The Derivative of a Function

7

Starting from the problem to define the tangent to the graph of a function, we introduce the derivative of a function. Two points on the graph can always be joined by a secant, which is a good model for the tangent whenever these points are close to each other. In a limiting process, the secant (discrete model) is replaced by the tangent (continuous model). Differential calculus, which is based on this limiting process, has become one of the most important building blocks of mathematical modelling.

In this section we discuss the derivative of important elementary functions as well as general differentiation rules. Thanks to the meticulous implementation of these rules, expert systems such as maple have become helpful tools in mathematical analysis. Furthermore, we will discuss the interpretation of the derivative as linear approximation and as rate of change. These interpretations form the basis of numerous applications in science and engineering.

The concept of the numerical derivative follows the opposite direction. The continuous model is discretised, and the derivative is replaced by a difference quotient. We carry out a detailed error analysis which allows us to find an optimal approximation. Further, we will illustrate the relevance of symmetry in numerical procedures.

7.1 Motivation

Example 7.1 (The free fall according to Galilei[1]) Imagine an object, which released at time $t = 0$, falls down under the influence of gravity. We are interested in the position $s(t)$ of the object at time $t \geq 0$ as well as in its velocity $v(t)$, see Fig. 7.1. Due to the definition of velocity as change in travelled distance divided by change

[1] G. Galilei, 1564–1642.

M. Oberguggenberger and A. Ostermann, *Analysis for Computer Scientists*,
Undergraduate Topics in Computer Science,
https://doi.org/10.1007/978-3-319-91155-7_7

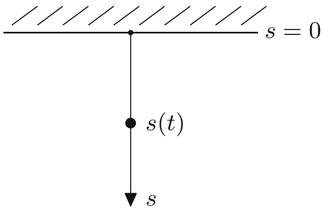

Fig. 7.1 The free fall

in time, the object has the *average velocity*

$$v_{\text{average}} = \frac{s(t + \Delta t) - s(t)}{\Delta t}$$

in the time interval $[t, t + \Delta t]$. In order to obtain the *instantaneous velocity* $v = v(t)$ we take the limit $\Delta t \to 0$ in the above formula and arrive at

$$v(t) = \lim_{\Delta t \to 0} \frac{s(t + \Delta t) - s(t)}{\Delta t}.$$

Galilei discovered through his experiments that the travelled distance in free fall increases quadratically with the time passed, i.e. the law

$$s(t) = \frac{g}{2} t^2$$

with $g \approx 9.81$ m/s^2 holds. Thus we obtain the expression

$$v(t) = \lim_{\Delta t \to 0} \frac{\frac{g}{2}(t + \Delta t)^2 - \frac{g}{2}t^2}{\Delta t} = \frac{g}{2} \lim_{\Delta t \to 0} \big(2t + \Delta t\big) = gt$$

for the instantaneous velocity. The velocity is hence proportional to the time passed.

Example 7.2 (The tangent problem) Consider a real function f and two different points $P = (x_0, f(x_0))$ and $Q = (x, f(x))$ on the graph of the function. The uniquely defined straight line through these two points is called *secant* of the function f through P and Q, see Fig. 7.2. The slope of the secant is given by the *difference quotient*

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

As x tends to x_0, the secant graphically turns into the tangent, provided the limit exists. Motivated by this idea we define the slope

$$k = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

of the function f at x_0. If this limit exists, we call the straight line

$$y = k \cdot (x - x_0) + f(x_0)$$

the *tangent* to the graph of the function at the point $(x_0, f(x_0))$.

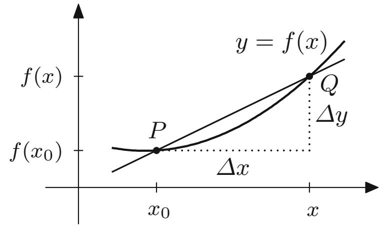

Fig. 7.2 Slope of the secant

Experiment 7.3 Plot the function $f(x) = x^2$ on the interval $[0, 2]$ in MATLAB. Draw the straight lines through the points $(1, 1)$, $(2, z)$ for various values of z. Adjust z until you find the tangent to the graph of the function f at $(1, 1)$ and read off its slope.

7.2 The Derivative

Motivated by the above applications we are going to define the derivative of a real-valued function.

Definition 7.4 (Derivative) Let $I \subset \mathbb{R}$ be an open interval, $f : I \to \mathbb{R}$ a real-valued function and $x_0 \in I$.

(a) The function f is called *differentiable* at x_0 if the difference quotient

$$\frac{\Delta y}{\Delta x} = \frac{f(x) - f(x_0)}{x - x_0}$$

has a (finite) limit for $x \to x_0$. In this case one writes

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

and calls the limit *derivative of f at the point x_0*.

(b) The function f is called *differentiable* (in the interval I) if $f'(x)$ exists for all $x \in I$. In this case the function

$$f' : I \to \mathbb{R} : x \mapsto f'(x)$$

is called the *derivative of f*. The process of computing f' from f is called *differentiation*.

In place of $f'(x)$ one often writes $\frac{\mathrm{d}f}{\mathrm{d}x}(x)$ or $\frac{\mathrm{d}}{\mathrm{d}x} f(x)$, respectively. The following examples show how the derivative of a function is obtained by means of the limiting process above.

Example 7.5 (The constant function $f(x) = c$)

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

The derivative of a constant function is zero.

Example 7.6 (The affine function $g(x) = ax + b$)

$$g'(x) = \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} = \lim_{h \to 0} \frac{ax + ah + b - ax - b}{h} = \lim_{h \to 0} a = a.$$

The derivative is the slope a of the straight line $y = ax + b$.

Example 7.7 (The derivative of the quadratic function $y = x^2$)

$$y' = \lim_{h \to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h \to 0} \frac{2hx + h^2}{h} = \lim_{h \to 0} (2x + h) = 2x.$$

Similarly, one can show for the power function (with $n \in \mathbb{N}$):

$$f(x) = x^n \quad \Rightarrow \quad f'(x) = n \cdot x^{n-1}.$$

Example 7.8 (The derivative of the square root function $y = \sqrt{x}$ for $x > 0$)

$$y' = \lim_{\xi \to x} \frac{\sqrt{\xi} - \sqrt{x}}{\xi - x} = \lim_{\xi \to x} \frac{\sqrt{\xi} - \sqrt{x}}{(\sqrt{\xi} - \sqrt{x})(\sqrt{\xi} + \sqrt{x})} = \lim_{\xi \to x} \frac{1}{\sqrt{\xi} + \sqrt{x}} = \frac{1}{2\sqrt{x}}.$$

Example 7.9 (Derivatives of the sine and cosine functions) We first recall from Proposition 6.10 that

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

Due to

$$(\cos t - 1)(\cos t + 1) = -\sin^2 t$$

it also holds that

$$\frac{\cos t - 1}{t} = -\underbrace{\sin t}_{\to 0} \cdot \underbrace{\frac{\sin t}{t}}_{\to 1} \cdot \underbrace{\frac{1}{\cos t + 1}}_{\to 1/2} \to 0 \qquad \text{for } t \to 0,$$

and thus

$$\lim_{t \to 0} \frac{\cos t - 1}{t} = 0.$$

Due to the addition theorems (Proposition 3.3) we get with the preparations from above

$$\begin{aligned}
\sin' x &= \lim_{h\to 0} \frac{\sin(x+h)-\sin x}{h} = \lim_{h\to 0} \frac{\sin x\cos h + \cos x \sin h - \sin x}{h} \\
&= \lim_{h\to 0} \sin x \cdot \frac{\cos h - 1}{h} + \lim_{h\to 0} \cos x \cdot \frac{\sin h}{h} \\
&= \sin x \cdot \underbrace{\lim_{h\to 0} \frac{\cos h - 1}{h}}_{=0} + \cos x \cdot \underbrace{\lim_{h\to 0} \frac{\sin h}{h}}_{=1} \\
&= \cos x.
\end{aligned}$$

This proves the formula $\sin' x = \cos x$. Likewise it can be shown that $\cos' x = -\sin x$.

Example 7.10 (The derivative of the exponential function with base e) Rearranging terms in the series expansion of the exponential function (Proposition C.12) we obtain

$$\frac{e^h - 1}{h} = \sum_{k=0}^{\infty} \frac{h^k}{(k+1)!} = 1 + \frac{h}{2} + \frac{h^2}{6} + \frac{h^3}{24} + \cdots$$

From that one infers

$$\left| \frac{e^h - 1}{h} - 1 \right| \le |h| \left(\frac{1}{2} + \frac{|h|}{6} + \frac{|h|^3}{24} + \cdots \right) \le |h| e^{|h|}.$$

Letting $h \to 0$ hence gives the important limit

$$\lim_{h\to 0} \frac{e^h - 1}{h} = 1.$$

The existence of the limit

$$\lim_{h\to 0} \frac{e^{x+h} - e^x}{h} = e^x \cdot \lim_{h\to 0} \frac{e^h - 1}{h} = e^x$$

shows that the exponential function is differentiable and that $(e^x)' = e^x$.

Example 7.11 (New representation of Euler's number) By substituting $y = e^h - 1$, $h = \log(y+1)$ in the above limit one obtains

$$\lim_{y\to 0} \frac{y}{\log(y+1)} = 1$$

and in this way

$$\lim_{y\to 0} \log(1+\alpha y)^{1/y} = \lim_{y\to 0} \frac{\log(1+\alpha y)}{y} = \alpha \lim_{y\to 0} \frac{\log(1+\alpha y)}{\alpha y} = \alpha.$$

Due to the continuity of the exponential function it further follows that

$$\lim_{y \to 0} (1 + \alpha y)^{1/y} = \mathrm{e}^{\alpha}.$$

In particular, for $y = 1/n$, we obtain a new representation of the exponential function

$$\mathrm{e}^{\alpha} = \lim_{n \to \infty} \left(1 + \frac{\alpha}{n}\right)^{n}.$$

For $\alpha = 1$ the identity

$$\mathrm{e} = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^{n} = \sum_{k=0}^{\infty} \frac{1}{k!} = 2.718281828459...$$

follows.

Example 7.12 Not every continuous function is differentiable. For instance, the function

$$f(x) = |x| = \begin{cases} x, & x \ge 0 \\ -x, & x \le 0 \end{cases}$$

is not differentiable at the vertex $x = 0$, see Fig. 7.3, left picture. However, it is differentiable for $x \neq 0$ with

$$(|x|)' = \begin{cases} 1, & \text{if } x > 0 \\ -1, & \text{if } x < 0. \end{cases}$$

The function $g(x) = \sqrt[3]{x}$ is not differentiable at $x = 0$ either. The reason for that is the vertical tangent, see Fig. 7.3, right picture.

There are even continuous functions that are nowhere differentiable. It is possible to write down such functions in the form of certain intricate infinite series. However, an analogous example of a (continuous) *curve in the plane* which is nowhere differentiable is the boundary of *Koch's snowflake*, which can be constructed in a simple geometric manner, see Examples 9.9 and 14.17.

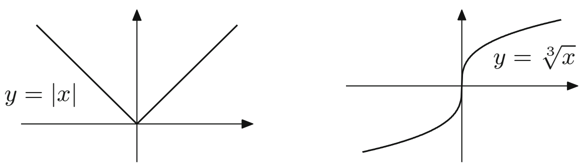

Fig. 7.3 Functions that are not differentiable at $x = 0$

Definition 7.13 If the function f' is again differentiable then

$$f''(x) = \frac{d^2}{dx^2} f(x) = \frac{d^2 f}{dx^2}(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h}$$

is called the *second derivative* of f with respect to x. Likewise higher derivatives are defined recursively as

$$f'''(x) = \bigl(f''(x)\bigr)' \quad \text{or} \quad \frac{d^3}{dx^3} f(x) = \frac{d}{dx}\left(\frac{d^2}{dx^2} f(x)\right), \quad \text{etc.}$$

Differentiating with maple. Using maple one can differentiate expressions as well as functions. If the expression g is of the form

```
g := x^2 - a*x;
```

then the corresponding function f is defined by

```
f := x -> x^2 - a*x;
```

The evaluation of functions generates expressions, for example `f(t)` produces the expression $t^2 - at$. Conversely, expressions can be converted to functions using `unapply`

```
h := unapply(g,x);
```

The derivative of expressions can be obtained using `diff`, those of functions using `D`. Examples can be found in the maple worksheet `mp07_1.mws`.

7.3 Interpretations of the Derivative

We introduced the derivative geometrically as the slope of the tangent, and we saw that the tangent to a graph of a differentiable function f at the point $(x_0, f(x_0))$ is given by

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

Example 7.14 Let $f(x) = x^4 + 1$ with derivative $f'(x) = 4x^3$.

(i) The tangent to the graph of f at the point $(0, 1)$ is

$$y = f'(0) \cdot (x - 0) + f(0) = 1$$

and thus horizontal.

(ii) The tangent to the graph of f at the point $(1, 2)$ is

$$y = f'(1)(x - 1) + 2 = 4(x - 1) + 2 = 4x - 2.$$

The derivative allows further interpretations.

Interpretation as linear approximation. We start off by emphasising that every differentiable function f can be written in the form

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + R(x, x_0),$$

where the remainder $R(x, x_0)$ has the property

$$\lim_{x \to x_0} \frac{R(x, x_0)}{x - x_0} = 0.$$

This follows immediately from

$$R(x, x_0) = f(x) - f(x_0) - f'(x_0)(x - x_0)$$

by dividing by $x - x_0$, since

$$\frac{f(x) - f(x_0)}{x - x_0} \to f'(x_0) \qquad \text{as } x \to x_0.$$

Application 7.15 As we have just seen, a differentiable function f is characterised by the property that

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + R(x, x_0),$$

where the remainder term $R(x, x_0)$ tends faster to zero than $x - x_0$. Taking the limit $x \to x_0$ in this equation shows in particular that *every differentiable function is continuous*.

Application 7.16 Let g be the function given by

$$g(x) = k \cdot (x - x_0) + f(x_0).$$

Its graph is the straight line with slope k passing through the point $(x_0, f(x_0))$. Since

$$\frac{f(x) - g(x)}{x - x_0} = \frac{f(x) - f(x_0) - k \cdot (x - x_0)}{x - x_0} = f'(x_0) - k + \underbrace{\frac{R(x, x_0)}{x - x_0}}_{\to 0}$$

as $x \to x_0$, the tangent with $k = f'(x_0)$ is the straight line which approximates the graph best. One therefore calls

$$g(x) = f(x_0) + f'(x_0) \cdot (x - x_0)$$

the *linear approximation* to f at x_0. For x close to x_0 one can consider $g(x)$ as a good approximation to $f(x)$. In applications the (possibly complicated) function f is often replaced by its linear approximation g which is easier to handle.

Example 7.17 Let $f(x) = \sqrt{x} = x^{1/2}$. Consequently,

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

We want to find the linear approximation to the function f at $x_0 = a$. According to the formula above it holds that

$$\sqrt{x} \approx g(x) = \sqrt{a} + \frac{1}{2\sqrt{a}}(x - a)$$

for x close to a, or, alternatively with $h = x - a$,

$$\sqrt{a + h} \approx \sqrt{a} + \frac{1}{2\sqrt{a}} h \qquad \text{for small } h.$$

If we now substitute $a = 1$ and $h = 0.1$, we obtain the approximation

$$\sqrt{1.1} \approx 1 + \frac{0.1}{2} = 1.05.$$

The first digits of the actual value are 1.0488...

Physical interpretation as rate of change. In physical applications the derivative often plays the role of a rate of change. A well-known example from everyday life is the *velocity*, see Sect. 7.1. Consider a particle which is moving along a straight line. Let $s(t)$ be the position where the particle is at time t. The average velocity is given by the quotient

$$\frac{s(t) - s(t_0)}{t - t_0} \qquad \text{(difference in displacement divided by difference in time).}$$

In the limit $t \to t_0$ the average velocity turns into the *instantaneous velocity*

$$v(t_0) = \frac{\mathrm{d}s}{\mathrm{d}t}(t_0) = \dot{s}(t_0) = \lim_{t \to t_0} \frac{s(t) - s(t_0)}{t - t_0}.$$

Note that one often writes $\dot{f}(t)$ instead of $f'(t)$ if the time t is the argument of the function f. In particular, in physics the *dot notation* is most commonly used.

Likewise one obtains the acceleration by differentiating the velocity

$$a(t) = \dot{v}(t) = \ddot{s}(t).$$

The notion of velocity is also used in the modelling of other processes that vary over time, e.g. for growth or decay.

7.4 Differentiation Rules

In this section $I \subset \mathbb{R}$ denotes an open interval. We first note that differentiation is a *linear* process.

Proposition 7.18 (Linearity of the derivative) *Let $f, g : I \to \mathbb{R}$ be two functions which are differentiable at $x \in I$ and take $c \in \mathbb{R}$. Then the functions $f + g$ and $c \cdot f$ are differentiable at x as well and*

$$\big(f(x) + g(x)\big)' = f'(x) + g'(x),$$
$$\big(cf(x)\big)' = cf'(x).$$

Proof The result follows from the corresponding rules for limits. The first statement is true because

$$\frac{f(x+h) + g(x+h) - (f(x) + g(x))}{h} = \underbrace{\frac{f(x+h) - f(x)}{h}}_{\to f'(x)} + \underbrace{\frac{g(x+h) - g(x)}{h}}_{\to g'(x)}$$

as $h \to 0$. The second statement follows similarly. □

Linearity together with the differentiation rule $(x^m)' = m\,x^{m-1}$ for powers implies that every polynomial is differentiable. Let

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0.$$

Then its derivative has the form

$$p'(x) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + a_1.$$

For example, $(3x^7 - 4x^2 + 5x - 1)' = 21x^6 - 8x + 5$.

The following two rules allow one to determine the derivative of products and quotients of functions from their factors.

Proposition 7.19 (Product rule) *Let $f, g : I \to \mathbb{R}$ be two functions which are differentiable at $x \in I$. Then the function $f \cdot g$ is differentiable at x and*

$$\big(f(x) \cdot g(x)\big)' = f'(x) \cdot g(x) + f(x) \cdot g'(x).$$

Proof This fact follows again from the corresponding rules for limits

$$\begin{aligned}
&\frac{f(x+h)\cdot g(x+h) - f(x)\cdot g(x)}{h} \\
&\quad = \frac{f(x+h)\cdot g(x+h) - f(x)\cdot g(x+h)}{h} + \frac{f(x)\cdot g(x+h) - f(x)\cdot g(x)}{h} \\
&\quad = \underbrace{\frac{f(x+h) - f(x)}{h}}_{\to f'(x)} \cdot \underbrace{g(x+h)}_{\to g(x)} + f(x)\cdot \underbrace{\frac{g(x+h) - g(x)}{h}}_{\to g'(x)}
\end{aligned}$$

as $h \to 0$. The required continuity of g at x is a consequence of Application 7.15. □

Proposition 7.20 (Quotient rule) *Let $f, g : I \to \mathbb{R}$ be two functions differentiable at $x \in I$ and $g(x) \neq 0$. Then the quotient $\frac{f}{g}$ is differentiable at the point x and*

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

In particular,

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

The proof is similar to the one for the product rule and can be found in [3, Chap. 3.1], for example.

Example 7.21 An application of the quotient rule to $\tan x = \dfrac{\sin x}{\cos x}$ shows that

$$\tan' x = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} = 1 + \tan^2 x.$$

Complicated functions can often be written as a composition of simpler functions. For example, the function

$$h : [2, \infty) \to \mathbb{R} : x \mapsto h(x) = \sqrt{\log(x-1)}$$

can be interpreted as $h(x) = f(g(x))$ with

$$f : [0, \infty) \to \mathbb{R} : y \mapsto \sqrt{y}, \qquad g : [2, \infty) \to [0, \infty) : x \mapsto \log(x-1).$$

One denotes the composition of the functions f and g by $h = f \circ g$. The following proposition shows how such compound functions can be differentiated.

Proposition 7.22 (Chain rule) *The composition of two differentiable functions $g : I \to B$ and $f : B \to \mathbb{R}$ is also differentiable and*

$$\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) = f'(g(x)) \cdot g'(x).$$

In shorthand notation the rule is

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

Proof We write

$$\begin{aligned}\frac{1}{h}\Big(f(g(x+h)) - f(g(x))\Big) &= \frac{f(g(x+h)) - f(g(x))}{g(x+h) - g(x)} \cdot \frac{g(x+h) - g(x)}{h} \\ &= \frac{f(g(x)+k) - f(g(x))}{k} \cdot \frac{g(x+h) - g(x)}{h},\end{aligned}$$

where, due to the interpretation as a linear approximation (see Sect. 7.3), the expression

$$k = g(x+h) - g(x)$$

is of the form

$$k = g'(x)h + R(x+h, x)$$

and tends to zero itself as $h \to 0$. It follows that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x} f(g(x)) &= \lim_{h\to 0} \frac{1}{h}\Big(f(g(x+h)) - f(g(x))\Big) \\ &= \lim_{h\to 0}\left(\frac{f(g(x)+k) - f(g(x))}{k} \cdot \frac{g(x+h) - g(x)}{h}\right) \\ &= f'(g(x)) \cdot g'(x)\end{aligned}$$

and hence the assertion of the proposition. □

The differentiation of a composite function $h(x) = f(g(x))$ is consequently performed in three steps:

1. Identify the *outer* function f and the *inner* function g with $h(x) = f(g(x))$.
2. Differentiate the outer function f at the point $g(x)$, i.e. compute $f'(y)$ and then substitute $y = g(x)$. The result is $f'(g(x))$.
3. Inner derivative: Differentiate the inner function g and multiply it with the result of step 2. One obtains $h'(x) = f'(g(x)) \cdot g'(x)$.

In the case of three or more compositions, the above rules have to be applied recursively.

Example 7.23 (a) Let $h(x) = (\sin x)^3$. We identify the outer function $f(y) = y^3$ and the inner function $g(x) = \sin x$. Then

$$h'(x) = 3\,(\sin x)^2 \cdot \cos x.$$

(b) Let $h(x) = \mathrm{e}^{-x^2}$. We identify $f(y) = \mathrm{e}^y$ and $g(x) = -x^2$. Thus

$$h'(x) = \mathrm{e}^{-x^2} \cdot (-2x).$$

The last rule that we will discuss concerns the differentiation of the inverse of a differentiable function.

Proposition 7.24 (Inverse function rule) *Let $f : I \to J$ be bijective, differentiable and $f'(y) \neq 0$ for all $y \in I$. Then $f^{-1} : J \to I$ is also differentiable and*

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

In shorthand notation this rule is

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

Proof We set $y = f^{-1}(x)$ and $\eta = f^{-1}(\xi)$. Due to the continuity of the inverse function (see Proposition C.3) we have that $\eta \to y$ as $\xi \to x$. It thus follows that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x} f^{-1}(x) &= \lim_{\xi \to x} \frac{f^{-1}(\xi) - f^{-1}(x)}{\xi - x} = \lim_{\eta \to y} \frac{\eta - y}{f(\eta) - f(y)} \\ &= \lim_{\eta \to y} \left(\frac{f(\eta) - f(y)}{\eta - y}\right)^{-1} = \frac{1}{f'(y)} = \frac{1}{f'(f^{-1}(x))}\end{aligned}$$

and hence the statement of the proposition. □

Figure 7.4 shows the geometric background of the inverse function rule: The slope of a straight line in x-direction is the inverse of the slope in y-direction.

If it is known beforehand that the inverse function is differentiable then its derivative can also be obtained in the following way. One differentiates the identity

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

with respect to x using the chain rule. This yields

$$1 = f'(f^{-1}(x)) \cdot (f^{-1})'(x)$$

and one obtains the inverse rule by division by $f'(f^{-1}(x))$.

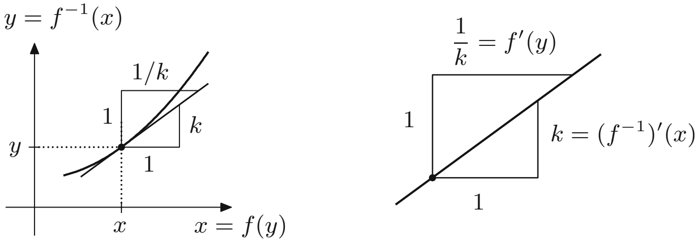

Fig. 7.4 Derivative of the inverse function with detailed view of the slopes

Example 7.25 (Derivative of the logarithm) Since $y = \log x$ is the inverse function to $x = \mathrm{e}^y$, it follows from the inverse function rule that

$$\big(\log x\big)' = \frac{1}{\mathrm{e}^{\log x}} = \frac{1}{x}$$

for $x > 0$. Furthermore

$$\log|x| = \begin{cases} \log x, & x > 0, \\ \log(-x), & x < 0, \end{cases}$$

and thus

$$\big(\log|x|\big)' = \begin{cases} \big(\log x\big)' = \dfrac{1}{x}, & x > 0, \\ \big(\log(-x)\big)' = \dfrac{1}{(-x)} \cdot (-1) = \dfrac{1}{x}, & x < 0. \end{cases}$$

Altogether one obtains the formula

$$\big(\log|x|\big)' = \frac{1}{x} \quad \text{for} \quad x \neq 0.$$

For logarithms to the base a one has

$$\log_a x = \frac{\log x}{\log a}, \quad \text{thus} \quad \big(\log_a x\big)' = \frac{1}{x \log a}.$$

Example 7.26 (Derivatives of general power functions) From $x^\alpha = \mathrm{e}^{\alpha \log x}$ we infer by the chain rule that

$$\big(x^\alpha\big)' = \mathrm{e}^{\alpha \log x} \cdot \frac{\alpha}{x} = x^\alpha \cdot \frac{\alpha}{x} = \alpha\ x^{\alpha-1}.$$

Example 7.27 (Derivative of the general exponential function) For $a > 0$ we have $a^x = \mathrm{e}^{x \log a}$. An application of the chain rule shows that

$$\big(a^x\big)' = \big(\mathrm{e}^{x \log a}\big)' = \mathrm{e}^{x \log a} \cdot \log a = a^x \log a.$$

Example 7.28 For $x > 0$ we have $x^x = \mathrm{e}^{x \log x}$ and thus

$$\left(x^x\right)' = \mathrm{e}^{x \log x}\left(\log x + \frac{x}{x}\right) = x^x \left(\log x + 1\right).$$

Example 7.29 (Derivatives of cyclometric functions) We recall the differentiation rules for the trigonometric functions on their principal branches:

$$\begin{aligned}
(\sin x)' &= \cos x = \sqrt{1 - \sin^2 x}, & -\tfrac{\pi}{2} &\le x \le \tfrac{\pi}{2},\\
(\cos x)' &= -\sin x = -\sqrt{1 - \cos^2 x}, & 0 &\le x \le \pi,\\
(\tan x)' &= 1 + \tan^2 x, & -\tfrac{\pi}{2} &< x < \tfrac{\pi}{2}.
\end{aligned}$$

The inverse function rule thus yields

$$\begin{aligned}
(\arcsin x)' &= \frac{1}{\sqrt{1 - \sin^2(\arcsin x)}} = \frac{1}{\sqrt{1 - x^2}}, & -1 < x < 1,\\
(\arccos x)' &= \frac{-1}{\sqrt{1 - \cos^2(\arccos x)}} = -\frac{1}{\sqrt{1 - x^2}}, & -1 < x < 1,\\
(\arctan x)' &= \frac{1}{1 + \tan^2(\arctan x)} = \frac{1}{1 + x^2}, & -\infty < x < \infty.
\end{aligned}$$

Example 7.30 (Derivatives of hyperbolic and inverse hyperbolic functions) The derivative of the hyperbolic sine is readily computed by invoking the defining formula:

$$(\sinh x)' = \left(\frac{1}{2}\left(\mathrm{e}^x - \mathrm{e}^{-x}\right)\right)' = \frac{1}{2}\left(\mathrm{e}^x + \mathrm{e}^{-x}\right) = \cosh x.$$

The derivative of the hyperbolic cosine is obtained in the same way; for differentiating the hyperbolic tangent, the quotient rule is to be applied (see Exercise 3):

$$(\cosh x)' = \sinh x, \quad (\tanh x)' = 1 - \tanh^2 x.$$

The derivative of the inverse hyperbolic sine can be computed by means of the inverse function rule:

$$(\operatorname{arsinh} x)' = \frac{1}{\cosh(\operatorname{arsinh} x)} = \frac{1}{\sqrt{1 + \sinh^2(\operatorname{arsinh} x)}} = \frac{1}{\sqrt{1 + x^2}}$$

for $x \in \mathbb{R}$, where we have used the identity $\cosh^2 x - \sinh^2 x = 1$. In a similar way, the derivatives of the other inverse hyperbolic functions can be computed on their respective domains (Exercise 3):

$$\begin{aligned}
(\operatorname{arcosh} x)' &= \frac{1}{\sqrt{x^2 - 1}}, & x &> 1,\\
(\operatorname{artanh} x)' &= \frac{1}{1 - x^2}, & -1 &< x < 1.
\end{aligned}$$

Table 7.1 Derivatives of the elementary functions ($\alpha \in \mathbb{R}$, $a > 0$)

$f(x)$	1	x^α	e^x	a^x	$\log\|x\|$	$\log_a x$
$f'(x)$	0	$\alpha x^{\alpha-1}$	e^x	$a^x \log a$	$\dfrac{1}{x}$	$\dfrac{1}{x \log a}$
$f(x)$	$\sin x$	$\cos x$	$\tan x$	$\arcsin x$	$\arccos x$	$\arctan x$
$f'(x)$	$\cos x$	$-\sin x$	$1 + \tan^2 x$	$\dfrac{1}{\sqrt{1-x^2}}$	$\dfrac{-1}{\sqrt{1-x^2}}$	$\dfrac{1}{1+x^2}$
$f(x)$	$\sinh x$	$\cosh x$	$\tanh x$	$\operatorname{arsinh} x$	$\operatorname{arcosh} x$	$\operatorname{artanh} x$
$f'(x)$	$\cosh x$	$\sinh x$	$1 - \tanh^2 x$	$\dfrac{1}{\sqrt{1+x^2}}$	$\dfrac{1}{\sqrt{x^2-1}}$	$\dfrac{1}{1-x^2}$

The derivatives of the most important elementary functions are collected in Table 7.1. The formulas are valid on the respective domains.

7.5 Numerical Differentiation

In applications it often happens that a function can be evaluated for arbitrary arguments, but no analytic formula is known which represents the function. This situation, for example, arises if the dependent variable is determined using a measuring instrument, e.g. the temperature at a given point as a function of time.

The definition of the derivative as a limit of difference quotients suggests that the derivative of such functions can be approximated by an appropriate difference quotient

$$f'(a) \approx \frac{f(a+h) - f(a)}{h}.$$

The question is how small h should be chosen. In order to decide this we will first carry out a numerical experiment.

Experiment 7.31 Use the above formula to approximate the derivative $f'(a)$ of $f(x) = \mathrm{e}^x$ at $a = 1$. Consider different values of h, for example for $h = 10^{-j}$ with $j = 0, 1, \dots, 16$. One expects a value close to $\mathrm{e} = 2.71828...$ as result. Typical outcomes of such an experiment are listed in Table 7.2.

One sees that the error initially decreases with h, but increases again for smaller h. The reason lies in the representation of numbers on a computer. The experiment was carried out in `IEEE double precision` which corresponds to a relative machine accuracy of $\texttt{eps} \approx 10^{-16}$. The experiment shows that the best result is obtained for

$$h \approx \sqrt{\texttt{eps}} \approx 10^{-8}.$$

Table 7.2 Numerical differentiation of the exponential function at $a = 1$ using a *one-sided* difference quotient. The numerical results and errors are given as functions of h

h	Value	Error
1.000E-000	4.67077427047160	1.95249244201256E-000
1.000E-001	2.85884195487388	1.40560126414838E-001
1.000E-002	2.73191865578714	1.36368273280976E-002
1.000E-003	2.71964142253338	1.35959407433051E-003
1.000E-004	2.71841774708220	1.35918623152431E-004
1.000E-005	2.71829541994577	1.35914867218645E-005
1.000E-006	2.71828318752147	1.35906242526573E-006
1.000E-007	2.71828196740610	1.38947053418548E-007
1.000E-008	2.71828183998415	1.15251088672608E-008
1.000E-009	2.71828219937549	3.70916445113778E-007
1.000E-010	2.71828349976758	1.67130853068187E-006
1.000E-011	2.71829650802524	1.46795661959409E-005
1.000E-012	2.71866817252997	3.86344070924416E-004
1.000E-013	2.71755491373926	-7.26914719783700E-004
1.000E-014	2.73058485544819	1.23030269891471E-002
1.000E-015	3.16240089670572	4.44119068246674E-001
1.000E-016	1.44632569809566	-1.27195613036338E-000

This behaviour can be explained by using *Taylor expansion*. In Chap. 12 we will derive the formula

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2} f''(\xi),$$

where ξ denotes an appropriate point between a and $a+h$. (The value of ξ is usually not known.) Thus, after rearranging, we get

$$f'(a) = \frac{f(a+h) - f(a)}{h} - \frac{h}{2} f''(\xi).$$

The *discretisation error*, i.e. the error which arises from replacing the derivative by the difference quotient, is proportional to h and decreases *linearly* with h. This behaviour can also be seen in the numerical experiment for h between 10^{-2} and 10^{-8}.

For very small h *rounding errors* additionally come into play. As we have seen in Sect. 1.4 the calculation of $f(a)$ on a computer yields

$$\mathrm{rd}(f(a)) = f(a) \cdot (1+\varepsilon) = f(a) + \varepsilon f(a)$$

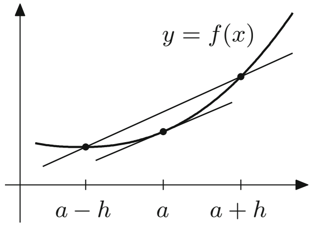

Fig. 7.5 Approximation of the tangent by a symmetric secant

with $|\varepsilon| \le \texttt{eps}$. The rounding error turns out to be proportional to $\texttt{eps}/h$ and increases dramatically for small h. This behaviour can be seen in the numerical experiment for h between 10^{-8} and 10^{-16}.

The result of the numerical derivative using the *one-sided difference quotient*

$$f'(a) \approx \frac{f(a+h) - f(a)}{h}$$

is then most precise if discretisation and rounding error have approximately the same magnitude, so if

$$h \approx \frac{\texttt{eps}}{h} \quad \text{or} \quad h \approx \sqrt{\texttt{eps}} \approx 10^{-8}.$$

In order to calculate the derivative of $f'(a)$ one can also use a secant placed *symmetrically* around $\big(a, f(a)\big)$, i.e.

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a-h)}{2h}$$

This suggests the *symmetric* formula

$$f'(a) \approx \frac{f(a+h) - f(a-h)}{2h}.$$

This approximation is called *symmetric difference quotient* (Fig. 7.5).

To analyse the accuracy of the approximation, we need the Taylor series from Chap. 12:

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2} f''(a) + \frac{h^3}{6} f'''(a) + \cdots$$

If one replaces h by $-h$ in this formula

$$f(a-h) = f(a) - hf'(a) + \frac{h^2}{2} f''(a) - \frac{h^3}{6} f'''(a) + \cdots$$

Table 7.3 Numerical differentiation of the exponential function at $a = 1$ using a *symmetric* difference quotient. The numerical results and errors are given as functions of h

h	Value	Error
1.000E-000	3.19452804946533	4.76246221006280E-001
1.000E-001	2.72281456394742	4.53273548837307E-003
1.000E-002	2.71832713338270	4.53049236583958E-005
1.000E-003	2.71828228150582	4.53046770765297E-007
1.000E-004	2.71828183298958	4.53053283777649E-009
1.000E-005	2.71828182851255	5.35020916458961E-011
1.000E-006	2.71828182834134	-1.17704512803130E-010
1.000E-007	2.71828182903696	5.77919490041268E-010
1.000E-008	2.71828181795317	-1.05058792776447E-008
1.000E-009	2.71828182478364	-3.67540575751946E-009
1.000E-010	2.71828199164235	1.63183308643511E-007
1.000E-011	2.71829103280427	9.20434522511116E-006
1.000E-012	2.71839560410381	1.13775644761560E-004

and takes the difference, one obtains

$$f(a+h) - f(a-h) = 2hf'(a) + 2\frac{h^3}{6}f'''(a) + \cdots$$

and furthermore

$$f'(a) = \frac{f(a+h) - f(a-h)}{2h} - \frac{h^2}{6}f'''(a) + \cdots$$

In this case the discretisation error is hence proportional to h^2, while the rounding error is still proportional to $\texttt{eps}/h$.

The symmetric procedure thus delivers the best results for

$$h^2 \approx \frac{\texttt{eps}}{h} \quad \text{or} \quad h \approx \sqrt[3]{\texttt{eps}},$$

respectively. We repeat Experiment 7.31 with $f(x) = \mathrm{e}^x$, $a = 1$ and $h = 10^{-j}$ for $j = 0, \ldots, 12$. The results are listed in Table 7.3.

As expected one obtains the best result for $h \approx 10^{-5}$. The obtained approximation is more precise than that of Table 7.2. Since symmetric procedures generally give better results, *symmetry* is an important concept in numerical mathematics.

Numerical differentiation of noisy functions. In practice it often occurs that a function which has to be differentiated consists of *discrete* data that are additionally perturbed by a noise. The noise represents small measuring errors and behaves statistically like random numbers.

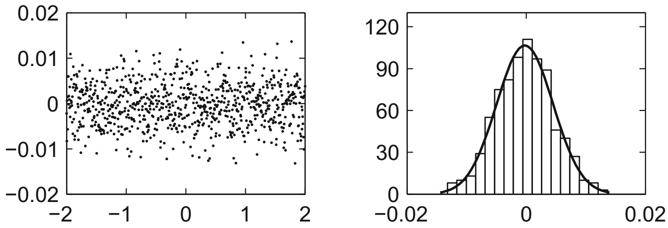

Fig. 7.6 The left picture shows random noise which masks the data. The noise is modelled by 801 normally distributed random numbers. The frequencies of the chosen random numbers can be seen in the histogram in the right picture. For comparison, the (scaled) density of the corresponding normal distribution is given there as well

Example 7.32 Digitising a line of a picture by $J + 1$ pixels produces a function

$$f : \{0, 1, \dots, J\} \to \mathbb{R} : j \mapsto f(j) = f_j = \text{brightness of the } j\text{th pixel.}$$

In order to find an edge in the picture, where the brightness locally changes very rapidly, this function has to be differentiated.

We consider a concrete example. Suppose that the picture information consists of the function

$$g : [a, b] \to \mathbb{R} : x \mapsto g(x) = -2x^3 + 4x$$

with $a = -2$ and $b = 2$. Let $\varDelta x$ be the distance between two pixels and

$$J = \frac{b - a}{\varDelta x}$$

denote the total number of pixels minus 1. We choose $\varDelta x = 1/200$ and thus obtain $J = 800$. The actual brightness of the jth pixel would then be

$$g_j = g(a + j\varDelta x), \qquad 0 \le j \le J.$$

However, due to measuring errors the measuring instrument supplies

$$f_j = g_j + \varepsilon_j,$$

where ε_j are random numbers. We choose normally distributed random numbers with expected value 0 and variance $2.5 \cdot 10^{-5}$ for ε_j, see Fig. 7.6. For an exact definition of the notions of expected value and variance we refer to the literature, for instance [18].

These random numbers can be generated in MATLAB using the command

```
randn(1,801)*sqrt(2.5e-5).
```

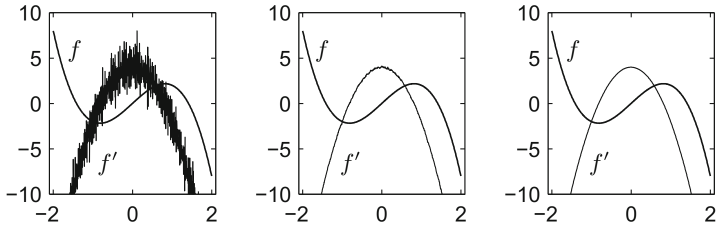

Fig. 7.7 Numerically obtained derivative of a noisy function f, consisting of 801 data values (left); derivative of the same function after filtering using a Gaussian filter (middle) and after smoothing using splines (right)

Differentiating f using the previous rules generates

$$f_j' \approx \frac{f_j - f_{j-1}}{\Delta x} = \frac{g_j - g_{j-1}}{\Delta x} + \frac{\varepsilon_j - \varepsilon_{j-1}}{\Delta x}$$

and the part with g gives the desired value of the derivative, namely

$$\frac{g_j - g_{j-1}}{\Delta x} = \frac{g(a + j\Delta x) - g(a + j\Delta x - \Delta x)}{\Delta x} \approx g'(a + j\Delta x).$$

The sequence of random numbers results in a *non-differentiable* graph. The expression

$$\frac{\varepsilon_j - \varepsilon_{j-1}}{\Delta x}$$

is proportional to $J \cdot \max_{0 \le j \le J} |\varepsilon_j|$. The errors become dominant for large J, see Fig. 7.7, left picture.

To still obtain reliable results, the data have to be smoothed before differentiating. The simplest method is a so-called *convolution* with a *Gaussian filter* which amounts to a weighted averaging of the data (Fig. 7.7, middle). Alternatively one can also use *splines* for smoothing, for example the routine `csaps` in MATLAB. For the right picture in Fig. 7.7 this method has been used.

Experiment 7.33 Generate Fig. 7.7 using the MATLAB program `mat07_1.m` and investigate the influence of the choice of random numbers and the smoothing parameter in `csaps` on the result.

7.6 Exercises

1. Compute the first derivative of the functions

$$f(x) = x^3, \qquad g(t) = \frac{1}{t^2}, \qquad h(x) = \cos x, \qquad k(x) = \frac{1}{\sqrt{x}}, \qquad \ell(t) = \tan t$$

using the definition of the derivative as a limit.

2. Compute the first derivative of the functions

$$a(x) = \frac{x^2-1}{x^2+2x+1}, \qquad b(x) = (x^3-1)\sin^2 x, \quad c(t) = \sqrt{1+t^2}\arctan t,$$
$$d(t) = t^2 \mathrm{e}^{\cos(t^2+1)}, \quad e(x) = x^{2\sin x}, \qquad f(s) = \log\left(s + \sqrt{1+s^2}\right).$$

Check your results with maple.

3. Derive the remaining formulas in Example 7.30. Start by computing the derivatives of the hyperbolic cosine and hyperbolic tangent. Use the inverse function rule to differentiate the inverse hyperbolic cosine and inverse hyperbolic tangent.

4. Compute an approximation of $\sqrt{34}$ by replacing the function $f(x) = \sqrt{x}$ at $x = 36$ by its linear approximation. How accurate is your result?

5. Find the equation of the tangent line to the graph of the function $y = f(x)$ through the point $(x_0, f(x_0))$, where

$$f(x) = \frac{x}{2} + \frac{x}{\log x} \quad \text{and} \quad \text{(a) } x_0 = \mathrm{e}; \quad \text{(b) } x_0 = \mathrm{e}^2.$$

6. Sand runs from a conveyor belt onto a heap with a velocity of 2 m^3/min. The sand forms a cone-shaped pile whose height equals $\frac{4}{3}$ of the radius. With which velocity does the radius grow if the sand cone has a diameter of 6 m?
Hint. Determine the volume V as a function of the radius r, consider V and r as functions of time t and differentiate the equation with respect to t. Compute $\dot{r}$.

7. Use the Taylor series

$$y(x+h) = y(x) + hy'(x) + \frac{h^2}{2}y''(x) + \frac{h^3}{6}y'''(x) + \frac{h^4}{24}y^{(4)}(x) + \cdots$$

to derive the formula

$$y''(x) = \frac{y(x+h) - 2y(x) + y(x-h)}{h^2} - \frac{h^2}{12}y^{(4)}(x) + \cdots$$

and read off from that a numerical method for calculating the second derivative. The discretisation error is proportional to h^2, and the rounding error is proportional to $\texttt{eps}/h^2$. By equating the discretisation and the rounding error deduce the optimal step size h. Check your considerations by performing a numerical experiment in MATLAB, computing the second derivative of $y(x) = \mathrm{e}^{2x}$ at the point $x = 1$.

8. Write a MATLAB program which numerically differentiates a given function on a given interval and plots the function and its first derivative. Test your program on the functions

$$f(x) = \cos x, \qquad 0 \le x \le 6\pi,$$

and

$$g(x) = \mathrm{e}^{-\cos(3x)}, \qquad 0 \le x \le 2.$$

9. Show that the nth derivative of the power function $y = x^n$ equals $n!$ for $n \geq 1$. Verify that the derivative of order $n + 1$ of a polynomial $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ of degree n equals zero.

10. Compute the second derivative of the functions

$$f(x) = \mathrm{e}^{-x^2}, \quad g(x) = \log\left(x + \sqrt{1 + x^2}\right), \quad h(x) = \log\frac{x + 1}{x - 1}.$$