Chapter 8

Differentiable Functions

A differentiable function is a function that can be approximated locally by a linear function.

8.1. The derivative

Definition 8.1. Suppose that $f : (a, b) \to \mathbb{R}$ and $a < c < b$. Then f is differentiable at c with derivative $f'(c)$ if

$$\lim_{h \to 0} \left[\frac{f(c+h) - f(c)}{h} \right] = f'(c).$$

The domain of f' is the set of points $c \in (a, b)$ for which this limit exists. If the limit exists for every $c \in (a, b)$ then we say that f is differentiable on (a, b).

Graphically, this definition says that the derivative of f at c is the slope of the tangent line to $y = f(x)$ at c, which is the limit as $h \to 0$ of the slopes of the lines through $(c, f(c))$ and $(c + h, f(c + h))$.

We can also write

$$f'(c) = \lim_{x \to c} \left[\frac{f(x) - f(c)}{x - c} \right],$$

since if $x = c + h$, the conditions $0 < |x - c| < \delta$ and $0 < |h| < \delta$ in the definitions of the limits are equivalent. The ratio

$$\frac{f(x) - f(c)}{x - c}$$

is undefined $(0/0)$ at $x = c$, but it doesn't have to be defined in order for the limit as $x \to c$ to exist.

Like continuity, differentiability is a local property. That is, the differentiability of a function f at c and the value of the derivative, if it exists, depend only the values of f in a arbitrarily small neighborhood of c. In particular if $f : A \to \mathbb{R}$

where $A \subset \mathbb{R}$, then we can define the differentiability of f at any interior point $c \in A$ since there is an open interval $(a, b) \subset A$ with $c \in (a, b)$.

8.1.1. Examples of derivatives.

Let us give a number of examples that illustrate differentiable and non-differentiable functions.

Example 8.2. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is differentiable on $\mathbb{R}$ with derivative $f'(x) = 2x$ since

$$\lim_{h\to 0}\left[\frac{(c+h)^2 - c^2}{h}\right] = \lim_{h\to 0}\frac{h(2c+h)}{h} = \lim_{h\to 0}(2c+h) = 2c.$$

Note that in computing the derivative, we first cancel by h, which is valid since $h \neq 0$ in the definition of the limit, and then set $h = 0$ to evaluate the limit. This procedure would be inconsistent if we didn't use limits.

Example 8.3. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x^2 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

is differentiable on $\mathbb{R}$ with derivative

$$f'(x) = \begin{cases} 2x & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

For $x > 0$, the derivative is $f'(x) = 2x$ as above, and for $x < 0$, we have $f'(x) = 0$. For 0, we consider the limit

$$\lim_{h\to 0}\left[\frac{f(h) - f(0)}{h}\right] = \lim_{h\to 0}\frac{f(h)}{h}.$$

The right limit is

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

and the left limit is

$$\lim_{h\to 0^-}\frac{f(h)}{h} = 0.$$

Since the left and right limits exist and are equal, the limit also exists, and f is differentiable at 0 with $f'(0) = 0$.

Next, we consider some examples of non-differentiability at discontinuities, corners, and cusps.

Example 8.4. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1/x & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases}$$

is differentiable at $x \neq 0$ with derivative $f'(x) = -1/x^2$ since

$$\begin{aligned}\lim_{h\to 0}\left[\frac{f(c+h)-f(c)}{h}\right] &= \lim_{h\to 0}\left[\frac{1/(c+h)-1/c}{h}\right]\\ &= \lim_{h\to 0}\left[\frac{c-(c+h)}{hc(c+h)}\right]\\ &= -\lim_{h\to 0}\frac{1}{c(c+h)}\\ &= -\frac{1}{c^2}.\end{aligned}$$

However, f is not differentiable at 0 since the limit

$$\lim_{h\to 0}\left[\frac{f(h)-f(0)}{h}\right] = \lim_{h\to 0}\left[\frac{1/h-0}{h}\right] = \lim_{h\to 0}\frac{1}{h^2}$$

does not exist.

Example 8.5. The sign function $f(x) = \operatorname{sgn} x$, defined in Example 6.8, is differentiable at $x \neq 0$ with $f'(x) = 0$, since in that case $f(x+h) - f(x) = 0$ for all sufficiently small h. The sign function is not differentiable at 0 since

$$\lim_{h\to 0}\left[\frac{\operatorname{sgn} h - \operatorname{sgn} 0}{h}\right] = \lim_{h\to 0}\frac{\operatorname{sgn} h}{h}$$

and

$$\frac{\operatorname{sgn} h}{h} = \begin{cases} 1/h & \text{if } h > 0\\ -1/h & \text{if } h < 0\end{cases}$$

is unbounded in every neighborhood of 0, so its limit does not exist.

Example 8.6. The absolute value function $f(x) = |x|$ is differentiable at $x \neq 0$ with derivative $f'(x) = \operatorname{sgn} x$. It is not differentiable at 0, however, since

$$\lim_{h\to 0}\frac{f(h)-f(0)}{h} = \lim_{h\to 0}\frac{|h|}{h} = \lim_{h\to 0}\operatorname{sgn} h$$

does not exist. (The right limit is 1 and the left limit is -1.)

Example 8.7. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|^{1/2}$ is differentiable at $x \neq 0$ with

$$f'(x) = \frac{\operatorname{sgn} x}{2|x|^{1/2}}.$$

If $c > 0$, then using the difference of two square to rationalize the numerator, we get

$$\begin{aligned}\lim_{h\to 0}\left[\frac{f(c+h)-f(c)}{h}\right] &= \lim_{h\to 0}\frac{(c+h)^{1/2}-c^{1/2}}{h}\\ &= \lim_{h\to 0}\frac{(c+h)-c}{h\left[(c+h)^{1/2}+c^{1/2}\right]}\\ &= \lim_{h\to 0}\frac{1}{(c+h)^{1/2}+c^{1/2}}\\ &= \frac{1}{2c^{1/2}}.\end{aligned}$$

If $c < 0$, we get the analogous result with a negative sign. However, f is not differentiable at 0, since

$$\lim_{h\to 0^+} \frac{f(h) - f(0)}{h} = \lim_{h\to 0^+} \frac{1}{h^{1/2}}$$

does not exist.

Example 8.8. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^{1/3}$ is differentiable at $x \neq 0$ with

$$f'(x) = \frac{1}{3x^{2/3}}.$$

To prove this result, we use the identity for the difference of cubes,

$$a^3 - b^3 = (a - b)(a^2 + ab + b^2),$$

and get for $c \neq 0$ that

$$\begin{aligned}
\lim_{h\to 0} \left[\frac{f(c+h) - f(c)}{h}\right] &= \lim_{h\to 0} \frac{(c+h)^{1/3} - c^{1/3}}{h} \\
&= \lim_{h\to 0} \frac{(c+h) - c}{h\left[(c+h)^{2/3} + (c+h)^{1/3}c^{1/3} + c^{2/3}\right]} \\
&= \lim_{h\to 0} \frac{1}{(c+h)^{2/3} + (c+h)^{1/3}c^{1/3} + c^{2/3}} \\
&= \frac{1}{3c^{2/3}}.
\end{aligned}$$

However, f is not differentiable at 0, since

$$\lim_{h\to 0} \frac{f(h) - f(0)}{h} = \lim_{h\to 0} \frac{1}{h^{2/3}}$$

does not exist.

Finally, we consider some examples of highly oscillatory functions.

Example 8.9. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

It follows from the product and chain rules proved below that f is differentiable at $x \neq 0$ with derivative

$$f'(x) = \sin\frac{1}{x} - \frac{1}{x}\cos\frac{1}{x}.$$

However, f is not differentiable at 0, since

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

which does not exist.

Example 8.10. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

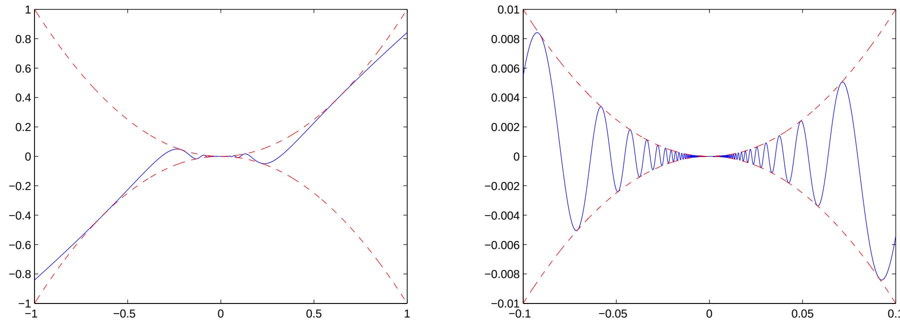

Figure 1. A plot of the function $y = x^2 \sin(1/x)$ and a detail near the origin with the parabolas $y = \pm x^2$ shown in red.

Then f is differentiable on $\mathbb{R}$. (See Figure 1.) It follows from the product and chain rules proved below that f is differentiable at $x \neq 0$ with derivative

$$f'(x) = 2x \sin\frac{1}{x} - \cos\frac{1}{x}.$$

Moreover, f is differentiable at 0 with $f'(0) = 0$, since

$$\lim_{h\to 0} \frac{f(h) - f(0)}{h} = \lim_{h\to 0} h \sin\frac{1}{h} = 0.$$

In this example, $\lim_{x\to 0} f'(x)$ does not exist, so although f is differentiable on $\mathbb{R}$, its derivative f' is not continuous at 0.

8.1.2. Derivatives as linear approximations. Another way to view Definition 8.1 is to write

$$f(c + h) = f(c) + f'(c)h + r(h)$$

as the sum of a linear (or, strictly speaking, affine) approximation $f(c) + f'(c)h$ of $f(c + h)$ and a remainder $r(h)$. In general, the remainder also depends on c, but we don't show this explicitly since we're regarding c as fixed.

As we prove in the following proposition, the differentiability of f at c is equivalent to the condition

$$\lim_{h\to 0} \frac{r(h)}{h} = 0.$$

That is, the remainder $r(h)$ approaches 0 faster than h, so the linear terms in h provide a leading order approximation to $f(c + h)$ when h is small. We also write this condition on the remainder as

$$r(h) = o(h) \qquad \text{as } h \to 0,$$

pronounced "r is little-oh of h as $h \to 0$."

Graphically, this condition means that the graph of f near c is close the line through the point $(c, f(c))$ with slope $f'(c)$. Analytically, it means that the function

$$h \mapsto f(c + h) - f(c)$$

is approximated near c by the linear function

$$h \mapsto f'(c)h.$$

Thus, $f'(c)$ may be interpreted as a scaling factor by which a differentiable function f shrinks or stretches lengths near c.

If $|f'(c)| < 1$, then f shrinks the length of a small interval about c by (approximately) this factor; if $|f'(c)| > 1$, then f stretches the length of an interval by (approximately) this factor; if $f'(c) > 0$, then f preserves the orientation of the interval, meaning that it maps the left endpoint to the left endpoint of the image and the right endpoint to the right endpoints; if $f'(c) < 0$, then f reverses the orientation of the interval, meaning that it maps the left endpoint to the right endpoint of the image and visa-versa.

We can use this description as a definition of the derivative.

Proposition 8.11. Suppose that $f : (a, b) \to \mathbb{R}$. Then f is differentiable at $c \in (a, b)$ if and only if there exists a constant $A \in \mathbb{R}$ and a function $r : (a-c, b-c) \to \mathbb{R}$ such that

$$f(c+h) = f(c) + Ah + r(h), \qquad \lim_{h \to 0} \frac{r(h)}{h} = 0.$$

In that case, $A = f'(c)$.

Proof. First suppose that f is differentiable at c according to Definition 8.1, and define

$$r(h) = f(c+h) - f(c) - f'(c)h.$$

Then

$$\lim_{h \to 0} \frac{r(h)}{h} = \lim_{h \to 0} \left[\frac{f(c+h) - f(c)}{h} - f'(c) \right] = 0,$$

so the condition in the proposition holds with $A = f'(c)$.

Conversely, suppose that $f(c+h) = f(c) + Ah + r(h)$ where $r(h)/h \to 0$ as $h \to 0$. Then

$$\lim_{h \to 0} \left[\frac{f(c+h) - f(c)}{h} \right] = \lim_{h \to 0} \left[A + \frac{r(h)}{h} \right] = A,$$

so f is differentiable at c with $f'(c) = A$. □

Example 8.12. For Example 8.2 with $f(x) = x^2$, we get

$$(c+h)^2 = c^2 + 2ch + h^2,$$

and $r(h) = h^2$, which goes to zero at a quadratic rate as $h \to 0$.

Example 8.13. For Example 8.4 with $f(x) = 1/x$, we get

$$\frac{1}{c+h} = \frac{1}{c} - \frac{1}{c^2} h + r(h),$$

for $c \neq 0$, where the quadratically small remainder is

$$r(h) = \frac{h^2}{c^2(c+h)}.$$

8.1.3. Left and right derivatives. For the most part, we will use derivatives that are defined only at the interior points of the domain of a function. Sometimes, however, it is convenient to use one-sided left or right derivatives that are defined at the endpoint of an interval.

Definition 8.14. Suppose that $f : [a, b] \to \mathbb{R}$. Then f is right-differentiable at $a \le c < b$ with right derivative $f'(c^+)$ if

$$\lim_{h \to 0^+} \left[\frac{f(c+h) - f(c)}{h} \right] = f'(c^+)$$

exists, and f is left-differentiable at $a < c \le b$ with left derivative $f'(c^-)$ if

$$\lim_{h \to 0^-} \left[\frac{f(c+h) - f(c)}{h} \right] = \lim_{h \to 0^+} \left[\frac{f(c) - f(c-h)}{h} \right] = f'(c^-).$$

A function is differentiable at $a < c < b$ if and only if the left and right derivatives at c both exist and are equal.

Example 8.15. If $f : [0, 1] \to \mathbb{R}$ is defined by $f(x) = x^2$, then

$$f'(0^+) = 0, \qquad f'(1^-) = 2.$$

These left and right derivatives remain the same if f is extended to a function defined on a larger domain, say

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

For this extended function we have $f'(1^+) = 0$, which is not equal to $f'(1^-)$, and $f'(0^-)$ does not exist, so the extended function is not differentiable at either 0 or 1.

Example 8.16. The absolute value function $f(x) = |x|$ in Example 8.6 is left and right differentiable at 0 with left and right derivatives

$$f'(0^+) = 1, \qquad f'(0^-) = -1.$$

These are not equal, and f is not differentiable at 0.

8.2. Properties of the derivative

In this section, we prove some basic properties of differentiable functions.

8.2.1. Differentiability and continuity. First we discuss the relation between differentiability and continuity.

Theorem 8.17. If $f : (a, b) \to \mathbb{R}$ is differentiable at at $c \in (a, b)$, then f is continuous at c.

Proof. If f is differentiable at c, then

$$\begin{aligned}\lim_{h\to 0} f(c+h) - f(c) &= \lim_{h\to 0}\left[\frac{f(c+h)-f(c)}{h}\cdot h\right]\\ &= \lim_{h\to 0}\left[\frac{f(c+h)-f(c)}{h}\right]\cdot \lim_{h\to 0} h\\ &= f'(c)\cdot 0\\ &= 0,\end{aligned}$$

which implies that f is continuous at c. □

For example, the sign function in Example 8.5 has a jump discontinuity at 0 so it cannot be differentiable at 0. The converse does not hold, and a continuous function needn't be differentiable. The functions in Examples 8.6, 8.8, 8.9 are continuous but not differentiable at 0. Example 9.24 describes a function that is continuous on $\mathbb{R}$ but not differentiable anywhere.

In Example 8.10, the function is differentiable on $\mathbb{R}$, but the derivative f' is not continuous at 0. Thus, while a function f has to be continuous to be differentiable, if f is differentiable its derivative f' need not be continuous. This leads to the following definition.

Definition 8.18. A function $f : (a, b) \to \mathbb{R}$ is continuously differentiable on (a, b), written $f \in C^1(a, b)$, if it is differentiable on (a, b) and $f' : (a, b) \to \mathbb{R}$ is continuous.

For example, the function $f(x) = x^2$ with derivative $f'(x) = 2x$ is continuously differentiable on $\mathbb{R}$, whereas the function in Example 8.10 is not continuously differentiable at 0. As this example illustrates, functions that are differentiable but not continuously differentiable may behave in rather pathological ways. On the other hand, the behavior of continuously differentiable functions, whose graphs have continuously varying tangent lines, is more-or-less consistent with what one expects.

8.2.2. Algebraic properties of the derivative.

A fundamental property of the derivative is that it is a linear operation. In addition, we have the following product and quotient rules.

Theorem 8.19. If $f, g : (a, b) \to \mathbb{R}$ are differentiable at $c \in (a, b)$ and $k \in \mathbb{R}$, then kf, $f + g$, and fg are differentiable at c with

$$(kf)'(c) = kf'(c), \quad (f+g)'(c) = f'(c) + g'(c), \quad (fg)'(c) = f'(c)g(c) + f(c)g'(c).$$

Furthermore, if $g(c) \neq 0$, then f/g is differentiable at c with

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

Proof. The first two properties follow immediately from the linearity of limits stated in Theorem 6.34. For the product rule, we write

$$\begin{aligned}(fg)'(c) &= \lim_{h\to 0}\left[\frac{f(c+h)g(c+h)-f(c)g(c)}{h}\right]\\ &= \lim_{h\to 0}\left[\frac{(f(c+h)-f(c))\,g(c+h)+f(c)\,(g(c+h)-g(c))}{h}\right]\\ &= \lim_{h\to 0}\left[\frac{f(c+h)-f(c)}{h}\right]\lim_{h\to 0} g(c+h)+f(c)\lim_{h\to 0}\left[\frac{g(c+h)-g(c)}{h}\right]\\ &= f'(c)g(c)+f(c)g'(c),\end{aligned}$$

where we have used the properties of limits in Theorem 6.34 and Theorem 8.19, which implies that g is continuous at c. The quotient rule follows by a similar argument, or by combining the product rule with the chain rule, which implies that $(1/g)' = -g'/g^2$. (See Example 8.22 below.) □

Example 8.20. We have $1' = 0$ and $x' = 1$. Repeated application of the product rule implies that x^n is differentiable on $\mathbb{R}$ for every $n \in \mathbb{N}$ with

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

Alternatively, we can prove this result by induction: The formula holds for $n = 1$. Assuming that it holds for some $n \in \mathbb{N}$, we get from the product rule that

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

and the result follows. It also follows by linearity that every polynomial function is differentiable on $\mathbb{R}$, and from the quotient rule that every rational function is differentiable at every point where its denominator is nonzero. The derivatives are given by their usual formulae.

8.3. The chain rule

The chain rule states that the composition of differentiable functions is differentiable. The result is quite natural if one thinks in terms of derivatives as linear maps. If f is differentiable at c, it scales lengths by a factor $f'(c)$, and if g is differentiable at $f(c)$, it scales lengths by a factor $g'(f(c))$. Thus, the composition $g \circ f$ scales lengths at c by a factor $g'(f(c)) \cdot f'(c)$. Equivalently, the derivative of a composition is the composition of the derivatives (regarded as linear maps).

We will prove the chain rule by showing that the composition of remainder terms in the linear approximations of f and g leads to a similar remainder term in the linear approximation of $g \circ f$. The argument is complicated by the fact that we have to evaluate the remainder of g at a point that depends on the remainder of f, but this complication should not obscure the simplicity of the final result.

Theorem 8.21 (Chain rule)**.** Let $f : A \to \mathbb{R}$ and $g : B \to \mathbb{R}$ where $A \subset \mathbb{R}$ and $f(A) \subset B$, and suppose that c is an interior point of A and $f(c)$ is an interior point of B. If f is differentiable at c and g is differentiable at $f(c)$, then $g \circ f : A \to \mathbb{R}$ is differentiable at c and

$$(g \circ f)'(c) = g'(f(c))\, f'(c).$$

Proof. Since f is differentiable at c, there is a function $r(h)$ such that

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

and since g is differentiable at $f(c)$, there is a function $s(k)$ such that

$$g\left(f(c)+k\right) = g\left(f(c)\right) + g'\left(f(c)\right)k + s(k), \qquad \lim_{k\to 0} \frac{s(k)}{k} = 0.$$

It follows that

$$\begin{aligned}(g\circ f)(c+h) &= g\left(f(c) + f'(c)h + r(h)\right)\\ &= g\left(f(c)\right) + g'\left(f(c)\right)\cdot\left(f'(c)h + r(h)\right) + s\left(f'(c)h + r(h)\right)\\ &= g\left(f(c)\right) + g'\left(f(c)\right)f'(c)\cdot h + t(h)\end{aligned}$$

where

$$t(h) = g'\left(f(c)\right)\cdot r(h) + s\left(\phi(h)\right), \qquad \phi(h) = f'(c)h + r(h).$$

Since $r(h)/h \to 0$ as $h \to 0$, we have

$$\lim_{h\to 0} \frac{t(h)}{h} = \lim_{h\to 0} \frac{s\left(\phi(h)\right)}{h}.$$

We claim that this limit exists and is zero, and then it follows from Proposition 8.11 that $g\circ f$ is differentiable at c with

$$(g\circ f)'(c) = g'\left(f(c)\right)f'(c).$$

To prove the claim, we use the facts that

$$\frac{\phi(h)}{h} \to f'(c) \quad \text{as } h\to 0, \qquad \frac{s(k)}{k} \to 0 \quad \text{as } k \to 0.$$

Roughly speaking, we have $\phi(h) \sim f'(c)h$ when h is small and therefore

$$\frac{s\left(\phi(h)\right)}{h} \sim \frac{s\left(f'(c)h\right)}{h} \to 0 \qquad \text{as } h \to 0.$$

In detail, let $\epsilon > 0$ be given. We want to show that there exists $\delta > 0$ such that

$$\left|\frac{s\left(\phi(h)\right)}{h}\right| < \epsilon \qquad \text{if } 0 < |h| < \delta.$$

First, choose $\delta_1 > 0$ such that

$$\left|\frac{r(h)}{h}\right| < |f'(c)| + 1 \qquad \text{if } 0 < |h| < \delta_1.$$

If $0 < |h| < \delta_1$, then

$$\begin{aligned}|\phi(h)| &\le |f'(c)|\,|h| + |r(h)|\\ &< |f'(c)|\,|h| + (|f'(c)|+1)|h|\\ &< \left(2|f'(c)|+1\right)|h|.\end{aligned}$$

Next, choose $\eta > 0$ so that

$$\left|\frac{s(k)}{k}\right| < \frac{\epsilon}{2|f'(c)|+1} \qquad \text{if } 0 < |k| < \eta.$$

(We include a "1" in the denominator on the right-hand side to avoid a division by zero if $f'(c) = 0$.) Finally, define $\delta_2 > 0$ by

$$\delta_2 = \frac{\eta}{2|f'(c)| + 1},$$

and let $\delta = \min(\delta_1, \delta_2) > 0$.

If $0 < |h| < \delta$ and $\phi(h) \neq 0$, then $0 < |\phi(h)| < \eta$, so

$$|s(\phi(h))| \leq \frac{\epsilon|\phi(h)|}{2|f'(c)| + 1} < \epsilon|h|.$$

If $\phi(h) = 0$, then $s(\phi(h)) = 0$, so the inequality holds in that case also. This proves that

$$\lim_{h\to 0} \frac{s(\phi(h))}{h} = 0.$$

□

Example 8.22. Suppose that f is differentiable at c and $f(c) \neq 0$. Then $g(y) = 1/y$ is differentiable at $f(c)$, with $g'(y) = -1/y^2$ (see Example 8.4). It follows that the reciprocal function $1/f = g \circ f$ is differentiable at c with

$$\left(\frac{1}{f}\right)'(c) = g'(f(c))f'(c) = -\frac{f'(c)}{f(c)^2}.$$

The chain rule gives an expression for the derivative of an inverse function. In terms of linear approximations, it states that if f scales lengths at c by a nonzero factor $f'(c)$, then f^{-1} scales lengths at $f(c)$ by the factor $1/f'(c)$.

Proposition 8.23. Suppose that $f : A \to \mathbb{R}$ is a one-to-one function on $A \subset \mathbb{R}$ with inverse $f^{-1} : B \to \mathbb{R}$ where $B = f(A)$. Assume that f is differentiable at an interior point $c \in A$ and f^{-1} is differentiable at $f(c)$, where $f(c)$ is an interior point of B. Then $f'(c) \neq 0$ and

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

Proof. The definition of the inverse implies that

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

Since f is differentiable at c and f^{-1} is differentiable at $f(c)$, the chain rule implies that

$$\left(f^{-1}\right)'(f(c))\, f'(c) = 1.$$

Dividing this equation by $f'(c) \neq 0$, we get the result. Moreover, it follows that f^{-1} cannot be differentiable at $f(c)$ if $f'(c) = 0$. □

Alternatively, setting $d = f(c)$, we can write the result as

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

Proposition 8.23 is not entirely satisfactory because it assumes the existence and differentiability of an inverse function. We will return to this question in Section 8.7 below, but we end this section with some examples that illustrate the necessity of the condition $f'(c) \neq 0$ for the existence and differentiability of the inverse.

Example 8.24. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Then $f'(0) = 0$ and f is not invertible on any neighborhood of the origin, since it is non-monotone and not one-to-one. On the other hand, if $f : (0,\infty) \to (0,\infty)$ is defined by $f(x) = x^2$, then $f'(x) = 2x \neq 0$ and the inverse function $f^{-1} : (0,\infty) \to (0,\infty)$ is given by

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

The formula for the inverse of the derivative gives

$$(f^{-1})'(x^2) = \frac{1}{f'(x)} = \frac{1}{2x},$$

or, writing $x = f^{-1}(y)$,

$$(f^{-1})'(y) = \frac{1}{2\sqrt{y}},$$

in agreement with Example 8.7.

Example 8.25. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^3$. Then f is strictly increasing, one-to-one, and onto. The inverse function $f^{-1} : \mathbb{R} \to \mathbb{R}$ is given by

$$f^{-1}(y) = y^{1/3}.$$

Then $f'(0) = 0$ and f^{-1} is not differentiable at $f(0)= 0$. On the other hand, f^{-1} is differentiable at non-zero points of $\mathbb{R}$, with

$$(f^{-1})'(x^3) = \frac{1}{f'(x)} = \frac{1}{3x^2},$$

or, writing $x = y^{1/3}$,

$$(f^{-1})'(y) = \frac{1}{3y^{2/3}},$$

in agreement with Example 8.8.

8.4. Extreme values

One of the most useful applications of the derivative is in locating the maxima and minima of functions.

Definition 8.26. Suppose that $f : A \to \mathbb{R}$. Then f has a global (or absolute) maximum at $c \in A$ if

$$f(x) \leq f(c) \qquad \text{for all } x \in A,$$

and f has a local (or relative) maximum at $c \in A$ if there is a neighborhood U of c such that

$$f(x) \leq f(c) \qquad \text{for all } x \in A \cap U.$$

Similarly, f has a global (or absolute) minimum at $c \in A$ if

$$f(x) \geq f(c) \qquad \text{for all } x \in A,$$

and f has a local (or relative) minimum at $c \in A$ if there is a neighborhood U of c such that

$$f(x) \geq f(c) \qquad \text{for all } x \in A \cap U.$$

If f has a (local or global) maximum or minimum at $c \in A$, then f is said to have a (local or global) extreme value at c.

Theorem 7.37 states that a continuous function on a compact set has a global maximum and minimum but does not say how to find them. The following fundamental result goes back to Fermat.

Theorem 8.27. If $f : A \subset \mathbb{R} \to \mathbb{R}$ has a local extreme value at an interior point $c \in A$ and f is differentiable at c, then $f'(c) = 0$.

Proof. If f has a local maximum at c, then $f(x) \le f(c)$ for all x in a δ-neighborhood $(c - \delta, c + \delta)$ of c, so

$$\frac{f(c+h) - f(c)}{h} \le 0 \qquad \text{for all } 0 < h < \delta,$$

which implies that

$$f'(c) = \lim_{h \to 0^+} \left[\frac{f(c+h) - f(c)}{h}\right] \le 0.$$

Moreover,

$$\frac{f(c+h) - f(c)}{h} \ge 0 \qquad \text{for all } -\delta < h < 0,$$

which implies that

$$f'(c) = \lim_{h \to 0^-} \left[\frac{f(c+h) - f(c)}{h}\right] \ge 0.$$

It follows that $f'(c) = 0$. If f has a local minimum at c, then the signs in these inequalities are reversed, and we also conclude that $f'(c) = 0$. $\square$

For this result to hold, it is crucial that c is an interior point, since we look at the sign of the difference quotient of f on both sides of c. At an endpoint, we get the following inequality condition on the derivative. (Draw a graph!)

Proposition 8.28. Let $f : [a, b] \to \mathbb{R}$. If the right derivative of f exists at a, then: $f'(a^+) \le 0$ if f has a local maximum at a; and $f'(a^+) \ge 0$ if f has a local minimum at a. Similarly, if the left derivative of f exists at b, then: $f'(b^-) \ge 0$ if f has a local maximum at b; and $f'(b^-) \le 0$ if f has a local minimum at b.

Proof. If the right derivative of f exists at a, and f has a local maximum at a, then there exists $\delta > 0$ such that $f(x) \le f(a)$ for $a \le x < a + \delta$, so

$$f'(a^+) = \lim_{h \to 0^+} \left[\frac{f(a+h) - f(a)}{h}\right] \le 0.$$

Similarly, if the left derivative of f exists at b, and f has a local maximum at b, then $f(x) \le f(b)$ for $b - \delta < x \le b$, so $f'(b^-) \ge 0$. The signs are reversed for local minima at the endpoints. $\square$

In searching for extreme values of a function, it is convenient to introduce the following classification of points in the domain of the function.

Definition 8.29. Suppose that $f : A \subset \mathbb{R} \to \mathbb{R}$. An interior point $c \in A$ such that f is not differentiable at c or $f'(c) = 0$ is called a critical point of f. An interior point where $f'(c) = 0$ is called a stationary point of f.

Theorem 8.27 limits the search for maxima or minima of a function f on A to the following points.

(1) Boundary points of A.

(2) Critical points of f:
  (a) interior points where f is not differentiable;
  (b) stationary points where $f'(c) = 0$.

Additional tests are required to determine which of these points gives local or global extreme values of f. In particular, a function need not attain an extreme value at a critical point.

Example 8.30. If $f : [-1, 1] \to \mathbb{R}$ is the function

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

then $x = 0$ is a critical point since f is not differentiable at 0, but f does not attain a local extreme value at 0. The global maximum and minimum of f are attained at the endpoints $x = 1$ and $x = -1$, respectively, and f has no other local extreme values.

Example 8.31. If $f : [-1, 1] \to \mathbb{R}$ is the function $f(x) = x^3$, then $x = 0$ is a critical point since $f'(0) = 0$, but f does not attain a local extreme value at 0. The global maximum and minimum of f are attained at the endpoints $x = 1$ and $x = -1$, respectively, and f has no other local extreme values.

8.5. The mean value theorem

The mean value theorem is a key result that connects the global behavior of a function $f : [a, b] \to \mathbb{R}$, described by the difference $f(b) - f(a)$, to its local behavior, described by the derivative $f' : (a, b) \to \mathbb{R}$. We begin by proving a special case.

Theorem 8.32 (Rolle)**.** Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on the closed, bounded interval $[a, b]$, differentiable on the open interval (a, b), and $f(a) = f(b)$. Then there exists $a < c < b$ such that $f'(c) = 0$.

Proof. By the Weierstrass extreme value theorem, Theorem 7.37, f attains its global maximum and minimum values on $[a, b]$. If these are both attained at the endpoints, then f is constant, and $f'(c) = 0$ for every $a < c < b$. Otherwise, f attains at least one of its global maximum or minimum values at an interior point $a < c < b$. Theorem 8.27 implies that $f'(c) = 0$. □

Note that we require continuity on the closed interval $[a, b]$ but differentiability only on the open interval (a, b). This proof is deceptively simple, but the result is not trivial. It relies on the extreme value theorem, which in turn relies on the completeness of $\mathbb{R}$. The theorem would not be true if we restricted attention to functions defined on the rationals $\mathbb{Q}$.

The mean value theorem is an immediate consequence of Rolle's theorem: for a general function f with $f(a) \neq f(b)$, we subtract off a linear function to make the values of the resulting function equal at the endpoints.

Theorem 8.33 (Mean value)**.** Suppose that $f : [a, b] \to \mathbb{R}$ is continuous on the closed, bounded interval $[a, b]$ and differentiable on the open interval (a, b). Then there exists $a < c < b$ such that

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

Proof. The function $g : [a, b] \to \mathbb{R}$ defined by

$$g(x) = f(x) - f(a) - \left[\frac{f(b) - f(a)}{b - a}\right](x - a)$$

is continuous on $[a, b]$ and differentiable on (a, b) with

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$

Moreover, $g(a) = g(b) = 0$. Rolle's Theorem implies that there exists $a < c < b$ such that $g'(c) = 0$, which proves the result. □

Graphically, this result says that there is point $a < c < b$ at which the slope of the tangent line to the graph $y = f(x)$ is equal to the slope of the chord between the endpoints $(a, f(a))$ and $(b, f(b))$.

As a first application, we prove a converse to the obvious fact that the derivative of a constant functions is zero.

Theorem 8.34. If $f : (a, b) \to \mathbb{R}$ is differentiable on (a, b) and $f'(x) = 0$ for every $a < x < b$, then f is constant on (a, b).

Proof. Fix $x_0 \in (a, b)$. The mean value theorem implies that for all $x \in (a, b)$ with $x \neq x_0$

$$f'(c) = \frac{f(x) - f(x_0)}{x - x_0}$$

for some c between x_0 and x. Since $f'(c) = 0$, it follows that $f(x) = f(x_0)$ for all $x \in (a, b)$, meaning that f is constant on (a, b). □

Corollary 8.35. If $f, g : (a, b) \to \mathbb{R}$ are differentiable on (a, b) and $f'(x) = g'(x)$ for every $a < x < b$, then $f(x) = g(x) + C$ for some constant C.

Proof. This follows from the previous theorem since $(f - g)' = 0$. □

We can also use the mean value theorem to relate the monotonicity of a differentiable function with the sign of its derivative. (See Definition 7.54 for our terminology for increasing and decreasing functions.)

Theorem 8.36. Suppose that $f : (a, b) \to \mathbb{R}$ is differentiable on (a, b). Then f is increasing if and only if $f'(x) \geq 0$ for every $a < x < b$, and decreasing if and only if $f'(x) \leq 0$ for every $a < x < b$. Furthermore, if $f'(x) > 0$ for every $a < x < b$ then f is strictly increasing, and if $f'(x) < 0$ for every $a < x < b$ then f is strictly decreasing.

Proof. If f is increasing and $a < x < b$, then

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

for all sufficiently small h (positive or negative), so

$$f'(x) = \lim_{h \to 0} \left[\frac{f(x+h) - f(x)}{h} \right] \geq 0.$$

Conversely if $f' \geq 0$ and $a < x < y < b$, then by the mean value theorem there exists $x < c < y$ such that

$$\frac{f(y) - f(x)}{y - x} = f'(c) \geq 0,$$

which implies that $f(x) \leq f(y)$, so f is increasing. Moreover, if $f'(c) > 0$, we get $f(x) < f(y)$, so f is strictly increasing.

The results for a decreasing function f follow in a similar way, or we can apply of the previous results to the increasing function $-f$. □

Note that although $f' > 0$ implies that f is strictly increasing, f is strictly increasing does not imply that $f' > 0$.

Example 8.37. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is strictly increasing on $\mathbb{R}$, but $f'(0) = 0$.

If f is continuously differentiable and $f'(c) > 0$, then $f'(x) > 0$ for all x in a neighborhood of c and Theorem 8.36 implies that f is strictly increasing near c. This conclusion may fail if f is not continuously differentiable at c.

Example 8.38. Define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} x/2 + x^2 \sin(1/x) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Then f is differentiable on $\mathbb{R}$ with

$$f'(x) = \begin{cases} 1/2 - \cos(1/x) + 2x \sin(1/x) & \text{if } x \neq 0, \\ 1/2 & \text{if } x = 0. \end{cases}$$

Every neighborhood of 0 includes intervals where $f' < 0$ or $f' > 0$, in which f is strictly decreasing or strictly increasing, respectively. Thus, despite the fact that $f'(0) > 0$, the function f is not strictly increasing in any neighborhood of 0. As a result, no local inverse of the function f exists on any neighborhood of 0.

8.6. Taylor's theorem

If $f : (a, b) \to \mathbb{R}$ is differentiable on (a, b) and $f' : (a, b) \to \mathbb{R}$ is differentiable, then we define the second derivative $f'' : (a, b) \to \mathbb{R}$ of f as the derivative of f'. We define higher-order derivatives similarly. If f has derivatives $f^{(n)} : (a, b) \to \mathbb{R}$ of all orders $n \in \mathbb{N}$, then we say that f is infinitely differentiable on (a, b).

Taylor's theorem gives an approximation for an $(n + 1)$-times differentiable function in terms of its Taylor polynomial of degree n.

Definition 8.39. Let $f : (a, b) \to \mathbb{R}$ and suppose that f has n derivatives

$$f', f'', \dots f^{(n)} : (a, b) \to \mathbb{R}$$

on (a, b). The Taylor polynomial of degree n of f at $a < c < b$ is

$$P_n(x) = f(c) + f'(c)(x - c) + \frac{1}{2!} f''(c)(x - c)^2 + \cdots + \frac{1}{n!} f^{(n)}(c)(x - c)^n.$$

Equivalently,

$$P_n(x) = \sum_{k=0}^{n} a_k (x - c)^k, \qquad a_k = \frac{1}{k!} f^{(k)}(c).$$

We call a_k the kth Taylor coefficient of f at c. The computation of the Taylor polynomials in the following examples are left as an exercise.

Example 8.40. If $P(x)$ is a polynomial of degree n, then $P_n(x) = P(x)$.

Example 8.41. The Taylor polynomial of degree n of e^x at $x = 0$ is

$$P_n(x) = 1 + x + \frac{1}{2!} x^2 \cdots + \frac{1}{n!} x^n.$$

Example 8.42. The Taylor polynomial of degree $2n$ of $\cos x$ at $x = 0$ is

$$P_{2n}(x) = 1 - \frac{1}{2!} x^2 + \frac{1}{4!} x^4 - \cdots + (-1)^n \frac{1}{(2n)!} x^{2n}.$$

We also have $P_{2n+1} = P_{2n}$ since the Tayor coefficients of odd order are zero.

Example 8.43. The Taylor polynomial of degree $2n + 1$ of $\sin x$ at $x = 0$ is

$$P_{2n+1}(x) = x - \frac{1}{3!} x^3 + \frac{1}{5!} x^5 - \cdots + (-1)^n \frac{1}{(2n + 1)!} x^{2n+1}.$$

We also have $P_{2n+2} = P_{2n+1}$.

Example 8.44. The Taylor polynomial of degree n of $1/x$ at $x = 1$ is

$$P_n(x) = 1 - (x - 1) + (x - 1)^2 - \cdots + (-1)^n (x - 1)^n.$$

Example 8.45. The Taylor polynomial of degree n of $\log x$ at $x = 1$ is

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

We write

$$f(x) = P_n(x) + R_n(x).$$

where R_n is the error, or remainder, between f and its Taylor polynomial P_n. The next theorem is one version of Taylor's theorem, which gives an expression for the remainder due to Lagrange. It can be regarded as a generalization of the mean value theorem, which corresponds to the case $n = 1$. The idea of the proof is to subtract a suitable polynomial from the function and apply Rolle's theorem, just as we proved the mean value theorem by subtracting a suitable linear function.

Theorem 8.46 (Taylor with Lagrange Remainder)**.** Suppose that $f : (a, b) \to \mathbb{R}$ has $n+1$ derivatives on (a, b) and let $a < c < b$. For every $a < x < b$, there exists ξ between c and x such that

$$f(x) = f(c) + f'(c)(x-c) + \frac{1}{2!}f''(c)(x-c)^2 + \cdots + \frac{1}{n!}f^{(n)}(c)(x-c)^n + R_n(x)$$

where

$$R_n(x) = \frac{1}{(n+1)!}f^{(n+1)}(\xi)(x-c)^{n+1}.$$

Proof. Fix $x, c \in (a, b)$. For $t \in (a, b)$, let

$$g(t) = f(x) - f(t) - f'(t)(x-t) - \frac{1}{2!}f''(t)(x-t)^2 - \cdots - \frac{1}{n!}f^{(n)}(t)(x-t)^n.$$

Then $g(x) = 0$ and

$$g'(t) = -\frac{1}{n!}f^{(n+1)}(t)(x-t)^n.$$

Define

$$h(t) = g(t) - \left(\frac{x-t}{x-c}\right)^{n+1} g(c).$$

Then $h(c) = h(x) = 0$, so by Rolle's theorem, there exists a point ξ between c and x such that $h'(\xi) = 0$, which implies that

$$g'(\xi) + (n+1)\frac{(x-\xi)^n}{(x-c)^{n+1}}g(c) = 0.$$

It follows from the expression for g' that

$$\frac{1}{n!}f^{(n+1)}(\xi)(x-\xi)^n = (n+1)\frac{(x-\xi)^n}{(x-c)^{n+1}}g(c),$$

and using the expression for g in this equation, we get the result. $\square$

Note that the remainder term

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

has the same form as the $(n+1)$th-term in the Taylor polynomial of f, except that the derivative is evaluated at a (typically unknown) intermediate point ξ between c and x, instead of at c.

Example 8.47. Let us prove that

$$\lim_{x\to 0}\left(\frac{1-\cos x}{x^2}\right) = \frac{1}{2}.$$

By Taylor's theorem,

$$\cos x = 1 - \frac{1}{2}x^2 + \frac{1}{4!}(\cos\xi)x^4$$

for some ξ between 0 and x. It follows that for $x \neq 0$,

$$\frac{1-\cos x}{x^2} - \frac{1}{2} = -\frac{1}{4!}(\cos\xi)x^2.$$

Since $|\cos\xi| \le 1$, we get

$$\left|\frac{1-\cos x}{x^2} - \frac{1}{2}\right| \le \frac{1}{4!}x^2,$$

which implies that

$$\lim_{x\to 0}\left|\frac{1-\cos x}{x^2}-\frac{1}{2}\right| = 0.$$

Note that as well as proving the limit, Taylor's theorem gives an explicit upper bound for the difference between $(1-\cos x)/x^2$ and its limit $1/2$. For example,

$$\left|\frac{1-\cos(0.1)}{(0.1)^2}-\frac{1}{2}\right| \le \frac{1}{2400}.$$

Numerically, we have

$$\frac{1}{2}-\frac{1-\cos(0.1)}{(0.1)^2} \approx 0.00041653, \qquad \frac{1}{2400} \approx 0.00041667.$$

In Section 12.7, we derive an alternative expression for the remainder R_n as an integral.

8.7. * The inverse function theorem

The inverse function theorem gives a sufficient condition for a differentiable function f to be locally invertible at a point c with differentiable inverse: namely, that f is *continuously* differentiable at c and $f'(c) \neq 0$. Example 8.24 shows that one cannot expect the inverse of a differentiable function f to exist locally at c if $f'(c) = 0$, while Example 8.38 shows that the condition $f'(c) \neq 0$ is not, on its own, sufficient to imply the existence of a local inverse.

Before stating the theorem, we give a precise definition of local invertibility.

Definition 8.48. A function $f : A \to \mathbb{R}$ is locally invertible at an interior point $c \in A$ if there exist open neighborhoods U of c and V of $f(c)$ such that $f|_U : U \to V$ is one-to-one and onto, in which case f has a local inverse $(f|_U)^{-1} : V \to U$.

The following examples illustrate the definition.

Example 8.49. If $f : \mathbb{R} \to \mathbb{R}$ is the square function $f(x) = x^2$, then a local inverse at $c = 2$ with $U = (1, 3)$ and $V = (1, 9)$ is defined by

$$(f|_U)^{-1}(y) = \sqrt{y}.$$

Similarly, a local inverse at $c = -2$ with $U = (-3, -1)$ and $V = (1, 9)$ is defined by

$$(f|_U)^{-1}(y) = -\sqrt{y}.$$

In defining a local inverse at c, we require that it maps an open neighborhood V of $f(c)$ onto an open neighborhood U of c; that is, we want $(f|_U)^{-1}(y)$ to be "close" to c when y is "close" to $f(c)$, not some more distant point that f also maps "close" to $f(c)$. Thus, the one-to-one, onto function g defined by

$$g : (1, 9) \to (-2, -1) \cup [2, 3), \qquad g(y) = \begin{cases} -\sqrt{y} & \text{if } 1 < y < 4 \\ \sqrt{y} & \text{if } 4 \le y < 9 \end{cases}$$

is not a local inverse of f at $c = 2$ in the sense of Definition 8.48, even though $g(f(2)) = 2$ and both compositions

$$f \circ g : (1, 9) \to (1, 9), \qquad g \circ f : (-2, -1) \cup [2, 3) \to (-2, -1) \cup [2, 3)$$

are identity maps, since $U = (-2, -1) \cup [2, 3)$ is not a neighborhood of 2.

Example 8.50. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \cos(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is locally invertible at every $c \in \mathbb{R}$ with $c \neq 0$ or $c \neq 1/(n\pi)$ for some $n \in \mathbb{Z}$.

Theorem 8.51 (Inverse function)**.** Suppose that $f : A \subset \mathbb{R} \to \mathbb{R}$ and $c \in A$ is an interior point of A. If f is differentiable in a neighborhood of c, $f'(c) \neq 0$, and f' is continuous at c, then there are open neighborhoods U of c and V of $f(c)$ such that f has a local inverse $(f|_U)^{-1} : V \to U$. Furthermore, the local inverse function is differentiable at $f(c)$ with derivative

$$[(f|_U)^{-1}]'(f(c)) = \frac{1}{f'(c)}.$$

.

Proof. Suppose, for definiteness, that $f'(c) > 0$ (otherwise, consider $-f$). By the continuity of f', there exists an open interval $U = (a, b)$ containing c on which $f' > 0$. It follows from Theorem 8.36 that f is strictly increasing on U. Writing

$$V = f(U) = (f(a), f(b)),$$

we see that $f|_U : U \to V$ is one-to-one and onto, so f has a local inverse on V, which proves the first part of the theorem.

It remains to prove that the local inverse $(f|_U)^{-1}$, which we denote by f^{-1} for short, is differentiable. First, since f is differentiable at c, we have

$$f(c+h) = f(c) + f'(c)h + r(h)$$

where the remainder r satisfies

$$\lim_{h \to 0} \frac{r(h)}{h} = 0.$$

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

$$|r(h)| \leq \frac{1}{2} f'(c)|h| \qquad \text{for } |h| < \delta.$$

It follows from the differentiability of f that, if $|h| < \delta$,

$$\begin{aligned} f'(c)|h| &= |f(c+h) - f(c) - r(h)| \\ &\leq |f(c+h) - f(c)| + |r(h)| \\ &\leq |f(c+h) - f(c)| + \frac{1}{2} f'(c)|h|. \end{aligned}$$

Absorbing the term proportional to $|h|$ on the right hand side of this inequality into the left hand side and writing

$$f(c+h) = f(c) + k,$$

we find that

$$\frac{1}{2} f'(c)|h| \leq |k| \qquad \text{for } |h| < \delta.$$

Choosing $\delta > 0$ small enough that $(c - \delta, c + \delta) \subset U$, we can express h in terms of k as

$$h = f^{-1}(f(c) + k) - f^{-1}(f(c)).$$

Using this expression in the expansion of f evaluated at $c+h$,

$$f(c+h) = f(c) + f'(c)h + r(h),$$

we get that

$$f(c) + k = f(c) + f'(c)\left[f^{-1}(f(c)+k) - f^{-1}(f(c))\right] + r(h).$$

Simplifying and rearranging this equation, we obtain the corresponding expansion for f^{-1} evaluated at $f(c)+k$,

$$f^{-1}(f(c)+k) = f^{-1}(f(c)) + \frac{1}{f'(c)}k + s(k),$$

where the remainder s is given by

$$s(k) = -\frac{1}{f'(c)}r\left(h\right) = -\frac{1}{f'(c)}r\left(f^{-1}(f(c)+k) - f^{-1}(f(c))\right).$$

Since $f'(c)|h|/2 \leq |k|$, it follows that

$$\frac{|s(k)|}{|k|} \leq \frac{2}{f'(c)^2}\frac{|r(h)|}{|h|}.$$

Therefore, by the "sandwich" theorem and the fact that $h \to 0$ as $k \to 0$,

$$\lim_{k\to 0}\frac{|s(k)|}{|k|} = 0.$$

This result proves that f^{-1} is differentiable at $f(c)$ with

$$\left[f^{-1}\left(f(c)\right)\right]' = \frac{1}{f'(c)}.$$

The expression for the derivative of the inverse also follows from Proposition 8.23, but only once we know that f^{-1} is dfferentiable at $f(c)$. $\square$

One can show that Theorem 8.51 remains true under the weaker hypothesis that the derivative exists and is nonzero in an open neighborhood of c, but in practise, we almost always apply the theorem to continuously differentiable functions.

The inverse function theorem generalizes to functions of several variables, $f : A \subset \mathbb{R}^n \to \mathbb{R}^n$, with a suitable generalization of the derivative of f at c as the linear map $f'(c) : \mathbb{R}^n \to \mathbb{R}^n$ that approximates f near c. A different proof of the existence of a local inverse is required in that case, since one cannot use monotonicity arguments.

As an example of the application of the inverse function theorem, we consider a simple problem from bifurcation theory.

Example 8.52. Consider the transcendental equation

$$y = x - k\left(e^x - 1\right)$$

where $k \in \mathbb{R}$ is a constant parameter. Suppose that we want to solve for $x \in \mathbb{R}$ given $y \in \mathbb{R}$. If $y = 0$, then an obvious solution is $x = 0$. The inverse function theorem applied to the continuously differentiable function $f(x;k) = x - k(e^x - 1)$ implies that there are neighborhoods U, V of 0 (depending on k) such that the equation has a unique solution $x \in U$ for every $y \in V$ provided that the derivative

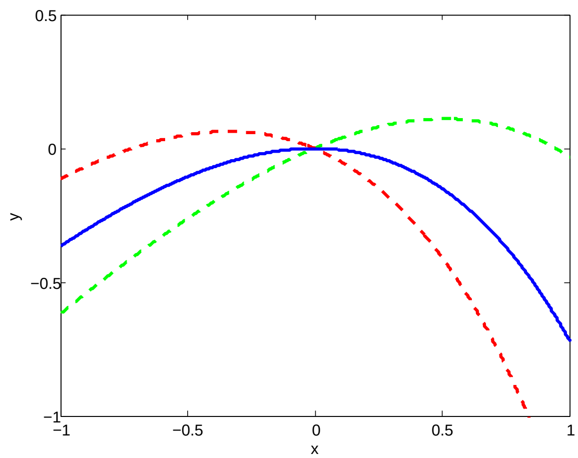

Figure 2. Graph of $y = f(x; k)$ for the function in Example 8.52: (a) $k = 0.5$ (green); (b) $k = 1$ (blue); (c) $k = 1.5$ (red). When y is sufficiently close to zero, there is a unique solution for x in some neighborhood of zero unless $k = 1$.

of f with respect to x at 0, given by $f_x(0; k) = 1 - k$ is non-zero i.e., provided that $k \neq 1$ (see Figure 2).

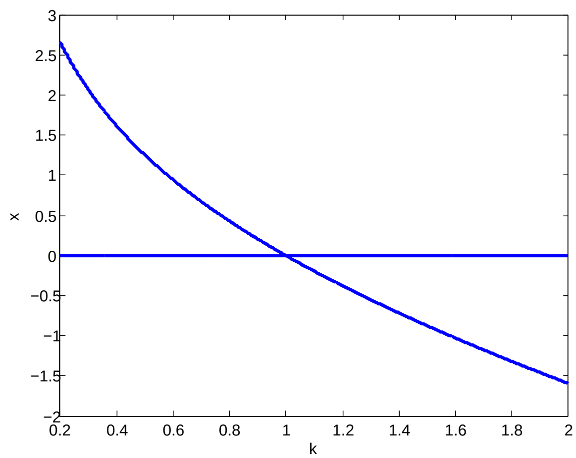

Figure 3. Plot of the solutions for x of the nonlinear equation $x = k(e^x - 1)$ as a function of the parameter k (see Example 8.52). The point $(x, k) = (0, 1)$ where the two solution branches cross is called a bifurcation point.

Alternatively, we can fix a value of y, say $y = 0$, and ask how the solutions of the corresponding equation for x,

$$x - k\,(e^x - 1) = 0,$$

depend on the parameter k. Figure 2 plots the solutions for x as a function of k for $0.2 \leq k \leq 2$. The equation has two different solutions for x unless $k = 1$. The branch of nonzero solutions crosses the branch of zero solution at the point $(x, k) = (0, 1)$, called a bifurcation point. The implicit function theorem, which is a generalization of the inverse function theorem, implies that a necessary condition for a solution (x_0, k_0) of the equation $f(x; k) = 0$ to be a bifurcation point, meaning that the equation fails to have a unique solution branch $x = g(k)$ in some neighborhood of (x_0, k_0), is that $f_x(x_0; k_0) = 0$.

8.8. * L'Hôpital's rule

In this section, we prove a rule (much beloved by calculus students) for the evaluation of inderminate limits of the form $0/0$ or ∞/∞. Our proof uses the following generalization of the mean value theorem.

Theorem 8.53 (Cauchy mean value)**.** Suppose that $f, g : [a, b] \to \mathbb{R}$ are continuous on the closed, bounded interval $[a, b]$ and differentiable on the open interval (a, b). Then there exists $a < c < b$ such that

$$f'(c)\,[g(b) - g(a)] = [f(b) - f(a)]\,g'(c).$$

Proof. The function $h : [a, b] \to \mathbb{R}$ defined by

$$h(x) = [f(x) - f(a)]\,[g(b) - g(a)] - [f(b) - f(a)]\,[g(x) - g(a)]$$

is continuous on $[a, b]$ and differentiable on (a, b) with

$$h'(x) = f'(x)\,[g(b) - g(a)] - [f(b) - f(a)]\,g'(x).$$

Moreover, $h(a) = h(b) = 0$. Rolle's Theorem implies that there exists $a < c < b$ such that $h'(c) = 0$, which proves the result. □

If $g(x) = x$, then this theorem reduces to the usual mean value theorem (Theorem 8.33). Next, we state one form of l'Hôpital's rule.

Theorem 8.54 (l'Hôpital's rule: 0/0)**.** Suppose that $f, g : (a, b) \to \mathbb{R}$ are differentiable functions on a bounded open interval (a, b) such that $g'(x) \neq 0$ for $x \in (a, b)$ and

$$\lim_{x \to a^+} f(x) = 0, \qquad \lim_{x \to a^+} g(x) = 0.$$

Then

$$\lim_{x \to a^+} \frac{f'(x)}{g'(x)} = L \quad \text{implies that} \quad \lim_{x \to a^+} \frac{f(x)}{g(x)} = L.$$

Proof. We may extend $f, g : [a, b) \to \mathbb{R}$ to continuous functions on $[a, b)$ by defining $f(a) = g(a) = 0$. If $a < x < b$, then by the mean value theorem, there exists $a < c < x$ such that

$$g(x) = g(x) - g(a) = g'(c)(x - a) \neq 0,$$

so $g \neq 0$ on (a, b). Moreover, by the Cauchy mean value theorem (Theorem 8.53), there exists $a < c < x$ such that

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(c)}.$$

Since $c \to a^+$ as $x \to a^+$, the result follows. (In fact, since $a < c < x$, the δ that "works" for f'/g' also "works" for f/g.) □

Example 8.55. Using l'Hôpital's rule twice (verify that all of the hypotheses are satisfied!), we find that

$$\lim_{x\to 0^+} \frac{1-\cos x}{x^2} = \lim_{x\to 0^+} \frac{\sin x}{2x} = \lim_{x\to 0^+} \frac{\cos x}{2} = \frac{1}{2}.$$

Analogous results and proofs apply to left limits ($x \to a^-$), two-sided limits ($x \to a$), and infinite limits ($x \to \infty$ or $x \to -\infty$). Alternatively, one can reduce these limits to the left limit considered in Theorem 8.54.

For example, suppose that $f, g : (a,\infty) \to \mathbb{R}$ are differentiable, $g' \neq 0$, and $f(x) \to 0$, $g(x) \to 0$ as $x \to \infty$. Assuming that $a > 0$ without loss of generality, we define $F, G : (0, 1/a) \to \mathbb{R}$ by

$$F(t) = f\left(\frac{1}{t}\right), \qquad G(t) = g\left(\frac{1}{t}\right).$$

The chain rule implies that

$$F'(t) = -\frac{1}{t^2} f'\left(\frac{1}{t}\right), \qquad G'(t) = -\frac{1}{t^2} g'\left(\frac{1}{t}\right).$$

Replacing limits as $x \to \infty$ by equivalent limits as $t \to 0^+$ and applying Theorem 8.54 to F, G, all of whose hypothesis are satisfied if the limit of $f'(x)/g'(x)$ as $x \to \infty$ exists, we get

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{t\to 0^+} \frac{F(t)}{G(t)} = \lim_{t\to 0^+} \frac{F'(t)}{G'(t)} = \lim_{x\to\infty} \frac{f'(x)}{g'(x)}.$$

A less straightforward generalization is to the case when g and possibly f have infinite limits as $x \to a^+$. In that case, we cannot simply extend f and g by continuity to the point a. Instead, we introduce two points $a < x < y < b$ and consider the limits $x \to a^+$ followed by $y \to a^+$.

Theorem 8.56 (l'Hôpital's rule: ∞/∞)**.** Suppose that $f, g : (a, b) \to \mathbb{R}$ are differentiable functions on a bounded open interval (a, b) such that $g'(x) \neq 0$ for $x \in (a, b)$ and

$$\lim_{x\to a^+} |g(x)| = \infty.$$

Then

$$\lim_{x\to a^+} \frac{f'(x)}{g'(x)} = L \quad \text{implies that} \quad \lim_{x\to a^+} \frac{f(x)}{g(x)} = L.$$

Proof. Since $|g(x)| \to \infty$ as $x \to a^+$, we have $g \neq 0$ near a, and we may assume without loss of generality that $g \neq 0$ on (a, b). If $a < x < y < b$, then the mean value theorem implies that $g(x) - g(y) \neq 0$, since $g' \neq 0$, and the Cauchy mean value theorem implies that there exists $x < c < y$ such that

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(c)}{g'(c)}.$$

We may therefore write

$$\begin{aligned}\frac{f(x)}{g(x)} &= \left[\frac{f(x)-f(y)}{g(x)-g(y)}\right]\left[\frac{g(x)-g(y)}{g(x)}\right] + \frac{f(y)}{g(x)}\\ &= \frac{f'(c)}{g'(c)}\left[1-\frac{g(y)}{g(x)}\right] + \frac{f(y)}{g(x)}.\end{aligned}$$

It follows that

$$\left|\frac{f(x)}{g(x)} - L\right| \le \left|\frac{f'(c)}{g'(c)} - L\right| + \left|\frac{f'(c)}{g'(c)}\right|\left|\frac{g(y)}{g(x)}\right| + \left|\frac{f(y)}{g(x)}\right|.$$

Given $\epsilon > 0$, choose $\delta > 0$ such that

$$\left|\frac{f'(c)}{g'(c)} - L\right| < \epsilon \qquad \text{for } a < c < a+\delta.$$

Then, since $a < c < y$, we have for all $a < x < y < a + \delta$ that

$$\left|\frac{f(x)}{g(x)} - L\right| < \epsilon + (|L| + \epsilon)\left|\frac{g(y)}{g(x)}\right| + \left|\frac{f(y)}{g(x)}\right|.$$

Fixing y, taking the lim sup of this inequality as $x \to a^+$, and using the assumption that $|g(x)| \to \infty$, we find that

$$\limsup_{x\to a^+}\left|\frac{f(x)}{g(x)} - L\right| \le \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have

$$\limsup_{x\to a^+}\left|\frac{f(x)}{g(x)} - L\right| = 0,$$

which proves the result.

Alternatively, instead of using the lim sup, we can verify the limit explicitly by an "$\epsilon/3$"-argument. Given $\epsilon > 0$, choose $\eta > 0$ such that

$$\left|\frac{f'(c)}{g'(c)} - L\right| < \frac{\epsilon}{3} \qquad \text{for } a < c < a+\eta,$$

choose $a < y < a + \eta$, and let $\delta_1 = y - a > 0$. Next, choose $\delta_2 > 0$ such that

$$|g(x)| > \frac{3}{\epsilon}\left(|L| + \frac{\epsilon}{3}\right)|g(y)| \qquad \text{for } a < x < a+\delta_2,$$

and choose $\delta_3 > 0$ such that

$$|g(x)| > \frac{3}{\epsilon}|f(y)| \qquad \text{for } a < x < a+\delta_3.$$

Let $\delta = \min(\delta_1, \delta_2, \delta_3) > 0$. Then for $a < x < a + \delta$, we have

$$\left|\frac{f(x)}{g(x)} - L\right| \le \left|\frac{f'(c)}{g'(c)} - L\right| + \left|\frac{f'(c)}{g'(c)}\right|\left|\frac{g(y)}{g(x)}\right| + \left|\frac{f(y)}{g(x)}\right| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3},$$

which proves the result. □

We often use this result when both $f(x)$ and $g(x)$ diverge to infinity as $x \to a^+$, but no assumption on the behavior of $f(x)$ is required.

As for the previous theorem, analogous results and proofs apply to other limits ($x \to a^-$, $x \to a$, or $x \to \pm\infty$). There are also versions of l'Hôpital's rule that imply the divergence of $f(x)/g(x)$ to $\pm\infty$, but we consider here only the case of a finite limit L.

Example 8.57. Since $e^x \to \infty$ as $x \to \infty$, we get by l'Hôpital's rule that

$$\lim_{x\to\infty} \frac{x}{e^x} = \lim_{x\to\infty} \frac{1}{e^x} = 0.$$

Similarly, since $x \to \infty$ as as $x \to \infty$, we get by l'Hôpital's rule that

$$\lim_{x\to\infty} \frac{\log x}{x} = \lim_{x\to\infty} \frac{1/x}{1} = 0.$$

That is, e^x grows faster than x and $\log x$ grows slower than x as $x \to \infty$. We also write these limits using "little oh" notation as $x = o(e^x)$ and $\log x = o(x)$ as $x \to \infty$.

Finally, we note that one cannot use l'Hôpital's rule "in reverse" to deduce that f'/g' has a limit if f/g has a limit.

Example 8.58. Let $f(x) = x + \sin x$ and $g(x) = x$. Then $f(x), g(x) \to \infty$ as $x \to \infty$ and

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \left(1 + \frac{\sin x}{x}\right) = 1,$$

but the limit

$$\lim_{x\to\infty} \frac{f'(x)}{g'(x)} = \lim_{x\to\infty} (1 + \cos x)$$

does not exist.