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.

The first factor on the right-hand side, that is dz/dy, must be interpreted with some care. We first differentiate z with respect to y, but then express this as a function of x, using the relationship between y and x.

We illustrate these steps by differentiating $\sqrt{1-x^2}$. We set $y = 1 - x^2$ and $z = \sqrt{y}$. Then

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} = \frac{1}{2\sqrt{y}}(-2x) = -\frac{x}{\sqrt{1-x^2}}.$$

Consider next inverse functions. If y is a function of x, namely $y = f(x)$, we can turn this round and look at x as a function of y, namely $x = f^{-1}(y)$. The rule for differentiating f^{-1} takes the memorable form

$$\frac{dx}{dy} = 1\bigg/\frac{dy}{dx}.$$

This is the same formula as the less intuitive

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

As an example we shall differentiate the function $x^{1/n}$. Let $y = x^{1/n}$ and turn it around giving $x = y^n$. Then

$$\frac{dx}{dy} = ny^{n-1}$$

so that by the rule we find

$$\frac{dy}{dx} = \frac{1}{ny^{n-1}} = \frac{1}{nx^{\frac{n-1}{n}}} = \frac{1}{n}x^{\frac{1}{n}-1}.$$

5.3.4 Tangents to Plane Curves

In analytic geometry, the simplest way to represent a circle with centre (a, b) and radius r is by means of the equation $(x-a)^2 + (y-b)^2 = r^2$. Here the curve is not seen as a graph; in order to do so we must solve for y as a function of x, or for x as a function of y. To represent a curve in analytic geometry as a graph, we usually have to break it into pieces.

A simple example is that of the unit circle $x^2 + y^2 = 1$. Solving for y we obtain two solutions, and two graphs:

$$y = \sqrt{1-x^2}, \quad (-1 \le x \le 1) \quad \text{the upper semicircle}$$

$$y = -\sqrt{1-x^2}, \quad (-1 \le x \le 1) \quad \text{the lower semicircle.}$$

Now we can differentiate these formulas in order to compute the tangents to the circle, using the appropriate formula for each semicircle.

However, there is another way to calculate the tangent at a point (x_0, y_0) on the curve without solving for y as a function of x. Suppose that we are looking at a part of the circle that can be represented as a graph $y = f(x)$, where f is differentiable, and contains the point (x_0, y_0). Then $y_0 = f(x_0)$ and the equation

$$x^2 + \big(f(x)\big)^2 = 1$$

holds for all x in some interval containing x_0. We may differentiate with respect to x, using the differentiation rules, and obtain

$$2x + 2f(x)f'(x) = 0.$$

In particular $f'(x_0) = -x_0/f(x_0) = -x_0/y_0$.

The calculation just given would normally be done without introducing a function symbol, using Leibniz's notation

$$x^2 + y^2 = 1 \;\Rightarrow\; 2x + 2y\frac{dy}{dx} = 0 \;\Rightarrow\; \frac{dy}{dx} = -\frac{x}{y}$$

or else a form of Newton's notation

$$x^2 + y^2 = 1 \;\Rightarrow\; 2x + 2yy' = 0 \;\Rightarrow\; y' = -\frac{x}{y}.$$

This procedure is called implicit differentiation. The differentiation proceeds with respect to x, but y is thought of as a function of x, the exact form of which is not required. We obtain dy/dx, and express it as a function of x and y, without knowing the function $y = f(x)$. Logically, we only need to know that the function $f(x)$ exists, and is differentiable. This can usually be guaranteed by a theorem of multivariate calculus, the implicit function theorem, which is beyond the scope of this text.

Example The equation $2y^5 - xy - x^4 = 0$ defines some kind of curve in the coordinate plane. We observe that it contains the point $(1, 1)$. To solve for y as a function of x, or for x as a function of y, is difficult (although some algebraic arguments show that there is a unique positive y for each positive x; see the nugget "Multiplicity"). Nevertheless, we can calculate the tangent to the curve at the point $(1, 1)$. Assuming that we can represent the curve around the point $(1, 1)$ as a graph $y = f(x)$ with differentiable f (a fact that can be justified using the implicit function theorem), implicit differentiation gives

$$10y^4y' - y - xy' - 4x^3 = 0 \;\Rightarrow\; y' = \frac{y + 4x^3}{10y^4 - x}$$

and therefore at the point $(1, 1)$ we have $y' = \frac{5}{9}$. Note how differentiating the middle term xy gives rise to $y + xy'$ because we are thinking of y as a function of x. The equation of the tangent line is therefore $y - 1 = \frac{5}{9}(x - 1)$, or more simply, $5x - 9y + 4 = 0$.

5.3.5 Exercises

1. (a) Determine the equation of the tangent to the parabola $y = x^2$ at the point (t, t^2).
   (b) Show that the line perpendicular to the tangent of item (a) and intersecting it on the x-axis, passes through the point $(0, \frac{1}{4})$, independently of t.
2. Show that the equation of the tangent to the ellipse

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

   at the point (x_0, y_0), assumed to be on the ellipse, is

$$\frac{x_0 x}{a^2} + \frac{y_0 y}{b^2} = 1.$$

3. Give an example of a graph $y = f(x)$ (with differentiable f) and a point $(a, f(a))$ on the graph, such that the tangent at $(a, f(a))$ crosses the graph at $(a, f(a))$.
4. A vessel has the shape of a right circular cone standing on its apex. Let h be the height of the cone and let r be the radius of its base. Mercury is poured into the vessel, not necessarily at a constant rate. Introduce variables: t for the time, v for the volume of mercury in the vessel and y for the height of the mercury in the vessel.
   (a) Find the relationship between $\dfrac{dv}{dt}$ and $\dfrac{dy}{dt}$.
   (b) Suppose that $h = 1$ m, $r = 1$ m, $y = 0.5$ m and the mercury is poured at a constant rate of 1 litre per second. Approximately, how much time is needed to raise the surface level by 1 cm?
   Note. Physics and engineering abound with problems like this one. A bunch of variables are connected by a constitutive relation. In this problem the relation between v and y is geometric. Examples from physics are pressure, volume and temperature connected by the ideal gas equation; or stress and strain connected by the law of elasticity. If the variables change with time, then the constitutive relation implies a linear connection between their derivatives with respect to time. If the variables are three or more then the problem really requires multivariate calculus, in particular partial derivatives. With two variables we can just about get by without them.

5.4 Higher Order Derivatives

If the function $f :]a, b[\to \mathbb{R}$ is differentiable, its derivative $f' :]a, b[\to \mathbb{R}$ is a new function. Now it could happen that f' is differentiable. If so, we can differentiate and produce the function $f'' :]a, b[\to \mathbb{R}$, called the second derivative of f. Continuing in this way as far as is allowable, we can define a whole sequence: second, third, fourth, fifth, ..., n^{th} derivatives of f. They are denoted by f'', f''', f''', f'''' and so on, but around the fourth it becomes more practical to write instead $f^{(4)}$, $f^{(5)}$, ..., $f^{(n)}$... as counting those little dashes becomes tiresome and irritating. When using this notation it is often convenient to allow $n = 0$ and interpret $f^{(0)}$ to be the same as f.

The differentiation can be continued beyond $f^{(n)}$ when the latter is differentiable on the interval $]a, b[$, where f was defined. It could happen that every function produced in this way is differentiable. Then we say that f is infinitely often differentiable. If the process can be continued at least as far as $f^{(n)}$ we say that f is n-times differentiable, or that f is differentiable to order n, or that f has derivatives to order n (none of which precludes going further).

Can we give any sense to the statement that f is n-times differentiable at the point c? For a function to be differentiable at a given point it must be defined on an interval that contains that point. Therefore the meaning to be attached to this phrase is the following. There exists $\delta > 0$, such that f is $(n-1)$-times differentiable in the interval $]c - \delta, c + \delta[$ and $f^{(n-1)}$ is differentiable at c. We sometimes say in this case that the derivatives $f^{(k)}(c)$ exist up to $k = n$; or, most briefly: f has n derivatives at c.

Leibniz's notation for the higher derivatives is

$$y = f(x), \quad \frac{dy}{dx} = f'(x), \quad \frac{d^2y}{dx^2} = f''(x), \quad \dots \quad \frac{d^my}{dx^m} = f^{(m)}(x).$$

5.4.1 Exercises

1. Let f be a polynomial of degree m. Show that $f^{(k)} = 0$ for all $k > m$.
2. Let g be a function having derivatives of all orders and let a be a real number, such that $g(a) \neq 0$. Set $f(x) = (x - a)^m g(x)$, where m is a positive integer. Show that $f^{(k)}(a) = 0$ for $k = 0, 1, ..., m - 1$, but that $f^{(m)}(a) \neq 0$.
3. Show that

$$\frac{d^k}{dx^k} x^a = a(a-1)...(a-k+1)x^{a-k}.$$

Here you may assume that a is rational (pending the rigorous definition of irrational powers in Chap. 7). Also you may assume that $x > 0$ if a is not an integer. If a is a positive integer show that

$$\frac{d^k}{dx^k}x^a = \frac{a!}{(a-k)!}x^{a-k}$$

for $k \le a$. One can even allow $k > a$ if we interpret $1/m!$ as 0 if m is a negative integer.

4. We assume some knowledge of the exponential function e^x, namely, that its derivative is again e^x. Let $f(x) = e^{-1/x}$ for $x \neq 0$. Show that for all natural numbers n we have

$$f^{(n)}(x) = P_n\left(\frac{1}{x}\right)e^{-1/x}, \quad (x \neq 0),$$

where for each n, $P_n(t)$ is a polynomial in the variable t of degree $2n$. Find a recurrence formula for $P_n(t)$.

5. Prove Leibniz's formula for the n^{th} derivative of a product. If u and v are functions with derivatives up to order n, then uv has derivatives to order n and

$$(uv)^{(n)} = \sum_{k=0}^{n}\binom{n}{k}u^{(k)}v^{(n-k)}.$$

6. Calculate some higher derivatives of the composite function $y = g(f(x))$, as far as your patience allows.
7. A function is defined by

$$f(x) = \begin{cases} -x^5, & \text{if } x < 0 \\ x^5, & \text{if } x \ge 0. \end{cases}$$

How many derivatives does f possess at $x = 0$?

8. A function with domain $\mathbb{R}$ is called an *even function* if it satisfies $f(-x) = f(x)$ for all x. It is called an *odd function* if it satisfies $f(-x) = -f(x)$ for all x.

   (a) Show that every function f with domain $\mathbb{R}$ has a unique decomposition $f = g + h$ where g is even and h is odd.
   (b) Suppose that f has m derivatives at $x = 0$. Show that if f is even, then all derivatives $f^{(k)}(0)$ with odd $k \le m$ are zero. Show, on the other hand, that if f is odd, then all derivatives $f^{(k)}(0)$ with even $k \le m$ are zero.

9. How many derivatives does the function $|x|^{7/2}$ possess at $x = 0$?
10. Define a function f on the domain $]-\infty, 1[$ by

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

Show that f is differentiable at all points of its domain, that f' is continuous, and that f is twice differentiable at all points except at $x = 0$. At $x = 0$ the second derivative does not exist; but calculate its "jump", the quantity

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

Note. Where a straight section of rail track joins a curved section, it is safer if the curve is designed so that the second derivative is continuous and is 0 at the join. This is to avoid discontinuities in the acceleration normal to the track. The graph in the exercise typifies the join in a model railway, where the curves are usually arcs of circles, and that is where the model train is most likely to leave the track.

11. Variables x and y are connected by the equation $2y^5 - xy - x^4 = 0$. Calculate the second derivative d^2y/dx^2 when $x = 1$ and $y = 1$.
12. In this exercise we assume some acquaintance with determinants. Let u_1 and u_2 be differentiable functions in an interval A.

    (a) Suppose that the functions u_1 and u_2 are linearly dependent in A; by this is meant that there exist constants λ_1 and λ_2, not both 0, such that

    $$\lambda_1 u_1(x) + \lambda_2 u_2(x) = 0$$

    for all x in A. Show that, for all x in A:

    $$\begin{vmatrix} u_1(x) & u_2(x) \\ u_1'(x) & u_2'(x) \end{vmatrix} = 0.$$

    (b) The example $u_1(x) = u_2(x) = 0$ for $x < 0$ and $u_1(x) = x^2$, $u_2(x) = 2x^2$ for $x \geq 0$ shows that the converse is false.
    (c) Extend the result of item (a) to the case of m functions $u_1,\ldots, u_m$, each $m - 1$ times differentiable. Show that a necessary condition for their linear dependence in A is that

    $$\begin{vmatrix} u_1(x) & u_2(x) & \ldots & u_m(x) \\ u_1'(x) & u_2'(x) & \ldots & u_m'(x) \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{(m-1)}(x) & u_2^{(m-1)}(x) & \ldots & u_m^{(m-1)}(x) \end{vmatrix} = 0$$

    for all x in A.

5.5 Significance of the Derivative

In this section we begin to extract useful information about a function from knowledge of its derivative.

If A is an interval we call a point c in A an interior point if c is not an endpoint of the interval. In the next paragraphs a set denoted by A will always be an interval with distinct endpoints.

Proposition 5.5 *Let $f : A \to \mathbb{R}$, let c be an interior point of A and let f be differentiable at c. Then the following hold:*

(1) If $f'(c) > 0$ there exists $\delta > 0$, such that

$$f(x) < f(c) \ \textit{if} \ c - \delta < x < c, \quad \textit{and} \quad f(x) > f(c) \ \textit{if} \ c < x < c + \delta.$$

(2) If $f'(c) < 0$, then there exists $\delta > 0$, such that

$$f(x) > f(c) \ \textit{if} \ c - \delta < x < c, \quad \textit{and} \quad f(x) < f(c) \ \textit{if} \ c < x < c + \delta.$$

Proof Let $f'(c) > 0$. Now

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

and taking ε to be $\frac{1}{2} f'(c)$ in the definition of limit we find that there exists $\delta > 0$, such that

$$\frac{f(c+h) - f(c)}{h} > \frac{f'(c)}{2}$$

for all h that satisfy $0 < |h| < \delta$. For such h that are negative we have

$$f(c+h) - f(c) < \frac{hf'(c)}{2} < 0$$

and for such h that are positive we have

$$f(c+h) - f(c) > \frac{hf'(c)}{2} > 0.$$

The case when $f'(c) < 0$ is treated similarly. □

We did not assume that f was differentiable at points other than c. But even if it is, the assumption that $f'(c) > 0$ tells us little about the derivative $f'(x)$, for x near to c. We could have points x, arbitrarily near to c, at which $f'(x) < 0$, for example. Or even points at which $f'(x)$ is arbitrarily large.

5.5.1 Maxima and Minima

One of the main applications of the last paragraph is to the problem, familiar from applied mathematics, of finding maxima and minima. Problems of this nature are generally called extremal problems.

Let $f : A \to \mathbb{R}$. Recall that A denotes an interval, with or without endpoints, though the latter must be distinct. The status of a point c in A regarding the local extremal behaviour of f can be usefully, if somewhat pedantically, classified as follows:

(a) The point c is called a *local minimum point for* f if there exists $\delta > 0$, such that $f(x) \geq f(c)$ for all x in A that satisfy $|x - c| < \delta$.
(b) The point c is called a *local maximum point for* f if there exists $\delta > 0$, such that $f(x) \leq f(c)$ for all x in A that satisfy $|x - c| < \delta$.
(c) The point c is called a *strict local minimum point for* f if there exists $\delta > 0$, such that $f(x) > f(c)$ for all x in A that satisfy $0 < |x - c| < \delta$.
(d) The point c is called a *strict local maximum point for* f if there exists $\delta > 0$, such that $f(x) < f(c)$ for all x in A that satisfy $0 < |x - c| < \delta$.

Note that c could be an endpoint of the interval A in these definitions. Moreover c could belong to none of the above four classes, in which case it is of no interest as regards the extremal problem for f.

The next proposition defines precisely the notion, loosely expressed, that the derivative vanishes at a maximum or minimum.

Proposition 5.6 *Let $f : A \to \mathbb{R}$, let c be a point in A and assume that c is either a local minimum point, or a local maximum point, of f. If, in addition, c is an interior point of A and f is differentiable at c, then $f'(c) = 0$.*

Proof Consider the case when c is a local minimum point. If $f'(c) < 0$ then, by Proposition 5.5, there exists $\delta > 0$, such that $f(x) < f(c)$ if $c < x < c + \delta$. If $f'(c) > 0$ then there exists $\delta > 0$, such that $f(x) < f(c)$ if $c - \delta < x < c$. In neither case can c be a local minimum point, so we have a contradiction. We conclude that $f'(c) = 0$. A similar argument is used for the case when c is a local maximum point. □

That the derivative is 0, given that c is an interior point and f is differentiable at c, is only a necessary condition for c to be a local minimum or maximum point. It is not sufficient. There is a need for a term to cover the case that $f'(c) = 0$, irrespective of whether c is a local maximum or minimum point. The terms extreme point, extremal point, stationary point and critical point have been used (and there are probably others). The last two should be preferred as they do not suggest a maximum or minimum.

5.5.2 Finding Maxima and Minima in Practice

Let $f : [a, b] \to \mathbb{R}$ be a continuous function. The domain $[a, b]$ is a bounded and closed interval. We know by the extreme value theorem (Proposition 4.11) that f attains both a maximum and a minimum value in $[a, b]$. The problem of maxima and minima is to find the points where these are attained, as well as the maximum and minimum values.

Suppose that the maximum is attained at a point $x = c$ (there could be more than one such point). There are three possibilities (exclusive; each excludes the other two):

(a) The point c is either a or b.
(b) The point c is an interior point, that is, a point of the open interval $]a, b[$, f is differentiable at c and (by Proposition 5.6) $f'(c) = 0$.
(c) The point c is an interior point at which f is not differentiable.

The most usual situation is that there are only a finite number of points c in $[a, b]$ that satisfy any one of these three conditions. It may be feasible to find them, and once found, to arrange them in a list. This might begin with the endpoints a and b, continue with the points in $]a, b[$ at which f is not differentiable (if finitely many) and conclude with all the solutions of $f'(x) = 0$ in $]a, b[$ (if finitely many). Now it only remains to calculate f at each of the points in the list and find the highest and lowest of these values.

5.5.3 Exercises

1. In each of the following cases determine the maximum and minimum of the function f over the interval A:

   (a) $f(x) = x^3 - 3x^2 + x, \quad A = [1, 3].$
   (b) $f(x) = \max\left(1 - 2x - x^2,\ 2 + x - x^2,\ 1 + 3x - x^2\right), \quad A = [-1, 2].$
   (c) $f(x) = \dfrac{5}{1 + |x - 4|} + \dfrac{4}{1 + |x - 5|}, \quad A = [-6, 6].$

   Hint. In items (b) and (c) it helps to express the functions by cases. For the numerical work in these exercises it makes sense to use a calculator and state the answers with a certain number of decimal digits, say, three.
2. Determine the minimum of the function

$$f(x) = x + \frac{1}{x^3}$$

   in the interval $]0, \infty[$.
3. 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. Determine the minimum of f over the whole real line.

4. Define the function f with domain $\mathbb{R}$ by $f(x) = x^2 \sin(1/x)$ for $x \neq 0$, and $f(0) = 0$. Show that f is everywhere differentiable, including at $x = 0$, but that f' is discontinuous at $x = 0$.
5. Find an example of a continuous function f that has a strict local minimum at $x = 0$, and for every $\delta > 0$ has also a strict local minimum in the interval $]0, \delta[$.
Hint. To help you think about it, note that there would have to be infinitely many minima in $]0, \delta[$, each with a value higher than $f(0)$.
6. Give an example of a differentiable function f, such that $f'(0) > 0$, but there exists no $\delta > 0$ such that f is strictly increasing in the interval $]-\delta, \delta[$.
Hint. Try to exploit the wildly oscillating function $\sin(1/x)$.
7. Give an example of a differentiable function f such that $f'(0) = 0$ and in every interval $]-\delta, \delta[$ (with $\delta > 0$) the derivative f' takes arbitrarily large positive values and arbitrarily large negative values.

5.6 The Mean Value Theorem

It has been called the most useful theorem in analysis (notably by the influential French mathematician and Bourbakiste, Jean Dieudonné, but he was probably echoing G. H. Hardy). We leave it to the reader to judge the truth or otherwise of this claim. It might seem more logical to write "mean-value theorem", as there is nothing mean about it, nor is it one of a collection of value theorems. The lack of a hyphen is sanctioned by usage, as it is in the names of other theorems with compound qualifiers (as in "small oscillation theorem").

In the following, the interval $[a, b]$ has distinct endpoints, and is manifestly bounded and closed.

Proposition 5.7 (Rolle's theorem) *Let $f : [a, b] \to \mathbb{R}$ be a continuous function that is differentiable for $a < x < b$. Assume that $f(a) = f(b)$. Then there exists c, such that $a < c < b$ and $f'(c) = 0$.*

Proof Let $m = \inf_{[a,b]} f$ and $M = \sup_{[a,b]} f$ (both m and M are attained by the extreme value theorem, so they are minimum and maximum). If $m = M$ then f is a constant and so $f'(x) = 0$ for all x in $]a, b[$ and we are done.

Assume next that $m < M$. If these values are attained at the endpoints, then, since $f(a) = f(b)$, we again have $m = M$. So at least one of them is attained at an interior point. Either there exists c in $]a, b[$ such that $f(c) = m$ or there exists c in $]a, b[$ such that $f(c) = M$. In both these cases we have $f'(c) = 0$. □

Proposition 5.8 (Mean value theorem) *Let $f : [a, b] \to \mathbb{R}$ be a continuous function that is differentiable for $a < x < b$. Then there exists c in the open interval $]a, b[$, such that*

$$f(b) - f(a) = f'(c)(b - a).$$