$ (rather than $\cos^2 \theta + \sin^2 \theta = 1$). We can check out this formula this way

$$\cosh^{2} x - \sinh^{2} x = (\cosh x)^{2} - (\sinh x)^{2}$$

$$= \left(\frac{e^{x} + e^{-x}}{2}\right)^{2} - \left(\frac{e^{x} - e^{-x}}{2}\right)^{2}$$

$$= \frac{(e^{x})^{2} + (e^{-x})^{2} + 2e^{x}e^{-x}}{4} - \frac{(e^{x})^{2} + (e^{-x})^{2} - 2e^{x}e^{-x}}{4}$$

$$= \frac{4e^{x}e^{-x}}{4} = e^{x}e^{-x} = e^{x-x} = e^{0} = 1$$

Changing x to t we get $\cosh^2 t - \sinh^2 t = 1$ and this means that the parametric curve described by

$$\begin{cases} x = \cosh t \\ y = \sinh t \end{cases} \quad (-\infty < t < \infty)$$

consists of points lying on the hyperbola

$$\frac{x^2}{1} - \frac{y^2}{1} = 1$$

(or more simply $x^2 - y^2 = 1$). In fact the parametric curve has $x = \cosh t > 0$ always and so the points are all on the right hand section of the hyperbola.

In 2.12 we had the example

which means plotting

$$\begin{cases} x = 2\cosh t \\ y = \sinh t \end{cases} \quad (-2 \le t \le 2)$$

(and the PlotRange instruction is to show the region where x and y each range -5 to 5).

In general, for a standard hyperbola

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

we can parametrise the right hand section as

$$\begin{cases} x = a \cosh t \\ y = b \sinh t \end{cases} \quad (-\infty < t < \infty).$$

2.26 Superposition of plots. Sometimes it is of interest to show two graphs together on the one plot. The Plot [] command in Mathematica has a facility for doing two plots together. For example

```
Plot[ { Cosh[x], Exp[x]/2 }, {x, -2, 2} ]
```

is a way to plot the two graphs $y = \cosh x$ and $y = e^x/2$ on the same diagram over the range $-2 \le x \le 2$. (If you try it out, you should see that they are very close to each other around x = 2. This is because the difference $\cosh x - e^x/2 = e^{-x}/2$ and $e^{-2}/2 = 0.0676676$ is quite small. In fact $\lim_{x\to\infty} e^{-x}/2 = 0$ and so the difference between the two graphs would be very hard to notice at all if we go to larger x. For x < 0 though, these graphs are quite different.)

One thing that might be of interest is to look at a graph y = f(x) and its tangent line at a specific point x = a. You know from course 1S1 that the slope of the tangent line to y = f(x) at the point on the graph where x = a is the derivative (or rate of change) f'(a). The tangent line is the line with slope f'(a) that goes through (a, f(a)) (which is the point on the graph at x = a, the point where the line should be tangent.

You can work out the equation of the tangent line by taking a line y = mx + c with slope m = f'(a) and then fixing c so that (x, y) = (a, f(a)) satisfies the equation. In fact there is a tidy way to write the equation of the tangent line as

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

and we will use this again quite often. You should see that x = a, y = f(a) satisfies this equation and it is the equation of a line. (If you insist on writing it in the form y = mx + c, you can do that and you get y = f'(a)x + (f(a) - f'(a)a), but that is not such a neat formula as the one displayed above.)

The sequence of 3 mathematica commands

```
a = 1.2;
f[x_] = Cosh[x];
Plot[ { f[x], f[a] + f'[a](x-a) }, {x, -2, 2} ]
```

will then show the graph $y = \cos x$ together with its tagent line at the point where x = 1.2.

Probably it is rather unnecessary to use any 'programming' here as the function has a very simple formula, and indeed Cosh[] is already known to Mathematica so that introducing a new shorter name f[] for the same thing is hardly worth it. But you can see that we can use the mathematical formula y = f(a) + f'(a)(x - a) for the tangent line with almost no change.

You could work out $\cosh 1.2 = 1.81066$ (so that (1.2, 1.81066) has to be a point on the tangent line) and figure out that the slope is $1.50946 = \frac{d}{dx} \cosh x \mid_{x=1.2} = \sinh x \mid_{x=1.2} = \sinh 1.2$. Then the equation of the tangent line is y = 1.81066 + 1.50946(x-1.2) or y = 1.50946x - 0.000698059 and we could plot the two things together by

```
Plot[ { Cosh[x], 1.50946 x -0.000698059 }, {x, -2, 2} ]
```

but the first way is easier to follow even if it uses some greater understanding of how Mathematica works.