Chapter 3

Derivatives

So far, continuity has given the existence of extreme values on a bounded, closed interval. But how do we actually find these extremas? Obviously, this question occurs in plenty of important problems in "real life": can you maximise a profit? can you minimise material used? can you minimise energy? The answer involves derivatives.

3.1 Definition of the derivative

If we have an extreme value in the interior of an interval then its tangent line has to be horizontal. How do we find such points? How do we find the tangent line? Do all functions have tangent lines at all points? These are the mathematical questions we will have to answer.

We first observe that the *slope of the secant* through $f(x)$ and $f(x+h)$, $h \neq 0$, is given by the *difference quotient*

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

Now, for h approaching 0, ideally $f(x+h)$ slides towards $f(x)$, and the slope of the secant approaches the slope of the tangent line.

Definition 3.1. A function $f : (a,b) \to \mathbb{R}$ is said to be *differentiable* at $x \in (a,b)$ if

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

exists. In this case, the limit is called the *derivative* of f at x, and is denoted by

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

Moreover, the line through $f(x)$ with slope $f'(x)$ is called the *tangent line* of f at x, and the line through $f(x)$ with slope $-\frac{1}{f'(x)}$ is called the *normal line* of f at x if $f'(x) \neq 0$.

Example 3.2. (i) $f(x) = mx + b$, $m, b \in \mathbb{R}$ has $f'(x) = m$:

$$\frac{f(x+h) - f(x)}{h} = \frac{m(x+h) + b - (mx+b)}{h} = \frac{mh}{h} \overset{h \neq 0}{=} m$$

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

(ii) $f(x) = x^2$ has $f'(x) = 2x$:

$$\frac{f(x+h) - f(x)}{h} = \frac{(x+h)^2 - x^2}{h} = \frac{2xh + h^2}{h} \overset{h \neq 0}{=} 2x + h$$

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

(iii) $f(x) = \sqrt{x}$, $x \in [0, \infty)$, is differentiable for $x \neq 0$ with $f'(x) = \frac{1}{2\sqrt{x}}$:

$$\begin{aligned} \frac{f(x+h) - f(x)}{h} &= \frac{\sqrt{x+h} - \sqrt{x}}{h} = \frac{\sqrt{x+h} - \sqrt{x}}{h} \cdot \frac{\sqrt{x+h} + \sqrt{x}}{\sqrt{x+h} + \sqrt{x}} \\ &= \frac{x + h - x}{h(\sqrt{x+h} + \sqrt{x})} \overset{h \neq 0}{=} \frac{1}{\sqrt{x+h} + \sqrt{x}} \end{aligned}$$

so that $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \to 0} \frac{1}{\sqrt{x+h} + \sqrt{x}} = \frac{1}{2\sqrt{x}}$, provided $x \neq 0$.

(iv) $f(x) = \sin x$ has $f'(x) = \cos x$:

$$\begin{aligned} \frac{f(x+h) - f(x)}{h} &= \frac{\sin(x+h) - \sin x}{h} = \frac{\sin x \cos h + \sin h \cos x - \sin x}{h} \\ &= \sin x \left(\frac{\cos h - 1}{h} \right) + \left(\frac{\sin h}{h} \right) \cos x \end{aligned}$$

so that by limit laws $f'(x) = \lim_{h \to 0} \sin x \left(\frac{\cos h - 1}{h} \right) + \lim_{h \to 0} \left(\frac{\sin h}{h} \right) \cos x = \cos x$.

(v) $f(x) = \cos x$ has $f'(x) = -\sin x$.

(vi) $f(x) = \begin{cases} x + 1 & x \geq 0 \\ x - 1 & x < 0 \end{cases}$ is not differentiable at $x = 0$:

For $x = 0$ we have $f(0) = 1$ so that

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

whereas the limit

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

does not exist. Thus, $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ does not exist either! (The function jumps at $x=0$: the secants from $f(h)=h-1$ to $f(0)=1$ have for $h<0$, and near 0, large slope).

You cannot compute the derivative of a function which is defined by cases by computing $f'(x)$ for $x<c$, $f'(x)$ for $x>c$ and showing $\lim_{x\to c-} f'(x) = \lim_{x\to c+} f'(x)$. The above example shows that in this case $\lim_{x\to c-} f'(x) = \lim_{x\to c+} f'(x) = 1$ but f is still not differentiable! Zero marks if you use this wrong argument.

If f is differentiable at c, then the equation of the tangent line at c is given by

$$y = f(c) + f'(c)(x-c)$$

and the equation of the normal line, if $f'(c) \neq 0$, by

$$y = f(c) - \frac{1}{f'(c)}(x-c)\,.$$

(If $f'(c)=0$ then the normal line is the vertical line $x = f(c)$).

The tangent line approximates the graph of f near c. In other words, the function

$$g(x) = f(c) + f'(c)(x-c)$$

approximates the function f near c. We say that g *approximates* f *linearly*.

Note that a function can be continuous at some x but not differentiable: this could happen if the difference quotient diverges to ∞, or if this limit does not exist.

Example 3.3. $f(x)=|x|$ is continuous at $x=0$ but not differentiable at $x=0$.

However, if a function is differentiable then it must be continuous:

Theorem 3.4. *If f is differentiable at x then f is continuous at x.*

Proof. If $h \neq 0$ then

$$f(x+h)-f(x) = \frac{f(x+h)-f(x)}{h} \cdot h\,.$$

By Limit Laws:

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

which shows that $\lim_{h\to 0} f(x+h) = f(x)$ and f is continuous at x. □

Example 3.5. If a point moves on a straight line with constant speed v, then after time Δt the travelled distance is $\Delta s = v\Delta t$. Conversely, the average speed of a point is computed by $v = \frac{\Delta s}{\Delta t}$. If the point moves with variable speed then by considering very small time intervals, one can still assume that the speed in the small interval is roughly constant, and $v(t) = \frac{s(t+\Delta t)-s(t)}{\Delta t}$ in the small time period. Taking the limit, we indeed see that *speed* at time t is given by the derivative $v(t) = s'(t)$.

3.2 Rules of differentiation

As before, we want to use known derivatives to compute new ones.

Theorem 3.6. *Let $\alpha \in \mathbb{R}$, and let $f, g : (a, b) \to \mathbb{R}$ be differentiable at $x \in (a, b)$. Then $f \pm g, \alpha f$ are differentiable at x and*

$$(f \pm g)'(x) = f'(x) \pm g'(x), \quad (\alpha f)'(x) = \alpha f'(x) .$$

Proof. Left for you to do: use the definition of the derivative using limits of difference quotients, and use the limit laws. □

By induction we also have

$$\left(\sum_{k=1}^{n} \alpha_l f_k\right)' = \sum_{k=1}^{n} \alpha_k f_k' .$$

This says that differentiation is a "linear map". However, things become more complicated when considering products of two functions:

Theorem 3.7 (Product rule)**.** *If f, g are differentiable at x then so is $f \cdot g$ and*

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x) .$$

Proof. We have

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

Now, g is continuous (since g is differentiable) and we can apply the Limit Laws:

$$\begin{aligned}
(fg)'(x) &= \lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \cdot \lim_{h\to 0} g(x+h) + f(x) \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} \\
&= f'(x)g(x) + f(x)g'(x) .
\end{aligned}$$

□

Example 3.8. For all $n \in \mathbb{N}$: $f(x) = x^n$ has derivative $f'(x) = nx^{n-1}$.

Corollary 3.9. *A polynomial $P(x) = a_n x^n + \ldots + a_0$ is differentiable for all $x \in \mathbb{R}$ and*

$$P'(x) = na_n x^{n-1} + \ldots + a_1 .$$

Example 3.10. Assume that g is a differentiable function and

$$F(x) = (x^3 - 5x)g(x) .$$

Compute $F'(2)$ provided that $g(2) = 3, g'(2) = -1$:

$$F'(x) = (3x^2 - 5)g(x) + (x^3 - 5x)g'(x)$$

so that $F'(2) = 23$.

Next, we investigate the reciprocal of a differentiable function:

Theorem 3.11 (Reciprocal rule)**.** *Let g be differentiable at x and $g(x) \neq 0$. Then $\frac{1}{g}$ is differentiable at x and*

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

Proof. Since g is differentiable, it is continuous. Since $g(x) \neq 0$ this shows that $g(x+h) \neq 0$ for small h. Thus,

$$\lim_{h\to 0} \frac{1}{g(x+h)} = \frac{1}{g(x)} .$$

Now

$$\frac{\frac{1}{g(x+h)} - \frac{1}{g(x)}}{h} = \frac{g(x) - g(x+h)}{hg(x)g(x+h)} = \frac{g(x) - g(x+h)}{h} \cdot \frac{1}{g(x)g(x+h)}$$

so that by Limit Laws

$$\lim_{h\to 0} \frac{\frac{1}{g(x+h)} - \frac{1}{g(x)}}{h} = -\lim_{h\to 0} \frac{g(x+h) - g(x)}{h} \cdot \lim_{h\to 0} \frac{1}{g(x)g(x+h)} = -\frac{g'(x)}{g^2(x)}$$

□

Example 3.12. For all $n \in \mathbb{Z}$: $(x^n)' = nx^{n-1}$.

Combining the product and the reciprocal rule we obtain:

Theorem 3.13 (Quotient rule)**.** *If f, g are differentiable at x and $g(x) \neq 0$ then $\frac{f}{g}$ is differentiable at x and*

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

Proof.

$$\left(\frac{f}{g}\right)' = (f \cdot \frac{1}{g})' = f'\frac{1}{g} + f(-\frac{g'}{g^2}) = \frac{f'g - fg'}{g^2} \, .$$

□

Example 3.14. $(\tan x)' = \frac{1}{\cos^2 x} = 1 + \tan^2 x$ for $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$.

In various source, the "Leibniz notation" is used: if y is a function of x then the derivative of y is denoted by $\frac{dy}{dx}$. If y is a function of t then the derivative of y is denoted by $\frac{dy}{dt}$. If we want to evaluate the derivative at a particular point x_0 then we write

$$y'(x_0) = \frac{dy}{dx}|_{x=x_0} \, .$$

With this notation, we can look at higher derivatives in the following way: if $f : (a, b) \to \mathbb{R}$ is differentiable on (a, b) then the derivative is a function $f' : (a, b) \to \mathbb{R}$. If this function is differentiable, we denote by

$$f''(x) = \frac{d}{dx}(f'(x)) = \frac{d^2}{dx^2}(f(x))$$

the *second derivative* of f and by $f'''(x) = \frac{d}{dx}(f''(x))$ the *third derivative* of f. We denote by $f^{(n)}(x)$ the n^{th}–derivative of f.

Example 3.15. The *acceleration* a is the rate of change of the speed v. Since the speed is the rate of change of location s, we have

$$a = \frac{dv}{dt} = \frac{d^2 s}{dt^2} \, .$$

We can now look at the composition of functions. That is, if y is a function of u and if u is a function of x then the chain rule states in Leibniz notation that

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} \, .$$

Written in function notation this reads as

Theorem 3.16 (Chain rule)**.** *If $F = f \circ g$ and g is differentiable at x and f differentiable at $g(x)$ then F is differentiable at x and*

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

"Proof:" Cancel du in the Leibniz notation. Done.

Of course, this is **not** a proof. Think about this: what would you want to cancel in $(f \circ g)' = (f' \circ g)g'$?

In this sense, the Leibniz notation is dangerous: it is tempting to consider dy and dx as independent objects (and indeed, there is a mathematical way to make this precise); however, we are not allowed to separate dy and dx (yet) but we read $\frac{dy}{dx}$ as "differentiate y with respect to the variable x".

Proof. We first recall

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

Thus, we consider the difference quotient

$$\frac{f(g(t))-f(g(x))}{t-x}=\underbrace{\frac{f(g(t))-f(g(x))}{g(t)-g(x)}}_{\to f'(g(x))}\underbrace{\frac{g(t)-g(x)}{t-x}}_{\to g'(x)}$$

What is wrong at this point? We cannot guarantee that $g(t)\neq g(x)$ near x, so on the right hand side we might divide by 0 — which is never a good idea!

To avoid this, we try to do something similar: we define a function which is the desired quotient as long as the denominator is not zero, and the predicted derivative otherwise:

$$H(y)=\begin{cases}\frac{f(y)-f(g(x))}{y-g(x)}, & y\neq g(x)\\ f'(g(x)), & y=g(x)\end{cases}\,.$$

This function H is continuous: the only critical point is $y=g(x)$ but there we have

$$\lim_{y\to g(x)}H(y)=\lim_{y\to g(x)}\frac{f(y)-f(g(x))}{y-g(x)}=f'(g(x))=H(g(x))\,.$$

For $t\neq x$ we have

$$\frac{f(g(t))-f(g(x))}{t-x}=H(g(t))\frac{g(t)-g(x)}{t-x}\,;$$

this holds since for $g(t)=g(x)$ both sides are 0, and otherwise by the definition of H.

Now, H and g are continuous (g is differentiable) so that the composition $F\circ g$ is continuous. Therefore, $\lim_{t\to x}H(g(t))=H(g(x))=f'(g(x))$ and the result follows from the limit laws.

□

The chain rule seems to have first been used by Leibniz. He used it to calculate the derivative of $\sqrt{a+bz+cz^2}$ as the composite of the square root function and the function $a+bz+cz^2$. He first mentioned it in a memoir with various mistakes in it. L'Hôpital uses the chain rule implicitly but also does not state it explicitly. The chain rule does not appear in any of Leonhard Euler's analysis books, even though they were written over a hundred years after Leibniz's discovery!

Example 3.17. Gravel is being poured by a conveyor onto a conical pile at the constant rate of 60π cubic feet per minute. Frictional forces within the pile are such that the height is always two–thirds of the radius. How fast is the radius of the pile changing at the instant the radius is 5 feet?

3.3 Derivatives of Inverse Functions

Contemplating a geometric picture, we observe that the tangent line of the inverse function f^{-1} should be well–defined wherever the tangent line of f is not horizontal.

Theorem 3.18. *Let $f : I \to \mathbb{R}$ be one–to–one and differentiable on an open interval I. Let $a \in I$ and $f(a) = b$. If $f'(a) \neq 0$ then f^{-1} is differentiable at b and*

$$(f^{-1})'(b) = \frac{1}{f'(a)} = \frac{1}{f'(f^{-1}(b))}\,.$$

What is wrong with the following "proof": Since $f \circ f^{-1}(x) = x$ we have by chain rule that $f'(f^{-1}(x)) \cdot (f^{-1}(x))' = 1$. Thus, if $f'(f^{-1}(x)) \neq 0$ then $(f^{-1}(x))' = \frac{1}{f'(f^{-1}(x))}$.

Of course, this argument doesn't work since we can only apply the chain rule, if we already know that f^{-1} is differentiable (but we don't!).

Even though the above argument NOT a proof, it is still a good way how to derive the formula for the derivative of the inverse in case you have forgotten it!

Obviously, the trigonometric functions are not one–to–one on their whole domain, so that we have to restrict the domain and consider

$$\begin{aligned} f(x) &= \sin(x), && x \in [-\tfrac{\pi}{2}, \tfrac{\pi}{2}]\,, \\ g(x) &= \cos(x), && x \in [0, \pi]\,, \\ h(x) &= \tan(x), && x \in (-\tfrac{\pi}{2}, \tfrac{\pi}{2})\,. \end{aligned}$$

Restricting to the above domains the trigonometric functions are one–to–one, and have inverse functions

$$\begin{aligned} f^{-1}(x) &= \arcsin(x), && x \in [-1, 1]\,, \\ g^{-1}(x) &= \arccos(x), && x \in [-1, 1]\,, \\ h^{-1}(x) &= \arctan(x), && x \in \mathbb{R}\,, \end{aligned}$$

with derivatives

$$\begin{aligned} \frac{d}{dx}\arcsin(x) &= \frac{1}{\sqrt{1-x^2}}, && x \in (-1, 1)\,, \\ \frac{d}{dx}\arccos(x) &= -\frac{1}{\sqrt{1-x^2}}, && x \in (-1, 1)\,, \\ \frac{d}{dx}\arctan(x) &= \frac{1}{1+x^2}, && x \in \mathbb{R}\,. \end{aligned}$$

3.4 Newton's method

Here we just give a very short introduction.

The Intermediate–Value Theorem can be used to determine whether a continuous function has a zero. But how do we find it? The bisection method is one way to compute a zero.

A more efficient way is to use Newton's method: Given a differentiable function f, our goal is to find zeros of f on an interval $[a, b]$. Start with some $x_1 \in [a, b]$. If $f(x_1) = 0$ you were lucky and you are done. If not, since the tangent line describes the local shape of the function, we may try to find a better approximation of the zero by using the intersection of the tangent line with the x–axis instead of x_1: call this point x_2. Continue this process.

Let $x_n \in [a, b]$. The tangent line at x_n is given by $y = f(x_n) + f'(x_n)(x - x_n)$. The intersection x_{n+1} with the x–axis is given by $0 = f(x_n) + f'(x_n)(x_{n+1} - x_n)$. Solving for x_{n+1} we obtain

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}\,.$$

Of course, this recipe does not work in all cases. For example, the tangent line could be horizontal at some point x_n (and then we cannot divide by its slope!). Or in the process you may just jump between two different values, or the successive values may not converge to a zero.

However, the method will usually converge, provided the initial guess of x_1 is close enough to the unknown zero, and that $f'(x_1) \neq 0$.

Newton's method was described by Isaac Newton (1669), however, his description differs substantially from the modern description given above: Newton applies the method only to polynomials. Isaac Newton probably derived his method from a similar but less precise method by Vieta. The essence of Vieta's method can be found in the work of the Persian mathematician, Sharaf al-Din al-Tusi. A special case of Newton's method for calculating square roots was known much earlier and is often called the Babylonian method.

3.5 The Mean–Value Theorem

For a differentiable function we see geometrically that for any secant there is a line parallel to the secant which is tangent to the graph of the function f.

A special case of this theorem was first described by Parameshvara (1370–1460) from the Kerala school of astronomy and mathematics. The mean value theorem in its modern form was later stated by Augustin Louis Cauchy (1789–1857).

The mathematical precise statement is:

Theorem 3.19 (Mean–Value Theorem)**.** *Let f be differentiable on (a, b), continuous on $[a, b]$. Then there exists at least one number $c \in (a, b)$ for which*

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

As before, stating that $c \in [a, b]$ will give mark reduction: not all relevant information is given. In the use of the Theorem, many students forget to check that f is continuous on $[a, b]$.

Although the statement of the Mean–Value Theorem may seem unimpressive at first sight, the theorem has a myriad of application in studying the shape of functions and thereby give solutions to various type of extremal questions: how large can a profit be? How much energy can we save by using a particular parameter?

The proof is done in various steps, each interesting in themselves. First we show that the sign of the derivative at one point x_0 prescribes the shape of the graph near x_0.

Theorem 3.20. *Let f be differentiable at x_0.*

(i) If $f'(x_0) > 0$ then $f(x_0 - h) < f(x_0) < f(x_0 + h)$ for small enough $h > 0$.

(ii) If $f'(x_0) < 0$ then $f(x_0 + h) < f(x_0) < f(x_0 - h)$ for small enough $h > 0$.

Proof. Assume $f'(x_0) > 0$. Put $\varepsilon = f'(x_0)$. Since f is differentiable at x_0 there exists $\delta > 0$ such that

$$0 < |x - x_0| < \delta \implies \left| \frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right| < \varepsilon = f'(x_0)\,.$$

This implies $0 < \frac{f(x)-f(x_0)}{x-x_0} < 2f'(x_0)$.
Now, if $0 < h < \delta$ then $x = x_0 \pm h$ satisfies

$$0 < |x - x_0| = |\pm h| = h < \delta\,.$$

Thus,

$$0 < \frac{f(x_0 \pm h) - f(x_0)}{x_0 \pm h - x_0}\,.$$

The + inequality implies $0 < f(x_0 + h) - f(x_0)$, the other one $0 > f(x_0 - h) - f(x_0)$. This shows the claim in the case $f'(x_0) > 0$. The other case can be done similarly.

□

Further important ingredients for the Mean-Value Theorem are the Extreme Value Theorem and Rolle's theorem. The first known formal proof was offered by Michel Rolle in 1691, which used the methods of differential calculus. The name "Rolle's theorem" was first used by Moritz Wilhelm Drobisch of Germany in 1834 and by Giusto Bellavitis of Italy in 1846.

Theorem 3.21 (Rolle's theorem)**.** *Suppose f is differentiable on (a, b) and continuous on $[a, b]$. If $f(a) = f(b) = 0$ then there exists $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. By the Extreme-Value Theorem there must be a maximum and a minimum of the function on $[a, b]$. The aim is to show that this extremum is attained in the interval (a, b): then we show that f has to have horizontal tangent line at this point.

Of course, we can assume that f is not vanishing identically on $[a, b]$: if that was the case then $f(x) = 0$ for all x implies $f'(x) = 0$ for all x and we are done.

Thus, we can assume that there exists at least one point with $f(x) > 0$ (or $f(x) < 0$) with $x \in (a, b)$. Take the first case (the other one works similarly). By the Extreme-Value Theorem, f attains its maximum on $[a, b]$ since it is a continous function. Since $f(x) > 0$, the maximum will be attained at some c with $f(c) > 0$ and thus $c \neq a, b$. Thus, $c \in (a, b)$ and f is differentiable at c.

If $f'(c) > 0$ the previous theorem gives points nearby with $f(c - h) < f(c) < f(c + h)$ with $h > 0$ small, which contradicts the fact that f attains a maximum at c. Similarly, if $f'(c) < 0$ then $f(c - h) > f(c) > f(c + h)$ with $h > 0$ small, which again contradicts the fact that f attains a maximum at c. Thus, $f'(c) = 0$.

$\square$

Example 3.22. The polynomial $p(x) = 2x^3 + 5x - 1$ has exactly one root: Since p is a polynomial of degree 3 it has at least one zero (use the Intermediate-Value Theorem: $p(0) = -1 < 0$, $p(1) = 6 > 0$). Assume that there exist $x_1 < x_2$ with $p(x_1) = p(x_2) = 0$. Since p is a polynomial, it is continous on $[x_1, x_2]$ and differentiable on (x_1, x_2). We can now apply Rolle's Theorem: there exists $c \in (x_1, x_2)$ with $p'(c) = 0$.

But $p'(x) = 6x^2 + 5 > 0$ for all $x \in \mathbb{R}$! Contradiction.

Proof of the Mean–Value Theorem. We will use Rolle's theorem. To apply Rolle's theorem, we want to find a function g which is zero at a and b, such that its derivative at c is the difference between $f'(c)$ and the difference quotient $\frac{f(b)-f(a)}{b-a}$. Our first attempt is to write

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

Then $g'(c) = f'(c) - \frac{f(b)-f(a)}{b-a}x$. However, at a and b this function is not zero. Can we repair this? Second attempt

$$g(x) = f(x) - \left(\frac{f(b) - f(a)}{b - a}(x - a) + f(a)\right)\,.$$

We haven't changed the derivative (we only changed the function by constants) and have now also $g(a) = g(b) = 0$. Thus, Rolle's theorem gives $c \in (a, b)$ with $g'(c) = 0$. Thus $f'(c) = \frac{f(b)-f(a)}{b-a}$. $\square$

Example 3.23. Suppose f is differentiable on $(1, 4)$, continuous on $[1, 4]$ and $f(1) = 2$. Given that $2 \leq f'(x) \leq 3$ for all $x \in (1, 4)$ what is the least/greatest value f can take at $x = 4$?

We can apply the Mean-Value Theorem since f is differentiable on $(1, 4)$ and continuous on $[1, 4]$. Thus, there exists $c \in (1, 4)$ with

$$f'(c) = \frac{f(4) - f(1)}{4 - 1} = \frac{f(4) - 2}{3}\,.$$

Since $2 \leq f'(x) \leq 3$ we thus have

$$2 \leq \frac{f(4) - 2}{3} \leq 3$$

that is $8 \leq f(40) \leq 11$.

Since the derivative gives the slope of tangent lines, we can connect information given by the derivative to the shape of graph of the function.

Theorem 3.24 (Increasing/Decreasing Theorem)**.** *Suppose $f : I \to \mathbb{R}$ is differentiable on (a, b) and continuous on $[a, b]$. Then*

- *If $f'(x) > 0$ for all $x \in (a, b)$ then f is increasing on $[a, b]$.*
- *If $f'(x) < 0$ for all $x \in (a, b)$ then f is decreasing on $[a, b]$.*
- *If $f'(x) = 0$ for all $x \in (a, b)$ then f is constant on $[a, b]$.*

Proof. Let $x_1, x_2 \in [a, b]$, $x_1 < x_2$. Since f is differentiable on $(x_1, x_2) \subset (a, b)$ and f is continuous on $[x_1, x_2] \subset [a, b]$, the Mean–Value Theorem gives a $c \in (x_1, x_2)$ with

$$0 < f'(c) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

Since $x_1 < x_2$ this implies $0 < f(x_2) - f(x_1)$ and f is increasing.
Similar arguments work for $f'(c) = 0$ or $f'(c) < 0$. □

Corollary 3.25. *If f is differentiable on an open interval (a, b) then: $f'(x) = 0$ for all $x \in (a, b)$ if and only if f is constant.*

From this we easily see a theorem which we will use frequently (in particular, for integrals):

Theorem 3.26. *Let $f, g : (a, b) \to \mathbb{R}$ be differentiable functions on an interval (a, b). If $f'(x) = g'(x)$ for all $x \in (a, b)$, then $f(x) = g(x) + C$ for all $x \in (a, b)$ where $C \in \mathbb{R}$ is a constant.*

Proof. Consider $h = f - g$. Then $h'(x) = 0$ for all x, thus h is constant. □

3.6 Local extrema

Definition 3.27. Suppose f is a function on an interval $[a, b]$ and c an interior point of (a, b). Then f is said to have a

local maximum at c if $f(c) \geq f(x)$ for all x sufficiently close to c, that is, if there exists $\delta > 0$ such that $0 < |x - c| < \delta \implies f(c) \geq f(x)$.

local minimum at c if $f(c) \leq f(x)$ for all x sufficiently close to c, that is, if there exists $\delta > 0$ such that $0 < |x - c| < \delta \implies f(c) \leq f(x)$.

Thus, at a local extremum of a differentiable function we expect to have a horizontal tangent line.

Definition 3.28. Let $f : [a, b] \to \mathbb{R}$. If $c \in (a, b)$ and

$$f'(c) = 0 \quad \text{or} \quad f'(c) \text{ does not exist}$$

then c is called a *critical point* of f.

Thus, if a function has a local extremum at a point, then either it is not differentiable at the point or has horizontal tangent line.

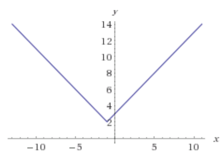
Theorem 3.29. *Let $f : [a, b] \to \mathbb{R}$. If f has a local extremum at $c \in (a, b)$ then c is a critical point of f.*

Proof. If f is not differentiable at c we are done. Consider now the case that $f'(c)$ exists but is not equal to 0. But then $f'(c) > 0$ or $f'(c) < 0$ and in either case there are x_1, x_2 arbitrarily close to c with $f(x_1) < f(c) < f(x_2)$. Contradiction.

□

A function f may not have a local extremum at c in the interior of its domain but still c may be a critical point of f, e.g. $f(x) = x^3$ has a critical point at 0 but not a local maximum or maximum.

Example 3.30. Find the critical points of $f(x) = |x + 1| + 2$.



We can write

$$f(x) = \begin{cases} -x + 1 & x < -1 \\ x + 3 & x \geq -1 \end{cases}.$$

Since f is given by the absolute value function, we know that f is not differentiable at $x = -1$. Thus

$$f(x) = \begin{cases} -1 & x < -1 \\ DNE & x = -1 \\ 1 & x > -1 \end{cases},$$

and $x = -1$ is the only critical point. Indeed, f has a local minimum at $x = -1$.

So far, we found a necessary condition for c to be a local extremum: c has to be a critical point. Can we give sufficient conditions which imply that c is indeed a local extremum?

Theorem 3.31 (First–derivative test)**.** *Suppose c is a critical point of f and f is continuous at c. If there exists $\delta > 0$ such that*

(i) $f'(x) > 0$ for all $x \in (c-\delta, c)$ and $f'(x) < 0$ for all $x \in (c, c+\delta)$ then f has a local maximum at c.

(ii) $f'(x) < 0$ for all $x \in (c-\delta, c)$ and $f'(x) > 0$ for all $x \in (c, c+\delta)$ then f has a local minimum at c.

(iii) If f' keeps constant sign on $(c-\delta, c) \cup (c, c+\delta)$ then $f(c)$ is not a local extreme value.

Proof. Follows from the Increasing/Decreasing Theorem 3.24. □

The assumption that f is continuous at c is important. E.g., the function

$$f(x) = \begin{cases} 1 + 2x, & x \leq 1 \\ 5 - x, & x > 1 \end{cases}$$

does not attain at local extremum at $c = 1$ even though 1 is a critical point and the derivative of f changes sign.

The following second test is easier to apply, however, it works in even fewer cases than the First–Derivative test: we need to assume that the second derivative exists.

Theorem 3.32 (Second–Derivative test)**.** *Suppose $f'(c) = 0$ and $f''(c)$ exists.*

(i) If $f''(c) < 0$ then f has a local maximum at c.

(ii) If $f''(c) > 0$ then f has a local minimum at c.

Proof. Assume $f''(c) > 0$ (the other case is similar). Since f'' is the derivative of f' there exists $\delta > 0$ such that $f'(x) < f'(c) = 0$ for all $x \in (c-\delta, c)$ and $f'(x) > f'(c) = 0$ for all $x \in (c, c+\delta)$. Thus, the First-Derivative Test shows that f attains a local minimum at c.

□

Note that the Second–Derivative test does not give any information if $f'(c) = 0$ and $f''(c) = 0$. Consider $f(x) = x^3$ and $f(x) = x^4$ at $x = 0$.

Example 3.33. The light reflection at a mirror satisfies the Fermat principle of shoterst distance. We consider two points a and b above a mirror whose horizontal distance is l. For which l is

$$L(x) = \sqrt{a^2 + x^2} + \sqrt{b^2 + (l - x)^2}$$

minimal?

If $x \to \pm\infty$ then $L(x) \to \infty$. Hence, there is a local minimum in $(-\infty, \infty)$ since L is differentiable (and thus continuous). We find the critical points by computing

$$L'(x) = \frac{x}{\sqrt{a^2 + x^2}} + \frac{-(l - x)}{\sqrt{b^2 + (l - x)^2}}\,.$$

Therefore, if $L'(c) = 0$ then c and $l - c$ have the same sign (thus are positive), and

$$\frac{c}{\sqrt{a^2 + c^2}} = \frac{(l - c)}{\sqrt{b^2 + (l - c)^2}}$$

is equivalent to

$$\frac{c^2}{a^2 + c^2} = \frac{(l - c)^2}{b^2 + (l - c)^2}\,.$$

This shows that $c^2b^2 + c^2(l-c)^2 = a^2(l-c)^2 + c^2(l-c)^2$ and $bc = a(l-c)$ since $a, b, c, l-c > 0$. Therefore, the only critical point is $c = \frac{al}{b+a}$, and thus L has an absolute minimum at $c = \frac{al}{b+a}$.

In particular, $\frac{l-c}{b} = \frac{l(1-\frac{al}{b+a})}{b} = \frac{l}{b+a} = \frac{c}{a}$. In other words, denoting by α the angle of incidence and by β the angle of reflection, then

$$\cot\alpha = \frac{c}{a} = \frac{l - c}{b} = \cot\beta\,,$$

so we have shown the *law of reflection*.

3.7 Global extreme values

So far, we have only looked at local extreme values in the interior of the domain of the function. Can we now find the global extreme values? (We insert the word “global” to emphasise that we are looking for the overall maximal/minimal values, and not local extrema).

Recall our previous definition:

Definition 3.34. The function f is said to have an

(global) maximum at $d \in \mathrm{dom}(f)$ if $f(d) \geq f(x)$ for all $x \in \mathrm{dom}(f)$;

(global) minimum at $d \in \mathrm{dom}(f)$ if $f(d) \leq f(x)$ for all $x \in \mathrm{dom}(f)$.

Of course, a function has to be at least bounded to attain both global extreme values but could have a minimum or maximum when it's unbounded (e.g., $f(x) = x^2$ has a minimum even though it is not a bounded function). However, for a continuous function on a bounded closed interval we know that the function is bounded. Thus, we can apply the following strategy:

Strategy to find global extremas of a continuous function f on a closed bounded interval $[a, b]$:

(i) Find critical points of f.

(ii) Compare $f(a), f(b)$ and $f(c)$ for all critical points c of f.

Example 3.35. The function $f(x) = 1 + 4x^2 - \frac{1}{2}x^4$, $x \in [-1, 3]$, has critical points $x = 0, x = \pm 2$. Comparing the values of f at the critical points and the end points $x = -1, x = 3$ we see that f attains a (global) maximum at $x = 2$ and a (global) minimum at $x = 3$.

Common errors in finding the global values are the obvious ones:

- Using the above strategy even though the interval is not closed or not bounded.
- Forgetting to consider the critical points (including points where the derivatives does not exist).
- Forgetting to consider the end points of the interval.

3.8 Concavity and points of inflection

A further question about the general shape of a function is whether it is curving in a particular direction.

Definition 3.36. Let f be a function differentiable on an open interval I. The graph of f is said to be

concave up if f' increases on I, and

concave down if f' decreases on I.

Of course, the points where we change the type of concavity are important!

Definition 3.37. Let f be continuous at c, differentiable near c. Then $(c, f(c))$ is called a *point of inflection* if the graph of f changes its type of concavity at c, that is, if there exists $\delta > 0$ such that graph(f) is concave up (or concave down) on $(c-\delta, c)$ and the opposite concave down (or concave up) on $(c, c+\delta)$.

Since the second derivative describes the local shape of the graph of the first derivative, we can find the concavity type by examining the second derivative:

Theorem 3.38. *Let f be twice differentiable on an open interval I.*

(i) If $f''(x) > 0$ for all $x \in I$ then the graph of f is concave up.

(ii) If $f''(x) < 0$ for all $x \in I$ then the graph of f is concave down.

Proof. Increasing/Decreasing Theorem 3.24. □

Theorem 3.39. *If $(c, f(c))$ is a point of inflection then $f''(c) = 0$ or $f''(c)$ does not exist.*

Proof. First Derivative Test 3.31. □

3.9 Indeterminate forms and the Cauchy–Mean–Value Theorem

A modified version of the Mean–Value theorem can be used to compute indeterminate forms of type "$\frac{0}{0}$", that is, limits of fractions $\frac{f(x)}{g(x)}$ for $x \to c$ where $\lim\limits_{x\to c} f(x) = \lim\limits_{x\to c} g(x) = 0$.

Theorem 3.40 (Cauchy–Mean–Value Theorem)**.** *Let f, g be differentiable on (a, b), continuous on $[a, b]$. If $g'(x) \neq 0$ for all $x \in (a, b)$ then there exists $r \in (a, b)$ with*

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

Why is this a generalisation of the Mean–Value theorem? Consider $g(x) = x$.

We now apply this to functions with zeros at a:

Theorem 3.41 (L'Hôpital's Rule)**.** *Suppose that* $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$ *with* $g(x) \neq 0$ *and* $g'(x) \neq 0$ *for all* x *near* c*. Then if*

$$\lim_{x\to c} \frac{f'(x)}{g'(x)} = L$$

exists, then $\lim_{x\to c} \frac{f(x)}{g(x)}$ *exists and equals* L*.*

A common error is to apply L'Hôpital's rule even though not both limits of f and g are zero. Moreover, if the limit of the quotient of the derivatives does not exist, the rule does not give any information about the original limit.

3.10 Taylor polynomials

Another important application of the Mean–Value Theorem are the Taylor polynomials. They are used to approximate a function.

Consider first the function $f(x) = (x-c)^n$, $c \in \mathbb{R}, n \in \mathbb{N}$.

Then we can approximate f linearly at c by its tangent

$$f(x) \sim f(c) + f'(c)(x-c) = P_1(x)\,.$$

Indeed, we have $f(c) = P_1(c)$ and $f'(c) = P_1'(c)$. However, the n^{th} derivative of f and P_1 do not coincide at c if $n > 1$, and P_1 is not a very good approximation in this case: $P_1(x) = 0$ is the constant zero function, whereas $f(x)$ is not constant.

Therefore, to approximate a function f we want to use a polynomial P_n of degree n instead of a polynomial of degree 1, and we require that the derivatives up to the n^{th} order of f and P_n agree.

Is there such a polynomial, and how do we find it?

Our example indicates that we want to write our polynomial in the form

$$P_n(x) = \sum_{l=0}^{n} a_l (x-c)^l$$

for some $a_l \in \mathbb{R}$, $l = 0, \ldots, n$. Then the requirement $P_n(c) = f(c)$ gives $a_0 = f(c)$. Furthermore, $P_n'(x) = \sum_{l=1}^{n} l a_l (x-c)^{l-1}$ gives with $P_n'(c) = f(c)$ that $a_1 = f'(c)$.

Using $P_n^{(k)}(x) = \sum_{l=k}^{n} l(l-1)\ldots(l-k+1)a_l(x-c)^{l-k}$ we obtain with $P_n^{(k)}(c) = f^{(k)}(c)$:

$$f^{(k)}(c) = k! a_k\,.$$

Definition 3.42. Let f be a function with n derivatives on an open interval I. Then for $c \in I$ the polynomial

$$P_n(x) = \sum_{l=0}^{n} \frac{f^{(l)}(c)}{l!}(x-c)^l = f(c) + f'(c)(x-c) + \frac{f''(c)}{2}(x-c)^2 + \ldots + \frac{f^{(n)}(c)}{n!}(x-c)^n$$

is called the n^{th} *Taylor polynomial* of f in powers of $(x-c)$.

For some reason, when computing the coefficients in the Taylor polynomial, it seems to be tempting to differentiate $f^{(n-1)}(c)$ to obtain $f^{(n)}(c)$: so the coefficients become zero quite quickly but unfortunately not rightfully. Instead you should take all needed higher derivatives first, before evaluating at c.

The concept of a Taylor polynomials was formally introduced by the English mathematician Brook Taylor in 1715. If the Taylor series is centred at $c = 0$, then that series is also called a Maclaurin polynomial, named after the Scottish mathematician Colin Maclaurin, who made extensive use of this special case of Taylor polynomial in the 18th century.

For every function with n derivatives, we can find the corresponding Taylor polynomial of order n. But how good is this approximation of the function f? A (partial) answer is the following Taylor formula. More results about Taylor approximation will be discussed in the second semester of Calculus and Analysis.

Theorem 3.43 (Taylor formula)**.** *If f has $n+1$ derivatives on an open interval I with $c \in I$ then*

$$f(x) = \sum_{l=0}^{n} \frac{f^{(l)}(c)}{l!}(x-c)^l + R_n(x)\,,$$

where the Lagrangian remainder R_n *satisfies*

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

for some ξ between x and c. In particular,

$$\lim_{x \to c} \frac{R_n(x)}{(x-c)^n} = 0\,.$$

Thus, the remainder goes faster to zero then $(x-c)^n$ when x approaches c. However, this does not mean that f is well–approximated for all $x \in \mathrm{dom}(f)$.

Proof. Fix $x \in I$. Put

$$F(t) = \sum_{l=0}^{n} \frac{f^{(l)}(t)}{l!}(x-t)^l \quad \text{and} \quad G(t) = (x-t)^n\,.$$

Note first, that F is not the Taylor polynomial P_n of f: instead of considering the function depending on x, we consider the function for fixed x but varying c.

However, for $t = c$ we obtain $F(c) = P_n(x)$ and

$$F(x) - F(c) = f(x) - P_n(x) = R_n(x)\,.$$

The Cauchy Mean–Value theorem gives ξ between c and x with

$$\frac{F'(\xi)}{G'(\xi)} = \frac{F(x) - F(c)}{G(x) - G(c)} = \frac{R_n(x)}{-(x-c)^{n+1}}\,,$$

where we used that $G(x) = 0$ and $G(c) = (x - c)^{n+1}$. Now the product rule and chain rule give

$$\begin{aligned}
\frac{d}{dt}F(t) &= \sum_{l=0}^{n} \frac{f^{(l+1)}(t)}{l!}(x-t)^l - \sum_{l=0}^{n} \frac{f^{(l)}(t)}{l!} l(x-t)^{l-1} \\
&= \sum_{l=0}^{n} \frac{f^{(l+1)}(t)}{l!}(x-t)^l - \sum_{l=1}^{n} \frac{f^{(l)}(t)}{(l-1)!}(x-t)^{l-1} \\
&= \sum_{l=0}^{n} \frac{f^{(l+1)}(t)}{l!}(x-t)^l - \sum_{l=0}^{n-1} \frac{f^{(l+1)}(t)}{l!}(x-t)^{l} \\
&= \frac{f^{(n+1)}(t)}{n!}(x-t)^n\,,
\end{aligned}$$

and $\frac{d}{dt}G(t) = -(n+1)(x-t)^n$. Plugging this in, we see

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

$\square$

3.11 Study guide

Go through the material of chapter 3, and identify the main topics and statements.

To deepen your understanding, think about the following:

- Draw continuous functions whose tangent line at the point 2 is
  - horizontal,
  - vertical,
  - does not exist, or
  - has slope -10.

  Try in each case to give a formula for the function. Verify that it is continuous, and satisfies the condition.
- Give examples of functions where $(fg)' \neq f'g'$.

- Give examples of functions where $\left(\frac{f}{g}\right)' \neq \frac{f'}{g'}$.
- Find two differentiable functions such that $(f \circ g)'(x) \neq f'(x)g'(x)$. Convince yourself that the correct formula is instead given by the chain rule.
- Discuss the difference between f^{-1} and $\frac{1}{f}$. What is the meaning of "inverse function"?
- Find situations in which Newton's method works/does not work.
- Does $f(x) = \sin(x^3) - 1, x \in \mathbb{R}$, attain the value -1.5? If so, find c with $f(c) = -1.5$ (use Newton's method). Compare your efficiency in solving with your attempt in Chapter 2.
- Find an example of a function which is differentiable on an open interval (a, b) but not continuous on $[a, b]$. Find such a function such that there is no tangent to the graph which is parallel to the secant between $f(a)$ and $f(b)$.
- Is the c in the Mean–Value Theorem unique? Prove your conjecture, or give an example in which it is not unique.
- What is the difference between Rolle's theorem and the Mean–Value Theorem?
- Why do we need the continuity assumption in the Increasing/Decreasing Theorem?
- Draw the graph of a function which is differentiable, has a local maximum at 1 and no minimum. Is the function concave up or down near 1 (or neither nor)? Is it possible that such a function has more than 2 inflection points?
- Find functions for which one cannot use the Second–Derivative test but the First Derivative test shows the existence of a local extremum.
- What is the difference between a local and an global extremum? Are there examples which are both local and global extremums (if so, give one!)
- Find a function which has no global maximum on $(0, 1)$.
- Why did we assume our function to be continuous in the "Finding global values"–strategy? What would you do if the function is not continuous?