Chapter 5
Derivatives and Differentiation

Big fleas have little fleas upon their backs to bite 'em, And little fleas have lesser fleas, and so, ad infinitum.

Augustus de Morgan

There's a problem with continuity. Suppose that f is continuous at a point x_0. Suppose we want to compute $f(x_0)$ with an error less than ε, for example $\varepsilon = 10^{-5}$, but we do not know x_0 exactly. We know that there exists δ, such that if $|x - x_0| < \delta$ then $|f(x) - f(x_0)| < 10^{-5}$. We do not therefore have to know x_0 exactly; a certain number of decimal places will suffice.

But what if δ needed to be uncomfortably small compared to ε in order to achieve the desired accuracy? What, for example, if δ was 10^{-10}, or 10^{-100} or even less....? The function may be continuous but continuity does not seem so useful here.

The problem is that f could be increasing or decreasing very rapidly at the point x_0. But what does that mean—the rate of increase or decrease of a function at a point?

The concept of the rate of growth of a function at a point is the key to the calculus of Newton and Leibniz and is what we call the derivative. As soon as it was introduced it became possible to solve important problems in geometry and physics with the new calculus, in spite of the fact that an acceptable definition of derivative was not given for some 200 years.

5.1 The Definition of Derivative

The average rate of growth of a function f between distinct points x_0 and x is the quotient

$$\frac{f(x) - f(x_0)}{x - x_0}.$$

R. Magnus, *Fundamental Mathematical Analysis*, Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-030-46321-2_5

The rate of growth at the point x_0 is defined as the limit.

Definition Let $f :]a, b[\to \mathbb{R}$ and $a < x_0 < b$. If the limit

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = A$$

exists and is a finite number we say that f is differentiable at x_0 and call A, *the derivative of f at* x_0.

We emphasise that A is a finite number; if the limit is ∞ or $-\infty$ we may sometimes say that the derivative is ∞ or $-\infty$ respectively, but we will never say that f is differentiable at x_0.

Another version of the definition of derivative, that arises by replacing $x - x_0$ by h, is

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h},$$

provided that the limit exists and is a finite number. The quotient appearing here is called a difference quotient. It is defined for both positive and negative values of h, though not for $h = 0$, but $|h|$ should not be so big that $x_0 + h$ falls outside the domain of f.

If f is differentiable at x_0 we denote its derivative at x_0 by $f'(x_0)$. We say that the function f is differentiable in the interval $]a, b[$ if f is differentiable at every point of the interval.

The definition of derivative follows a pattern that we have set in defining limit and the sum of an infinite series, and will continue in defining integral. The quantity in question that we wish to define does not necessarily exist. The definition of the quantity states when it exists, and given that it exists defines its value. Just as it is illogical to write $\lim_{x \to a} f(x)$ without first ascertaining whether the limit exists (though we often do this), we should not write $f'(c)$ without first ascertaining whether f is differentiable at c.

If f is differentiable in the interval $]a, b[$ we get a new function

$$f' :]a, b[\to \mathbb{R}, \quad f'(x) = \text{ derivative of } f \text{ at } x.$$

The operation of creating f' from f is called differentiation of the function f.

5.1.1 Differentiability and Continuity

Proposition 5.1 *Let the function f be differentiable at the point x_0. Then f is continuous at x_0.*

Proof We have, for $x \neq x_0$,

$$f(x) - f(x_0) = \left(\frac{f(x) - f(x_0)}{x - x_0} \right) . (x - x_0)$$

so that, by the rule for the limit of a product,

$$\lim_{x \to x_0} (f(x) - f(x_0)) = \lim_{x \to x_0} \left(\frac{f(x) - f(x_0)}{x - x_0} \right) . \lim_{x \to x_0} (x - x_0) = f'(x_0) . 0 = 0.$$

In other words

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

which says that f is continuous at x_0. □

Continuity is therefore necessary for differentiability, but it is far from being sufficient.

5.1.2 Derivatives of Some Basic Functions

Now we can begin to differentiate functions *from first principles*, that is, by applying the definition of derivative as the limit of the difference quotient.

(a) Let f be the constant function, $f(x) = C$ for all real numbers x. Then

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

and so $f'(x) = 0$.

(b) Next let f be the so-called identity function, defined by $f(x) = x$ for all real numbers x. Then

$$\frac{f(x+h) - f(x)}{h} = \frac{x + h - x}{h} = 1$$

and so $f'(x) = 1$.

(c) Next let f be the function $f(x) = x^2$. Then

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

and so, by the rule for the limit of a sum,

$$f'(x) = \lim_{h \to 0} (2x + h) = 2x.$$

We could go on, but it is far better to use the differentiation rules, as set out in the next paragraphs. These allow one to differentiate without considering difference quotients and limits. They make differentiation an almost mechanical procedure, and are of immense practical and historical importance. Without them there would be no calculus justifying the name.

5.1.3 Exercises

1. Differentiate from first principles, that is, using the definition of derivative as limit of the difference quotient:

   (a) $ax + b$, where a and b are constants.
   (b) x^3
   (c) x^n (n a natural number)
   (d) $\sqrt{x}$
   (e) $\sqrt[3]{x}$.

   Hint. Use algebraic properties of these functions. The only analytic input needed is their continuity.

2. Differentiate the circular functions $\sin x$ and $\cos x$ from first principles (that is, by calculating the limit of the difference quotient). You will need algebraic input in the form of the addition formulas

$$\sin(u + v) = \sin u \cos v + \cos u \sin v,$$
$$\cos(u + v) = \cos u \cos v - \sin u \sin v,$$

   and two facts of analysis to be proved later: the continuity of both functions, and the limit

$$\lim_{x\to 0} \frac{\sin x}{x} = 1.$$

   Note. The circular functions will be defined analytically in a later chapter. The reader has doubtlessly been introduced to them through school mathematics, in which it is usual to obtain the addition formulas by geometry and the limit of $\sin x/x$ by geometric intuition.

3. Differentiate the exponential function e^x from first principles. You will need the algebraic input that e^x satisfies the first law of exponents:

$$e^{x+y} = e^x . e^y,$$

   and the analytic input that

$$\lim_{x\to 0} \frac{e^x - 1}{x} = 1,$$

   equivalent to giving the derivative of e^x at $x = 0$; this essentially pins down the special base e.

Note. Just like the circular functions the exponential function and its inverse the natural logarithm will be defined analytically in a later chapter. We do not have to take for granted the existence of a function with these properties.

4. An exponential function a^x can be defined for any positive base a. It satisfies the law of exponents $a^{x+y} = a^x a^y$. For the sake of this exercise we shall adopt the notation $E_a(x) = a^x$. Assuming that E_a is differentiable derive the formula

$$E_a'(x) = k_a E_a(x)$$

where $k_a = E_a'(0)$.
Note. The special base e could be defined as the number that satisfies $k_e = 1$, though there would be difficulties involved—for example, why does such a number exist and why is it unique? Compare the previous exercise.

5. The natural logarithm $\ln x$ is the inverse function to the exponential function (Exercise 3) and from it $\ln x$ inherits the law of logarithms: $\ln(xy) = \ln x + \ln y$. Differentiate $\ln x$ from first principles.
Hint. You may need to figure out first why $\lim_{h\to 0} \ln(1+h)/h = 1$.
6. Let $f(x) = |x|$. Show that f is differentiable at all points except $x = 0$. Show that $f'(x) = x/|x|$ if $x \neq 0$.
7. Let a_1, a_2,..., a_n be a strictly increasing sequence of real numbers. Let $f(x) = \sum_{j=1}^{n} |x - a_j|$ for each real x.

   (a) Show that f is continuous at every point x, whereas it is differentiable everywhere except at the points a_j, ($j = 1, ..., n$).
   (b) Show that the derivative is constant in each of the open intervals $]a_k, a_{k+1}[$, as well as in $]-\infty, a_1[$ and in $]a_n, \infty[$, and find a formula for it.
   (c) Sketch the graphs in the cases

$$y = |x+1| + |x| + |x-1|$$

   and

$$y = |x+2| + |x+1| + |x-1| + |x-2|.$$

8. Let f be the function with domain $\mathbb{R}$ defined by letting $f(x) = x$ if x is rational and $f(x) = 0$ if x is irrational.

   (a) Are there any points at which f is differentiable?
   (b) Are there any points at which the function $g(x) := xf(x)$ is differentiable?

9. Let $f :]0, 1[\to \mathbb{R}$ be the function defined in Sect. 4.2, Exercise 18. Recall that $f(x) = 0$ if x is irrational and $f(x) = 1/b$ if x is the fraction a/b expressed in lowest terms. Show that f is nowhere differentiable.
Hint. Show that if x is irrational then there exist arbitrarily small h such that $\left|\big(f(x+h) - f(x)\big)/h\right| > 1$.

5.2 Differentiation Rules

The elementary differentiation rules put the calculus into analysis. There are two groups of rules. The first deals with functions constructed by algebraic operations, addition, multiplication and division, from other functions. The second comprises the rule for differentiating composite functions (the chain rule) and the rule for differentiating inverse functions.

Let $f :]a, b[\to \mathbb{R}$, $g :]a, b[\to \mathbb{R}$. Sum, product and quotient of functions are defined pointwise:

$$(f+g)(x) = f(x) + g(x), \quad (fg)(x) = f(x)g(x), \quad \left(\frac{f}{g}\right)(x) = \frac{f(x)}{g(x)}.$$

Take care not to confuse product fg and composition $f \circ g$.

We state the first group of differentiation rules in the following lengthy proposition. In the proofs we rely entirely on the limit rules of Sect. 4.2; we never have to say "Let $\varepsilon > 0$".

Proposition 5.2 *Let $f :]a, b[\to \mathbb{R}$, $g :]a, b[\to \mathbb{R}$. Let c be in the interval $]a, b[$ and assume that both f and g are differentiable at the point c. Let α be a numerical constant. Then αf, $f + g$ and fg are differentiable at c and we have*

(1) $(\alpha f)'(c) = \alpha f'(c)$ *(Multiplication by a constant)*
(2) $(f+g)'(c) = f'(c) + g'(c)$ *(Sum of functions)*
(3) $(fg)'(c) = f'(c)g(c) + f(c)g'(c)$ *(Product of functions; Leibniz's rule).*

If moreover $g(c) \neq 0$, then $1/g$ and f/g are differentiable at c, and we have the further rules:

(4) $\left(\dfrac{f}{g}\right)'(c) = \dfrac{g(c)f'(c) - g'(c)f(c)}{(g(c))^2}$ *(Quotient rule)*
(5) $\left(\dfrac{1}{g}\right)'(c) = -\dfrac{g'(c)}{(g(c))^2}$ *(Reciprocal rule).*

Proof The rule for the derivative of αf is a special case of the rule for fg and that for $1/g$ a special case of that for f/g (left to the reader to see why).

Now for the proofs of rules 2, 3 and 4. Firstly the sum. We examine the difference quotient:

$$\begin{aligned}\frac{(f+g)(c+h) - (f+g)(c)}{h} &= \frac{f(c+h) - f(c) + g(c+h) - g(c)}{h} \\ &= \frac{f(c+h) - f(c)}{h} + \frac{g(c+h) - g(c)}{h}\end{aligned}$$

and taking the limit we obtain the rule $(f+g)'(c) = f'(c) + g'(c)$.

Secondly the product. We transform the difference quotient:

$$\begin{aligned}\frac{(fg)(c+h)-(fg)(c)}{h} &= \frac{f(c+h)g(c+h)-f(c)g(c)}{h}\\ &= \frac{f(c+h)g(c+h)-f(c)g(c+h)+f(c)g(c+h)-f(c)g(c)}{h}\\ &= \frac{f(c+h)-f(c)}{h}g(c+h)+f(c)\frac{g(c+h)-g(c)}{h}.\end{aligned}$$

We take the limit as $h \to 0$, using the rules for the limits of sums and products, and remembering that $\lim_{h\to 0} g(c+h) = g(c)$ since g, being differentiable, is continuous at c. Thus we obtain the rule for the product, $(fg)'(c) = f'(c)g(c) + f(c)g'(c)$.

Next the quotient. Again we transform the difference quotient by algebra:

$$\begin{aligned}\frac{\left(\frac{f}{g}\right)(c+h)-\left(\frac{f}{g}\right)(c)}{h} &= \frac{1}{h}\left(\frac{f(c+h)}{g(c+h)}-\frac{f(c)}{g(c)}\right)\\ &= \frac{1}{h}\left(\frac{f(c+h)g(c)-f(c)g(c+h)}{g(c+h)g(c)}\right)\\ &= \frac{\left(\frac{f(c+h)-f(c)}{h}\right)g(c)-f(c)\left(\frac{g(c+h)-g(c)}{h}\right)}{g(c+h)g(c)}.\end{aligned}$$

We let $h \to 0$, use the rules for limits of sums, products and quotients, remember that $\lim_{h\to 0} g(c+h) = g(c)$ and obtain the limit

$$\frac{g(c)f'(c)-g'(c)f(c)}{(g(c))^2}.$$

□

5.2.1 Differentiation of the Power Function

If n is a positive integer and f is the function x^n then we have

$$f'(x) = nx^{n-1}.$$

This is now easy to prove without considering the limit of a difference quotient. We use induction. The rule is known for $n = 1$. Let us assume it holds for a particular integer n and write $x^{n+1} = x \cdot x^n$. Using the rule for differentiating a product we obtain for the derivative of x^{n+1} the formula

$$1 \cdot x^n + x \cdot nx^{n-1} = (n+1)x^n.$$

This proves the rule generally for positive integers n.

Next we consider $f(x) = x^{-n} = 1/x^n$. The rule for the derivative of a quotient gives the formula

$$f'(x) = -\frac{nx^{n-1}}{x^{2n}} = -nx^{-n-1}.$$

So we have now shown that the derivative of x^a is ax^{a-1} in all cases when a is an integer (positive or negative).

What about the power $x^{1/n}$, which denotes the n^{th} root $\sqrt[n]{x}$, or the fractional power $x^{m/n} = \sqrt[n]{x^m}$? For these we need the rule for differentiating inverse functions, and the celebrated chain rule, often referred to somewhat misleadingly as the rule for functions of a function. The latter rule, which we take first, is used to differentiate composite functions and is perhaps the most remarkable of the differentiation rules.

5.2.2 The Chain Rule

Proposition 5.3 *Let $f : A \to \mathbb{R}$, $g : B \to \mathbb{R}$, where A and B are open intervals and $f(A) \subset B$. Form the composition $g \circ f : A \to \mathbb{R}$,*

$$(g \circ f)(x) = g(f(x)), \quad (x \in A).$$

Let $x_0 \in A$, assume that f is differentiable at x_0 and g is differentiable at $f(x_0)$. Then $g \circ f$ is differentiable at x_0 and

$$(g \circ f)'(x_0) = g'((f(x_0))f'(x_0).$$

The chain rule is illustrated in Fig. 5.1

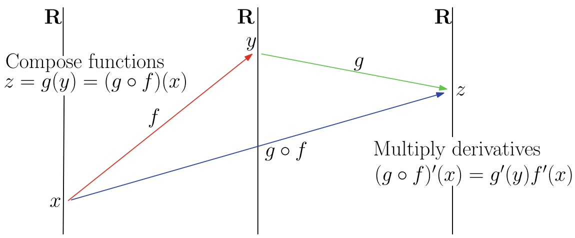

Fig. 5.1 A view of the chain rule

Proof of the Chain Rule Let $y_0 = f(x_0)$. As for the previous rules we start by applying some algebra to the difference quotient:

$$\frac{(g \circ f)(x_0+h) - (g \circ f)(x_0)}{h} = \frac{g\big(f(x_0+h)\big) - g\big(f(x_0)\big)}{f(x_0+h) - f(x_0)} \cdot \frac{f(x_0+h) - f(x_0)}{h}. \tag{5.1}$$

The second factor on the right-hand side has the limit $f'(x_0)$. As for the first factor it looks as if it should have the correct limit $g'(y_0)$. For we can think of $f(x_0+h)$ as $y_0 + k$ (effectively defining the new quantity k) and then the first factor is the quotient

$$\frac{g(y_0+k) - y_0}{k}.$$

As $h \to 0$ we have $k \to 0$ also and we seem to have a proof.

But there is a problem here. Although h is not 0 (as befits a correctly formed difference quotient) the denominator k, defined to be the difference $f(x_0+h) - f(x_0)$, can be 0, and the first factor is then not defined for such values of h. There could even exist such values of h that are arbitrarily small which are then impossible to escape.

To save the proof we shall define a function R, the domain of which is a suitably small interval $]-\alpha, \alpha[$, in such a way that formula (5.1) for the difference quotient is correct if $R\big(f(x_0+h) - f(x_0)\big)$ replaces the first factor.

For $\alpha > 0$ and suitably small (the reader should try to figure out what "suitably small" means in this context and why we have to say it) we set

$$R(t) = \begin{cases} \dfrac{g(y_0+t) - g(y_0)}{t} & \text{if } 0 < |t| < \alpha \\ g'(y_0) & \text{if } t = 0. \end{cases}$$

Note that R is continuous at the point $t = 0$ because

$$\lim_{t \to 0} \frac{g(y_0+t) - g(y_0)}{t} = g'(y_0).$$

Moreover

$$g(y_0+t) - g(y_0) = R(t)t$$

both when $t \neq 0$ and when $t = 0$. In this equation we replace t by the difference $f(x_0+h) - f(x_0)$. This is allowed if $|h|$ is sufficiently small and then we have

$$g\big(f(x_0+h)\big) - g\big(f(x_0)\big) = R\big(f(x_0+h) - f(x_0)\big)\big(f(x_0+h) - f(x_0)\big).$$

Division by h when the latter is not 0, but still sufficiently small, gives

$$\frac{g\big(f(x_0+h)\big)-g\big(f(x_0)\big)}{h}=R\big(f(x_0+h)-f(x_0)\big)\left(\frac{f(x_0+h)-f(x_0)}{h}\right).$$

Now we may let h tend to 0 and by the limit rules the right-hand side has the limit

$$\Big(\lim_{h\to 0}R\big(f(x_0+h)-f(x_0)\big)\Big)\,f'(x_0)=R(0)f'(x_0)=g'\big(f(x_0)\big)f'(x_0).$$

In slightly more detail (we seriously want *this* proof to be correct) we can introduce the function $\phi(h):=f(x_0+h)-f(x_0)$. Then ϕ is continuous at $h=0$, and $\phi(0)=0$. Moreover the function R is continuous at 0 as we saw. Hence the composition $R\circ\phi$ is continuous at 0 and so $\lim_{h\to 0}R(\phi(h))=R(\phi(0))=R(0)$, as we wrote above. □

5.2.3 Differentiation of Inverse Functions

This is the last of the elementary differentiation rules. The lengthy preamble repeats the conditions (see Proposition 4.12) under which the inverse function exists and should not distract the reader from the extraordinary simplicity of the formula that is the conclusion.

Proposition 5.4

Preamble. *Let* $f:]a,b[\to\mathbb{R}$ *be continuous and strictly increasing (the point a may be* $-\infty$ *and b may be* ∞*). Let* $c=\lim_{x\to a+}f(x)$ *and* $d=\lim_{x\to b-}f(x)$ *(the limits exist if we allow* $c=-\infty$ *and* $d=\infty$*). The inverse function* $g:]c,d[\to\mathbb{R}$ *therefore exists, is continuous, and maps the interval* $]c,d[$ *on to the interval* $]a,b[$.

Conclusion. *Let* $a<x_0<b$ *and assume that* f *is differentiable at* x_0*, and that* $f'(x_0)\neq 0$*. Then g is differentiable at* $f(x_0)$ *and*

$$g'(f(x_0))=\frac{1}{f'(x_0)}.$$

A similar conclusion holds if f is strictly decreasing; the only difference is that $d<c$ *and g has the domain* $]d,c[$.

Proof Let $y_0=f(x_0)$. We have to show that $g'(y_0)=f'(x_0)^{-1}$. Connect the variables h and k by the equation

$$y_0+k=f(x_0+h),\quad \text{equivalently}\quad h=g(y_0+k)-g(y_0)$$

(recall that $g(y_0)=x_0$). The second equation here shows h as a function of k; it is a continuous, injective function of k, defined when k is sufficiently small. Moreover $h=0$ when $k=0$.

We also have

$$\frac{g(y_0+k)-g(y_0)}{k}=\frac{h}{k}=\frac{h}{f(x_0+h)-f(x_0)}.$$

Let $k \to 0$ and think of h as a function of k, as defined above. Then h tends to 0, but is not 0 as long as $k \neq 0$. By the rule for the limit of a reciprocal, the right-hand side has the limit $f(x_0)^{-1}$. □

Assuming that $f'(x) \neq 0$ for all x in the interval $]a, b[$, we have the conclusion that

$$(f^{-1})'(y)=\frac{1}{f'(f^{-1}(y))} \tag{5.2}$$

for all y in the interval $]c, d[$ (or in $]d, c[$ if f is decreasing).

We shall see later (Sect. 5.6) that if $f'(x) > 0$ for all x in the open interval A, then f is strictly increasing in A, so that Proposition 5.4 is immediately applicable.

5.2.4 *Differentiation of Fractional Powers*

Let $f(x) = x^{1/n}$, where n is a positive natural number. We have here the inverse function of the function $g(x) = x^n$. The domain is the interval $]0, \infty[$. By the rule for differentiating an inverse function (that is, we apply (5.2) to the function g with x instead of y) we have

$$f'(x)=(g^{-1})'(x)=\frac{1}{g'(g^{-1}(x))}=\frac{1}{n(x^{\frac{1}{n}})^{n-1}}=n^{-1}x^{\frac{1-n}{n}}=\frac{1}{n}x^{\frac{1}{n}-1}.$$

Next we consider the function $f(x) = x^{m/n}$, where m is an integer, positive or negative. This is the composition $(x^m)^{1/n}$. By the chain rule we have

$$f'(x)=\frac{1}{n}(x^m)^{\frac{1}{n}-1}mx^{m-1}=\frac{m}{n}x^{\frac{m}{n}-1}.$$

The conclusion is striking. The derivative of the power function x^a is ax^{a-1} for every rational power a.

It is a further task to define the power function x^a for irrational powers and prove that the same differentiation formula continues to be valid.

5.2.5 *Exercises*

1. Differentiate the following functions. You may assume that the domain of each function is the set of all x for which the formula makes sense.

   (a) $\dfrac{1}{x^2+2}$

(b) $\dfrac{x^2 - x + 1}{x^2 + x - 1}$

(c) $\sqrt{\dfrac{x^2 - x + 1}{x^2 + x - 1}}$

(d) $\sqrt[4]{\dfrac{x^2 - x + 1}{x^2 + x - 1}}$

(e) $\sqrt{1 + \sqrt{x}}$

(f) $\sqrt{1 + \sqrt{1 + \sqrt{x}}}$

(g) $\sqrt{1 + \sqrt{1 + \sqrt{1 + \sqrt{x}}}}.$

2. Define the function f on the whole real line by

$$f(x) = \begin{cases} x, & (x \le 0) \\ 1 - 2\sqrt{1-x}, & (0 < x < 1) \\ x, & (1 \le x \le 2) \\ \sqrt{x^2+5} - 1, & (2 < x). \end{cases}$$

Determine where f is differentiable, and where it is, find its derivative.
Hint. In this, and in similar examples where a function is defined by cases, the differentiation rules are only useful in the open intervals between the partition points. At the partition points something else is required, such as arguing by examination of the difference quotient.

3. For this exercise we assume some knowledge of the circular functions $\sin x$ and $\cos x$, including their derivatives (see Sect. 5.1, Exercise 2). Determine where the following functions are differentiable and calculate the derivative when it exists:

(a) $f(x) = \begin{cases} \sin \frac{1}{x}, & (x > 0) \\ 0, & (x \le 0). \end{cases}$

(b) $f(x) = \begin{cases} x \sin \frac{1}{x}, & (x > 0) \\ 0, & (x \le 0). \end{cases}$

(c) $f(x) = \begin{cases} x^2 \sin \frac{1}{x}, & (x > 0) \\ 0, & (x \le 0). \end{cases}$

4. Show that the function $f(x) = x^5 + x$ is strictly increasing on the whole real line and calculate $(f^{-1})'(2)$.

5. Let f_k, $(k = 1, 2, ..., n)$ be differentiable functions. Let g be their product, $f_1 f_2 ... f_n$. Show that

$$\frac{g'(x)}{g(x)} = \sum_{k=1}^{n} \frac{f_k'(x)}{f_k(x)}$$

at every point x at which none of the denominators is 0.

5.3 Leibniz's Notation

There are several notational systems in use for derivatives. They reflect the differing views of Newton and of Leibniz. Newton used dots to signify the derivative, as in $\dot{x}$ and $\dot{y}$. One might say that the various dashes, as in f' and f'', popularised by Lagrange, reflect Newton's notation. Leibniz introduced the expressions dx and dy, signifying in his view infinitesimal changes in the variables x and y ("*d*" for Latin "*differentia*"), and leading to the differential quotient dy/dx. He also introduced the integral sign "$\int$", an elongated "*S*" (for Latin "*Summa*"). Each notation has its advantages and it is best to learn how to use both.

5.3.1 Tangent Lines

We often think of a function f as a curve in the (x, y)-plane. The curve in question is the set of all points (x, y) that satisfy $y = f(x)$, in other words the graph of f. Leibniz's notation reflects the geometric intuition behind the idea of a tangent line to a curve.

A line is a curve of the form $y = mx + c$ with constants m (the slope) and c (the intercept). The curve $x = a$ is also a line but it is not a graph of a function in the above sense. It is though a graph if we think of x as a function of y (in this case a constant function).

We could ask whether every curve in the plane can be described (perhaps locally; in small sections at a time) as a graph, in which y is a function of x, or x is a function of y. The question arises even for familiar everyday curves like the circle and shows the limitation of thinking of a curve simply as a graph. This gets us into the area of differential geometry. We would have to give a general definition of curve, a task that is not so straightforward.

The ancient Greek geometers tried to define a tangent line to a curve as a line that meets the curve in only one point. This works for circles (and more generally conic sections) but not for more complicated curves. Differential calculus allows us to give a correct definition of tangent line to a curve when that curve is a graph $y = f(x)$, and its extension to differential geometry does the job for more general curves. For this reason it is said that differentiation solved the problem of tangents.

Consider a differentiable function f. The tangent line to the curve $y = f(x)$, at a point (x_0, y_0) on the graph (that this point lies on the graph means that $y_0 = f(x_0)$), is the line through the point (x_0, y_0) that has the slope $f'(x_0)$. In other words it is the line

$$y - y_0 = f'(x_0)(x - x_0)$$

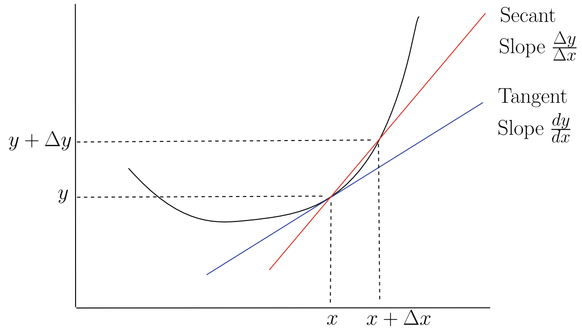

Fig. 5.2 Leibniz's differential quotient

or equivalently

$$y = f'(x_0)x + (y_0 - f'(x_0)x_0).$$

The intuitive thinking behind this is that the tangent line at (x_0, y_0) is the limit of a secant line through the two points, (x_0, y_0) and $(x_0 + \Delta x, y_0 + \Delta y)$, both on the graph $y = f(x)$, the limit being taken as $\Delta x \to 0$. The slope of the secant line is $\Delta y/\Delta x$ and we want to make Δx, and as a result Δy, tend to 0.

In the view of seventeenth century mathematicians, who did not possess a definition of limit, the quantity Δx was actually supposed to become infinitely small, the tangent being thought to intersect the curve at two distinct points infinitely close together. For the slope of the tangent we obtain a quotient of infinitely small quantities, or infinitesimals. This intuition lies behind Leibniz's notation for derivatives (Fig. 5.2).

5.3.2 *Differential Quotients*

Leibniz proposed setting an infinitesimal dx in place of Δx, as the notion of limit was not available to him. He would have said that y underwent a corresponding change, which was also an infinitesimal dy, and the derivative was the quotient dy/dx. Although dx and dy are infinitesimals (whatever that means) the quotient is an ordinary real number. He called the infinitesimals dx and dy differentials. The derivative was then the *differential quotient*.

According to the prevailing modern view the derivative is not a quotient; it is though the limit of a quotient, namely the limit of the difference quotient. In spite of this it is possible to define differentials, expressed in the classical notation dx and dy, without resorting to the mysterious infinitesimals. This is very useful for calculus in several variables and differential geometry of surfaces and their generalisations, manifolds. It means, for example, that classical formulas, such as $dy = f'(x)\,dx$, remain valid with an appropriate interpretation of their symbols. However, that is a whole new topic.[1]

Here are some examples of statements written using Leibniz's notation. It will be seen that they have certain advantages, notably brevity and flexibility, over their equivalents using function symbols:

(a) If $y = x^3$ then $\dfrac{dy}{dx} = 3x^2$.

That is, if f is the function $f(x) = x^3$ then $f'(x) = 3x^2$.

(b) $\dfrac{d}{dx}x^3 = 3x^2$.

Same meaning as the previous item. We again avoid using a symbol for the function, as well as mentioning the variable y.

(c) $\dfrac{d}{dx}x^3\Big|_{x=1} = 3$.

In other words if f is the function $f(x) = x^3$ then $f'(1) = 3$. The vertical stroke with the subscript "$x = 1$" means evaluate the preceding expression at $x = 1$.

5.3.3 The Chain Rule and Inverse Functions in Leibniz's Notation

Many calculations using the chain rule or the inverse-function rule are easier to carry out using Leibniz's notation. This makes it particularly useful for effecting a change of variables in a differential equation, a subject not covered in the present text.

Functions f and g are given and we wish to differentiate the composed function $g \circ f$. We consider that the function f sets up a relation between variables x and y, namely $y = f(x)$, whilst g sets up a relation between variables y and z, namely $z = g(y)$. Then the composition $g \circ f$ sets up the relation $z = (g \circ f)(x)$.

We can differentiate the composition $g \circ f$ using the chain rule. In Leibniz's notation we are finding the differential quotient dz/dx and this is given by the striking formula

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}.$$

This is of course just the formula

$$(g \circ f)'(x) = g'(f(x))f'(x).$$

[1]This has nothing to do with what is known as non-standard analysis. In the latter the real number system is extended by including infinitely small quantities and infinitely large quantities.