7 Differentiation and Differentials

7.1 Differentiation – Definition

The following derivation of the concept of differentiation is probably familiar. However, it is worth revisiting, as we will generalise this idea later to new contexts.

Often we want to know the rate at which a function's output varies with respect to its input. For sufficiently small δx, a good approximation to this rate of change is $\delta f(x_0)/\delta x$. Note that we normalise the rate by scaling down by δx. Without taking this ratio, the "rate of change" would depend on the particular δx we choose.

Choosing a large δx is not going to give us a very good estimate of the rate of change. This is because the rate of change can itself vary at different inputs. For example, if the function is increasing at x_0 but then starts to decrease for larger x, a small δx can give a positive rate of change, while a large δx may give a negative rate of change.

Our estimate of the rate of change improves as we make δx smaller. So if we allow δx to decrease and become arbitrarily close to 0, we will hone in on the exact value of the rate of change. We define the **derivative** of f at x_0 as:

$$\frac{\mathrm{d}f(x_0)}{\mathrm{d}x} = \lim_{\delta x \to 0} \frac{\delta f(x_0)}{\delta x} = \lim_{\delta x \to 0} \frac{f(x_0 + \delta x) - f(x_0)}{\delta x} \,. \tag{1}$$

This process of taking the limit of the ratio between the changes in output and input values of the function is differentiation. It is worth remembering that this is the **definition** of differentiation and that everything else related to the concept follows from here. Later, when we generalise differentiation to less familiar contexts, this is the key concept that we wish to keep.

Note that we have, so far, defined the derivative of a function at a point. Since the point x_0 was arbitrary, we could equally choose a different fixed point, x_1 and find the derivative of the function at this new point: $\mathrm{d}f(x_1)/\mathrm{d}x$. In fact, since we can choose any base point, the collection of such derivatives is itself a function: $f'(x) \equiv \mathrm{d}f(x)/\mathrm{d}x$.

7.2 Derivatives of Familiar Functions

Armed with the definition of the derivative, we can use some basic algebra to derive some very general results. In particular, for constants a and n, and functions f and g, we have

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\left(ax^n\right) &= nax^{n-1} \\
\frac{\mathrm{d}}{\mathrm{d}x}\left(\sin(x)\right) &= \cos(x) \\
\frac{\mathrm{d}}{\mathrm{d}x}\left(\cos(x)\right) &= -\sin(x) \\
\frac{\mathrm{d}}{\mathrm{d}x}\left(a^x\right) &= \left(\ln a\right) a^x \,,
\end{aligned}
\tag{2}
$$

which can all be derived from the definition above. The derivations of the trigonometric examples here rely on the Sandwich Theorem, which follows shortly.

We can also define an **anti-derivative** of the function f as a function F such that $F'(x) = f(x)$. Notice that there are infinitely many anti-derivatives of any function, that differ by a constant, since $\mathrm{d}\left(F(x)+c\right)/\mathrm{d}x = F'(x)$.

7.3 Example: $f(x) = x^n$

Let's calculate the derivative of x^n using the definition. We have

$$
\begin{aligned}
\frac{\mathrm{d}f}{\mathrm{d}x} &= \lim_{\delta x \to 0} \frac{(x+\delta)^n - x^n}{\delta x} \\
&= \lim_{\delta x \to 0} \frac{\sum_k = 0^n \begin{matrix} n \\ k \end{matrix} x^{n-k}\delta x^k - x^n}{\delta x} \\
&= \lim_{\delta x \to 0} \frac{\sum_k = 1^n \begin{matrix} n \\ k \end{matrix} x^{n-k}\delta x^k}{\delta x} \\
&= \lim_{\delta x \to 0} \sum_k = 1^n \begin{matrix} n \\ k \end{matrix} x^{n-k}\delta x^{k-1} \\
&= \lim_{\delta x \to 0} \sum_k = 1^n \begin{matrix} n \\ k \end{matrix} x^{n-k}\delta x^{k-1} \\
&= \begin{matrix} n \\ 1 \end{matrix} x^{n-1} \\
&= nx^{n-1} \,.
\end{aligned}
$$

7.4 Example: $f(x) = a^x$

Following the same steps again for a^x for some constant a, we have

$$\begin{aligned}
\frac{\mathrm{d}f}{\mathrm{d}x} &= \lim_{\delta x \to 0} \frac{a^{x+\delta x} - a^x}{\delta x} \\
&= \lim_{\delta x \to 0} \frac{a^x a^\delta x - a^x}{\delta x} \\
&= \lim_{\delta x \to 0} \frac{a^x \left(a^\delta x - 1\right)}{\delta x} \\
&= a^x \times \lim_{\delta x \to 0} \frac{a^\delta x - 1}{\delta x} \,,
\end{aligned}$$

and we find that the derivative is proportional to the function we started with. We can then manipulate the remaining limit to define a value, $a = e$, for which this limit is 1.

7.5 The Sandwich Theorem

The sandwich theorem is a very useful result that allows us to calculate some tricky limits. We will not give a proof of the theorem here, since it relies on other results from real analysis that we do not wish to explore. However, when given the statement of the theorem, hopefully it will seem intuitive. Consider three functions, f, g and h, such that $f(x) \leq g(x) \leq h(x)$ for all x in some interval. If there is some argument, x_0, such that

$$\lim_{x \to x_0} f(x) = \lim_{x \to x_0} h(x) = L \,,$$

then

$$\lim_{x \to x_0} g(x) \text{ exists and is equal to } L \,.$$

A good example of this theorem in action is in the calculation of the derivative of $\sin(x)$, since it is necessary to show

$$\lim_{\delta x \to 0} \frac{\sin(\delta x)}{\delta x} = 0 \,,$$

which follows from considering the area of the three regions shown in Fig. 1.

7.6 Differentials

Notice that the definition of the derivative does not require any concept of "infinitesimal quantities". We can work entirely with finite quantities while performing any necessary algebraic steps in the calculation, and then take the limit that the difference tends to 0. However, it is sometimes beneficial to work with "infinitesimal differences" in our quantities, while keeping at the back of our minds that these are really a shorthand for the steps listed. Such infinitesimal differences are called **differentials** and are typically denoted $\mathrm{d}x$, $\mathrm{d}f$, etc.

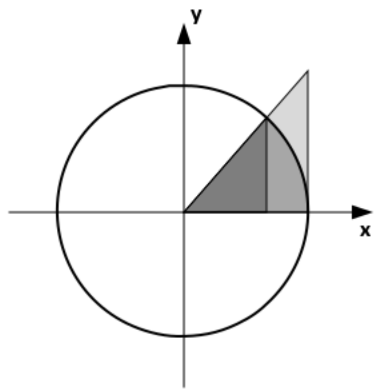

Figure 1: For acute angles, the area of the smaller triangle is always smaller than the area of the sector, which in turn is always smaller than the area of the larger triangle.

As an example, let's re-calculate the derivative of the cubic function using the differential approach instead. For an infinitesimal change in the function argument, $\mathrm{d}x$, there will be a corresponding infinitesimal change in the function's value, $\mathrm{d}f$, given by

$$\begin{aligned}\mathrm{d}f &= f(x+\mathrm{d}x) - f(x)\\ &= (x+\mathrm{d}x)^3 - x^3\\ &= x^3 + 3x^2\mathrm{d}x + 3x\mathrm{d}x^2 + \mathrm{d}x^3 - x^3\\ &= 3x^2\mathrm{d}x + 3x\mathrm{d}x^2 + \mathrm{d}x^3\,.\end{aligned}$$

At this point, we do not need to take any limits, since the quantities are already infinitesimally small. Instead, we simply recognise that $\mathrm{d}x^2$ is infinitesimal **on the scale of** $\mathbf{df}$. And $\mathrm{d}x^3$ is even smaller than that! So these quantities are 0 on the scale of the quantities in which we are interested. Therefore,

$$\mathrm{d}f = 3x^2\mathrm{d}x\,,$$

which we can rearrange to

$$\frac{\mathrm{d}f}{\mathrm{d}x} = 3x^2\,.$$

7.7 Second-Order Differentials

We can extend the notion of differentials to second- and higher-order differentials. Remember that $\mathrm{d}f$ means "the infinitesimal change in f at a point when the argument is changed by $\mathrm{d}x$". This requires comparing two points $\mathrm{d}x$ apart. We could also consider a point a further $\mathrm{d}x$ along to get the "next" value of $\mathrm{d}f$,

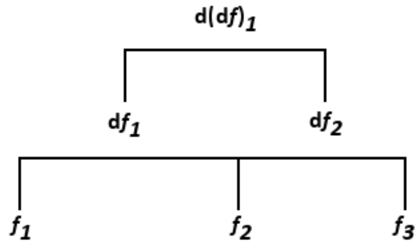

Figure 2: The second-order differential as a change of differentials

and then consider the change in $\mathrm{d}f$ between these two differences (see Fig. 2). This gives us the change in the differential, $\mathrm{dd}f$, or $\mathrm{d}^2 f$. Since this is a second-order differential, it is infinitesimal on the scale of first-order differentials, and so will be proportional to $\mathrm{d}x^2$, rather than $\mathrm{d}x$, and we must divide through by $\mathrm{d}x^2$ to find the second derivative.