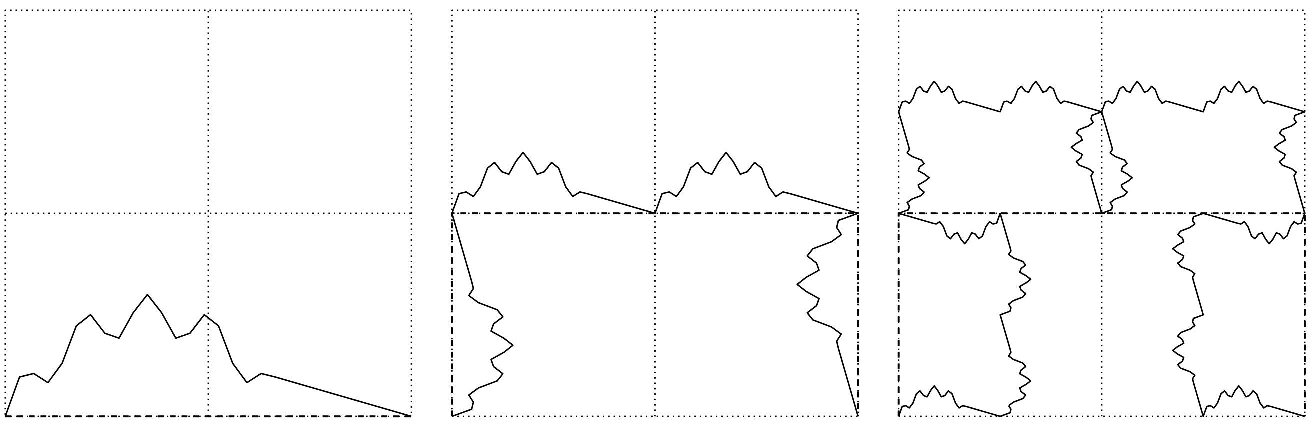
Analysis with Uniform Error Bounds

Hermann Karcher, Bonn University, 2002
Translated by Ursula Weiss and Michael Livshits, 2011

Abstract. Translation of notes for a first term analysis course that was twice taught at Bonn, winter 1999/2000 and 2002/2003. The material is drastically reorganized: Tangents of circles, derivative of polynomials, quadratic bounds for the deviation from the tangent, differentiation rules, rational functions, the **monotonicty theorem** (proved via uniform error bounds). Completeness to get limit functions. Complex functions, power series, integrals. Ending with a better understanding of continuity.

Contents



Preface to the Oct. 2002 Version

Over the years I have tried out many details from this manuscript in my Analysis I courses, but only in the winter 1999/2000 course for computer science students did I have a chance to teach completely along the following lines. In the first term I deviate substantially from how the Analysis material is usually organized. Nevertheless, starting from the second term, one can continue with a standard course.
A second preface is a 10 page summary to facilitate a comparison with the standard approach.
Since I treat continuity only at the end of the first term, people jump to the conclusion that I want to reduce the role of continuity. Exam results support my conviction that I actually achieve a better understanding of continuity, because it is only treated when the students have already achieved quite a lot of experience with the arguments and goals of Analysis.
One central theme is to make students observe that new objects in analysis are mostly constructed via approximation by already well understood objects. Irrational numbers are approximated by rational ones, new functions like exp, sin and cos are approximated by polynomials or rational functions. These "constructions by approximation" are on the one hand amazingly powerful, and on the other hand require careful arguments to figure out what the derivatives of these only approximately computable functions are. On top of that, the importance of almost all these new functions lies in the properties of their derivatives (e.g., $f' = f$), so that some preliminary understanding of differentiation is needed before one can turn to them. This preliminary understanding is acquired by working with polynomials.

This manuscript begins with two introductory chapters. First we discuss, with nine examples, absolute and relative differences and errors. The last example discusses tangents of the circle. The observed properties allow in the second chapter to define tangents (and slopes) for the quadratic parabola and to explain with the properties of these tangents some important applications, e.g. the concave mirror.
The developed comparison arguments can be generalized to define, in chapter 3, tangents and slopes of all the polynomial functions. An important detail is that polynomial functions deviate from their tangents by less than an explicit quadratic error term. These observations are the basis for a differentiability definition in terms of explicit quadratic error terms. One may think of this definition as "easy differentiability". All rational functions can immediately be seen to be easily differentiable. In chapter 4 we prove the differentiation rules. These rules do not change when the definition is developed until the final version is reached.
In chapter 5 we meet the absolutely central "Monotonicity Theorem" for the first time. In the present context it can be proved before discussing completeness, because all functions

known at this point satisfy the assumption of the uniform quadratic error bounds. Applications of this theorem, even when dealing with polynomials, quickly go beyond what can be computed explicitly. For example, it supplies the following fact which we will use a lot:

If one knows for a polynomial P in some interval $[r, R]$ a bound $|P''| \leq K$, then we have:

$$x, a \in [r, R] \;\Rightarrow\; |P(x) - P(a) - P'(a) \cdot (x - a)| \leq \frac{K}{2}|x - a|^2.$$

After these preparations we want to construct new functions with interesting derivative properties, construct them "by approximation". This requires convergent sequences and the completeness of the real numbers (chapter 6). We benefit from two mathematical facts. First, all the approximating functions we need to use are rational functions. Therefore, we can apply our Monotonicity Theorem and all the convergence estimates are derived with this one tool. Second, the Monotonicity Theorem implies estimates such as

$$|f_n(x) - f_n(y)| \leq L \cdot |x - y|, \; L \text{ independent of } n.$$

Already Archimedes has formalized, how this implies the same inequality for a limit function. In other words, the Monotonicity Theorem for rational functions supplies an alternative to treating limit functions via uniform convergence.

The complex numbers are fundamental within Mathematics and also in other disciplines. In chapter 7 we develop the content of chapters 3 - 5 for complex numbers and complex functions, enriched by now having the completeness of $\mathbb{R}$ available.

The second tool for convergence proofs, in addition to the Monotonicity Theorem, is majorisation, usually by a geometric series. We use this tool in chapter 8 in developing complex power series.

Simultanously with differentiation, the mathematicians of the 17th century invented a dramatic generalization of summing, namely "Integration". All functions treated so far can be approximated *uniformly* by piecewise linear functions, again a consequence of the monotonicity theorem. With this notion we prove in chapter 9 that Differentiation and Integration are inverses of each other.

In chapter 10 we introduce continuous functions via the property that they map convergent sequences to convergent sequences. The main theorems for continuous functions can, because of the preceeding preparation, now be proved with much less sweat and much more appreciation than at an earlier stage of the analysis education.

Communication Problems in Class

The following points regularly cause problems in mathematical education; maybe they can be reduced by being pointed out early.

NUMBER 1: Of course one has to learn fundamental definitions by heart. It is a bit grotesque, how many students can't bring themselves to do this during the first term.

NUMBER 2: A mathematical theorem, correctly learnt by heart and quoted, does not yet say anything about understanding, objections from lawyers notwithstanding. It is necessary

to use one's own words to make correct arguments with the new mathematical notions.
NUMBER 3: There are mathematical definitions which in daily language would not be accepted: We call a set of objects a "Vector Space" if certain relations between its elements hold true - but we do not say what the individual elements "are". Or, we call a function continuous, if an infinite set of implications is true - but we do not say how these implications are to be verified.
NUMBER 4: The syntax of mathematical statements is much more sensitive to minor changes than we are used to in non-mathematical contexts. In particular one needs to get used to the fact that a word that has been defined by mathematicians, can only be used *exactly as defined* - it is not enough to have some feeling for what it means to be able to talk mathematics.
NUMBER 5: There is a conflict between the understanding of details and the overall perception. When learning, a missing detail feels like a gap and distracts from getting an overall picture. But the usual attempts to illuminate the overall picture do ignore the details. So the feeling of being trapped between the gaps starts... turn to the details... miss the overall picture... ignore the details... feel the gaps... Scylla and Charybdis.

Added Feb. 2011: Many, many thanks to Ursula Weiss for translating and to Michael Livshits for starting the translation project and editing the English.

Hermann Karcher

Preface 2: Comparison with the Classical Approach

"Uniform error bounds and differentiation first, continuity and uniform convergence later"

The standard analysis route proceeds as follows:

1.) Convergent sequences and series, completeness of the reals,
2.) Continuity, main theorems, trivial examples,
3.) Differentiability via limits,
4. or 5.) Integrals,
5. or 4.) Properties of limit functions via uniform convergence.

I propose the following reorganization of the material:

1.) Differentiation of polynomials, emphasis on tangent approximation,
2.) Differentiation rules, differentiation of rational functions,
3.) The monotonicity theorem via uniform estimates, before completeness,
4.) Completeness and limit functions, proofs via uniform estimates,
5.) Integrals, defined as generalized sums, computed with antiderivatives,
6.) Continuity, with less trivial examples.

Arguments **why** one would **want** such a reorganization are: The start is closer to the background of the student, one spends more time on differentiation where technical skills have to be trained, and discusses continuity when logical skills are more developed.

The main reasons **why** such a reorganization **is possible** are: The properties of limit functions follow more easily from uniform error bounds than from uniform convergence (e.g. if one has a pointwise convergent sequence of functions $\{f_n\}$ with a uniform Lipschitz bound L, then the limit function also has L as Lipschitz bound). The first such uniform error bounds follow for polynomials in a pre-analysis fashion. These bounds suffice to prove the Monotonicity Theorem before completeness. Finally, the Monotonicity Theorem shows that many interesting converging sequences indeed have uniform error estimates.

1. Derivatives of Polynomials

The elementary formula

$$x^k - a^k = (x-a)\cdot(x^{k-1} + x^{k-2}a + \ldots + a^{k-1})$$

is good enough to replace all calls on continuity. The first consequence is

$$x, a \in [-R, R] \Rightarrow |x^k - a^k| \le kR^{k-1}\cdot|x-a|.$$

One only needs the triangle inequality to get for polynomials $P(X) := \sum_{k=0}^{n} a_k X^k$:

$$x, a \in [-R, R] \Rightarrow |P(x) - P(a)| \le \left(\sum |a_k| kR^{k-1}\right)\cdot|x-a| =: L\cdot|x-a|.$$

In other words: *From the data which give the polynomial and from the interval on which we want to study it we can explicitly compute a Lipschitz bound.* This conforms with the common expectation (which is disappointed by continuous functions): Differences between function values, $|P(x) - P(a)|$, increase no worse than proportionally to the difference of the arguments, $|x - a|$.

The derivative controls this rough proportionality more precisely. The term $(x^{k-1}+x^{k-2}a+\ldots+a^{k-1})$, which multiplies $(x-a)$ in the above elementary formula, differs "little" from ka^{k-1} on "small" intervals around a. The first inequality above makes this precise:

$$x \in [a-r, a+r], R := |a| + r \Rightarrow$$
$$\big|(x^{k-1} + x^{k-2}a + \ldots + a^{k-1}) - ka^{k-1}\big| \leq \frac{k(k-1)}{2} R^{k-2} \cdot |x-a|$$

hence

$$\big|x^k - a^k - ka^{k-1}(x-a)\big| \leq \frac{k(k-1)}{2} R^{k-2} \cdot |x-a|^2.$$

Observe that $|x-a|^2$ is less than 1% of the argument difference $|x-a|$ if $r < 0.01$. Again, the triangle inequality extends this to polynomials $P(X) := \sum_{k=0}^{n} a_k X^k$ and again *the error bound is explicitly computable from the data of the function and the interval in question*:

$$x \in [a-r, a+r], R := |a| + r,\ P'(a) := \sum_{k=1}^{n} a_k k a^{k-1} \Rightarrow$$
$$|P(x) - P(a) - P'(a)(x-a)| \leq \Big(\sum |a_k| k(k-1)/2R^{k-2}\Big) \cdot |x-a|^2 =: K \cdot |x-a|^2.$$

Application: At interior extremal points the derivative has to vanish. (If $P'(a) > 0$ and $0 < x - a < P'(a)/K$ then $P(x) > P(a)$, etc.)

Note on Completeness. Consider the step function, which jumps from 0 to 1 at $\sqrt{2}$, but consider as its domain only the rational numbers. This function is differentiable, but **not** uniformly. On the other hand, if a function is uniformly continuous (or uniformly differentiable) on a **dense** subset of an interval, then, using completeness, one can extend the function to the whole interval without loosing continuity (or differentiability) and without changing the visible behaviour of the function. This is clearly the case for polynomials. Moreover, the above estimates make sense in any field between $\mathbb{Q}$ and $\mathbb{R}$ which the student happens to know. Even in $\mathbb{C}$ they are useful. Therefore one can indeed discuss differentiability before completeness. – Of course completeness remains essential if one wants to define inverse functions or limit functions. I chose to discuss completeness immediately before constructing limit functions, because I view this as the more spectacular application, but I am not advertising that choice here, I am merely saying: *If one decides to work with uniform error bounds then this imposes fewer restrictions on where in* *the course one chooses to discuss completeness*, while uniform convergence does need prior knowledge of completeness.

2. Differentiation Rules, Derivatives of Rational Functions

Differentiation rules are meant to compute derivatives of "complicated" functions, built out of "simpler" ones, from the derivatives of the "simpler" functions. Since linear combinations, products and compositions of polynomials give again (easy to differentiate) polynomials the promise looks limited. However, there are some other functions which can be differentiated directly from their definitions, and this will broaden our possibilities considerably. We start with $x \to 1/x$ and observe $(1/x - 1/a) = (a-x)/(x \cdot a)$. While in the case of polynomials we could compute suitable error constants for *any* interval $[a-r, a+r]$ we now have to avoid division by zero. With this extra care we have:

$$0 < a/2 \leq x \Rightarrow |1/x - 1/a| = \frac{|x-a|}{x \cdot a} \leq \frac{2}{a^2} \cdot |x-a|$$

$$0 < a/2 \leq x \Rightarrow (1/x - 1/a + \frac{x-a}{a^2}) = \frac{(x-a)^2}{x \cdot a^2} \begin{cases} \geq 0 \\ \leq 2a^{-3} \cdot (x-a)^2. \end{cases}$$

Composition of polynomials with this one extra function gets us to all rational functions. Of course, the proof of the chain rule must pay attention to the distances from zeros which occur in the denominators.
In a similar way we handle the square root function. Here extra attention is needed for the domain of the function: As long as we only know rational numbers, the domain is rather thin, it contains only the squares of rational numbers (which, however, are still a dense subset). The following computation remains valid as more numbers become known.

$$0 < a/2 \leq x \Rightarrow |\sqrt{x} - \sqrt{a}| = \frac{|x-a|}{\sqrt{x} + \sqrt{a}} \leq \frac{1}{1.7} a^{-1/2} \cdot |x-a|$$

$$0 < a/2 \leq x \Rightarrow (\sqrt{x} - \sqrt{a} - \frac{x-a}{2\sqrt{a}}) = \frac{-(x-a)^2}{2\sqrt{a}(\sqrt{x}+\sqrt{a})^2} \begin{cases} \leq 0 \\ \geq 0.2a^{-3/2} \cdot (x-a)^2. \end{cases}$$

If we compose this function with polynomials then the discussion of the domain quickly becomes unmanagable; the computation shows that we know, even with error bounds, what the derivative of the square root function has to be, well before we know enough about numbers to be able to use this function freely.

All known proofs of the differentiation rules have the following property:
Given two differentiable functions f, g, we abbreviate their tangent functions as
$l_f(x) := f(a) + f'(a)(x-a),\ l_g(x) := g(a) + g'(a)(x-a)$.
Then, if we assume that the differences $f - l_f,\ g - l_g$ are "small", then also the differences
$(\alpha \cdot f + \beta \cdot g) - (\alpha \cdot l_f + \beta \cdot l_g), \quad f \cdot g - l_f \cdot l_g, \quad f \circ g - l_f \circ l_g$

are, in the same sense, "small".
To prove something we have to specify "small". Because of the functions known so far we may say: $f - l_f$ is "**small**" near a, means, there exists an interval $[a-r,\ a+r]$ and a constant K such that

$$x \in [a-r,\ a+r] \Rightarrow |f(x) - l_f(x)| \leq K \cdot |x-a|^2.$$

Later $f - l_f$ is "**small**" near a may have a more subtle interpretation:
For every $\epsilon > 0$ there exists $\delta > 0$ such that

$$x \in [a-\delta,\ a+\delta] \Rightarrow |f(x) - l_f(x)| \leq \epsilon \cdot |x-a|.$$

I find it important to emphasize that the form of the tangent approximation and the form of the differentiation rules **do not depend on the specification of "small"**. Moreover, even the strategy of the proofs does not depend on what exactly we mean by "small": if we change the definition of "small" in the assumptions, we can follow these changes through the proofs and end up with the changed conclusion. This possibility, to repeat the proofs under slightly changed assumptions, allows one to emphasize what is essential in these proofs.

3. Monotonicity Theorem and Related Results

A fundamental fact of analysis is that rough information about the derivative f' of a function f allows to deduce sharper information about f. As an example, take Lipschitz bounds:

$$\text{If} \quad x \in [\alpha, \omega] \Rightarrow |f'(x)| \leq L \quad \text{then}$$
$$a, b \in [\alpha, \omega] \Rightarrow |f(b) - f(a)| \leq L \cdot |b-a|.$$

For functions with uniform tangent approximations, i.e. for all the functions discussed so far, this fundamental theorem can be proved without invoking (even before discussing) the completeness of the reals. The only tool needed is Archimedes Principle, which is an obvious fact for the rationals and, later, an axiom for the reals:

Archimedes Principle

$$\text{If } 0 \leq r \text{ and } r \leq \frac{1}{n} \text{ for all } n \in \mathbb{N} \text{ then } r = 0.$$

The most intuitive result from the family of theorems that exploit derivative information probably is the

Monotonicity Theorem
Main assumption:

$$x \in [\alpha, \omega] \Rightarrow f'(x) \geq 0.$$

Technical assumption replacing completeness: The function f can be **uniformly** approximated by its derivatives, i.e. there exist positive constants r, K such that

$$x, a \in [\alpha, \omega], |x-a| \leq r \Rightarrow |f(x) - f(a) - f'(a)(x-a)| \leq K \cdot |x-a|^2.$$

Then f is nondecreasing:

$$a, b \in [\alpha, \omega], a < b \Rightarrow f(a) \leq f(b).$$

Note: The uniformity in the technical assumption is the crucial part, not the quadratic error bound; if one assumes

For each $\epsilon > 0$ exists a $\delta > 0$, which can be chosen **independently** of a, such that

$$x, a \in [\alpha, \omega], |x-a| \leq \delta \Rightarrow |f(x) - f(a) - f'(a)(x-a)| \leq \epsilon \cdot |x-a|,$$

then the strategy of the following proof also works.

Proof. For any $x, y \in [\alpha, \omega],\ x < y$ with $|x-y| \leq r$ we have (because of the two assumptions of the theorem):

$$-K \cdot |x-y|^2 \leq f(y) - f(x) - f'(x)(y-x) \leq f(y) - f(x).$$

Apply this to sufficiently short subintervals $[t_{j-1}, t_j]$ of the interval $[a, b]$, i.e. put $t_j := a + j/n \cdot (b-a), 0 \leq j \leq n$, with $(b-a)/n \leq r$ (Archimedes!) and have:

$$-K(b-a)^2/n^2 \leq f(t_j) - f(t_{j-1}),\ j = 1 \ldots n,$$

then sum for $j = 1 \ldots n$:

$$-K(b-a)^2/n \leq f(b) - f(a).$$

Archimedes Principle improves this to the desired claim $0 \leq f(b) - f(a)$. (The previous inequality changes under the ϵ-δ-assumption to $-K(b-a) \cdot \epsilon \leq f(b) - f(a)$, which also implies the theorem.)

Immediate consequences are:

Generalized monotonicity:

$$f' \leq g',\ a < b \Rightarrow f(b) - f(a) \leq g(b) - g(a)$$

Explicit bounds:

$$m \leq f' \leq M,\ a < b \Rightarrow m \cdot (b-a) \leq f(b) - f(a) \leq M \cdot (b-a)$$

Multiplicative version:

$$0 < f, g,\ \ \frac{f'}{f} \leq \frac{g'}{g},\ a < b \Rightarrow (\frac{g}{f})' = \frac{g}{f}(\frac{g'}{g} - \frac{f'}{f}) \geq 0 \Rightarrow \frac{f(b)}{f(a)} \leq \frac{g(b)}{g(a)}$$

Iterated application to second derivatives:

$$|f''| \leq B,\ a < x\ \Rightarrow -B(x-a) \leq f'(x) - f'(a) \leq B(x-a)$$

$$\Rightarrow \frac{-B}{2}(x-a)^2 \leq f(x) - f(a) - f'(a)(x-a) \leq \frac{B}{2}(x-a)^2.$$

These are strong improvements over the error bounds which were initially computed from the coefficients of polynomials. As illustration consider the Taylor polynomials $T_n(X)$ for a (not yet constructed) function with $f' = -f,\ f(0) = 1$:

$$T_n(x) := \sum_{k=0}^{n} (-1)^k \frac{x^k}{k!},\ 0 \le x \le 1,\ T_n'(x) = -T_{n-1}(x).$$

Since this is a Leibniz series we have the nested intervals:

$$[1-x, 1] \supset [T_1(x), T_2(x)] \supset [T_3(x), T_4(x)] \supset \ldots \supset [T_{2n-1}(x), T_{2n}(x)]$$

hence in particular: $|T_n''(x)| \le 1$. The "second derivative consequence" of the monotonicity theorem now implies uniform bounds for polynomials of **arbitrarily large** degree:

$$x, a \in [0,1] \Rightarrow |T_n(x) - T_n(a) + T_{n-1}(a)(x-a)| \le \frac{1}{2}(x-a)^2.$$

Clearly, if we only had the *existence* of a limit function T_∞ then Archimedes Principle would imply without further words differentiability and derivative of the limit function:

$$x, a \in [0,1] \Rightarrow |T_\infty(x) - T_\infty(a) + T_\infty(a)(x-a)| \le \frac{1}{2}(x-a)^2.$$

4. Completeness and Limits of Sequences of Functions

At most occasions I have prefered to begin the discussion of completeness with nested intervals; immediate applications are Leibniz series like the just mentioned Taylor polynomials. Thus we have interesting limit functions together with their derivatives right from the start. Cauchy sequences are the second step; the most important applications are sequences or series which are dominated by the geometric series (i.e. power series and the contraction lemma). Existence of sup and inf for bounded nonempty sets gives the most compact formulation for dealing with completeness. I have used the standard text book arguments, only the applications change since the results of the earlier sections are of significant technical help.

Note. Since so much emphasis was placed on computable error bounds I mention that the theorem "*Monotone increasing, bounded sequences converge*" is a very significant exception. In my opinion this intuitively desirable theorem is a major justification of the standard limit definition: Because the convergence speed of monotone increasing, bounded sequences can be slowed down arbitrarily (namely by repeating the elements of the sequence more and more often) one **cannot** prove the monotone sequence theorem with any version of a limit definition which requires more explicit error control than the standard definition. By contrast, in numerical analysis one tries to use sequences which converge at least as **fast** as some geometric sequence, i.e. one has more explicit control like $|a_{n+p} - a_n| \le C \cdot q^n$.

Next I illustrate how the monotonicity theorem often allows short replacements of standard induction proofs.

Bernoulli's inequality is a version of the monotonicity theorem:
$f(x) := (1+x)^n \geq 1 + n \cdot x = f(0) + f'(0) \cdot x$, since $f''(x) \geq 0$ for $x \geq -1$.

The monotonicity of $n \to f_n(x) := (1 + x/n)^n,\ 0 \leq x$ and the decreasing of $n \to g_n(x) := (1 - x/n)^{-n},\ 0 \leq x < n$ requires no technical skill since $(f_n'/f_n)(x) = 1/(1+x/n) \leq (f_{n+1}'/f_{n+1}(x) \leq 1 \leq 1/(1-x/n) = (g_n'/g_n)(x) \leq (g_{n-1}'/g_{n-1})(x)$.

The following are convenient estimates of the geometric series and its derivatives which give the desired uniform constants when dealing with power series:

$$0 \leq x < 1, \qquad \sum_{k=0}^{n} x^k = \frac{1 - x^{n+1}}{1-x} \leq \frac{1}{1-x}$$

$$\left(\sum_{k=0}^{n} x^k\right)' = \sum_{k=1}^{n} k x^k = \frac{-(n+1)x^n}{1-x} + \frac{1 - x^{n+1}}{(1-x)^2} \leq \frac{1}{(1-x)^2}$$

$$\left(\sum_{k=0}^{n} x^k\right)'' = \frac{-(n+1)n x^{n-1}}{1-x} - \frac{2(n+1)x^n}{(1-x)^2} + 2\frac{1 - x^{n+1}}{(1-x)^3} \leq \frac{2}{(1-x)^3}.$$

As mentioned, the standard proof of the Contraction Lemma uses explicit error bounds and our limit arguments are in the same spirit: Functions $f : M \to M$ with $|f'| \leq q < 1$ are contracting, $|f(x) - f(y)| \leq q \cdot |x - y|$. And for contracting maps sequences generated by iteration, $a_{n+1} := f(a_n)$, are geometrically dominated and hence Cauchy:

$$|a_{n+p} - a_n| \leq \frac{|a_1 - a_0|}{1-q} \cdot q^n.$$

An example of a contracting rational map with the irrational golden ratio as fixed point is

$$f(x) := 1/(1+x),\ f : [1/2, 1] \to [1/2, 1],\ |f'(x)| \leq |-(1+x)^{-2}| \leq 4/9.$$

The approximating sequence $a_0 = 1, a_1 = 1/(1+1), \ldots, a_n = 1/(1+1/(1+1/\ldots))$ consists of the optimal approximations from continued fractions. This is to show that sequences do appear in my approach, but they are handled with the monotonicity theorem and not presented as the road to differentiation.

The main question about limit functions is, of course, "What is their derivative?" Usually one employs uniform convergence of the differentiated sequence and the main theorem connecting differentiation and integration. I illustrate the use of uniform error bounds in the case of power series $P_n(X) := \sum_{k=0}^{n} a_k X^k$. The basic tool in both approaches is the comparison with a geometric series and it will better bring out the differences if I add an assumption which simplifies either approach: a bound for the coefficients, $|a_k| \leq C$.
Then we have for all x with $|x| \leq q < 1$

$$|P_{n+m}(x) - P_n(x)| \leq \sum_{n<k} |a_k| q^k \leq \frac{C}{1-q} q^{n+1},$$

which shows that $\{P_n(x)\}$ is a Cauchy sequence (in fact uniformly for $|x| \le q$). For the first derivatives we obtain uniform bounds by comparing with the geometric series

$$|P_n'(x)| \le \sum_{k\ge1} |a_k| k q^{k-1} \le \frac{C}{(1-q)^2} =: L.$$

The Monotonicity Theorem implies **for all** n the Lipschitz bound

$$|x|, |y| \le q \Rightarrow |P_n(y) - P_n(x)| \le L \cdot |y - x|$$

and Archimedes Principle extends this uniform estimate to the limit function

$$|x|, |y| \le q \Rightarrow |P_\infty(y) - P_\infty(x)| \le L \cdot |y - x|.$$

Similarly we employ our estimate of the geometric series to get uniform bounds for the second derivatives

$$|P_n''(x)| \le \sum_{k\ge2} |a_k| k(k-1) q^{k-2} \le \frac{C}{(1-q)^3} =: K,$$

we use the Monotonicity Theorem to get uniform tangent approximations

$$|x|, |y| \le q \Rightarrow |P_n(y) - P_n(x) - P_n'(x)(y-x)| \le K \cdot |y - x|^2,$$

and Archimedes Principle again extends these uniform bounds to the limit

$$|x|, |y| \le q \Rightarrow |P_\infty(y) - P_\infty(x) - \lim_{n\to\infty} P_n'(x)(y-x)| \le K \cdot |y - x|^2.$$

This proof of the differentiability and the determination of the derivative of the limit function clearly extends to complex power series, a first step to higher dimensional analysis. Another immediate extension is to the differentiation of curves $c := (c_1, c_2, c_3) : [a, b] \to \mathbb{R}^3$, which is important in itsself but also a prerequisite for analysis in $\mathbb{R}^n$.

5. Integrals, Riemann Sums, Antiderivatives

Some notion of tangent (and hence some version of derivative) was known centuries before Newton. Similarly, summation of infinitesimals had already troubled the Greeks, and Archimedes determination of the area bounded by a parabola and a secant was definite progress. Against the background of this early knowledge I find the conceptual progress achieved with the definition of the integral even more stunning than that achieved with differentiation. I try to teach integrals as a fantastic generalization of sums, they allow for example to "continuously sum" the velocity of an object in order to obtain the distance which it travelled. With this goal in mind I think it is fundamental that Riemann sums of a function f can be computed up to controlled errors if one knows an antiderivative F of the given f, i.e. $F' = f$. I use the standard definition of the integral in terms of Riemann sums. I add a construction of an antiderivative F for a continuous f which is in the spirit of uniform error control, but from now on the differences to the standard approach are not very pronounced.

Let $f : [a,b] \to \mathbb{R}^3$ be given. Consider a subdivision of $[a,b]$, i.e. $a = t_0 < t_1 < \ldots < t_n = b$ and choose intermediate points $\tau_j \in [t_{j-1}, t_j],\ j = 1 \ldots n$. For these data we **define** the

Riemann Sum

$$\mathfrak{RS}(f) := \sum_{j=1}^{n} f(\tau_j)(t_j - t_{j-1}).$$

Now, if f is at least continuous and $F' = f$ then the difference $|F(b) - F(a) - \mathfrak{RS}(f)|$ is "small". How "small" depends on the precise assumption for f. For example a Lipschitz bound, $x, y \in [a,b] \Rightarrow |f(y) - f(x)| \leq L \cdot |y - x|$, implies the

Error Bound

$$|F(b) - F(a) - \mathfrak{RS}(f)| \leq L \cdot (b-a) \cdot \max_{1 \leq j \leq n} |t_j - t_{j-1}|.$$

This estimate (or other versions) follows from the Monotonicity Theorem and the triangle inequality; first note the derivative bound:

$$x \in [t_{j-1}, t_j] \Rightarrow |(F(x) - f(\tau_j) \cdot x)'| = |f(x) - f(\tau_j)| \leq L \cdot |x - \tau_j| \leq L \cdot |t_j - t_{j-1}|,$$

hence

$$|F(t_j) - F(t_{j-1}) - f(\tau_j)(t_j - t_{j-1})| \leq L \cdot |t_j - t_{j-1}|^2,$$

so that summation over $j = 1 \ldots n$ gives the claimed error bound.

Continuity (see below) of f implies with the same arguments an ϵ-δ-error bound:

$$\max_{1 \leq j \leq n} |t_j - t_{j-1}| \leq \delta \Rightarrow |F(b) - F(a) - \mathfrak{RS}(f)| \leq \epsilon \cdot (b-a).$$

I prefer to define the integral for vector valued functions (for direct application to integrals of velocities). The notion of limit has to be generalized to cover "convergence" of Riemann sums: For every $\epsilon > 0$ there is a subdivison of $[a,b]$ such that **for all finer subdivisions** and all choices of intermediate points τ_j the corresponding Riemann sums differ by less than ϵ. Almost the same argument as used to prove the error bound gives: For continuous (or better) f the Riemann sums converge; the limit is called the integral of f over $[a,b]$, notation $\int_a^b f(x)dx$. Moreover, if $F' = f$ then the error bound proves

$$F(b) - F(a) = \int_a^b f(x)dx.$$

The triangle inequality for Riemann sums gives the **Triangle Inequality for Integrals**

$$a < b \Rightarrow \left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx.$$

Also, other analogies between sums and integrals need to be discussed.

I describe a construction of an antiderivative which is independent of the definition of the integral (its proof, based on uniformity, works in Banach spaces).

Theorem. Let f be Lipschitz continuous on $[a,b]$ (soon: uniformly continuous). Then one can approximate f uniformly by piecewise linear "secant" functions s_n. These have piecewise quadratic antiderivatives and a limit of these is an antiderivative of f. In more detail:

$$s_n(t_j) := f(t_j),\ j = 0,\dots,n,\ (a = t_0 < t_1 < \dots < t_n = b)$$
$$s_n(x) := \frac{f(t_{j-1})\cdot(t_j - x) + f(t_j)\cdot(x - t_{j-1})}{t_j - t_{j-1}} \text{ for } x \in [t_{j-1}, t_j]$$
$$|f(x) - s_n(x)| \le L \cdot \max|t_j - t_{j-1}| =: r_n$$

Obviously, s_n has a piecewise quadratic antiderivative $S_n, S_n'(x) = s_n(x), S_n(0) = 0$. From the Monotonicity Theorem and

$$|S_{n+m}'(x) - S_n'(x)| \le 2r_n$$

we have the (uniform) Cauchy property:

$$|S_{n+m}(x) - S_n(x)| \le 2(b-a)r_n,$$

which, by completeness, gives a limit function S_∞. But we again have uniform error bounds:

$$|(S_n(x) - S_n(c) - s_n(c)(x-c))'| = |s_n(x) - s_n(c)| \le L\cdot|x-c|,$$

hence

$$|S_n(x) - S_n(c) - s_n(c)(x-c)| \le L\cdot|x-c|^2.$$

Archimedes Principle gives the final result:

$$|S_\infty(x) - S_\infty(c) - f(c)(x-c)| \le L\cdot|x-c|^2,$$

which says that S_∞ is differentiable and $S_\infty' = f$, as claimed.

Note in this proof: The approximation $S_n(b) - S_n(a)$ for $S_\infty(b) - S_\infty(a) = \int_a^b f(x)dx$ is a frequently used numerical approximation for the integral of f.

6. Continuity, Theorems and Examples

The arguments in this last section are the standard arguments. My point is that with the experience of the previous sections one can achieve a better understanding of continuity and, moreover, this takes less time than a treatment of continuity in the early parts of a course. Since convergent sequences were such an essential tool (for getting limit functions) we ask: "What kind of functions are compatible with convergence?" We define:

$f : A \subset \mathbb{R}^d \to \mathbb{R}^e$ is called **sequence continuous** at $a \in A$
if **every** sequence $a_n \in A$ which converges to the limit $a \in A$ has its image sequence $f(a_n)$ converging to $f(a)$.

We see that linear combinations, products, compositions of sequence continuous functions are (directly from the definition) sequence continuous. But $1/f$ causes a problem:

If $f(a) \neq 0$ then we would like to find an interval $[a-\delta, a+\delta]$ on which f **is not zero**.

If such an interval could be found then on it $1/f$ would clearly be sequence continuous. It is well known that such a $\delta > 0$ can only be found with an indirect proof. But this proof is essential for understanding continuity. It is also very similar to the equivalence of sequence continuous and ϵ-δ-continuous.

Definition: f is ϵ-δ-continuous at a if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$x \in [a-\delta, a+\delta\,] \Rightarrow |f(x) - f(a)| \leq \epsilon.$$

It remains to prove the main theorems and to show examples. First I give short summaries of the proofs to recall what kind of arguments are involved. 1.) The intermediate value theorem: By interval halfing construct a Cauchy sequence which converges to a preimage of the given intermediate value. 2.) Boundedness on complete, bounded sets: In an indirect proof construct by interval halfing a Cauchy sequence on which the function f is unbounded, a contradiction to the continuity at the limit point of the Cauchy sequence. 3.) Extremal values are assumed on complete, bounded sets: Since the function is bounded by the previous result we have $\sup f$ and $\inf f$; with interval halfing we find Cauchy sequences $\{a_n\}$ such that the sequences of values, $\{f(a_n)\}$, converge to $\sup f$ resp. $\inf f$. 4.) Uniform continuity on complete, bounded sets: Necessarily indirect, if for some $\epsilon^* > 0$ no $\delta > 0$ is good enough then we find a pair of Cauchy sequences $\{a_n\}, \{b_n\}$ which have the **same** limit but $|f(a_n) - f(b_n)| \geq \epsilon^*$, a contradiction to the continuity at the common limit point. 5.) Uniformly convergent sequences of continuous functions have a continuous limit function: To given $\epsilon > 0$ choose an $\epsilon/3$-approximation from the sequence and for this (continuous!) approximation find $\delta > 0$ for $\epsilon/3$-deviations; this δ guarantees with the triangle inequality at most ϵ-deviations for the limit function.

From the construction of examples I recall that comparison with a geometric sequence is the main tool. I find it misleading to call the following examples "weird", as if continuity "really" were much more harmless. 1.) The polygonal approximations of Hilbert's cube filling curve are explicitly continuous: To guarantee value differences $\leq 2^{-n}$ the arguments have to be closer than 8^{-n}; Archimedes Principle concludes the same for the limit function. 2. Similarly for Cantor's staircase, a monotone increasing continuous function which is differentiable with derivative 0 except on a set of measure zero: The piecewise linear approximations satisfy: To guarantee value differences $\leq 2^{-n}$ the arguments have to be closer than 3^{-n} and Archimedes Principle does the rest. 3.) And continuous but nowhere differentiable functions can be obtained as obviously uniform limits of sums of faster and faster oscillating continuous functions like $\sum_k 2^{-k} \sin(8^k x)$.

In view of the much better-than-continuos properties of all the functions we have so far been concerned with, I find it not so obvious how to demonstrate the usefulness of the notion of continuity. The proof of the existence of solutions of ordinary differential equations based on the Contraction Lemma in the complete metric space of **continuous** functions (sup-norm) is my lowest level convincing example.

Since the Monotonicity Theorem and its consequences are usually derived from the (essentially one-dimensional) Mean Value Theorem of Differentiation, I finally show that the theorems which conclude from assuming derivative bounds can be obtained with shorter proofs (again valid in Banach spaces).

Theorem: Derivative bounds are Lipschitz bounds.

More precisely, let $A \subset \mathbb{R}^d$ be a convex subset and let $F : A \to \mathbb{R}^e$ be ϵ-δ-differentiable, and for emphasis: without any uniformity assumption. Assume further a bound on the derivative: $x \in A \Rightarrow |TF(x)| \leq L$. Then

$$a, b \in A \Rightarrow |F(a) - F(b)| \leq L \cdot |a - b|.$$

Indirect proof via halfing. If the inequality were not true then we had some $a, b \in A$ with $|F(a) - F(b)| > L \cdot |a - b|$, i.e. with some fixed $\eta > 0$ we had $|F(a) - F(b)| \geq (L + \eta) \cdot |a - b|$. Let $m := (a + b)/2$ be the midpoint. Then for either a, m or m, b the same inequality has to hold (since otherwise, by the triangle inequality, $|F(a) - F(b)| < (L + \eta) \cdot |a - b|$). In other words, we have a_1, b_1 with half the distance $|a_1 - b_1| = |a - b|/2$, but still

$$|F(a_1) - F(b_1)| \geq (L + \eta) \cdot |a_1 - b_1|.$$

This procedure can be repeated, we get a pair of Cauchy sequences $\{a_n\}, \{b_n\}$ with the same limit c between a_n and b_n on the closed segment from a to b, but with the inequalities:

$$|F(a_n) - F(b_n)| \geq (L + \eta) \cdot |a_n - b_n|.$$

By differentiability of F at $c \in A$ we have for $\epsilon = \eta/2$ a $\delta > 0$ such that

$$\begin{aligned} x \in A, |x - c| \leq \delta \Rightarrow |F(x) - F(c) - TF|_c \cdot (x - c)| &\leq \epsilon \cdot |x - c| \\ \Rightarrow |F(x) - F(c)| &\leq (L + \epsilon) \cdot |x - c|. \end{aligned}$$

Choose n so large that $|a_n - c|, |b_n - c| \leq \delta$ so that the last inequality holds for $x = a_n$ and $x = b_n$. Add the inequalities and observe $|a_n - c| + |b_n - c| = |a_n - b_n|$ to obtain the contradiction

$$(L + \eta) \cdot |a_n - b_n| \leq |F(a_n) - F(b_n)| \leq (L + \eta/2) \cdot |a_n - b_n| < (L + \eta) \cdot |a_n - b_n|.$$

Absolute and Relative Errors

In this section we collect prerequisites about inequalities. Simple calculus situations can already be described adequately using percentage calculation. Looking at the propagation of errors before the language of analysis is available prepares the way for a good intuitive understanding of the objectives of calculus.

Example 1. Absolute and relative differences and their behavior under adding or multiplying a constant.
The equivalent statements $x \in [0.9, 1.1]$ or $0.9 \le x \le 1.1$ can be put into words in the following two ways, which have the same meaning but different focus:

x differs from 1 by no more than 0.1 ("absolute difference"),
x differs from 1 by no more than 10% ("relative difference"),

Terminology: Clearly, the statement "The relative (or percentage) difference between x and a is $:= |x-a|/|a|$" will be used only if $a \neq 0$. Often we will also tacitly assume $0 < x, a$.

Under addition and multiplication by a constant, relative and absolute differences show different behavior. The first statement ($0.9 \le x \le 1.1$) implies:

$x+1$ differs from 2 by no more than 0.1 ,
$x+1$ differs from 2 by no more than 5% ,
$2x$ differs from 2 by no more than 0.2,
$2x$ differs from 2 by no more than 10%,
$x-0.7$ differs from 0.3 by no more than 0.1
$x-0.7$ differs from 0.3 by no more than 33% (!),
$0.5x$ differs from 0.5 by no more than 0.05,
$0.5x$ differs from 0.5 by no more than 10%.

Special attention deserves the enormous increase of *relative* differences when we *subtract* numbers close to the reference quantity.
When we square the numbers, it makes a difference whether we look at the deviations at the upper or the lower bound:

x^2 lies in the interval $[0.81, 1.21]$,

therefore x^2 differs from 1 by no more than 21% . Normally we don't say: x is less than 1 by no more than 19% , and bigger than 1 by no more than 21% . But it's common to say: If x differs from 1 by "approximately " 10%, then x^2 differs from 1 by "approximately" 20% .
Apply these examples to the initial situation $x \in [99, 101]$.

Example 2. The Sizes of Different Powers.
While we have verbally expressed the difference between approximate numbers and precise constants in example 1, we will now consider the *relative sizes* of different powers. Caveat: The behavior is opposite in intervals close to 0 and close to ∞!

For all $x \in (0,\ 0.1]$ we have:
x^2 is not exceeding 10% of x,
x^3 is not exceeding 1% of x.

For all $x \in (0,\ 0.02]$ we have:
x^2 is not exceeding 2% of x,
x^3 is not exceeding 0.04% of x.

In the interval $[20, \infty)$ we have:
x does not exceed 5% of x^2,
x does not exceed 0.25% of x^3.

In the interval $[100, \infty)$ we have:
x dos not exceed 1% of x^2,
x dos not exceed 0.01% of x^3.

See how the percentages depend on different intervals: e.g. in the first case: $x^2 = x \cdot x \leq 0.1 \cdot x$, and that is to say that x^2 is less or equal 10% of x, or $x^2/x \leq 0.1$.

Example 3. When we take reciprocals $x \to 1/x$, small relative differences change very little
We assume a relative difference of no more than 5%. Thus with a as 100%,

$$|x - a|/|a| = |(x/a) - 1| \leq 0.05.$$

Calculations are often simpler without the absolute value bars: $1 - 0.05 \leq x/a \leq 1 + 0.05$ implies

$$1/(1 + 0.05) \leq a/x \leq 1/(1 - 0.05) \text{ or } -0.05/(1 + 0.05) \leq (a/x) - 1 \leq 0.05/(1 - 0.05).$$

It gives us
$$\frac{|1/x - 1/a|}{|1/a|} = \frac{|x - a|}{|x|} = |1 - \frac{a}{x}| \leq \frac{0.05}{1 - 0.05}.$$

We notice that when we assume $|x/a|$ differs only "a little bit" ($\leq 5\%$) from 1 , almost the same holds for the reciprocal a/x, and the relative difference between $1/x$ and $1/a$ is just "a little bit" bigger than 0.05.
What changes when we sharpen the assumption from $\leq 5\%$ to $\leq 1\%$?

Example 4. Errors
So far we have talked about "differences". It's common to talk about "errors" when dealing with numerical approximations of theoretically precise numbers. Let me use two famous examples to explain.

(i) Archimedes' approximation for π, i.e. $3\frac{10}{71} < \pi < 3\frac{1}{7}$, can be expressed as follows. The numerical approximation $3\frac{1}{7}$ for π has an absolute error of at most $3\frac{10}{70} - 3\frac{10}{71} \sim 0.002$ and a relative error of at most $(\frac{10}{70} - \frac{10}{71})/3.14 \sim 0.064 \cdot 10^{-2}$, i.e. of at most 0.064%.
(ii) The inequality $1.4^2 = 1.96 < 2 < 2.25 = 1.5^2$ implies the inequality $1.4 < \sqrt{2} < 1.5$ for the number written by the symbol $\sqrt{2}$. This means that the numerical approximation 1.5 for $\sqrt{2}$ has an absolute error of at most 0.1 and a relative error of at most $0.1/1.4 \sim 0.072$, i.e., at most 7.2%.

Example 5. Improvement of approximations by reasoning with relative and absolute errors.
We combine examples 1, 3 and 4 in order to improve a given approximation for $\sqrt{2}$. Remembering the factoring formula $(a+b)(a-b) = a^2 - b^2$, we specialize it to $(\sqrt{2}+1)(\sqrt{2}-1) = 2 - 1 = 1$ and convert this to

$$\sqrt{2} = \frac{1}{(1+\sqrt{2})} + 1.$$

We read the right hand side as "the value of the function $x \mapsto 1/(1+x) + 1$ at $x = \sqrt{2}$". Now we will show that evaluating the function at an approximation x for $\sqrt{2}$ returns a *better* approximation: According to example 1, the relative error of $1+x$ (with respect to $1+\sqrt{2}$) is less than that of x (with respect to $\sqrt{2}$); according to example 3 the relative error of $1+x$ is almost the same as that of $1/(1+x)$. According to example 1, the relative error of $1/(1+x)+1$ is again less than that of $1/(1+x)$; hence $1/(1+x)+1$ is a better approximation for $\sqrt{2}$ than x.
Let's go through this again, now using our numerical values.
Since 1.5 is an approximation of $\sqrt{2}$ with an absolute error < 0.1, the number 2.5 is an approximation of $1+\sqrt{2}$ with the same absolute error and hence with a relative error less than $0.1/2.4 \sim 0.042$ or 4.2%. Therefore $1/2.5 = 0.4$ is an approximation of $1/(1+\sqrt{2})$ with almost the same relative error 4.2%, which means with an absolute error of $0.4 \cdot 4.2/100 < 0.017$. (To be sure, calculate this absolute error directly from the approximations 1.4 and 1.5 for $\sqrt{2}$ by plugging them into $1/(1+x)$.) Thus, $1 + 0.4$ is an approximation of $1 + 1/(1+\sqrt{2}) = \sqrt{2}$ with the same absolute error (< 0.017) and a relative error $0.017/1.4 < 0.0122$, i.e. is less than 1.22%.
Our skillful use of our factoring formula, i.e. converting $(\sqrt{2}+1)(\sqrt{2}-1) = 1$ *into* $\sqrt{2} = 1 + 1/(1+\sqrt{2})$, *helps us to get an approximation with the relative error of less than* 1.3% *from an approximation with the relative error of less than* 7.2%. *We have reduced the error to almost one sixth of the original.* Of course we can repeat this improvement. $1 + 1/(1+1.4) = 1 + 1/(1+7/5) = \frac{17}{12} \sim 1.4167$ and $1.4167 - \sqrt{2} \sim 0.0025$, which results in the relative error of $0.0025/1.4 < 0.0018$ or less than 0.18% (roughly one sixth of 1.3%).

Example 6. Irrationality of $\sqrt{2}$, an indirect proof.
I hope it causes some amazement to see these "self improving" formulas. Therefore we look at our factoring formula once again. The symbol $\sqrt{2}$ appears twice. Let us see what happens when we resolve it not as in example 5, but in a different way, i.e., to $\sqrt{2} = 1/(\sqrt{2}-1) - 1$. When we carefully reproduce the previous calculations, we observe the following. Evaluating the right hand side at an approximation for $\sqrt{2}$ results in a *worse* approximation! Who would ever want this? Indeed, nothing useful happens, when we see the approximation (for $\sqrt{2}$) as a decimal number. However, when we see the approxiation as a *reduced* fraction p/q with $1 < p/q < 2$, we obtain (when we expand the first double

fraction with the denominator of p/q):

$$1/(\frac{p}{q}-1)-1=\frac{q}{p-q}-1=\frac{2q-p}{p-q},$$

and because of $\quad q<p<2q \quad$ we get $\quad 0<p-q<q.$

Therefore we obtain a fraction with a *smaller denominator*! This observation allows us to find an *indirect proof* for the irrationality of $\sqrt{2}$ with the help of our formula useless for improvement of approximations.
Suppose $\sqrt{2}$ were rational, therefore equals a fraction, say, $\sqrt{2}=P/Q$. Since there are *only finitely many* fractions between 1 and 2 with a denominator $\leq Q$, we can pick the one with the *smallest denominator* that equals $\sqrt{2}$, say $\sqrt{2}=p/q$. But now, the previous calculation produces:

$$\frac{p}{q}=\sqrt{2}=\frac{1}{\sqrt{2}-1}-1=\frac{2q-p}{p-q},$$

which is a fraction with *denominator* $(p-q)$ *smaller* than the selected fraction p/q having the *smallest* denominator. This contradiction refutes our assumption that $\sqrt{2}$ were rational. — Of course we may think of p/q as the completely reduced fraction. But the proof is formulated such that we don't use the concept of a completely reduced fraction. Instead we only use the "principle of the smallest delinquent": If $\sqrt{2}$ were rational, there would be a smallest example for it. But a "smallest" example plugged into our formula produces smaller denominators and thus shows that such a smallest example cannot exist.

Example 7. Irrationality of $\sqrt{2}$, "direct" proof using absolute errors
We pose the question: How far away from 0 is the value of the polynomial $P(x):=x^2-2$ at some rational number $x=p/q$? We want to show $|P(p/q)|\geq 1/q^2$, in words: *The absolute value of the polynomial P at $x=p/q$ is at least* $1/q^2$. If we prove the statement for all reduced fractions, then it's even more true for the non-reduced ones since in this case the statement is weaker. Therefore we may assume that all the factors 2 have been cancelled, i.e

either p is odd, or if not then q is odd.

In the first case, the numerator of $P(p/q)=(p^2-2q^2)/q^2$ is **odd**, hence it's absolute value is at least 1 and therefore $|P(p/q)|\geq 1/q^2$. In the second case, p^2 is divisible at least by $2\cdot 2$, but $2q^2$ by only one factor 2; in this case the numerator of $P(p/q)$ is the double of an odd number, hence it's absolute value is at least 2. In any case we get: The polynomial $P(x):=x^2-2$, whose zeros are the square roots of 2, has an absolute value of at least $1/q^2$. Therefore the polynomial cannot have any rational zeros.
It is useful to develop some feeling for how good the derived inequalities are. Allthough we have used only a very simple argument for proving $|P(p/q)|\geq 1/q^2$ the equality sign holds in fact for infinitely many fractions: First, the equal sign in $|P(p/q)|\geq 1/q^2$ holds for $p/q:=3/2$; and then as well for the infinitely many fractions $7/5, 17/12, 41/29,\ldots$ which

are generated from $p/q := 3/2$ in example 5, since $|\text{numerator}^2 - 2\cdot\text{denominator}^2|$ doesn't change during improvement:

$$|(2q+p)^2 - 2(p+q)^2| = |2q^2 - p^2| = \ldots = |2 \cdot 2^2 - 3^2|.$$

Example 8. Mean Values and Errors

Mean values can often be used to make errors smaller. I am going to explain the three most important mean values and the inequalities between them. Then I will use them to improve approximations for square roots *faster* than in example 5.

Let $0 < a, b$ be positive numbers. Then

the Arithmetic Mean is the number $A(a,b) := (a+b)/2$,

the Geometric Mean is the number $G(a,b) := \sqrt{a \cdot b}$,

the Harmonic Mean is the number $H(a,b) := 1/((1/a + 1/b)/2) = 2ab/(a+b)$.

$A(a,b)$ is the center of the line segment with end points a, b. $G(a,b)$ ist the side length of a square which has the same area as the rectangle with side lengths a and b. Therefore $G(a,b)$ is as well the altidude of a right triangle where the measures of the hypotenuse segments are a and b (right triangle altitude theorem). The radius of the circumcircle of this triangle is $A(a,b)$, and therefore we have:

$G(a,b) \leq A(a,b)$.

This followos as well from the binomial formula $0 \leq (\sqrt{a} - \sqrt{b})^2 = a + b - 2\sqrt{a \cdot b}$.

We have the simple relationships between the three means:

$1/H(a,b) = A(1/a, 1/b)$ und $G(a,b)^2 = H(a,b) \cdot A(a,b)$. Since $G \leq A$, the latter implies

$H(a,b) \leq G(a,b)$.

The distance between harmonic and arithmetic mean of a and b can be a lot smaller than the distance between a and b, which equals

$$A(a,b) - H(a,b) = \frac{a+b}{2} - \frac{2ab}{a+b} = \frac{(a+b)^2 - 4ab}{2(a+b)} = \frac{a-b}{2} \cdot \frac{a-b}{a+b}.$$

We see immediately that the distance between the means is always smaller than half of the distance between a and b. But since we have seen in example 2 that small numbers get a lot smaller when we square them, we understand that $A(a,b) - H(a,b)$ is a lot smaller than $|a-b|$ in case $a-b$ is only trickles of $a+b$.

Now we apply this to $\sqrt{3}$. Let a be an approximation of $\sqrt{3}$ and $a > \sqrt{3}$. Then $b := 3/a$ is an approximation which is too small. The arithmetic mean $a_1 := A(a, 3/a)$ and the harmonic mean $b_1 := H(a, 3/a) = 3/a_1$ lie *closer* to each other having the unchanged geometric mean $G(a, 3/a) = \sqrt{3}$ in between. It is worthwhile to try out by calculator how rapidly this process of *quadratic error reduction* improves the approximation

$$a_1 - b_1 = A(a, 3/a) - H(a, 3/a) = \frac{a - 3/a}{2} \cdot \frac{a - 3/a}{a + 3/a} < \left(\frac{a - 3/a}{2}\right)^2 = \left(\frac{a-b}{2}\right)^2.$$

Example 9. Deviation of circle tangents from the circle.
Tangents will play a prominent role in differential calculus. Up to precalculus, "tangents" appear only as tangents to a circle. We want to look closer at the unit circle $K := \{(x,y);\ x^2+y^2=1\}$ and its upper tangent $T := \{(x,y);\ y=1\}$

If we increase x starting from 0 and move along the tangent, we move away from the tangent point B. Simultaneously the distance to the circle gets larger, but in the beginning apparently by a small amount only. Can we express this more clearly? Clear enough to be able to invent the definition of a tanget to more complicated curves but circles?

The distance between the point $P = (x,1) \in T$ and the tangent point $B = (0,1) \in T \cap K$ is $|x|$, compare the following picture. By the Pythagorean theorem, the distance between P and the origin, the center of the circle K, is $\sqrt{1+x^2}$. Therefore the distance between P and the unit circle is $\sqrt{1+x^2}-1$. Using example 2, we want to show that this distance is *a lot* smaller than $|x|$, at least when P is close to B. (In the following calculation, we first expand by $\sqrt{1+x^2}+1$ then we apply $(a-1)(a+1) = a^2-1$, and eventually we decrease the denominator under the square root by omitting x^2, i.e. we make the fraction larger):

$$0 \le \sqrt{1+x^2}-1 = (\sqrt{1+x^2}-1)\cdot\frac{\sqrt{1+x^2}+1}{\sqrt{1+x^2}+1} = \frac{x^2}{\sqrt{1+x^2}+1} \le \frac{x^2}{2}$$

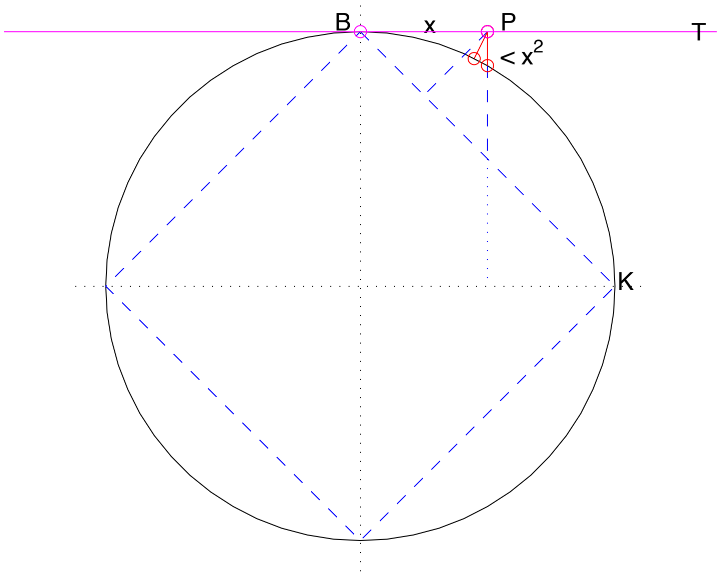
and therefore e.g. $|x| \le 0.1 \Rightarrow \sqrt{1+x^2}-1 \le \frac{|x|}{2}\cdot|x| \le 0.05\cdot|x|$.

I.e., close to B, more precisely for $|x| \le 0.1$, the distance between P and K is at most 5% of the distance $|x|$ between P and B; and for $|x| \le 0.01$ the distance between P and K is at most 0.5% of $|x|$, etc. as we have seen in example 2. This shows: The tangent moves away from the circle so slowly (when $|x|$ grows) that the formulation "the tangent touches the circle" is justified.

If we rather measure the distance between the circle and the tangent perpendicular to the tangent (instead of perpendicular to the circle as we've done right now), then we have to take the distance between $(x,1) \in T$ and $(x,\sqrt{1-x^2}) \in K$. Apart from the fact that we obviously have to assume $|x| \le 1$, almost the same as before happens (only the factor 1/2 disappears):

$$0 \le 1-\sqrt{1-x^2} = (1-\sqrt{1-x^2})\cdot\frac{1+\sqrt{1-x^2}}{1+\sqrt{1-x^2}} = \frac{x^2}{1+\sqrt{1-x^2}} \le x^2.$$

The distance between P and K is not bigger than the **square** of the distance between P and the tangent point B, hence by example 2 "a lot" smaller than that.



How much slower than from the square does the tangent move away from the circle?

If we think the straight line T and the square Q were made out of metal, we could still say: The straight line *touches* the square. But we do *not* mean this physical contact when we say "The tangent touches the circle" in mathematics. We can very well express the difference by means of inequalities: The distance between the point $P \in T$ and the square (measured in parallel to the y-axis) is the same as the distance between P and B, namely x, *no matter* how small x is. I repeat that the distance between P and K is smaller than x^2 and therefore, it becomes a *smaller and smaller* percentage of x the smaller x becomes. This big difference between the tangency of T and K on the one side and the tangency of T and Q on the other side doesn't disappear when we measure the distances from P perpendicular to Q and K (the perpendicular distance to the square ist $x/\sqrt{2}$, the perpendicular distance to the circle is smaller than $x^2/2$).

This is an important example where the meaning of a word as it has been defined in mathematics can be different from its meaning in colloquial language. It's completely legitimate that the word "to touch" has a completely different meaning for somebody who lays electricity cables than for somebody who says: A curve is touched by its tangent.

Tangents to the Parabola

Derivatives and tangents are central to calculus. Having looked at the circle only we haven't been able to notice that yet, because we call a line a tangent line at a point on the circle if it is perpendicular to its radius. Obviously we don't use any calculus notions in this case. Now we want to understand the case of a less symmetric curve, the graph of the quadratic function $x \mapsto x^2$. There is no doubt about what straight line we want to call a tangent line. The properties of these "naive" tangents which we can draw in the case of a circle or a parabola, will lead us to the *definition of a tangent* in the next chapter. Though this definition is not yet the definition that has been used *since the end of the 19th century*, it will take quite a long time until we are able to find functions that should be "differentiable" but our preliminary definition is too restrictive for these complicated functions. Our preliminary definition will allow us to treat many functions, including all the rational functions, using simpler arguments and more explicit inequalities than the final definition.

The tangent at the lowest point of a circle has the property, that all secant lines on the right hand side have a bigger (that is positive) slope than the (horizontal) tangent, while all the secant lines on the left hand side have a smaller slope than the tangent. Now, through every point (a, a^2) on the graph of the function $x \mapsto P(x) := x^2$ there is exactly one straight line having this property. In order to show this, we calculate for every $x \neq a$ the

Slope of the secant line: $$\frac{P(x) - P(a)}{x - a} = \frac{x^2 - a^2}{x - a} = x + a.$$

Left hand side: $x < a \Rightarrow x + a < 2a$, Right hand side: $a < x \Rightarrow 2a < x + a$.

Therefore, the graph of the linear function ℓ with slope $2a$ and the value $\ell(a) = a^2$, i.e. $\ell(x) := 2a \cdot (x - a) + a^2 = 2a \cdot x - a^2$ has the property we've noticed when we looked at the circle, namely, the slope of this straight line is smaller than the slope of any of the secant lines on the right hand side of the parabola and larger that the slope of any of the secant lines on its left hand side.
Furthermore, the Pythagoras theorem shows that any tangent to a circle doesn't meet the inside of this circle. We could say as well that the tangent keeps the circle to one of its sides. The straight line we have just found has the same property for the parabola, i.e. it is located entirely below the parabola since

$$P(x) - \ell(x) = x^2 - (2a \cdot x - a^2) = (x - a)^2 \geq 0.$$

These two properties can already motivate us to call the graph of the linear function ℓ a *tangent to the parabola*. But there are further properties that support this. As we did

for the circle, we will see now how slowly the tangent to the parabola moves away from the parabola. We consider a point $(x, \ell(x))$, $x \neq a$ (different from the tangency point) on the tangent. By the Pythagoras theorem, its distance from the tangency point is *at least* equal to the distance $|x-a|$ between the x-coordinates. The distance from the parabola is *at most* equal to the distance measured parallel to the y-axis, i.e. equals at most the distance $P(x) - \ell(x) = (x-a)^2$ between the points $(x, \ell(x))$ and $(x, P(x))$. (It looks like the distance parallel to the y-axis is very far from the shortest, but no other distance can be calculated that easily; and even this "unfavorable" distance is quadratically small and therefore"very" small.) Exactly as for the circle, the *ratio* of the distance between the point on the tangent and on the parabola, and the distance to the tangency point is $\leq (x-a)^2/|x-a|$. Hence, the tangent point $(x, \ell(x))$ is much closer to the parabola than it is to the tangency point. The tangent touches the parabola so well that we have a bound for the *ratio* of these distances which is *proportional* to the distance $|x-a|$ from the tangency point (along the x-axis).
We carry further the comparison between circle and parabola since the quantitative comparison between the new and the well-known old stuff is the most important working method of calculus. We will show the following. For any point (a, a^2) on the parabola there is a circle such that the parabola *lies between the circle and its tangent line*, see the following picture. Hence, the tangent line touches the parabola at least as well as this circle.

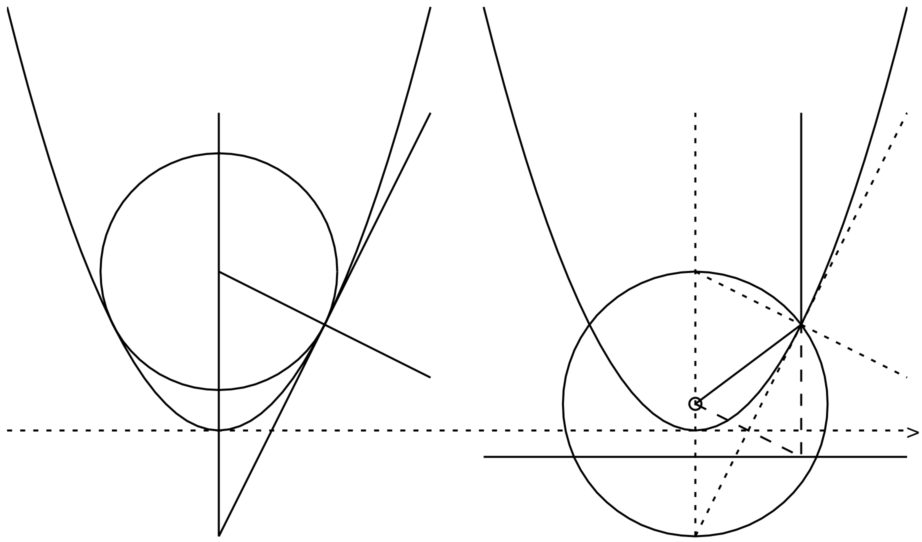
Proof. What circle should we take? The definition of a tangent to the circle says that the center of the circle lies on the straight line through (a, a^2) which is *perpendicular* to the tangent to the parabola. This straight line is called the *normal line* to the parabola and is the graph of the following function

$$n(x) := -\frac{1}{2a} \cdot (x-a) + a^2 = -\frac{1}{2a} \cdot x + \frac{1}{2} + a^2.$$

(Reminder: Perpendicular lines have slopes m and $-1/m$.)
When we imagine circles falling into the parabola from above, they stop when their center is on the y-axis. Therefore, we choose the intersection $(0, n(0))$ of the normal line with the y-axis as the center of the circle, and the distance from (a, a^2) as its radius r. By the Pythagoras theorem $r^2 = 1/4 + a^2$, and we get the promising

Equation of a Circle: $(x-0)^2 + (y - \frac{1}{2} - a^2)^2 = r^2 = \frac{1}{4} + a^2.$

How can we show that the graph of the parabola does *not* meet the inside of this circle? When we evaluate the left hand side at $(x, y) = (x, x^2)$, the value must be $\geq r^2$! It is a good idea to substract r^2, since it's often easier to recognize that an inequality is correct if it is in the form $\ldots \geq 0$. Further, we use the binomial formula



Left hand side: Parabola between circle and tangent.
Right hand side: The parabola as a concave mirror, its focal point and focal line, a Thales' circle.

$(u-v)^2 = u^2 - 2uv + v^2$ with $u = x^2$ and $v = \frac{1}{2} + a^2$. Since squares are always ≥ 0, we get the desired inequality:

$$\begin{aligned} x^2 + (x^2 - \frac{1}{2} - a^2)^2 - r^2 &= x^2 + x^4 - 2x^2(\frac{1}{2} + a^2) + (\frac{1}{2} + a^2)^2 - (\frac{1}{4} + a^2) \\ &= x^4 - 2x^2a^2 + a^4 = (x^2 - a^2)^2 \\ &\geq 0. \end{aligned}$$

Now we have so many analogies between a circle and a parabola that we can **define**:

The straight line $\ell(x) := a^2 + 2a(x-a)$ is called the tangent to the parabola $x \to x^2$ at $x = a$. The slope $2a$ of this straight line is called the slope of the parabola at a.

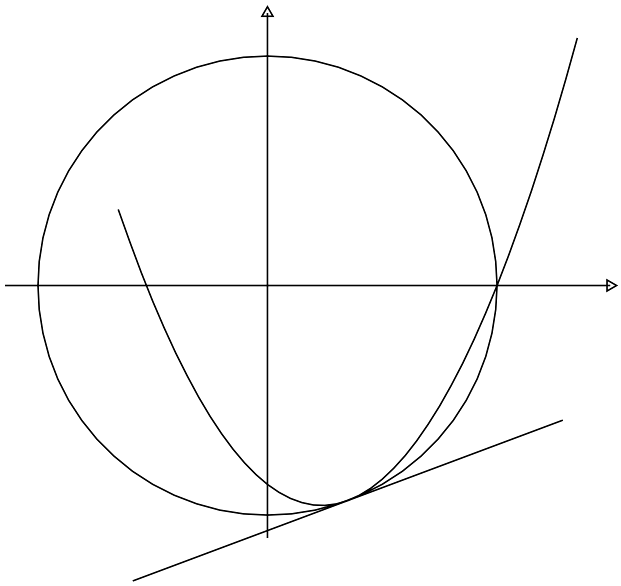
Application:

I find it important that we can use the new definition immediately to explain new types of things, e.g. the perfect behavior of a parabola as a concave mirror. Having defined the normal line to the parabola together with its tangent line, we can generalize the law of reflection: Light beams are reflected from curves in such a way that at the point of incidence the angle which the incident ray makes with the normal is equal to the angle which the reflected ray makes with the same normal. Now, when a beam that enters parallel to the y-axis (picture to the right) reflects off the parabola (x, x^2), the reflected beam pass through the point $(0, \frac{1}{4})$. This is because this point is the center of Thales' circle for the (dotted) right triangle built by the y-axis, the normal and the tangent line. Therefore this point is called the *focus point* of the parabola. Having discovered this point, we can easily see another geometric property of the parabola. Any point (x, x^2) on the parabola is equidistant from both the focus point and the so-called "focus line" $y = -\frac{1}{4}$ since

$$\text{Distance to the focus point} = \sqrt{(x-0)^2 + (x^2 - \frac{1}{4})^2}$$
$$= |x^2 + \frac{1}{4}| = \text{Distance to the focus line.}$$

Now we have *two different defintions* of the parabola, and I want to underline their difference once again. First, it is the graph of the function $x \mapsto x^2$, and second it is the set of all points in the Euclidean plane which are equidistant from a given point (the "focus point") and a given line (the "focus line"). The second property is obviously more similar to the definition of a circle than the first one. We can also adapt the description of tangents to a parabola to both defintions. When we use the geometric definition of a parabola, the tangent at a point P of the parabola is defined to be the *bisecting line* of the line segments from P to the focus and from P to the focus line. When we consider the parabola as the graph of the function $x \mapsto x^2$, the tangent line is the *graph of a linear function*: The tangent at $x = a$ is the graph of the linear function $\ell(x) := a^2 + 2a(x - a)$. In the next section we will start with the definition of tangents to the graph of a function. Those will be again the graphs of linear functions. After that we will get back to the tangents to more general curves.

Excercise. Find the equation of the parabola that touches the unit circle from inside at the point (a, b), with $a^2 + b^2 = 1$, $b \neq 0$ and passes through $(1, 0)$, as in the picture to the right.
In other words, for any point (a, b) of the circle, where the tangent is *not parallel to the y-axis*, there is a parabola such that in the interval $[a - l, a + l]$, with $l := 1 - |a|$ the circle lies between the parabola and its tangent.



Therefore we can say: Circles and parabola are touched equally well by their tangents: First we have seen, that at any point the parabola lies between a circle and its tangent. The excercise shows the inverse: At any point (where the tangent is not vertical) a circle lies between a (vertical) parabola and its tangent. – The geometric definition of a parabola allows of course to consider non-vertical parabolas. Then the circle points with vertical tangents don't play a special role. I preferred vertical parabolas since we wanted to concentrate on the graph of a function first.

Exercises on the triangle inequality

a) Show the so-called triangle inequality $|a+b| \leq |a| + |b|$ for $a, b \in \mathbb{R}$.

b) Under what conditions on a, b do we get equality $|a+b| = |a| + |b|$?

c) Conclude from a) that $|a+b+c| \leq |a| + |b| + |c|$ for $a, b, c \in \mathbb{R}$.

d) Conclude in a similar way that $a_k \in \mathbb{R}, k = 1 \dots n \;\Rightarrow\; |\sum_{k=1}^{n} a_k| \leq \sum_{k=1}^{n} |a_k|$.

The name "traingle inequality" is more suitable for the two-dimensional generalization of this inequality.

Derivatives and Tangents to Polynomials

It is perfectly O.K. to call appropriate straight lines *tangents* of particularly simple curves like circles and parabolas, even though we don't have a general definition yet. Now we want to define the concept of a tangent for a class of curves, namely the graphs of *rather special functions*, the graphs of polynomials; and then we want to prove differentiation rules. We need *definitions* for that. The properties used for the definition will be motivated by what we have observed for tangents to the graph of the quadratic function $x \mapsto x^2$.

It is easy to write down a linear function ℓ with slope m whose value at a is equal to b; it is

$$\ell(x) := m \cdot (x - a) + b.$$

From now on we agree that: If ℓ is the tangent to the polynomial P at a, we define the *slope of the polynomial* P at a to be the slope of this tangent ℓ . And vice versa, if we succeed in defining the slope m of a polynomial P at a, we call the straight line $\ell(x) := m \cdot (x-a) + P(a)$ the tangent of P at a. Also, we can easily express in formulas the difference between the polynomial P and the linear function ℓ by means of the difference of slopes:

$$\begin{aligned} P(x) - \ell(x) = P(x) - P(a) - m \cdot (x - a) &= \left(\frac{P(x) - P(a)}{x - a} - m \right) \cdot (x - a) \\ &= (\text{Slope of the secant line} - \text{Slope of the straight line } \ell) \cdot (x - a). \end{aligned}$$

In other words: We only have to define either the term *slope* (of a function at a), or the term *tangent* (of a function at a); the other term is then defined easily. For illustration we always draw the graph $(x, f(x))$, where x is in the domain of the function f. The tangent (at a) is the graph of a linear function $x \mapsto f(a) + m \cdot (x - a)$.

First, we consider the power functions $P(x) := x^n$ and show that we can define the slope m at a in such a way that the difference between the values of the polynomial P and the linear function $x \mapsto P(a) + m \cdot (x - a)$ is as "small" as we have observed for the quadratic function $x \mapsto x^2$. Then we turn this approximation property into our definition of a tangent and show that all polynomials have computable tangents.

We start with computing the **slope of the secant line:**

$$\frac{P(x) - P(a)}{x - a} := \frac{x^n - a^n}{x - a} = x^{n-1} + x^{n-2}a + \ldots + xa^{n-2} + a^{n-1}.$$

Proof. To prove the last equation, we multiply it by $(x - a)$. On the left hand side we can cancel the denominator $(x - a)$. On the right hand side after multiplying out, the terms x^n und $-a^n$ appear only once, all the other terms twice, one of them with the plus sign and

the other one with the minus sign, so that all of these cancel out.
First we assume $0 < a, x$ to make the situation as clear as possible, since we don't really know yet what to expect. Then the slopes of the secants on the right side of a (i.e. when $a < x$) are *larger* than na^{n-1}, and the slopes of the secants on the left side of a ($x < a$) are *smaller* than na^{n-1}. Therefore we should expect that na^{n-1} is the slope of P at a. The *expected tangent* with this slope is

$$T_a(x) := na^{n-1} \cdot (x-a) + a^n.$$

We will regard this expectation as legitimate if we can prove that this linear function doesn't deviate from P more than what we have observed for $x \mapsto x^2$. We also note that it doesn't bother us at all that *not all* of the points of the tangent to a circle are close to this circle. Therefore we want to require from a tangent only that it stays very close to the curve as long as we are not too far away from the tangency point. For this reason we make the following assumption: The distance between x and a is not bigger than some $r > 0$. i.e. $|x-a| \leq r$. (In the case of polynomials we could always take $r = 1$, but we have already seen in the case of circles that this is not practical, e.g. since the radius of the circle can be much smaller than 1. Further, we will encounter other functions like $x \mapsto 1/x$ where only sufficiently small intervals $[a-r, a+r]$ around the tangent point $(a, f(a))$ make sense for approximations.)
Having agreed on this, we now derive an inequality from the somewhat lengthy formula for the slope of the secant line, by estimating each term and using the triangle inequality:

Assumption: $x \in [a-r, a+r]$ and $R := |a| + r$

Inequality of slopes: $\left|\dfrac{x^n - a^n}{x-a}\right| \leq n \cdot R^{n-1}.$

To work on the difference between x^n and the expected tangent $T_a(x)$, we have to apply this inequality for the slope of secants for all exponents between 1 and $n-1$:

$$\begin{aligned}
&|\text{Value of the function} - \text{Value of the tangent}| = |x-a| \cdot |\text{Secant slope} - \text{Tangent slope}| \\
&= |x^n - T_a(x)| = |x-a| \cdot \left|x^{n-1} + x^{n-2}a + \ldots + xa^{n-2} + a^{n-1} - n \cdot a^{n-1}\right| \\
&\qquad = |x-a| \cdot \left|(x^{n-1} - a^{n-1}) + (x^{n-2}a - a^{n-1}) + \ldots + (xa^{n-2} - a^{n-1})\right| \\
&\qquad \leq |x-a| \cdot \big(|x^{n-1} - a^{n-1}| + |a| \cdot |x^{n-2} - a^{n-2}| + \ldots + |a^{n-2}| \cdot |x-a|\big) \\
&\qquad \leq (x-a)^2 \cdot R^{n-2} \cdot \big((n-1) + (n-2) + \ldots + 1\big) \quad \text{(Inequality of slopes)} \\
&\qquad = \frac{n(n-1)}{2} R^{n-2} \cdot (x-a)^2.
\end{aligned}$$

This inequality is right on target, since first, it says that the graph of $P : x \mapsto x^n$ in the considered interval lies between two parabolas with the same tangent T_a, namely

$$x \in [a-r, a+r] \quad \Rightarrow \quad P_-(x) \leq x^n \leq P_+(x) \qquad \text{with}$$

$$P_-(x) := T_a(x) - \frac{n(n-1)}{2} R^{n-2} \cdot (x-a)^2, \; P_+(x) := T_a(x) + \frac{n(n-1)}{2} R^{n-2} \cdot (x-a)^2,$$

and second, it says that the slopes of the secant lines to P in the interval $[a-r, a+r]$ differ from na^{n-1}, i.e. the slope of $x \mapsto T_a(x)$, by no more than $\frac{n(n-1)}{2} R^{n-2} \cdot |x-a|$ – i.e., they differ the less the closer x is to a. This idea, to compare more complicated functions, namely power functions, with quadratic functions, has worked out and justifies the following definition for functions f that are not necessarily defined on all of $\mathbb{R}$ but only on an interval $[\alpha, \omega] \subset \mathbb{R}$:

Definition of the elementary differentiability:
A function $f : [\alpha, \omega] \to \mathbb{R}$ is called *differentiable* at $a \in [\alpha, \omega]$ with derivative (or slope) $f'(a) = m$, if the following holds:
There is an intervall $[a-r, a+r],\ r > 0$, and a constant K such that

$$x \in [a-r, a+r] \cap [\alpha, \omega] \Rightarrow \big|f(x) - f(a) - m \cdot (x-a)\big| \leq K \cdot |x-a|^2,$$

or equivalently formulated with the slopes of secant lines (i.e. $x \neq a$)

$$x \in [a-r, a+r] \cap [\alpha, \omega] \Rightarrow \left|\frac{f(x) - f(a)}{x-a} - m\right| \leq K \cdot |x-a|.$$

Remarks. (i) *Justifications.* With such an axiomatic defnition, we have to ask some questions. Are the slopes $f'(a)$ *uniquely* determined by this definition? The uniqueness lemma will answer this question with "yes". Are sufficiently many functions differentiable? The following proposition will show this for all polynomials, the next chapter for rational functions, and after having treated completeness we will have substantial possibilities for our construction. Of course, we also have to ask: Is the term "slope" appropriate? The answer "yes" has two parts: The definition implies immediately that increasing functions f have derivatives $f' \geq 0$ (what we will show after having proved that polynomials are differentiable); the inverse, the **monotonicity theorem**, is a main result of calculus: namely, functions with $f' \geq 0$ **are** (weekly) increasing.
(ii) *Other definitions.* With this definition only those functions are differentiable that can be approximated by their tangents as well as we've observed previously for circles and parabolas (or where the slope of the secant lines differ no more from the slope of the tangents than we've observed for $x \mapsto x^2$). Later we will learn to build much more complicated functions. These functions can be defined to be"differentiable" using a slightly weaker, less explicit property. This makes sense, since all the propositions we prove for differentiable functions, remain true when we use the weaker definition. In the beginning we encounter only functions that don't differ from their tangents by more than the quadratic function $x \mapsto x^2$ does.

Uniqueness lemma. If a function f is differentiable at a, then it's slope m is uniquely determined.

Indirect proof. If there were two different slopes $m_1 \neq m_2$, the definition would give us constants $r_i, K_i,\ i = 1, 2$, with

$$|x - a| \le r_i \Rightarrow |f(x) - f(a) - m_i(x - a)| \le K_i|x - a|^2, \quad \text{i=1,2}$$

Using the triangle inequality we get

$$|x - a| \le \min(r_1, r_2) \Rightarrow |(m_1 - m_2)(x - a)| \le (K_1 + K_2) \cdot |x - a|^2,$$

and therefore

$$0 < |x - a| < \min(r_1, r_2) \Rightarrow |m_1 - m_2| \le (K_1 + K_2) \cdot |x - a|.$$

But the last inequality is wrong if we choose x to be so close to a that

$$0 < |x - a| < \min(r_1, r_2,\ \frac{|m_1 - m_2|}{2(K_1 + K_2)}).$$

Proposition.
Polynomials $P : x \mapsto \sum_{k=0}^{n} a_k x^k$ are *differentiable* with $P'(a) = \sum_{k=1}^{n} a_k k a^{k-1}$.

Here, we can choose the length of the interval $[a - r, a + r]$ in the definition independently of the selected polynomial; and the constant K will be calculated explicitly from the coefficients of the polynomial and the bounds of the interval.

Exercise. Specialize every step of the following proof to polynomials of degree 3.

Proof. To give a similar proof as for power functions, one first has to extend the inequality of slopes to polynomials.

Let the interval be (arbitrarily chosen) $[a - r, a + r]$; define as above $R := |a| + r$. Then the inquality of slopes for powers gives us

$$x \in [a - r, a + r] \Rightarrow |a_k x^k - a_k a^k| \le |a_k| k R^{k-1} \cdot |x - a|,\ k = 1 \dots n.$$

We can add up these inequalities to get the one for polynomials, stated below. Reminder: The triangle inequality says $|a + b| \le |a| + |b|$, or for several summands $|\sum_k b_k| \le \sum_k |b_k|$.

Assumption: $x \in [a - r,\ a + r\,],\ R := |a| + r,$

Inequality of slopes: $$\left|\sum_{k=0}^{n} a_k x^k - \sum_{k=0}^{n} a_k a^k\right| \le \left(\sum_{k=1}^{n} |a_k| k R^{k-1}\right) \cdot |x - a|.$$

The proposition about differentiability of polynomials, is derived the same way, by adding up the inequalities that express the deviation of the individual powers from their tangents.

$$|a_k x^k - a_k \cdot (a^k + k a^{k-1}(x - a))| \le |a_k| \frac{k(k-1)}{2} R^{k-2} |x - a|^2.$$

Define a constant $K := \sum_{k=1}^{n} |a_k| \frac{k(k-1)}{2} R^{k-2}$, that combines all the deviations of the individual powers, and add up (for $x \in [a - r, a + r]$) to get the desired

Tangent deviation $$\left|\sum_{k=0}^{n} a_k x^k - \sum_{k=0}^{n} a_k a^k - (\sum_{k=1}^{n} a_k k a^{k-1}) \cdot (x - a)\right| \le K \cdot |x - a|^2.$$

Notation: The polynomial $P'(x) := \sum_{k=1}^{n} a_k k x^{k-1}$ from which the slopes can be calculated is called the *derivative* of the polynomial $P(x) := \sum_{k=0}^{n} a_k x^k$. The value $P'(a)$ is called slope of P at a. With this, we can rewrite the tangent deviation for polynonmials more clearly:

Differentiation of Polynomials

Assumption: $x \in [a-r,\, a+r\,],\ P(x) := \sum_{k=0}^{n} a_k x^k,\ P'(x) := \sum_{k=1}^{n} a_k k x^{k-1},$

Constants: $R := |a| + r,\ K := \sum_{k=1}^{n} |a_k| \frac{k(k-1)}{2} R^{k-2}.$

Tangent at a : $\ell_a(x) := P(a) + P'(a) \cdot (x-a).$

Tangent deviation: $|P(x) - P(a) - P'(a)(x-a)| \le K|x-a|^2,$

or for slopes ($x \ne a$): $\left| \frac{P(x) - P(a)}{x-a} - P'(a) \right| \quad \le \quad K \cdot |x-a|.$

The following proposition and its proof are part of the imminent understanding of the definition of differentiability. It also shows: if f has a local extreme value at c then $f'(c) = 0$.

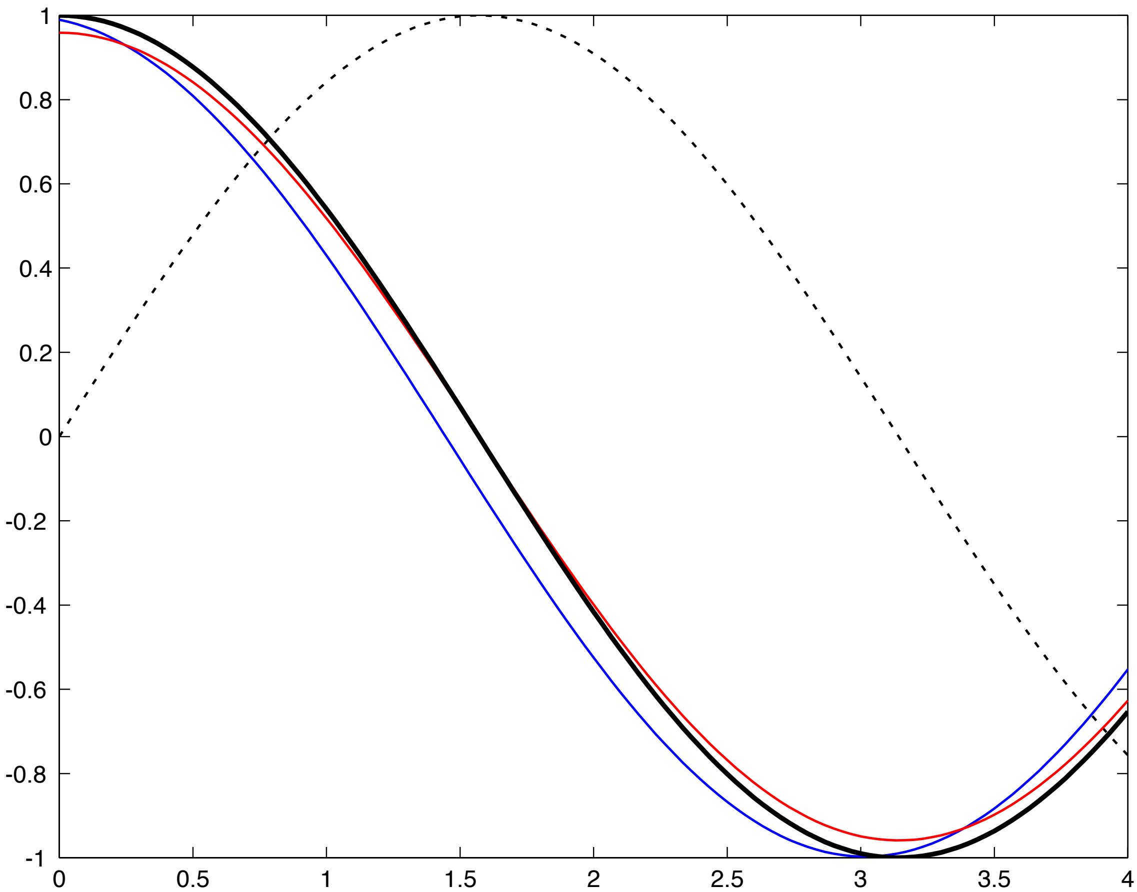
Proposition. Growth and non-negative slope.
Increasing functions f (i.e. $x \le y \Rightarrow f(x) \le f(y)$) have a non-negative slope $f'(a) \ge 0$ at any point a. – The inequalities $\le$ cannot be replaced by $<$ as can be seen from the strictly increasing function $x \mapsto f(x) := x^3$ where $f'(0) = 0$.
Indirect proof. Suppose the derivative is negative at a, $f'(a) < 0$. The definition gives us constants $r, K > 0$ with

$$|x-a| \le r \ \Rightarrow\ |f(x) - f(a) - f'(a)(x-a)| \le K \cdot |x-a|^2.$$

We make r smaller, namely to $\min(r, |f'(a)|/2K) > 0$, and conclude that

$$\begin{aligned} a-r \le x < a \ &\Rightarrow\ f(x) \ge f(a) + f'(a)(x-a) - K \cdot |x-a|^2 \ge f(a) + f'(a)(x-a)/2 > f(a), \\ a < x \le a+r \ &\Rightarrow\ f(x) \le f(a) + f'(a)(x-a) + K \cdot |x-a|^2 \le f(a) + f'(a)(x-a)/2 < f(a). \end{aligned}$$

In other words, for x close to a and $x < a$ we have $f(x) > f(a)$, while at the right side of a the values of the function are $f(x) < f(a)$ – as we expect it from negative derivatives, but which is a *contradiction* to the assumption that f is increasing.



Comparison of derivative and difference quotients.

The derivative f' of the dotted function f is drawn with a bold line. It is approximated by two difference quotients (with the relatively big step size $h = 0.5$),

by a one-sided one (blue): $(f(x+h) - f(x))/h$

and by a symmetric one (red): $(f(x+h) - f(x-h))/2h$.

The symmetric difference quotients are too small when we are close to the inflection points of f but otherwise a lot more precise than the one-sided ones. The smaller the step size, the more superior are the symmetric quotients; for quadratic functions, they even give the precise derivative, independent of the step size, because $((a+h)^2 - (a-h)^2)/2h = 2a$.

Rational Functions, Differentiation Rules.

We can apply the definition of differentiability to the function $x \mapsto f(x) := 1/x$ as well. It is differentiable in its domain $\mathbb{R} \setminus \{0\}$ with the derivative $f'(a) = -1/a^2$. All rational functions (i.e. fractions of polynomials) can be generated from this function and the polynomials. Here the differentiation rules become important. They say: If two functions f, g are differentiable, then $f+g, f\cdot g, f\circ g$ are differentiable as well and the derivatives of these composed functions can be calculated *explicitly* from the values and derivatives of f, g. Though the concept of a function changed a lot from Newton to Hilbert and consequently the definition of differentiability, the differentiation rules stayed **the same**. We will explain the reason for that.

The following example is important for two reasons. First, together with the chain rule, it extends very much the set of functions we can differentiate. Second, the interval $[a-r, a+r]$ where we compare the function and its tangent, cannot be chosen completely independent of a any longer, since of course 0 must **not** lie in this interval.

Example $f(x) = 1/x$. We assume $0 < a, x$. Then the secant slopes are:

$$\frac{1/x - 1/a}{x-a} = \frac{-1}{x \cdot a}$$

Obviously (as for the other expamples with power functions), the number $-1/a^2$ lies in between the secant slopes on the right hand side ($a < x$) and the secant slopes on the left hand side ($x < a$), so that only the value $-1/a^2$ can be expected as the derivative. Then $T_a(x) := 1/a - (x-a)/a^2$ is the expected tangent. We calculate the difference between $f(x)$ and the expected tangent:

$$f(x) - T_a(x) = x^{-1} - (a^{-1} - a^{-2}\cdot(x-a)) = (\frac{1}{a^2} - \frac{1}{x\cdot a})\cdot(x-a) = \frac{(x-a)^2}{x\cdot a^2}.$$

Choosing $r := |a|/2$ we avoid that 0 lies in the interval $[a-r, a+r]$. With $f'(a) = -1/a^2$ and the constant $K := 2/a^3$ this implies that *the inequality in the definition of differentiability holds*:

$$0 < \frac{a}{2} \le x \quad \Rightarrow \quad 0 \le f(x) - f(a) - f'(a)\cdot(x-a) \le +\frac{2}{a^3}(x-a)^2.$$

Here I assumed $0 < a$ additionally, since then the calculation shows that the graph of this function lies *above* any tangent and *below* some quadratic function with the same tangent. (Vice versa for $a < 0$.)

Excercise. Deduce for $f(x) = x^{-n}$ and $n > 1$ that $f'(x) = -nx^{-n-1}$ is expected and:

$$0 < \frac{a}{2} \le x \quad \Rightarrow \quad 0 \le f(x) - f(a) - f'(a)\cdot(x-a) \le \frac{n(n+1)}{(a/2)^{n+2}}\cdot(x-a)^2.$$

Differentiation rules. Obviously one wants to determine the derivative directly from the definition only for introductory examples where the definition is to be explained. Later one looks for less time-consuming methods, like rules that allow for calculating the derivative of $f \cdot g$ and $f \circ g$ from the derivatives of the two functions f and g. In general, for every new way to construct functions from simpler ones, one looks for an appropriate differentiation rule. First, we answer the following question: How must the rules that correspond to the sum, the product and the composition of functions look like? Then we prove these rules. Next, we observe that we can describe curves in the plain by *pairs of functions*; how can we differentiate those? Later we will construct new functions as the *inverse functions*. What differentiation rule can we use to calculate their derivatives? In later chapters it will be particularly important to construct new functions as limits of sequences of approximations by some simpler functions; are there differentiation rules for that as well?

Linear combinations.

We have already treated polynomials as linear combinations of power functions and can see that for two polynomials P and Q, their derivatives P' and Q', and $\alpha, \beta \in \mathbb{R}$ we have

$$(\alpha \cdot P + \beta \cdot Q)' = \alpha \cdot P' + \beta \cdot Q'.$$

This rule shouldn't be different for more complicated functions, thus

$$(\alpha \cdot f + \beta \cdot g)' = \alpha \cdot f' + \beta \cdot g'.$$

Excercise. The proof is a direct application of the triangle inequality.

Product rule

If differentiation rules exist at all, they have to be valid for the simplest functions as well. To see how a product rule could look like, we multiply linear functions first:

$$\ell_1(x) = w_1 + m_1(x-a), \quad \ell_2(x) = w_2 + m_2 \cdot (x-a)$$
$$(\ell_1\ell_2)(x) = w_1w_2 + (w_1m_2 + w_2 \cdot m_1) \cdot (x-a) + m_1m_2 \cdot (x-a)^2.$$

The quadratic function $\ell_1\ell_2$ has at a the factor of $(x-a)$ as its derivative: $w_1m_2 + w_2m_1 = (\ell_1 \cdot \ell_2' + \ell_2 \cdot \ell_1')(a)$. Thus, if there is a product rule, it should look as follows:

$$(f_1 \cdot f_2)'(a) = f_1'(a) \cdot f_2(a) + f_1(a) \cdot f_2'(a).$$

Chain rule, or Composition rule

Again we try to find out how the rule could look like. Let ℓ be a linear function

$\ell(x) = A + m \cdot (x-a),\ m \neq 0$

and f a function which is differentiable at A. Then there are numbers R, K such that

$Y \in [A-R, A+R] \ \Rightarrow\ |f(Y) - f(A) - f'(A) \cdot (Y-A)| \leq K \cdot |Y-A|^2.$

Plugging in $Y = \ell(x)$ etc. into the inequaltiy for f, we find:

$$|x-a| \le R/m \;\Rightarrow\; |f(\ell(x)) - f(\ell(a)) - f'(\ell(a)) \cdot m \cdot (x-a)| \le K \cdot m^2 \cdot |x-a|^2.$$

Therefore, the composition $f \circ \ell$ has the derivative $(f \circ \ell)'(a) = f'(\ell(a)) \cdot \ell'(a)$.
We expect this differentiation rule also to be true for non-linear ℓ.

Commentary. In the 19th century it became accepted to call a function differentiable when it can be approximated fairly well by linear functions (still called tangents), though not quite as well as we have observed for polynomials. The essential thing is, that the difference between $f(x)$ and the tangent $T_a(x)$ near a should decrease faster than proportionally to the distance $|x-a|$. As examples for such rougher but still suffiently good error bounds we can take const $\cdot\, |x-a|^{1+p}$ with $0 < p \le 1$ (only the linear functions will be "differentiable" if we take $p > 1$). These error bounds have the advantage of being explicit, but the deficiency is that it's not possible to prove such explicit approximations in all the interesting cases. The ultimate definition (that is 200 years younger than Newton!) works with even rougher errors. I will use this definition later in this text. The error bounds are expressed in the following way: the errors are $\le \epsilon \cdot |x-a|$; however, ϵ is not a given constant here, this would not be sufficient; we may choose an arbitrarily small $\epsilon > 0$ as a factor, but we have to pay for this by the fact that the approximation is valid only in a smaller interval for a smaller ϵ. We will work with the final definition later, but we mention it already now:

The ultimate definition of differentiability. For every error factor $\epsilon > 0$ there is a (possibly very small) $\delta(\epsilon) = \delta > 0$, such that the following inequalities hold at least in the interval $[a-\delta, a+\delta]$:

$$|x-a| \le \delta \Rightarrow |f(x) - f(a) - f'(a) \cdot (x-a)| \le \epsilon \cdot |x-a|.$$

There are still two alternatives for this formulation: For every ϵ we have to find the δ *independent* of $a \in [R, S]$, or weaker: we have to find the δ allowing dependence on a.

In spite of all these possible variations for formulating the definitions and the assumptions of theorems, the differentiation **rules** are always the same. Namely, if two functions f and g can be approximated by their tangents "well in some way", then $f \cdot g$ and $f \circ g$ have the derivatives, which our initial considerations make us expect, and they are approximated by their tangents "well in some way" in the **same sense**. We can express this "well in some way" in formulas:

$$f(x) - f(a) - f'(a) \cdot (x-a) = \phi(x,a) \cdot |x-a|,$$

and *depending on the assumptions required right now* we have:

$$x \in [a-r, a+r] \;\Rightarrow\; |\phi(x,a)| \le K \cdot |x-a|^p,\; (0 < p \le 1)$$

or, for each $\epsilon > 0$ there is a $\delta > 0$ such that

$$x \in [a-\delta, a+\delta] \;\Rightarrow\; |\phi(x,a)| \le \epsilon,$$

and where the constants r, K or δ either depend on a, or better, where one choice of constants holds "uniformly" for all a in the subset of the domain of f which is of interest right now. All these variations occur, but the proofs of the differentiation rules must be adapted only to the exact requirements on ϕ, otherwise they stay the same. During most of the following proof of the product rule we do not need to know how "small" is specified, only at the very end the definition of ϕ_i enters. – Most authors start their proofs by choosing r or δ so small that $|\phi(x,a)| \le 1$.

Proof of the product rule

Assumption: $f_i(x) = f_i(a) + f_i'(a)\cdot(x-a) + \phi_i(x,a)\cdot|x-a|,\ i = 1,2$ holds for functions ϕ_i, which desribe the"small" errors which we have previously discussed. Then

$$f_1(x)\cdot f_2(x) =$$
$$\Big(f_1(a) + f_1'(a)\cdot(x-a) + \phi_1(x,a)\cdot|x-a|\Big)\cdot\Big(f_2(a) + f_2'(a)\cdot(x-a) + \phi_2(x,a)\cdot|x-a|\Big)$$

gives us, after we expand and regroup, the difference between the product and the expected tangent:

$$\begin{aligned}(f_1\cdot f_2)(x) \ -\ &\big((f_1\cdot f_2)(a) + (f_1'\cdot f_2 + f_1\cdot f_2')(a)\cdot(x-a)\big) = \\ \Big[(f_1(a) + f_1'(a)\cdot(x-a))\cdot\phi_2(x,a) &+ (f_2(a) + f_2'(a)\cdot(x-a))\cdot\phi_1(x,a) \\ &+ (f_1'(a)\cdot f_2'(a) + \phi_1(x,a)\cdot\phi_2(x,a))\cdot|x-a|\Big]\cdot|x-a|.\end{aligned}$$

First, we simplify only under the assumption that a, x belong to a bounded (but maybe very big) interval, let's say $a, x \in [-R, R]$. Then the expression in the square brackets (using $|\phi_1\cdot\phi_2| \le 1$) becomes

$$[\ldots] \le const_1\cdot\phi_2 + const_2\cdot\phi_1 + (|(f_1'\cdot f_2')(a)| + 1)\cdot|x-a|,$$

where the constants $const_1, const_2$ can obviously be calculated explicitly from $f_i(a), f_i'(a)$ and the length $2R$ of the interval. Such a combination of ϕ_1 and ϕ_2 is now "small" in the same sense as required for ϕ_1 and ϕ_2, since the third term $(|(f_1'\cdot f_2')(a)|+1)\cdot|x-a|$ complies with the sharpest condition for being "small". For example:
The condition

$|x-a| \le r_1 \Rightarrow |\phi_1(x,a)| \le M_1\cdot|x-a|^p$, and $|x-a| \le r_2 \Rightarrow |\phi_2(x,a)| \le M_2\cdot|x-a|^p$

implies

$$\begin{aligned}&|x-a| \le \min(r_1, r_2, 1) \Rightarrow \\ &\big|(const_1\cdot\phi_2 + const_2\cdot\phi_1)(x,a) + (|(f_1'\cdot f_2')(a)| + 1)\cdot|x-a|\big| \le \\ &\qquad\qquad \le (const_1\cdot M_2 + const_2\cdot M_1 + const_3)\cdot|x-a|^p \\ &\qquad\qquad = M_3\cdot|x-a|^p.\end{aligned}$$

The difference between the product $f_1 \cdot f_2$ and the expected tangent is indeed a small error in the same sense as we have assumed for f_1 and f_2. In our simple example this meant $\leq M_3|x-a|^{p+1}$ in the interval $|x-a| \leq \min(r_1, r_2, 1)$. All the later proofs of the other product rules follow this pattern, in particular, the product written by $\cdot$ here, can be replaced by any bilinear product, such as the scalar product, the cross product, the matrix product, etc.

Proof of the Chain rule

First of all, we have to make sure that the values of the inner function fall into an interval where the assumptions on the argument of the outer function hold. In our calculation of $f \circ \ell$ in the beginning, this was guaranteed by the implication $|x-a| \leq R/m \Rightarrow |\ell(x) - \ell(a)| = |m(x-a)| \leq R$. For a non-linear inner function this requires more effort. Unlike for the product rule, I will not formulate the proof for all kinds of "small" errors at once, but restrict myself to the quadratic errors observed for polynomials. I postpone the discussion of ϵ-δ-errors since we have to come back to the proof of the chain rule in higher-dimensional situations anyway.

Assumptions:

$|x-a| \leq r_1 \Rightarrow |f(x) - f(a) - f'(a) \cdot (x-a)| \leq M_1(x-a)^2, \quad f(a) = A$ and

$|Y-A| \leq R_2 \Rightarrow |F(Y) - F(A) - F'(A) \cdot (Y-A)| \leq M_2 \cdot (Y-A)^2.$

Now, if we can guarantee $|f(x) - f(a)| \leq R_2$ (!!), we get our first important step,

$$(*) \qquad |F(f(x) - F(f(a)) - F'(f(a)) \cdot (f(x) - f(a))| \leq M_2 \cdot (f(x) - A)^2.$$

Applying the triangle inequality to the assumption on f, we get a Lipschitz bound for f:

$$(**) \qquad |x-a| \leq r_1 \Rightarrow |f(x) - f(a)| \leq (|f'(a)| + M_1 \cdot r_1) \cdot |x-a| =: L \cdot |x-a|.$$

Therefore

$$|x-a| \leq r := \min\big(r_1,\ R_2/L\big)$$

indeed implies $|f(x) - A| \leq R_2$ whenever $|x-a| \leq r$. This is an important step in the proof that makes (*) available.

Next, multiply the assumption on f by $|F'(f(a))|$) to get

$$|x-a| \leq r_1 \Rightarrow |F'(f(a)) \cdot (f(x) - f(a)) - F'(f(a)) \cdot f'(a) \cdot (x-a)| \leq |F'(f(a))| \cdot M_1 \cdot |x-a|^2.$$

The term $F'(f(a)) \cdot (f(x) - f(a))$ appears in this last inequaltiy and in inequaltiy (*). Hence, both these inequalties in combination with the triangle inequality can be used to eliminate this term, and we get the implication that finishes the proof:

$$|x-a| \leq r \Rightarrow |F(f(x)) - F(f(a)) - F'(f(a)) \cdot f'(a) \cdot (x-a)| \leq \text{Const} \cdot (x-a)^2,$$

where the constants have been combined to Const $:= M_2(|f'(a)| + M_1 \cdot r_1)^2 + |F'(f(a))| \cdot M_1$.

We repeat that it is an essential part of the proof to guarantee with assumptions on x that $Y := f(x)$ satisfies the assumptions made for F. We handled this part by deriving the Lipschitz constant L for f near a. Proofs of the chain rule that ignore this step, are incorrect.
Our proof with the ϵ-δ-errors will follow the same pattern, but neither will there be *explicit* intervals in the assumptions, nor will the proof generate any explicit intervals where the asserted inequalities hold.

Applications

Reciprocal Rule: $$(\frac{1}{f})' = -\frac{f'}{f^2}.$$

Proof. Use $F(x) = 1/x,\ F'(x) = -1/x^2$ in the chain rule we have just proven.
This rule also gives a

Quotient Rule: $$(\frac{f}{g})' = (f\frac{1}{g})' = \frac{f'}{g} - \frac{f \cdot g'}{g^2} = \frac{f'g - fg'}{g^2} = \frac{f}{g} \cdot (\frac{f'}{f} - \frac{g'}{g}).$$

The latter form of the quotient rule is useful, for example, in the discussion of percentage (or relative) errors. In natural sciences the squares of errors are often negligible, then the derivative tells us
what impact errors in the arguments will have on the values of a function:

Absolute error: $$\triangle f := f(x) - f(a) \approx f'(a) \cdot (x - a)$$

Often one is more interested in the so-called *relative error*, where one has to divide by $f(a) > 0$.

Relative error: $$\frac{\triangle f}{f} := \frac{f(x) - f(a)}{f(a)} \approx \frac{f'}{f}(a) \cdot (x - a).$$

Notation: For positive functions $f > 0$ we call f'/f its *growth rate*, e.g. the birth rate is such a quotient. These growth rates f'/f appear in the natural sciences certainly as often as the slopes f'.

Curves in the plane. With pairs of functions $f, g : [a, b] \to \mathbb{R}$ we can describe curves in the plane. E.g. $f(t) := 2t/(1 + t^2),\ g(t) = (1 - t^2)/(1 + t^2)$ gives us a map from $\mathbb{R}$ to the unit circle (except $(0, -1)$):

$$(f, g) : \mathbb{R} \to \mathbb{R}^2,\quad t \to \big(\frac{2t}{(1 + t^2)}, \frac{(1 - t^2)}{(1 + t^2)}\big), \qquad f(t)^2 + g(t)^2 = 1.$$

Remarkably, this map sends the rational points in $\mathbb{R}$ into the points on the circle with rational coordinates and vice versa, for rational p and q such that $p^2 + q^2 = 1$ we have the rational $t := (1+q)/p$. On the other hand, we will see that the circle is not traced out at a constant speed. Among pairs of rational functions there are none that can do that, trace a circle at constant speed. Only sin and cos, which we will construct at the end of the chapter on the completeness of the real numbers, can.
In order to come to the definition of derivatives and tangents, we consider the simplest case first. Let f and g be linear functions:

$$f(t) := p_1 + m_1 \cdot t, \qquad g(t) := p_2 + m_2 \cdot t.$$

Then we may write

$$p(t) := (f,g)(t) = (p_1,p_2) + (m_1,m_2) \cdot t, \qquad \frac{p(t_2) - p(t_1)}{t_2 - t_1} = (m1, m2).$$

Thus, all the average velocities of this movement are constant, $\vec{v} = (m_1, m_2)$. Of course we will agree that we define the instantaneous velocity in this clear case to be the vector $\vec{v}$.

Now we want to compare more complicated curves with these simplest linear curves.

Strategy: A curve $(f_1, f_2) : [a,b] \to \mathbb{R}^2$ has the "derivative" or "velocity" $\vec{v} = (f_1'(t_0), f_2'(t_0))$ at t_0, if $(f_1(t), f_2(t))$ differs, near $t = t_0$, sufficiently little from the following linear curve, the tangent movement:

$$t \to (f_1(t_0), f_2(t_0)) + (f_1'(t_0), f_2'(t_0)) \cdot (t - t_0).$$

We have to make more precise what we mean by "differs sufficiently little". For this purpose we generalize the "small errors", that we assume for f_1, f_2, using the Pythagoras theorem. Suppose

$$f_i(t) - f_i(t_0) - f_i'(0) \cdot (t - t_0) = \phi_i(t, t_0) \cdot |t - t_0|, \quad (i = 1, 2),$$

where $\phi_i(t, t_0)$ describe the currently used "small" errors, i.e., depending on the situation either explicit errors like $|\phi_i(t,t_0)| \le M_i \cdot |t - t_0|$ in intervals $[t_0 - r_i, t_0 + r_i]$ or ϵ-δ-errors $|\phi_i(t,t_0)| \le \epsilon$ in intervals $[t_0 - \delta, t_0 + \delta]$. Here M_i and $r := \min(r_1, r_2)$ are constants, which may depend on t_0 or be independent of t_0 in more favorable cases (we say in this case, that the constants can be choosen "uniformly"), while for ϵ-δ-errors we need a $\delta > 0$ for *every* $\epsilon > 0$, where again $\delta = \delta(\epsilon, t_0)$, or better $\delta = \delta(\epsilon)$ (uniformly) for all t_0 .
Indepenently of a precise meaning of "small" errors, the Pythagoras theorem will give us $t \in [t_0 - r, t_0 + r] \Rightarrow$

$$\left|(f_1, f_2)(t) - (f_1, f_2)(t_0) - (f_1', f_2')(t_0) \cdot (t - t_0)\right| \le \sqrt{\phi_1(t,t_0)^2 + \phi_2(t,t_0)^2} \cdot |t - t_0|.$$

And again we can see easily for any kind of small errors ϕ_1, ϕ_2, that $\sqrt{\phi_1^2+\phi_2^2}$ is a small error *of the same kind*. For example, if for some p with $0 < p \le 1$

$$|t-t_0| \le r \Rightarrow |\phi_i(t,t_0)| \le M_i \cdot |t-t_0|^p \quad (i=1,2),$$

then

$$|t-t_0| \le r \Rightarrow \sqrt{\phi_1(t,t_0)^2+\phi_2(t,t_0)^2} \le \sqrt{M_1^2+M_2^2} \cdot |t-t_0|^p =: M \cdot |t-t_0|^p.$$

In other words: *We can easily generalize the differentiation of functions* $f:[a,b]\to\mathbb{R}$ *to the differentiation of curves* $(f,g):[a,b]\to\mathbb{R}^2$*.*

The corresponding simple differentiation rule is called "componentwise differentiation."

The derivative of a curve is: $\quad (f,g)'(t) = \big(f'(t),\ g'(t)\big).$

So we are able to differentiate curves as easily as functions, and our initial considerations, when we were looking for tangents to curves other than circles without having any definitions, fit neatly into our theory now.

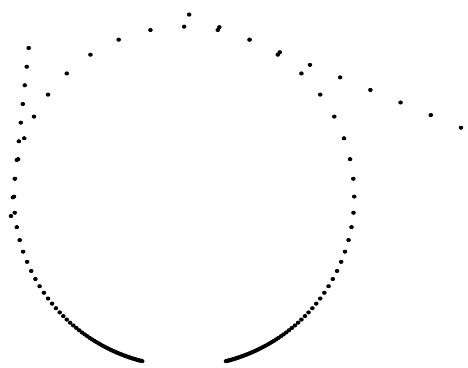
Example. For the circle we've just looked at,

$$k(t) := (\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}) \;=\; (\frac{2t}{1+t^2}, \frac{2}{1+t^2}-1)$$

we obtain

$$k'(t) = \frac{2}{(1+t^2)^2}\cdot(1-t^2,\ -2t), \qquad |k'(t)| = \frac{2}{1+t^2}.$$

Obviously the velocity (and therefore the tangent) is always perpendicular to the radius, but the absolute value of the velocity of this circular movement is not constant.



Image, not graph, of this parametrized circle, with two parametrized tangents.

Commentary. Now we can differentiate a big class of functions and curves, and do it quickly by using the differentiation rules. Yet, we don't know any theorems that would enable us to draw conclusions about a function from assumptions about its derivative. We will turn to this central subject of analysis in the next chapter.

The Monotonicity Theorem

> The monotonicity theorem and related results allow us to derive properties of functions from properties of their derivatives. First, we have to understand why such conclusions are central to calculus. Second, to prove the monotonicity theorem we need more subtle arguments when the assumptions on differentiability are weaker. Usually, the completeness of the real numbers is used in an indirect proof. However, the uniform estimates that are true for all the functions we have discussed so far, allow us to prove it directly without using completeness. Therefore these proofs work already if we know the rational numbers only. They allow us to develop an essential part of quantitative analysis before any use of completeness.

Of course, the rules for calculating the derivatives we have discussed, don't explain how *useful* the calculation of the derivatives might be. With the monotonicity theorem we turn to this question. While so far we have been able to obtain the treated properties of derivatives of polynomial functions by simple calculation, this won't be the case in most situations. General statements can rarely be verified by direct calculation, we have to argue from the definitions. And these arguments give better results than the earlier calculations, even for polynomials.
First, repeat the definition of (elementary) differentiability at one point, without mentioning uniformity.

Definition. A function $f\colon(\alpha,\omega)\to\mathbb{R}$ is called **differentiable** at $c\in(\alpha,\omega)$ with slope (or derivative) $m=f'(c)$ and tangent $T(x)=f(c)+m\cdot(x-c)$, if the following holds:
There is an interval $[c-r,c+r]\subset(\alpha,\omega),\ r>0$, and a constant K such that:

$$\text{(dif)}\qquad x\in[c-r,c+r]\Rightarrow|f(x)-f(c)-m\cdot(x-c)|\le K\cdot(x-c)^2.$$

We call $K\cdot(x-c)^2$ the quadratic deviation from the tangent. It is a smaller deviation than allowed by the final definition in chapter 10. Now for **"uniformity"**. Without better knowledge one would have to expect that the length $2r$ of the interval and the constant K depend on the point c. However, all functions we have met so far (and will meet until chapter 9) are better behaved: A function f is called **uniformly differentiable** on $[a,b]\subset(\alpha,\omega)$ if constants r,K exist that work in (dif) for **all** $c\in[a,b]$.

From now on, the following argument that goes back to Archimedes becomes important.

> **Archimedes Strategy**: In order to prove an inequality $a\le b$ it is enough to prove the weaker inequalities $a\le b+p$ **for all** $p>0$. Surprisingly, this is often much easier.

Proof: If $a\le b$ were wrong, we would have $a>b$, and we could take $p:=\frac12(a-b)>0$. By assumption we would have the inequality $a\le b+p$, which would give us $a\le b+\frac12(a-b)$, and $\frac12 a\le\frac12 b$, contradicting our assumption $a>b$.
I don't know any other proof which deserves the title *The Prototype of an Indirect Proof* more than this one. Archimedes strategy is used extremely often in analysis.

Archimedes himself worked with more particular assumptions, he used only unit fractions $p = 1/n$. We will explain this in more details in the next chapter after we formulate the Archimedes axiom for the real numbers.

After these preliminary remarks we now get to the core of the matter. The most basic example of a theorem that deduces properties of f from assumptions on its derivative f' is the

Monotonicity Theorem. Let f be differentiable with the derivative $f' \geq 0$, then f is weakly increasing:

$$x \leq y \Rightarrow f(x) \leq f(y).$$

Corollaries 1 - 6. The statement of the theorem is of course already suggested by the name “slope”. Nevertheless, the proof is not obvious. The theorem is very useful even for the treatment of polynomials.
As an advertisement for the Monotonicity Theorem we are first showing some of its immediate consequences.

1.) Generalzied Monotonicity Theorem
Functions with bigger derivative grow faster:
$$x \leq y \text{ and } f' \leq g' \text{ imply } f(y) - f(x) \leq g(y) - g(x),$$
since $(g - f)$ meets the assumptions of the monotonicity theorem.

2.) The (D⇒L)–Theorem : Derivative bounds imply expansion-, or Lipschitz, bounds.
Or: a bound for f' is also a bound for all the slopes of secant lines to f:
$$|f'| \leq L \Rightarrow |f(y) - f(x)| \leq L \cdot |y - x| :$$
Define the linear function $g(x) := L \cdot x$. Then $-g' \leq f' \leq g'$ and the first corollary implies
$$-L \cdot (y - x) \leq f(y) - f(x) \leq +L \cdot (y - x) \text{ for } x \leq y.$$

Functions with this property play such a big role that they have their own name:
Definition. Functions with $|f(y) - f(x)| \leq L \cdot |y - x|$ are called **Lipschitz continuous** (or **of bounded expansion**) with (global) Lipschitz bound (or expansion bound) L.

3.) Non-negative second derivatives imply convexity.
$f'' \geq 0$ implies that the graph of f lies above any of its tangents.
Proof. Let $T_c(X) := f(c) + f'(c) \cdot (x - c)$ be the tangent at c. For the auxiliary function $h := f - T_c$ we have first: $h'' \geq 0$. Further $h'(c) = 0$, which gives: $x \leq c \Rightarrow h'(x) \leq 0$ and: $c \leq x \Rightarrow h'(x) \geq 0$. Therefore, by the monotonicity theorem, h is increasing on the right hand side of c, and decreasing on the left hand side of c. Since $h(c) = 0$, this implies $h \geq 0$, or $f \geq T_c$.

Application. Bernoulli's inequality: $-1 \leq x, n \in \mathbb{N} \;\Rightarrow\; (1 + x)^n \geq (1 + n \cdot x)$ (tangent at $c = 0$) is an immediate consequence, since $((1 + x)^n)'' \geq 0$ if $-1 \leq x$.
Using 3) is a very convenient way to derive a lot of other estimates, even for polynomials.

4.) Second derivative and deviation from the tangent.
$|f''| \leq K$ implies that f differs from its tangent at c by no more than $\frac{1}{2}K \cdot (x-c)^2$. Indeed, for the auxiliary function $h(x){:=}\ f(x) - T_c(x)$ we have $h(c) = 0 = h'(c)$ and $-K \leq h'' \leq K$.
We apply corollary 1 twice, first

$$x \geq c \Rightarrow -K \cdot (x-c) \leq h'(x) - h'(c) = h'(x) \leq +K \cdot (x-c),$$
$$x \leq c \Rightarrow -K \cdot (c-x) \leq h'(x) - h'(c) = h'(x) \leq +K \cdot (c-x)$$

and second:

$$-\frac{1}{2}K \cdot (x-c)^2 \leq h(x) - h(c) = f(x) - T_c(x) \leq \frac{1}{2}K \cdot (x-c)^2.$$

This corollary explains the meaning of the constant in the quadratic deviation from the tangent: every bound for the second derivative is a suitable constant. This result is often much better than the "tangent deviation" we calculated from the absolute values of the coefficients in chapter 3.

5.) Second derivative and the deviation from a secant line, convexity again.
$0 \leq f''$ implies that in any intervall $[a, b]$ the graph of f lies below the secant line S_{ab}.
The secant line is the linear function $S_{ab}(x) := (f(a) \cdot (b-x) + f(b) \cdot (x-a))/(b-a)$.
For the auxiliary function $h(x) := f(x) - S_{ab}(x)$ we have $h'' \geq 0,\ h(a) = 0 = h(b)$.
First, h' is increasing since $h'' \geq 0$. Next, let $c \in (a, b)$ be arbitrary.
EITHER: $h'(c) \geq 0$, then $h' \geq 0$ in the interval $[c, b]$ (h' is increasing!) and therefore $h(c) \leq h(b) = 0$.
OR: $h'(c) < 0$, then $h' \leq h'(c) < 0$ in the interval $[a, c]$ (h' is increasing!) and therefore $0 = h(a) > h(c)$. Hence in both cases, we have $h(c) \leq 0$, and therefore the graph of f is below the secant line.

If $f'' \leq B$, then as above, we have $0 \leq h''$ for the auxiliary function
$$h(x) := -B(x-a)(b-x)/2 + S_{ab}(x) - f(x).$$
We proved $h \leq 0$, hence $S_{ab}(x) - B(x-a)(b-x)/2 \leq f(x)$. Similarly, if $-B \leq f''$.

6.) The multiplicative version of the monotonicity theorem for positive functions.
In connection with the relative errors we have to know the growth rates of functions.
Due to the quotient rule we have:

$$\frac{f'}{f} \leq \frac{g'}{g} \quad \Rightarrow \quad (\frac{g}{f})' = \frac{g}{f} \cdot (\frac{g'}{g} - \frac{f'}{f}) \geq 0.$$

By the monotonicity theorem g/f is increasing and g grows faster than f:

$$a \leq x \ \text{ and } \ \frac{f'}{f} \leq \frac{g'}{g} \Rightarrow \frac{g(a)}{f(a)} \leq \frac{g(x)}{f(x)} \Rightarrow \frac{f(x)}{f(a)} \leq \frac{g(x)}{g(a)}.$$

We illustrate this multiplicative version by a specific application, where I want to emphasize how easy the computations are.
Consider the polynomials $f_m(x) := (1 + x/m)^m$ for $0 \le x$ and $m \in \mathbb{N}$. For small x, their growth rates differ only little from 1, therefore these polynomials will turn out to be a good approximations of the exponential function. If $m < n$, we have

$$\frac{f_m'}{f_m}(x) = \frac{1}{1 + x/m} \le \frac{1}{1 + x/n} = \frac{f_n'}{f_n}(x).$$

Together with $f_m(0) = 1 = f_n(0)$ the multiplicative monotonicity theorem implies

$$0 \le x \text{ und } m < n \Rightarrow (1 + \frac{x}{m})^m \le (1 + \frac{x}{n})^n.$$

It is a lot more tedious to prove this monotonicity (of $n \to (1 + x/n)^n$ for fixed x) without our multiplicative monotonicity theorem.
After having available this increasing sequence of functions $\{f_n(x)\}$, the question arises: “How fast does it grow?” Is it bounded for a fixed x? In spite of the simple appearance of these polynomials, it is not easy to calculate an upper bound, but the multiplicative monotonicity theorem will do it fast. Our approach deserves yet another comment. If we regard the factor $1 + x$ as paying an interest rate of x with respect to the principal amount at the end of one year, then $(1 + x/2)^2$ means paying half of the interest twice a year, etc. Looking at this “backwards in time”, the final amount must be multiplied by $(1 - x)$ in order to calculate the initial amount that corresponds to an interest rate of x with respect to the final amount. With this backwards view, we multiply by $1/(1-x/k)^k$ for k payments of x/k .
Therefore we consider

$$h_j(x) := \frac{1}{(1 - x/j)^j} \quad \text{for} \quad x \in [0, j),$$

and we have

$$h_j(0) = 1, \text{ and } \frac{h_j'}{h_j}(x) = \frac{1}{1 - x/j} \ge 1 \ge \frac{1}{1 + x/n} = \frac{f_n'}{f_n}(x).$$

The multiplicative monotonicity theorem therefore implies without further calculations

$$0 \le x < j \Rightarrow (1 + x/n)^n \le 1/(1 - x/j)^j.$$

Thus, the compound interest interpretation of the first formula leads us to an increasing and a decreasing sequence of rational functions, which are convenient bounds for each other!

Enough of the corollaries of the monotonicity theorem for now, let us proceed towards its proof. The most common way is an indirect proof which uses the completeness of the real numbers (cf. next chapter) in an essential way. Since I have mangaged to get along without this completeness so far - we haven't needed any other numbers than the rational ones yet - I want to demonstrate this important theorem without completeness first. As an added benefit, this proof will be a step towards the integral calculus. On the other hand, waiving completeness has its price. Indeed, the final differentiability definition is too weak, so that we *cannot* choose independently of c the constant K and the length r of intervals where the quadratic inequality holds. In this case we will have to use completeness in order to find a point of the interval that will allow us to get a contradiction to finish the indirect proof.
It seems to me that the sophistication of these refined arguments shines more clearly if we get along with the simpler arguments in simpler situations first, and then see later what difficulties will make these simpler arguments fail.
At the moment, we have been dealing only with the rational functions, and these can be approximated particularly well by their tangents, as we have seen. Also the next classes of functions we are going to construct, sin, cos, exp and power series in general, do have this pleasant property. It will need a substantial effort to construct functions that are less well differentiable. Therefore we accept the stronger uniformity assumptions here.

Proof of the monotonicity theorem in case of uniform differentiability.

The main assumption: Let f be differentiable on $[\alpha, \omega]$ and $f' \geq 0$.
The additional assumption: There are positive constants r, K such that the function can be **uniformly** approximated by its tangents, i.e.,

$$c, x \in [\alpha, \omega],\ |c - x| \leq r \ \Rightarrow\ |f(x) - f(c) - f'(c) \cdot (x - c)| \leq K \cdot |x - c|^2.$$

The additional assumption allows us to waive completeness, hence we can, for example, consider rational functions defined only on $\mathbb{Q}$.

Statement: f is weakly increasing, i.e.

$$x \leq y \in [\alpha, \omega] \quad \Rightarrow \quad f(x) \leq f(y).$$

Remark: In our additional assumption it is essential that the constants r, K are indepedent of c, we say they apply "uniformly". The following proof works for all kinds of "small errors", as soon as the required constants can be choosen uniformly. A proof without uniformity follows after we discuss completeness in the next chapter.

Proof: Divide the subinterval $[x, y]$ into n equidistant parts, where n is at least so big that $(y - x)/n \leq r$, or $n \geq (\omega - \alpha)/r$. Set $t_k := x + \frac{k}{n} \cdot (y - x), k = 0, \ldots, n$.

Then, using the additional assumption, we find a constant K, independent of n and k, such that the second of the following inequalities holds, and then, because $f' \geq 0$, the first one follows as well.

$$-K \cdot (t_k - t_{k-1})^2 \leq -K \cdot (t_k - t_{k-1})^2 + f'(t_{k-1}) \cdot (t_k - t_{k-1}) \leq f(t_k) - f(t_{k-1}).$$

We plug $t_k - t_{k-1} = (y-x)/n$ into these inequalities and sum up over $k = 1, \ldots, n$.

$$x \leq y \quad \Rightarrow \quad -K \cdot (y-x)^2 \cdot \frac{1}{n} \leq f(y) - f(x).$$

But now, following Archimedes, we can improve this inequality to $0 \leq f(y) - f(x)$.
Let me summarize the proof. If we apply the definition of differentiability to "small" intervals and also use uniformity, then we obtain the statement for the subintervals, accurate to "very small" errors. Adding all these subresults, we obtain the statement accurate to an irrelevant $1/n$ Archimedes error.

Exercise. Let $(f,g) : [\alpha, \omega] \to \mathbb{R}^2$ be a uniformly differentiable curve with bounded derivative: $\big|(f,g)'\big| := \sqrt{f'^2 + g'^2} \leq L$. Modify the previous proof a little bit and show that L is an expansion bound (Lipschitz bound) to the curve, i.e.:
$x, y \in [\alpha, \omega] \;\Rightarrow\; \big|(f,g)(x) - (f,g)(y)\big| := \sqrt{(f(x) - f(y))^2 + (g(x) - g(y))^2} \leq L \cdot |x - y|.$
Similar arguments give us further properties of uniformly differentiable functions (thus, in particular of polynomials).

Theorem. Expansion bounds (Lipschitz bounds) for f'.
Assumption. Let f be uniformly differentiable with error constant K on all subintervals of length $2r$.
Statement. The derivative f' is Lipschitz bounded with expansion bound $2K$. In formulas:

Expansion bound for f': $\qquad a, b \in [\alpha, \omega] \Rightarrow |f'(a) - f'(b)| \leq 2K \cdot |b - a|.$

Note. A bound $|f''| \leq 2K$ is a Lipschitz bound for f' by corollary 2 to the monotonicity theorem and also controls the deviation of the graph from the tangent by corollary 4. Here we see the opposite conclusion: If K is a (uniform) bound for the deviation of f from its tangents, then $2K$ is a Lipschitz bound for f'. If f'' exists then also $|f''| \leq 2K$.

Proof: First, for every function h, the triangle inequality (reminder: $|C - A| \leq |C - B| + |B - A|$) implies that an expansion bound on $[a, b]$ and $[b, c]$ provides an expansion bound on $[a, c]$, for example

$$\begin{aligned} a < b < c, \quad |h(b) - h(a)| &\leq L \cdot |b - a|, \quad |h(c) - h(b)| \leq L \cdot |c - b| \\ \Rightarrow \qquad |h(c) - h(a)| &\leq L \cdot (|b - a| + |c - b|) = L \cdot |c - a|. \end{aligned}$$

Therefore it suffices to prove the existence of such expansion bounds on small subintervals. Thus, we may assume in addition $|b - a| \leq r$.

Then the assumption gives us first:

$$|f(b) - f(a) - f'(a)\cdot(b-a)| \le K\cdot(b-a)^2,$$
$$|f(a) - f(b) - f'(b)\cdot(a-b)| \le K\cdot(a-b)^2,$$

and now, the triangle inequality gives us the expansion bound on f' which we claimed:

$$|(f'(a) - f'(b))\cdot(b-a)| \le 2K\cdot(b-a)^2, \quad \text{or, for } a \ne b: \ |f'(a) - f'(b)| \le 2K\cdot|a-b|.$$

Information from Higher Derivatives, Taylor Approximation.

In the fourth corollary to the monotonicity theorem we got the quadratic deviation of tangents from bounds for the second derivative. This idea can be iterated. We will get higher and higher powers for the error terms. For small differences in the arguments, higher powers $|x-a|^k$ are much smaller than lower powers (chapter 1, example 2). In cases where the higher derivatives don't grow too fast, we can expect significantly better approximations than by the tangent. As an example we consider the following

Assumption on the 4th derivative: $\quad f : [-R, +R] \to \mathbb{R},\ |f^{(4)}| \le M.$

Statement: $\quad \big|f(x) - (f(0) + f'(0)\cdot x + f''(0)\cdot x^2/2 + f'''(0)\cdot x^3/6\ )\big| \le M\cdot x^4/24.$

Proof. The already proven corollary 4 implies for f'' due to $|(f'')''| \le M$:
$-M\cdot x^2/2 \ \le\ f''(x) - (f''(0) + f'''(0)\cdot x) \ \le\ M\cdot x^2/2 = \big(M\cdot x^3/6\big)'.$
By the generalized monotonicity theorem (1st corollary) we get from this for $x \ge 0$
$-M\cdot x^3/6 \ \le\ f'(x) - f'(0) - (f''(0)\cdot x + f'''(0)\cdot x^2/2) \ \le\ M\cdot x^3/6 = \big(M\cdot x^4/24\big)',$
and analogous inequalities for $x \le 0$. One more application of the generalized monotonicity theorem implies the assertion in both cases, $x \ge 0$ and $x \le 0$.

Remark. It is not difficult to generalize this statement to higher derivatives. Even when M is a lot bigger than $B := \max|f''|$, by chapter 1 there is still an interval around 0, where the error $M\cdot x^4/24$ is much smaller than the tangent deviation $B\cdot x^2/2$. This observation leads in favorable cases to a sequence of better and better approximations, called the Taylor approximations. We will get back to this when we define the functions sin and cos.

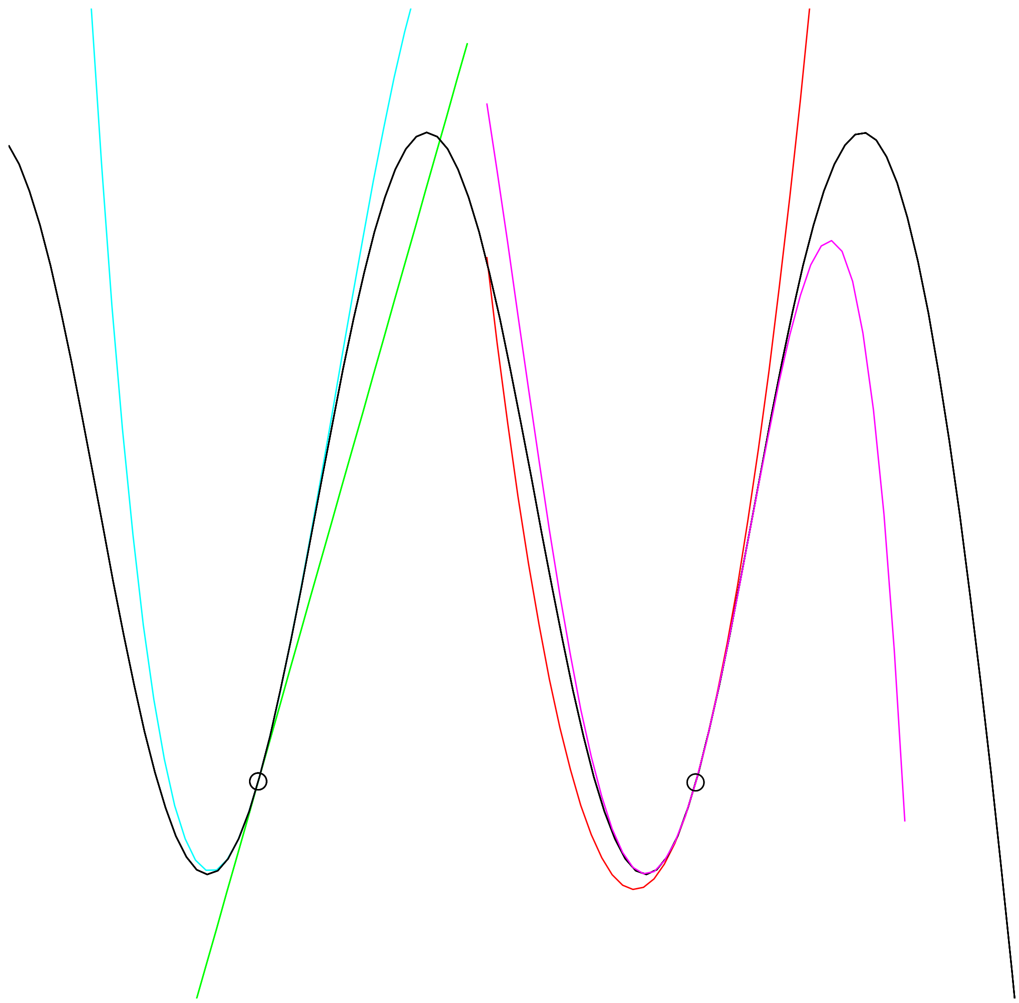


Illustration of Taylor approximations

Taylor polynomials of odd degree stay above or below the approximated function in general, i.e., if the next higher derivative is $\neq 0$. Taylor polynomials of even degree switch from one side to the other in general. At the point on the left, you see the tangent and the 3rd Taylor polyonmial, at the point on the right the 2nd and 4th Taylor polynomial – Since $f''(w) = 0$ at inflection points w, the 2nd Taylor polynomial coincides there with the tangent; therefore the graphs of cubic functions are intersected by their tangents at the inflection point.

Beginnings of integral calculus At the time of Newton, differentiation ("Tangent problem") and integration ("Generalized sums") had been developed independently of each other before it was discovered that they are inverse processes ("Fundamental theorem of calculus"). The completeness of the real numbers was formulated only 200 years later. The next application of the monotonicity theorem goes as closely in the direction of integral calculus as it is possible without completeness.

For this, I am explaining Riemann sums first. For their definition we need the following notations. Let $[\alpha, \omega]$ be contained in the domain of a function f and $[a,b] \subset [\alpha, \omega]$ a subinterval with a so-called partition $a = t_0 < t_1 < \ldots < t_n = b$. The length of the longest subinterval is called "the mesh size" of the partition, let's say $\delta := \max |t_j - t_{j-1}|$. In every subinterval we choose a sample point $\tau_j \in [t_{j-1}, t_j],\ j = 1 \ldots n$, the so-called "tag". Then we come to the following

Definition: Each sum of the type $\sum_{j=1}^{n} f(\tau_j) \cdot (t_j - t_{j-1})$ is called a **Riemann sum** of f for the interval $[a,b]$.

Let us take a look at the following example. Consider the sum $S_N := \sum_{n=1}^{N} n^k$ for fixed k, thus a Riemann sum of the function $x \to x^k$. This helps us to see how fast the S_N grow as a function of N. We compare the summands with values of the function x^k:
$x \in [n, n+1] \Rightarrow (x-1)^k \le n^k \le x^k$ (inequalities reversed if $k < 0$).
All of these three functions are derivatives, namly

$$x \in [n, n+1] \Rightarrow (\tfrac{1}{k+1}(x-1)^{k+1})' \le (n^k \cdot x)' \le (\tfrac{1}{k+1} \cdot x^{k+1})'.$$

Therefore we get from the corollary 1 of the monotonicity theorem:

$$\tfrac{1}{k+1}((n)^{k+1} - (n-1)^{k+1}) \le n^k \cdot (n+1-n) \le \tfrac{1}{k+1} \cdot ((n+1)^{k+1} - n^{k+1})$$

(The monotonicity theorem is in fact trivial for power functions, therefore these inequalities can be verified also directly). Next we sum up over n from 1 to N and get

$$\tfrac{1}{k+1} N^{k+1} \le \sum_{n=1}^{N} n^k \le \tfrac{1}{k+1}((N+1)^{k+1} - 1).$$

The lower and the upper bounds coincide in the highest power of N, and we get the leading term of the sum, $\sum_{n=1}^{N} n^k \sim \frac{1}{k+1} N^{k+1}$. The other terms contribute a percentage of this principal term that gets smaller when N gets bigger.

In the discussion of this example it was important that the summands were values of a function which we could write as the derivative of another function. This is the case for example for all polynomials; we also know that all the power functions $x \mapsto x^k$ with negative exponent $k < -1$ are derivatives. To express this relationship we introduce the following

Definition. A function F is called an **antiderivative** of a function f, if $F' = f$.

Proposition. Riemann sums and uniformly differentiable antiderivates.
Riemann sums of derivatives f' can be calculated by means of the antiderivative f up to "small" errors. More precisely, the following is true
Statement: $|f(b) - f(a) - \sum_{j=1}^{n} f'(\tau_j) \cdot (t_j - t_{j-1})| \le K \cdot |b-a| \cdot \delta$,
where K is the constant provided by uniform differentiability of f and $\delta := \max_j |t_{j+1} - t_j|$.
Note: If we have even $f'(t) \le f'(\tau_j)$, for $t \in [t_{j-1}, t_j]$ then the same proof gives us

$$f(b) - f(a) \le \sum_{k=1}^{n} f'(\tau_j) \cdot (t_j - t_{j-1}),$$

similarly for the opposite inequality, if $f'(t) \ge f'(\tau_j)$.

Proof. In every subinterval we have

$$\begin{aligned}
&|(f(t_j) - f(t_{j-1})) - f'(\tau_j) \cdot (t_j - t_{j-1})| \\
&= |f(t_j) - f(\tau_j) + f(\tau_j) - f(t_{j-1}) - f'(\tau_j) \cdot (t_j - \tau_j + \tau_j - t_{j-1})| \leq \\
&\leq K \cdot (|t_j - \tau_j|^2 + |\tau_j - t_{j-1}|^2) \leq K \cdot \delta \cdot |t_j - t_{j-1}|.
\end{aligned}$$

If we sum up these inequalities over $j = 1, \ldots, n$ and take into account $\sum |t_j - t_{j-1}| = b - a$, we obtain the statement.

Remarks. First, it's easy to believe that this proposition allows for a rather precise estimation of certain sums, like in the previous example. Second, it is import that the errors $K \cdot |b - a| \cdot \delta$ get smaller, when the length δ of the longest subinterval decreases. Visualizing each summand $f'(\tau_j) \cdot (t_j - t_{j-1})$ as the area of a rectangle of height $f'(\tau_j)$ over the subinterval $[t_{j-1}, t_j]$, we recognize that the Riemann sums approximate the "area" below the graph of f' over the interval $[a, b]$. The biggest and the smallest Riemann sum for a fixed partition differs at most by $K \cdot |b - a| \cdot \delta$ from the difference between the values of the antiderivative, $f(b) - f(a)$. Alhough we haven't provided a definition of the area under the graph of f' yet, our error estimate is so good that we can expect now from any definition of the area, as the Archimedes strategy demands, that

$f(b) - f(a)$ is the area under the graph of f'.

Certainly, any different number could **not** be a candidate for the area.

By the way, our estimation of the difference between $f(b) - f(a)$ and any Riemann sum is already one half of the fundamental theorem of calculus. The remaining difficulties come from the fact that for some functions, like $g(x) := 1/x$, we don't know in advance (as we do for polynomials) that they are derivatives of some other functions.

This is about as far as we can go without using the completeness of the real numbers.

Real numbers, Completeness

The rational numbers are not enough to treat inverse functions or the existence of limit functions like exp. They are not even enough to describe **all** the points of a line segment. For an axiomatic description of the real numbers we need the concepts of *convergent sequence, sequence converging to zero or null sequence, nested intervals, Cauchy sequence*. For their definition Archimedes' axiom comes into action. The completeness axiom allows for proving the theorem of Bolzano-Weierstrass, defining inverse functions, and constructing new functions like $\exp, \sin, \cos$ as limits of approximations. – In the standard approach, the difficulties of this chapter appear already at the beginning.

Problems, which cannot be treated with rational numbers only. For all the definitions and propositions discussed so far, it was enough to have the rational numbers in mind when we talk about "numbers". (Of course, it was allowed to know more already, since the statements and theorems of the first chapters remain true if the word "numbers" means either the rational or the real numbers.) I have to explain now, in which way the rational numbers don't suffice. It's probably known from school, that single numbers like $\sqrt{2}$ are irrational, but when we focus too much on single numbers, we misconceive the size of the problem. I will explain first, why we need more functions than the quotients of polynomials we've considered up to now.
The square root is a special example of an *inverse function*. To some given, strictly increasing (e.g. rational) function f it is easy to picture the graph of f together with the the graph of its inverse function g: Reflect the points $(x, f(x))$ of the graph of f at the bisecant of the first quadrant into the points $(f(x), x)$ to get the graph of the inverse function g. In this way we obtain a horizontal parabola as the graph of the square root function from the graph of $f(x) := x^2$. Under this reflection the straight lines with slope m are mapped to straight lines of slope $1/m$; therefore there is no doubt what the derivative of g should be:

$$\text{Derivative of the inverse function: } f'(a) = m,\ g(f(x)) = x \ \Rightarrow\ g'(f(a)) = 1/m,$$

and, this is in accordance with the chain rule: $1 = (g \circ f)'(x) = g'(f(x)) \cdot f'(x)$.
Against this easy looking visual description, it is annoying that we cannot do this with numbers. It wouldn't help though to add the square roots of rational numbers to the rational numbers. First, the roots of these roots would still be missing, and secondly, what about all the other "simple" inverse functions? Obviously, the rational numbers don't suffice when dealing with inverse functions.
Other desires come from the natural sciences. A function with the growth rate $f'/f = 1$ is more important than a function with slope $f' = 1$ in many situations. Functions with $f'/f = 1$ grow in intervals of length a by the *constant factor* $f(a)/f(1)$, since the function $h(x) := f(x+a)/f(x)$ has the derivative 0 by the quotient rule, i.e. has the value $h(x) = h(0) = f(a)/f(0)$. (Functions with $f' = 1$ grow in intervals of length a *additively* by

the constant $f(a) - f(0)$.) Yet, such "exponential functions" cannot be found among the rational functions: Suppose $f = P/Q = f'$, then the quotient rule for polynomials implies

$$\frac{P}{Q} = \left(\frac{P}{Q}\right)' = \frac{QP' - PQ'}{Q^2} \quad \Rightarrow \quad PQ = QP' - PQ',$$

but this is impossible, since the degree of $QP' - PQ'$ is for sure at least 1 less than the degree of PQ. But not only that, we cannot even define (with rational numbers only) exponential functions or the trigonometric functions $(\sin, \cos \ldots)$ which are necessary to describe oscillations or movements on the circle (rotations). The values of the exponential function for example, are not rational for all rational arguments but 0, so that the graph would have this single point $(0, 1) \in \mathbb{Q} \times \mathbb{Q}$ only.
How immensely more numbers we desire, will be clear by the following reasoning: Of course, we would want to be able to express the distance of any point on a straight line segment from its initial point in terms of a *number*. But the rational numbers are very far away from this. To see how far, we order the rational numbers so that we are able to count them: We associate any (reduced) fraction m/n with the point (m, n) in the first quadrant. Now we count these points by counting along the short diagonals one after the other:

$$(1, 1);\ (2, 1), (1, 2);\ (3, 1), (2, 2), (1, 3);\ (4, 1), (3, 2), (2, 3), (1, 4);\ \ldots.$$

Now either we don't count the non-reduced fractions or we ignore that every rational number is counted again every once a while. Having been able to count the rational numbers in the unit interval *one after the other*, we proceed to a second construction. We want to surround every rational number by some space, i.e. cover it by an interval. Surprisingly we can keep the total length ℓ of all the intervals arbitrarily small: We choose a number $\ell < 1$, then use half of the supply for covering the first rational number by an interval of length $\ell/2$. For the second rational number we use again half of the remaining supply, and cover it by an interval of lenght $\ell/2^2$. We continue this way, and after every step the last interval used is as long as the remaining supply. The nth rational number will be covered by an interval of length $\ell/2^n$ and this is just the remaining supply. Now, every rational number will eventually be covered by some nonzero interval, but the total lenght of the intervals is $\leq \ell$. Since we've choosen ℓ as small as we wanted, the rational numbers make up for only an arbitrarily small part of the unit interval, all the other points must be described by "new" numbers.

Indeed, the real numbers are enough to solve all these problems: Rational functions f which are strictly increasing in an interval $[\alpha, \omega]$ of real numbers, have an inverse function that is defined on the real image interval $[f(\alpha), f(\omega)]$; exp, sin, cos can be defined as real valued functions on $\mathbb{R}$; and every *point* of the unit interval can be described by exactly one real *number*, the distance from the initial point. But these successes have their price. We cannot write down the individual real numbers in a similarly explicit way as the rational numbers. We will have to consider a real number as given, when we have proved a sequence to be "convergent": its limit is the real number. Usually, we don't explicitly know

this limit, we only know the elements of the sequence as approximations. Thus, argueing with real numbers, means argueing via approximations.
Notation: The word "sequence" is used as name for a map from the natural numbers $\mathbb{N}$ to a set (of numbers, points, functions ...). We write sequences as $\{a_n\}, \{P_n\}, \{f_n\}, \ldots$.

The necessary notions.

We are going for an axiomatic description of the real numbers. For this, we need the four very closely related notions *sequence converging to zero = null sequence, convergent sequence, nested intervals* and *Cauchy sequence*. All four have a more complicated logical structure, than required in the previous chapters; they cannot be defined by *explicit* inequalities any more. But we only have to master this difficulty once, the definition of the other three notions can be formulated easily in terms of null sequences, we will discuss this notion in detail.

Necessary Defintions.

Convergent sequence: A sequence $\{a_n\}$ is called *convergent* to a, if the sequence $\{a_n - a\}$ converges to zero. The number a is called *limit* of the sequence $\{a_n\}$, $a = \lim_{n\to\infty} a_n$.
It is intended to describe non-rational numbers as limits of convergent sequences of rational numbers.

Nested intervals: A sequence of intervals, $\{[a_n, b_n]\}$ is called *(convergent) nested intervals* if first, they are decreasing, $[a_{n+1}, b_{n+1}] \subset [a_n, b_n]$, and second, $\{b_n - a_n\}$ is a null sequence. (This null sequence property is implied by the Germann "Intervallschachtelung"; we always want $\{b_n - a_n\}$ to be a null sequence and we occasionally emphasize this by adding the word "convergent".) If a number c is contained in all intervals of some nested intervals, $c \in [a_n, b_n]$, all $n \in \mathbb{N}$, then c is called *limit of the (convergent) nested intervals*.
It is intended to see the intervals $[a_n, b_n]$ as approximations of their limit.

Cauchy sequence: A sequence of numbers $\{a_n\}$ is called *Cauchy sequence* if there is a null sequence $\{r_n\}$ such that for all $m > n$ we have: $|a_m - a_n| \le r_n$.
We define Cauchy sequences in order to be able to talk about convergent sequences *without* having to mention their limit.

Discussion of the notion "sequence convergent to 0" or "null sequence".

Since we want to create something principally new, something very different from all previous notions, I can only explain by examples what the new definition has to accomplish. Therefore it will help, if you read the following reflections up to the definition of a null sequence more than once.

Example. The sequence $\{r_n := 1/n\}$ shall be a null sequence.

Comparison. If $\{r_n\}$ is a null sequence and $|a_n| \le r_n$, then $\{a_n\}$ should be a null sequence as well. And if c is a constant, then $\{c \cdot r_n\}$ should be a null sequence as well.

Consequences. We show that we can already argue with the three derived notions, if we

only use the just formulated properties of null sequences.
(i) If c is the limit of nested interals $\{[a_n, b_n]\}$, then c is also the limit of the sequences $\{a_n\}$ and $\{b_n\}$. Proof: By assumption, $\{r_n := b_n - a_n\}$ is a null sequence. It follows $|c - a_n| \le r_n,\ |b_n - c| \le r_n$, so that by definition $\lim a_n = c = \lim b_n$.
(ii) Let $\{[a_n, b_n]\}$ be nested intervals and $\{c_n\}$ a sequence with $c_n \in [a_n, b_n]$, then $\{c_n\}$ is a Cauchy sequence. (In particular, $\{a_n\}$ and $\{b_n\}$ are Cauchy sequences.)
Proof: First, $m \ge n$ implies by assumption that $c_m \in [a_m, b_m] \subset [a_n, b_n]$, hence due to $|c_m - c_n| \le |b_n - a_n| = r_n$ we get the Cauchy property.
(iii) If $0 \le q < 1$ then $\{a_n := q^n\}$ is a sequence with limit 0. Proof: Due to $0 \le q < 1$ and the summation formula for the geometric series we have a comparison ("majorization") with an already known sequence with limit 0:

$$n \cdot q^n \le \sum_{j=1}^{n} q^j = \frac{1 - q^{n+1}}{1 - q} \le \frac{1}{1 - q} \quad \text{hence} \quad q^n \le \frac{1}{1 - q} \cdot \frac{1}{n} = \frac{const}{n}.$$

(iv) If $0 \le q < 1$ then the geometric series $a_n := \sum_{j=0}^{n} q^j$ converges to $1/(1 - q)$.
Proof: The difference $1/(1 - q) - a_n = q^{n+1}/(1 - q)$ is a sequence with limit 0 by (iii).
(v) If $\{a_n\}$ converges to a and f is a function of bounded expansion, (i.e. $|f(x) - f(y)| \le L \cdot |x - y|$), then the sequence of images $\{f(a_n)\}$ converges to $f(a)$. – This result will allow us to see slopes of tangents as limits of the slopes of secant lines.
Proof: Plug in both definitions, $|f(a_n) - f(a)| \le L \cdot |a_n - a| \le L \cdot r_n$, and use the comparison properties of null sequences.
Propaganda. These examples are not at all "artfully" chosen. It is, indeed, only necessary to understand null sequences in order to understand limits.
Next, for the definition of null sequences the following property is important which has been defined already by Eudoxos and has been used by Archimedes very effectively:
If we have two lengths different from 0, then we can lay out the shorter one repeatedly until we have exceeded the longer one.
Since we want to describe the length of line segments by numbers, these numbers must have this Eudoxos' property as well. In our literature this axiom has two names:

Axom of Archimedes, Axiom of Eudoxos. For every real number R (no matter how big) there is a natural number $n \in \mathbb{N}$ which is bigger than R, i.e. $R < n \cdot 1 = n$.
Comments. (i) For the rational numbers the statement of this axiom is an obvious fact.
(ii) There is an equivalent formulation for "small" numbers: the axiom also says that there is no real number between 0 and the set of all unit fractions, $\{1/n;\ n \in \mathbb{N}\}$: For any $r > 0$ (no matter how small) there exists an $n \in \mathbb{N}$ with $n > R := 1/r$ or $0 < 1/n < r$. We formulate this property in a way that appears in many proofs, as a mnemonic:

Archimedes' strategy. In order to prove the inequality $a \le b$ it suffices to show the (infinitely many) weaker inequalities $a \le b + const/n$ for all $n \in \mathbb{N}$.
Archimedes' strategy is so important since is occurs astonishingly often that the weaker

inequalities $a \leq (b+const/n)$ can be proven **a lot** easier than the desired inequality $a \leq b$.
Indirect proof. Assume by contradiction that $b < a$. Then we set $r := const/(a-b) > 0$. By Archimedes' Axiom there is a natural number n with $n > r$. This implies $const/n < a-b$ and $b + const/n < a$, contradicting the assumption $a \leq b + const/n$ for all $n \in \mathbb{N}$.
Below I will describe an "ordered field", where Archimedes' axiom doesn't hold, where there are "infinitely small" elements between 0 and all the unit fractions $\{1/n;\ n \in \mathbb{N}\}$.

Now we can make a first *attempt of defintion*: A sequence $\{r_n\}$ is called a *null sequence* (limit 0), if there is a constant *const* such that we have $|r_n| \leq const/n$ for all $n \in \mathbb{N}$. The above listed properties of null sequences (and hence also the consequences (i)-(v)) are valid statements following this definition. It will turn out that this attempt of a defintion formulates too restrictive a condition. The majority of null sequences under the final definition would not be null sequences with the above attempted definition. What else do we want, what else have mathematicians found desirable? For example, the monotonically decreasing sequence $\{1/\sqrt{n};\ n \in \mathbb{N}\}$ is not a null sequence under our first attempted definition, although Archimedes' axiom implies that between 0 and the elements of this sequence there cannot be any real number. Our last request is motivated by such examples: A monotonically decreasing sequence $\{r_n\}$ shall be a null sequence if between 0 and the set of all r_n there is no other real number. The accepted formulation of this can be expressed by Archimedes' Axiom as follows

Defintion of a null sequence (limit 0). A sequence $\{r_n\}$ is called *null sequence* if for any natural number $k \in \mathbb{N}$ there is an index n_k with the following property:

$$n \geq n_k \quad \Rightarrow \quad |r_n| \leq \frac{1}{k}.$$

Variations. In the literatur we often find the seemingly stronger defintion: A sequence $\{r_n\}$ is called a *null sequence*, if for any positive number $\epsilon > 0$ there is an index n_ϵ with the following property:

$$n \geq n_\epsilon \quad \Rightarrow \quad |r_n| \leq \epsilon.$$

Proof of equivalence. Let $\{r_n\}$ be a null sequence by the first formulation and let $\epsilon > 0$. How can we find n_ϵ? First, by Archimedes' axiom we find a natural number $k > 1/\epsilon$ and by the definition an n_k such that $n \geq n_k \Rightarrow |r_n| \leq 1/k$. Since $1/k < \epsilon$ we can put $n_\epsilon := n_k$.
Of course, the final definition of a null sequence can be used in the definitions of the three derived notions in order to get alternative formulations where the term "null sequence" doesn't appear (as often found in the literature):

Expanded definition of convergence. A sequence $\{a_n\}$ is called **convergent to** a, if for every $\epsilon > 0$ there is an index n_ϵ, such that:

$$n \geq n_\epsilon \quad \Rightarrow \quad |a_n - a| \leq \epsilon.$$

Expanded definition of a Cauchy sequence. A sequence $\{a_n\}$ is called a **Cauchy sequence**, if for every $\epsilon > 0$ there is an index n_ϵ, such that

$$m, n \geq n_\epsilon \quad \Rightarrow \quad |a_m - a_n| \leq \epsilon.$$

These reformulations serve for a better comparison with the literature. They mean exactly the same as the definitions in terms of null sequences. The direct recourse of the definition to null sequences perhaps leads to proofs of convergence more concentrated on the essentials, since of course a proof of convergence is finished as soon as we have managed to find a majorization $|a_n - a| \leq r_n$ by *some already known* null sequence $\{r_n\}$.

Supplement: A non-Archimedean ordering.

Definition. A polynomial is called *positive by coefficients*, in formulas $P \overset{k}{>} 0$, if the smallest **nonzero** coefficient of P is positive. Based on this we compare polynomials:

P is called *bigger than Q by coefficients*, $P \overset{k}{>} Q$, if and only if $P - Q \overset{k}{>} 0$.

Excercise. As for real numbers we have: sum and product of polynomials positive by coefficents are positive by coefficients; for every $P \neq 0$ we have $P^2 \overset{k}{>} 0$. Unlike for real numbers we have for $P(x) := x$, that $P \overset{k}{>} 0$ – but there is no $n \in \mathbb{N}$ with $P \overset{k}{>} 1/n$.

Excersise. This ordering can be extended to rational functions P/Q . Assume that all common factors of P/Q have been cancelled, and define $P/Q \overset{k}{>} 0$ if and only if either $P, Q \overset{k}{>} 0$ or $-P, -Q \overset{k}{>} 0$. As for polynomials we define $P/Q \overset{k}{>} P_1/Q_1$ if and only if $P/Q - P_1/Q_1 \overset{k}{>} 0$. In this way, the field of rational functions becomes an ordered field, but the order is not Archimedean.

Real numbers. It is common in the introductory analysis literature to describe the real numbers axiomaticaly as an ordered field in which the Archimedean axiom and the axiom of completeness hold. The existence of such a field is only proved in advanced literature. Now that I have explained the role of the Archimedean axiom for the definition of the notion of convergence, we only need one more notion, that of completeness. I will discuss three common formulations. Here is the

First formulation of the axiom of completeness:

For all (convergent) nested intervals $[a_k, b_k]_{k \in \mathbb{N}}$ there is a (obviously then unique) real number r, which is contained in all the nested intervals, $r \in [a_k, b_k]$ for all $k \in \mathbb{N}$. r is called the limit of the (convergent) nested intervals.

Uniqueness lemma. If $r_1 \leq r_2$ are limits of the nested intervals $[a_k, b_k]_{k \in \mathbb{N}}$, then $r_1 = r_2$.

Proof. By assumption, r_1, r_2 are contained in all the intervals $[a_k, b_k]$, therefore $0 \leq r_2 - r_1 \leq b_k - a_k$ for all k. Now, since $0 \leq r_2 - r_1$ lies below a sequence with limit 0, Archimedes' axiom implies $r_2 - r_1 = 0$.

Now we have to understand that we get indeed sufficiently many real numbers from the completeness axiom. For this, I first show the theorem of Bolzano-Weierstrass which implies that we have indeed as many numbers as there are points on a straight line segment. It wouldn't be possbile to prove this important theorem if we had been content with less than the final defintion of null sequence. After that, we show the existence of inverse functions. Eventually, we will construct the important functions exp, sin, cos.

Theorem of Bolzano-Weierstrass

Every monotonically increasing and bounded sequence $\{a_k\}$ converges.

Proof. We construct nested intervals $\{[a_k, b_k]\}$, where the left interval bounds are the points of the given monotone sequence $\{a_k\}$, and where the right interval bounds become better and better upper bounds. The difficulty lies in choosing the b_k in such a way that $b_k - a_k$ is a sequence with limit 0. Since the sequence $\{a_k\}$ is bounded, we denote some given upper bound by b_1. A sequence of auxiliary indices starts with $e_1 := 1$.
Now, we repeat the following procedure:
Suppose we have defined already the indices of success $e_j \geq j$ for all $j,\ j \leq m$, and for all $k \leq e_m$ also the upper bounds b_k.
Induction start: This supposition is true for $m = 1$.
Consider the next index $k := e_m + 1$ together with the center $c_k := \frac{1}{2}(a_k + b_{k-1})$.
Either c_k is an upper bound of the sequence $\{a_k\}_{k\in\mathbb{N}}$, then set: $b_k := c_k$, $e_{m+1} := k > e_m$.
Or c_k is **not** an upper bound, then there is a next index (of success) $e_{m+1} > k$ with $a_{e_{m+1}} > c_k$. Since c_k isn't an upper bound, we continue to use the previous upper bound and set for all k with $e_m < k \leq e_{m+1}$: $b_k := b_{e_m}$.
Only in the Either-part of the Alternative do we get a new (smaller) upper bound by interval halfing; in the Or-part we achieve $|b_{e_{m+1}} - a_{e_{m+1}}| \leq 0.5 \cdot |b_{e_m} - a_{e_m}|$ after, possibly, many steps. But this inequality says that the sequence of interval lengths $\{b_k - a_k\}$ of the constructed nesting is indeed a null sequence. Using the axiom of completeness we find a limit $r \in [a_k, b_k]$ of this constructed nested intervals and therefore of the given monotonically increasing sequence $\{a_k\}$. Of course $r = \lim_{k\to\infty} a_k$.
Obviously we have $a_k \leq r$ for all $k \in \mathbb{N}$, and there is no smaller upper bound for a_k. We'll get back to this observation later.

Now we prove similarly the proposition that is suggested by reflecting the graph of f

Proposition on the Surjectivity of Monotone Functions. Let $f\colon [a, b] \to \mathbb{R}$ be monotone and expansion bounded. Then f is surjective onto the image $[A_1, B_1] := [f(a), f(b)]$. If f is strictly monotone, then there exists an inverse function defined on the real interval $[A_1, B_1]$.

Proof. Let $C \in [A_1, B_1]$ be arbitrary and L be the expansion bound of f. We construct nested intervals in $[a, b]$, whose image invervals converge to C:
Put $a_1 = a, b_1 = b$, also $f(a_1) = A_1, f(b_1) = B_1$.
Consider $D := f((a_1 + b_1)/2)$.

If $D \geq C$ then put $a_2 := a_1,\ b_2 := (a_1+b_1)/2$, otherwise $a_2 := (a_1+b_1)/2,\ b_2 := b_1$.
We repeat this procedure:
Let $[a_m, b_m]$ be the last defined interval – we know by construction that $C \in f([a_m, b_m])$.
Consider $D := f((a_m+b_m)/2)$.
If $D \geq C$, then put $a_{m+1} := a_m,\ b_{m+1} := (a_m+b_m)/2$,
otherwise $a_{m+1} := (a_m+b_m)/2,\ b_{m+1} := b_m$.
Obviously $\{[a_k, b_k]\}$ are (convergent) nested intervals with $C \in f([a_k, b_k])$ for all $k \in \mathbb{N}$. Since f has an expansion bound, this image nesting converges to C, by consequence (vi) in the discussion of null sequences. Now we quote the axiom of completeness for the *nested intervals in the domain* and find $c \in [a_k, b_k]$ for all $k \in \mathbb{N}$. Then $|f(c) - C| \leq L \cdot |b_k - a_k|$, hence $f(c) = C$ by Archimedes' axiom. – If we further assume strict monotonicity, then c is the only preimage of C, thus the inverse function can be defined: $f^{-1}(C) := c$.
Example: For $f : (0, \infty] \to (0, \infty],\ f(x) := x^2,\ \ (f'(x) = 2x)$ the inverse function, namely the square root, has been defined: $f^{-1} : (0, \infty] \to (0, \infty],\ f^{-1}(y) = \sqrt{y},\ f^{-1} \circ f(x) = x$.
Due to the chain rule we expect:

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

We can deduce this without any differentiation rule directly from the definition of differentiability by proving a quadratic approximation:

$$\sqrt{x} - \sqrt{a} - \frac{1}{2\sqrt{a}} \cdot (x-a) = (\frac{1}{\sqrt{x}+\sqrt{a}} - \frac{1}{2\sqrt{a}}) \cdot (x-a)$$
$$= \frac{\sqrt{a}-\sqrt{x}}{2\sqrt{a}(\sqrt{x}+\sqrt{a})} \cdot (x-a) = \frac{-(x-a)^2}{2\sqrt{a} \cdot (\sqrt{x}+\sqrt{a})^2}, \qquad \text{hence}$$
$$\frac{a}{2} \leq x \leq 2a \ \Rightarrow \ \frac{-1}{4\sqrt{a}^3} \cdot (x-a)^2 \leq \ \sqrt{x} - (\sqrt{a} + \frac{1}{2\sqrt{a}} \cdot (x-a)) \ \leq 0.$$

Remark. In order to emphasize that a differentiation rule belongs to every new construction of functions, we prove the differentiation rule for the inverse functions, which has been plausibilized earlier. This proof doesn't contribute to the discussion of completeness and can therefore be read later.
Proof of the differentiation rule for inverse functions.
Assumption: Let $f : [a, b] \to [A, B]$ be strictly monotone; The inverse function exists by the previous proposition and is denoted by $f^{-1} : [A, B] \to [a, b]$. Further let f be differentiable at $c \in [a, b],\ f(c) = C,\ f(x) = X$ and $f'(c) > 0$.
We use differentiability with quadratic error, i.e. the existence of constants r, K such that

$$|x - c| \leq r \Rightarrow |f(x) - f(c) - f'(c) \cdot (x-c)| \leq K \cdot (x-c)^2$$

Claim: $$(f^{-1})'(C) = 1/f'(c).$$

Proof. Plug the notations into the inequality assumed for f and divide by $f'(c)$:

$$(*) \qquad |x-c| \le r \Rightarrow \left|\frac{1}{f'(c)} \cdot (X-C) - f^{-1}(X) + f^{-1}(C)\right| \le \frac{K}{f'(c)} \cdot (x-c)^2$$

This inequality is almost the one we need; on the right hand side we only have to bound $(x-c)^2$ by $(X-C)^2$, and further, the assumptions on x have to be transformed into assumptions on X. As in the proof of the chain rule (please compare), the definition of differentiability together with the triangle inequality implies a two-sided comparison between $|x-c|$ and $|f(x)-f(c)| = |X-C|$:

$$|x-c| \le r \;\Rightarrow\; (f'(c) - K\cdot r)\cdot|x-c| \le \; |f(x)-f(c)| \; \le (f'(c) + K\cdot r)\cdot|x-c|.$$

To make the left inequaltiy useful, we decrease r to $r_- := \min(r, f'(c)/2K)$ and make the latter inequality rougher in order to get a good two-sided approximiation:

$$|x-c| \le r_- \;\Rightarrow\; \frac{1}{2}f'(c)\cdot|x-c| \le |X-C| \le \frac{3}{2}f'(c)\cdot|x-c|.$$

So far, the inequality $(*)$ holds under the assumption $|x-c| \le r$ and the lower estimate for $|X-C|$ under the assumption $|x-c| \le r_-$. We can formulate now an assuption on $X-C$ which implies $|x-c| \le r_-$, namely $|X-C| \le R := \frac{1}{2}f'(c)\cdot r_-$. Thus we have $(*)$ under an assumption on the arguments X of f^{-1}. Finally we make $(*)$ rougher by plugging in $|x-c| \le \frac{2}{f'(c)} \cdot |X-C|$ to get

$$|X-C| \le R \;\Rightarrow\; \left|f^{-1}(X) - f^{-1}(C) - \frac{1}{f'(c)}\cdot(X-C)\right| \le \frac{4K}{f'(c)^3}\cdot|X-C|^2.$$

This is the desired quadratic approximation by tangents which shows $(f^{-1})'(C) = 1/f'(c)$. The proof can be applied to ϵ-δ-errors: It is easy to replace $K\cdot(x-c)^2$ by $\epsilon\cdot|x-c|$; we get more difficulties in arguing that the *assumptions* on $|X-C|$ guarantee that the assumptions for the validity of the used estimates for f hold. As in the proof of the chain rule we have to get back to this detail only if we want to prove these propositions under the weaker final definition of differentiability, see the last chapter.

Construction of differentiable functions as limit functions. This is the third problem where the axiom of completeness has to succeed. Here we need a longer chain of arguments, since it cannot be too simple to determine the *derivative* of functions whose *values* we don't calculate but can only approximate.

First construction of the exponential function. We use the compound interest functions which we have discussed as an application of the multiplicative monotonicity theorem:

$$0 \le x < j \Rightarrow f_n := (1+x/n)^n \le 1/(1-x/j)^j =: h_j(x).$$

The growth rates of these function were

$$\frac{f_n'}{f_n}(x) = \frac{1}{1+x/n} \le 1 \le \frac{1}{1-x/j} = \frac{h_j'}{h_j}(x).$$

We see that the growth rates of the f_n increase monotonically, and that those of the h_j converge monotonically decreasing to 1. So we can hope that these compound interest functions converge to the exponential function. Due to the zeroes in the denominator we consider now a fixed interval $[0,4]$ and take only the functions with $n, j \ge 8$. Then the values of these functions are in nested intervals

$$x \in [0,4],\ n \ge 8 \quad \Rightarrow \quad [1, h_8(4)] \supset [f_8(x), h_8(x)] \supset [f_n(x), h_n(x)] \supset [f_{n+1}(x), h_{n+1}(x)].$$

Indeed, we even have convergent nested intervals, since the differences $h_n(x) - f_n(x)$ are majorized by a null sequence already known to us. Set $q := 1 - (x/n)^2 < 1$ and use $(1-q^n) = (1-q)(1+q+\ldots+q^{n-1}) \le (1-q)\cdot n = x^2/n$ to get:

$$0 \le h_n(x) - f_n(x) = h_n(x) \cdot \Big(1 - \big(1-(x/n)^2\big)^n\Big) \le h_8(4) \cdot x^2/n \le 256x^2/n\,.$$

Therefore for every $x \in [0,4]$, the nested intervals $[f_n(x), h_n(x)]$ converge *due to the axiom of completeness*. We call the limit $E(x)$. We expect that this limit function is differentiable and that $E'(x) = E(x)$ holds. The proof is surprisingly simple since the functions f_n comply with the assumptions of the monotonicity theorem and since obviously:

$$x \in [0,4],\ n \ge 8 \ \Rightarrow \ 0 \le f_n''(x) \le f_n'(x) \le f_n(x) \le h_8(4) = 256.$$

First, the monotonicity theorem implies the Lipschitz bound 256 for all f_n in $[0,4]$, hence

$$|f_n(x) - f_n(y)| \le 256|x-y|.$$

In our construction, the difference between the limit function and the approximation is already majorized by a sequence with limit 0:

$$|E(x) - f_n(x)| \le 256x^2/n$$

Therefore Archimedes' strategy implies **the same** Lipschitz bound for the limit function:

$$x, y \in [0,4] \Rightarrow |E(x) - E(y)| \le 256|x-y|.$$

This argumentation with bounds, that hold for all approximations in the same way, and which therefore can be extended immediately to the limit function, allows to prove differentiability and determine the derivative. The bound for the second derivative and the consequence 4 (deviation from the tanget) of the monotonicity theorem gives:

$$x, y \in [0,4],\ n \ge 8 \ \Rightarrow |f_n(y) - f_n(x) - f_n'(x)(y-x)| \le 128(y-x)^2.$$

Again, the differences $E(y) - f_n(y),\ E(x) - f_n(x),\ E(x) - f_n'(x)$ can be estimated by sequences with limit 0, and hence Archimedes' strategy implies

$$x, y \in [0,4] \Rightarrow |E(y) - E(x) - E(x)(y-x)| \le 128(y-x)^2.$$

This approximation means: The function $y \mapsto E(y)$ is so well approximated by the linear function $y \mapsto E(x) - E(x)(y - x)$, that this is the tangent of the function E at x, see the definition of differentiability. Therefore we have in particular $E'(x) = E(x)$.
This means that we have constructed our first limit function and determined its derivative.

Of course, we have to get rid of the restriction to the interval $[0, 4]$ now. For that we define functions $E_k(x) := E(x/k)^k : [0, 4k] \to \mathbb{R}$. We have $E_k(0) = E(0) = 1$ and the chain rule implies $E_k' = E_k$. Thus the quotient rule gives $(E_k/E_j)' = 0$, i.e. every two of these functions coincide on the common part of their domain. This trick expands the domain to $[0, \infty)$. The formulas are also valid for $k = -1$, hence with $E(-x) := 1/E(x)$ the domain gets extended to all of $\mathbb{R}$. In the same way we get the addition theorem from $h(x) := E(x)E(a)/E(a + x),\ h' = 0,\ h(0) = 1$.

Notation. We call the limit function which we have obtained in this way **exponential function**,

$$\exp : \mathbb{R} \to \mathbb{R},\ \exp(0) = 1,\ \exp' = \exp, \quad \exp(a + x) = \exp(a) \cdot \exp(x).$$

Taylor approximation of exp**.** For the construction of exp I have used the compound interest functions since they supply upper and lower bounds. Even better known are the Taylor approximations of exp. The n-th Taylor polynomial of exp is the polynomial T_n of degree n, whose derivatives up to the order n at 0 coincide with those of the exponential function. Due to

$$1 = \exp(0) = \exp'(0) = \exp''(0) = \ldots = \exp^{(n)}(0)$$

we can easily write down these Taylor polynomials.

Taylor polynomials of exp: $T_n(x) := \sum_{k=0}^{n} \frac{x^k}{k!}, \qquad T_n' = T_{n-1}.$

For any $x \in [0, \infty)$ the sequence $\{T_n(x)\}$ increases monotonically. It is also bounded: Due to $T_n'(x) \le T_n(x)$ and $\exp' = \exp$ we have that $h(x) := T_n(x)/\exp(x)$ satisfies $h' \le 0,\ 0 < h(x),\ h(0) = 1$ hence $T_n(x) \le \exp(x)$. Therefore $\{T_n(x)\}$ converges by the theorem of Bolzano-Weierstrass. On the other hand, we have $f_n^{(k)}(0) \le 1,\ k = 0 \ldots n$ for the lower compound interest approximations. Therefore all the coefficients of the polynomials f_n are at most as big as those of the Taylor Polynomials T_n, i.e.

$$x \in [0, \infty) \ \Rightarrow\ f_n(x) \le T_n(x) \le \exp(x).$$

Therefore, also the $T_n(x)$ converge to $\exp(x)$. – If we want to start the construction with the Taylor polynomials, then we have to provide primarily an upper bound for the $T_n(x)$, compare Leibniz series below.

Excercise. For every $n \in \mathbb{N}$ the sequence $\{a_k := k^n/\exp(k)\}$ has limit 0. (How big do we have to choose x so that we can guarantee $x^{n+1} \le \exp(x)$? It is not requested to get a good approximation. Cf. p.56.)

Irrationality of e**.** To demonstrate the quality of this Taylor approximation, I will show:

$$e := \exp(1) \text{ is irrational.}$$

Proof. For $1/e = \exp(-1) = \sum_{k=0}^{\infty} (-1)^k/k!$ we can immediately provide convergent nested intervals (by employing the alternating signs):

$$a_n := \sum_{k=0}^{2n-1} \frac{(-1)^k}{k!}, \;\; b_n := \sum_{k=0}^{2n} \frac{(-1)^k}{k!}, \;\; n = 1, 2, \dots$$

$$a_{n+1} - a_n = \frac{1}{(2n)!} - \frac{1}{(2n+1)!} > 0, \;\; b_n - b_{n+1} = \frac{1}{(2n+1)!} - \frac{1}{(2n+2)!} > 0$$

$$0 < b_n - a_n = \frac{1}{(2n)!}.$$

This implies for all N:

$$0 \neq |1/e - \sum_{k=0}^{N} (-1)^k/k! \;| < 1/(N+1)!$$

Now, if $1/e = P/Q$ were such that $P, Q \in \mathbb{N}$, then we select $N := Q$ in the last approximation and observe that the estimated difference is a rational number whose denominator is $Q!$:

$$0 \neq \frac{|P \cdot (Q-1)! - \text{ Some Integer } |}{Q!} < \frac{1}{(Q+1)!}.$$

Since a fraction $\neq 0$ with denominator $Q!$ has at least the size $1/Q!$, this last inequality is a contradiction to the assumption $1/e = P/Q$. (The "Some Integer" is $\sum_{k=0}^{Q}(-1)^k \cdot Q!/k!$.)

Leibniz series. The convenient nested intervals for $1/e$ just used is a special case of a situation, which appears frequently enough to get its own name; we describe the nested intervals for a *Leibniz series:*
Assumption: $\{r_k\}$ is a monotonically decreasing sequence with limit 0
Propostion: $a_n := \sum_{k=0}^{2n-1}(-1)^k r_k, \;\; b_n := \sum_{k=0}^{2n}(-1)^k r_k$ defines nested intervals.
Proof:

$$a_{n+1} - a_n = r_{2n} - r_{2n+1} \geq 0, \; b_n - b_{n+1} = r_{2n+1} - r_{2n+2} \geq 0, \; b_n - a_n = r_{2n}.$$

Examples . First, we can apply this on a negative interval $[-R, 0]$ to the Taylor polynomials T_n of exp, if we restrict ourselves to indices $n \geq R$. Next we show a

Construction of the trigonometric functions sin and cos. We want to try to construct functions $f, g : \mathbb{R} \to \mathbb{R}$ with $f^2(t) + g^2(t) = 1, f'^2(t) + g'^2(t) = $ const., since then the curve $t \to (f(t), g(t))$ parametizeses the unit circle with *constant* speed. Due to the chain rule, we may assume const $= 1$ (replace $f(t)$ by $f(t/\sqrt{\text{const.}})$).
Assume we could find functions f, g which are repeatedly differentiable with these properties. Then, by differentiation, we conclude from

$f^2 + g^2 = 1, \; f'^2 + g'^2 = 1$ that:

$2 \cdot f \cdot f' + 2 \cdot g \cdot g' = 0$ and $2f' \cdot f'' + 2g' \cdot g'' = 0$.

These two equalities can be read as scalar products: As expected, the velocity vector (f', g') is perpendicular to the position vector (f, g), and the acceleration vector (f'', g'') is

perpendicular to (f', g') hence proportional to (f, g), or $(f'', g'') = c \cdot (f, g)$. We obtain the factor $c = -1$ by differentiating $f \cdot f' + g \cdot g' = 0$ to get $f \cdot f'' + g \cdot g'' + f'^2 + g'^2 = 0$ or $c + 1 = 0$. Hence necessarily: $(f'', g'') + (f, g) = 0$.
Next we also agree that our parametrization starts at $(f(0), g(0)) = (1, 0)$ with the velocity vector $(f'(0), g'(0)) = (0, 1)$. Having chosen these initial values we get from $(f'', g'') + (f, g) = 0$ all derivatives of the functions f and g at $t = 0$ – provided they exist and are repeatedly differentiable:

$$f(0) = 1, f'(0) = 0, f''(0) = -1,\ \ f'''(0) = 0, f^{(2k)}(0) = (-1)^k, f^{(2k+1)}(0) = 0;$$
$$g(0) = 0, g'(0) = 1, g''(0) = 0,\ \ g'''(0) = -1,\ \ g^{(2k)}(0) = (0), g^{(2k+1)}(0) = (-1)^k.$$

Of couse, now it is easy to write down polynomial sequences which have derivatives at 0 coinciding with these derived values for higher and higher order:

$$P_n(t) = 1 - \frac{t^2}{2} + \frac{t^4}{4!} - + \ldots = \sum_{k=0}^{n} \frac{(-1)^k}{(2k)!} \cdot t^{2k},$$
$$Q_n(t) = t - \frac{t^3}{6} + \frac{t^5}{5!} - + \ldots = \sum_{k=0}^{n} \frac{(-1)^k}{(2k+1)!} \cdot t^{2k+1}.$$

They satisfy:

$$Q'_n = P_n,\quad P'_n = -Q_{n-1},\quad P''_n = -P_{n-1},\quad Q''_n = -Q_{n-1}.$$

I will look more closely at these polynomials

$$P_{n+1}(t) = P_n(t) - \frac{(-1)^n}{(2n+2)!} \cdot t^{2n+2},\qquad Q_{n+1}(t) = Q_n(t) - \frac{(-1)^n}{(2n+3)!} \cdot t^{2n+3}.$$

They are suitable for the application of the Leibniz criterion: Choose a fixed $R > 0$ and only look at the polynomials with $n \geq R$. Then we have:
for $n \geq R$ and $|t| \leq 2R$, the sequences $\{|t|^{2n}/(2n)!\}_{n\in\mathbb{N}}$ and $\{|t|^{2n+1}/(2n+1)!\}_{n\in\mathbb{N}}$ have limit 0 and are monotone since consecutive terms are decreased by factors $\leq |t|/2n \leq 1$.
Therefore, for $2m \geq R$ and $|t| \leq 2R$ we get the Leibniz nested intervals for the desired limit functions f and g

$$\{[P_{2m+1}(t), P_{2m}(t)]\},\quad \{[Q_{2m+1}(t), Q_{2m}(t)]\}.$$

Due to the axiom of completeness, the functions f and g *do exist* as limit functions of these convergent nested intervals.
Now we still need the differentiability properties of these limit functions. This is now particularly easy since the Leibniz' nested intervals already give bounds indepenent of n.

Archimedes' strategy gets the same bounds for the limit functions:

$$2m \ge 2, \quad 0 \le t \le 4 \quad \Rightarrow$$
$$-7 \le P_1(t) \le P_{2m+1}(t) \le \dots f(t) \dots \le P_{2m}(t) \le P_2(t) \le 4$$
$$-7 \le Q_1(t) \le Q_{2m+1}(t) \le \dots g(t) \dots \le Q_{2m}(t) \le Q_2(t) \le 4.$$

Proof: $0 \le t \le 4 \quad \Rightarrow$

$$P_2(t) = 1 - \frac{t^2}{2}(1 - \frac{t^2}{12}) \le 4 \qquad Q_2(t) = t \cdot (1 - \frac{t^2}{6} \cdot (1 - \frac{t^2}{20})) \le 4$$
$$-7 \le 1 - \frac{t^2}{2} = P_1(t) \qquad -7 \le t - \frac{t^3}{3!} = Q_1(t).$$

From these bounds for the polynomials and $Q'_n = P_n,\ P'_n = -Q_{n-1}$ we get expansion bounds from (corollaries of) the monotonicity theorem

$$n \ge 4, \quad 0 \le s, t \le 4 \ \Rightarrow \ |P_n(s) - P_n(t)| \le 7 \cdot |s - t|, \quad |Q_n(s) - Q_n(t)| \le 7 \cdot |s - t|.$$

Then the triangle inequality implies

$$|f(s) - f(t)| \le 7 \cdot |s - t| + \text{ null sequence}, \quad |g(s) - g(t)| \le 7 \cdot |s - t| + \text{ null sequence}.$$

and Archimedes' trick removes the null sequences. Hence we have found the expansion bound 7 (wich can be improved) for the limit functions in the interval $[0, 4]$.
In the same way we get bounds independet of n for the second derivatives

$$P''_n = -P_{n-1}, \quad Q''_n = -Q_{n-1}$$

and therefore also bounds independent of n for the tangent deviation. Again by Archimedes' trick we get the same inequalities for the two limit functions:

$$n \ge 4, \quad 0 \le s, t \le 4 \Rightarrow$$
$$|P_n(s) - P_n(t) + Q_{n-1}(t)(s - t)| \le 4 \cdot |s - t|^2 \Rightarrow |f(s) - f(t) + g(t)(s - t)| \le 4 \cdot |s - t|^2,$$
$$|Q_n(s) - Q_n(t) - P_n(t) \cdot (s - t)| \le 4 \cdot |s - t|^2 \Rightarrow |g(s) - g(t) - f(t)(s - t)| \le 4 \cdot |s - t|^2.$$

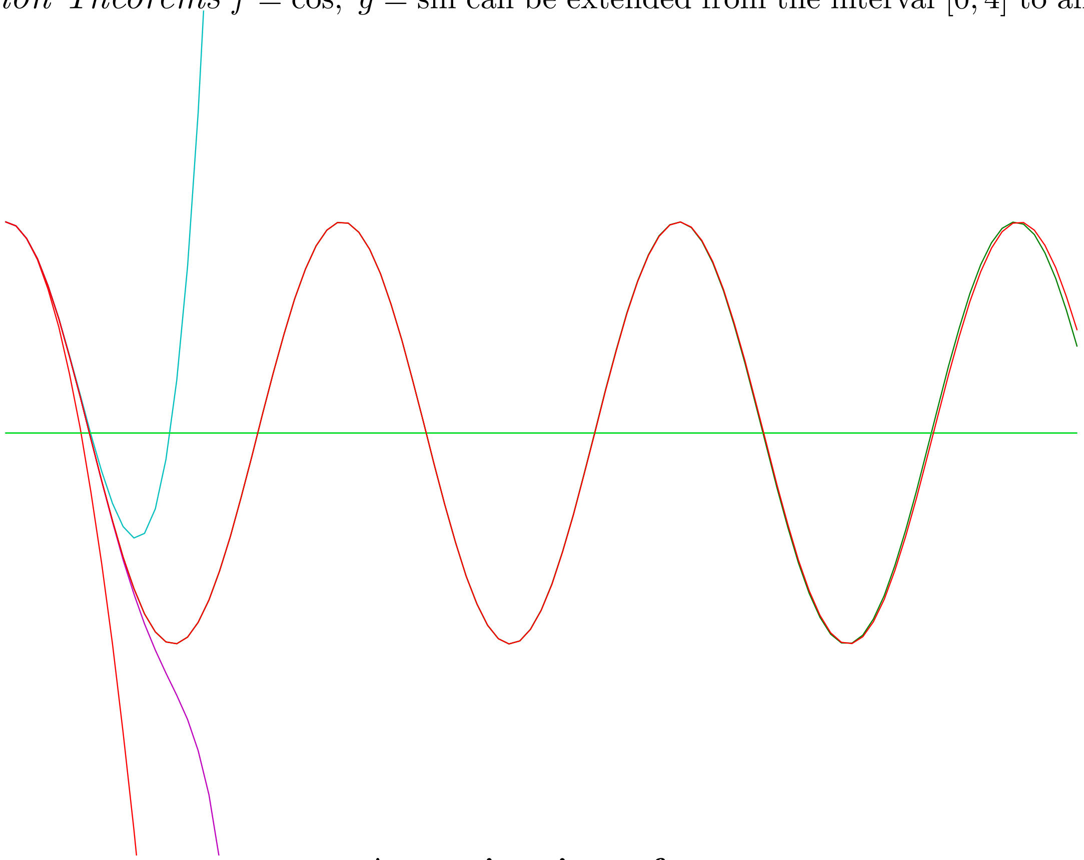
The latter two inequalitites say that the limit functions are differentiable (more precisely even uniformly differentiable with quadratic tangent approximation).

We have: $f' = -g, \quad g' = f, \quad f(0) = 1, \ g(0) = 0.$

These equalities show that f' and g' are differentiable, $f'' = -f,\ g'' = -g$, etc.: The limit functions are infinitely many times differentiable. (We can now easily provide a better error bound: From $f'^2 + g'^2 = 1$ and the monotonicity theorem we get that 1, instead of 7, is an expansion bound for f and g. In the same way we have $|f''| \le 1, |g''| \le 1$, and therefore the tangent deviation is $\le 0.5|s - t|^2$.)
Notations: $f = \cos,\ g = \sin,\ \sin' = \cos,\ \cos' = -\sin, \quad \tan := \frac{\sin}{\cos}, \quad \tan' = 1 + \tan^2.$

The first positive zero of cos is called $\pi/2$; due to $P_2(t) \leq P_6(t) \leq \cos(t) \leq P_4(t)$ and $P_4(\sqrt{3}) = -1/8, P_6(1.5) \geq 0.07$ we have: $\sqrt{2} \leq 1.5 + 0.07 \leq \pi/2 \leq \sqrt{3} - 1/8 \sim 1.607$.

Addition Theorems. For functions h with $h'' = -h$ the function $K(t) := h(t)^2 + h'(t)^2$ has the derivative $K'(t) = 2h'(t)(h(t) + h''(t)) = 0$, hence is constant. Therefore, two functions h_1, h_2 with the *same* initial values $h_1(a) = h_2(a)$, $h_1'(a) = h_2'(a)$ have a difference h with initial values $h(a) = 0 = h'(a)$, thus $0 = K(a) = K(t)$, and $h = 0$. Therefore the functions $h_1(t) := \sin(a + t)$ and $h_2(t) := \sin(a)\cos(t) + \cos(a)\sin(t)$ coincide and their derivatives $h_1'(t) := \cos(a + t)$ and $h_2'(t) := -\sin(a)\sin(t) + \cos(a)\cos(t)$ coincide as well. – With these *Addition Theorems* $f = \cos$, $g = \sin$ can be extended from the interval $[0, 4]$ to all of $\mathbb{R}$.



Approximations of cos.

The figure shows the graph of cos and the graphs of the 2nd, 4th and 6th Taylor polynomials T_n. In addition, a much better approximation is shown:
From $\cos(2t) = \cos^2(t) - \sin^2(t) = 2\cos^2(t) - 1$ we get $\cos(4t) = 8\cos^4(t) - 8\cos^2(t) + 1$. Therefore one can compute $\cos(t)$ from $\cos(t/4)$ with the help of the polynomial $P4(x) := 8x^2(x^2 - 1) + 1$ as follows: $\cos(t) = P4(\cos(t/4))$. The Taylor polynomials are for *small arguments much better* approximations, so that we get improved approximations if we combine Taylor polynomials on small arguments with this addition theorem. The figure shows $\cos(t) \sim P4 \circ P4 \circ T_4(t/16)$, and the extension of the graph over several periods is to illustrate how good an approximation (more so on the smaller period interval $[0, \pi/2]$) we obtained.

Inverse Functions of $\exp, \tan, \sin$**.** Since $\exp : \mathbb{R} \to (0, \infty)$ is strictly monotone and has on each finite interval a bounded derivative we have a differentiable inverse function, called *Logarithm Function*, $\log \colon (0,\infty) \to \mathbb{R}$ with $\log(\exp(x)) = x$. The chain rule gives:

$$\log'(\exp(x)) = \frac{1}{\exp'(x)} \quad \text{or with } \; y := \exp(x)\colon \; \log'(y) = \frac{1}{y}.$$

From $\exp(x)^n = \exp(nx)$ we conclude $n \log(\sqrt[n]{y}) = \log(y)$. This is combined with a simple application of the monotonicity theorem:

$$y \geq 1 \;\Rightarrow\; 0 < \log'(y) \leq 1 \;\Rightarrow\; \log(y) \leq 1 \cdot (y-1) \;\Rightarrow\; \log(y) \leq n \cdot (\sqrt[n]{y} - 1).$$

This inequality shows, that the logarithm is an extremely slowly increasing function. The earlier question " How large do we have to choose x so that $x^m \leq \exp(x)$ holds?" now has a simple answer: The inequality is equivalent to $m \cdot \log(x) \leq x$ and therefore – because of $m \cdot \log(x) \leq m \cdot 2(\sqrt{x} - 1) \leq 2m\sqrt{x}$ – the claim is implied by $2m\sqrt{x} \leq x$, i.e.by $4m^2 \leq x$. Exponential functions for an arbitrary base $a > 0$ are defined with the help of the logarithm:

$$\exp_a(x) := a^x := \exp(x \log a), \quad \text{then} \quad \exp_a' = \log(a) \exp_a = \exp_a'(0) \exp_a . \text{Consequently}$$

$$a^{x+h} = a^x \cdot a^h \;\text{ and }\; \frac{\exp_a(x+h) - \exp_a(x)}{h} = \frac{\exp_a(h) - 1}{h} \cdot \exp_a(x) \geq \exp_a'(0) \exp_a(x).$$

We see: the addition theorem implies that we know the derivatives at all x, i.e. all limits of difference quotients of exponential functions, if we only know the limit at $x = 0$.

Also $\tan \colon (-\pi/2, +\pi/2) \to \mathbb{R}$ is a strictly monotone differentiable function with bounded expansion on every closed subinterval $[-a, b]$. Again we have a differentiable inverse function, called *arc tangent function*, $\arctan \colon \mathbb{R} \to (-\pi/2, +\pi/2)$ with $\arctan(\tan(x)) = x$ and, by the chain rule

$$\arctan'(\tan(x)) = \frac{1}{\tan'(x)} \quad \text{or with } \; y := \tan(x)\colon \; \arctan'(y) = \frac{1}{1+y^2}.$$

We emphasize, that the differential equations $\exp' = \exp, \; \tan' = 1 + \tan^2$ imply that the derivatives of the inverse functions are *rational* functions. In particular one can easily compute higher derivatives of these inverse functions.

Finally, also $\sin \colon (-\pi/2, +\pi/2) \to (-1, +1)$ is a strictly monotone differentiable function with bounded nonzero derivative on closed subintervals $[-a, b]$. Again we have a differentiable inverse function, called *arc sin function*, $\arcsin \colon (-1, +1) \to (-\pi/2, +\pi/2)$ with $\arcsin(\sin(x)) = x$ and by the chain rule

$$\arcsin'(\sin(x)) = \frac{1}{\sin'(x)} \quad \text{or with } \; y := \sin(x)\colon \; \arcsin'(y) = \frac{1}{\sqrt{1-y^2}}.$$

With sin also arcsin is an *odd function*, so that all even derivatives vanish at 0.

$$\arcsin''(y) = y(1-y^2)^{-3/2},\ \arcsin'''(0) = 1.$$

The third Taylor polynomial of arcsin at 0 therefore is $T_3(y) = y + y^3/6$. We illustrate with an iteration algorithm to compute $\pi/2$ how well these Taylor polynomials approximate these new functions. Assume that a_1 is an approximation for $\pi/2$. We have seen above that $b_1 := \cos(a_1)$ can easily be computed acurately. Since $\cos(\pi/2 - x) = \sin(x)$ we have $\pi/2 = a_1 + \arcsin(b_1)$. Hence put $a_2 := a_1 + (b_1 + b_1^3/6),\ b_2 := \cos(a_2)$, etc. Example:
$a_1 = 1, b_1 = 0.5403, \quad a_2 = 1.5666, b_2 = 0.0042059, \quad a_3 = 1.570796\ldots, b_3 = 9.881 \cdot 10^{-14}$.

Other Formulations of Completeness. Frequently one has some sequence which one suspects to be convergent but which one cannot explicitly connect to (convergent) nested intervals (e.g. the sequence $T_n(x)$ of Taylor polynomials of exp). It would be nice to be able to judge a sequence as convergent without having to know its limit. A necessary condition comes from applying the triangle inequality to the definition of a sequence $\{a_k\}$ converging to a :
For every $n \in \mathbb{N}$ there exists an index k_n such that

$$k, l \geq k_n \Rightarrow |a_k - a| \leq \frac{1}{n}, \quad |a_\ell - a| \leq \frac{1}{n}, \quad \text{hence:} \quad |a_k - a_\ell| \leq \frac{2}{n}.$$

The last inequality is formulated without the limit a of the sequence! That suggests the definition of a Cauchy sequence, already met earlier, but repeated:
A sequence $\{a_k\}$ is called **Cauchy sequence** if it satisfies: For every $n \in \mathbb{N}$ there exists some k_n such that

$$k, \ell \geq k_n \Rightarrow |a_k - a_\ell| \leq \frac{1}{n}.$$

With this notion we have the

Second Formulation of the Axiom of Completeness:
Every real Cauchy sequence has some real number r as limit.

This second formulation of completeness implies the first:
Consider given nested intervals $\{[a_k, b_k]\}$ (with $\lim(b_k - a_k) = 0$). We have for $k \leq \ell :\ a_k \leq a_\ell \leq b_k$, consequently $|a_\ell - a_k| \leq |b_k - a_k|$ and $\{|b_k - a_k|\}$ converges by assumption to 0. Therefore $\{a_k\}$ and similarly $\{b_k\}$ are Cauchy sequences. These have limits by assumption (2nd formulation of completeness), in fact the same limit r because of the triangle inequality. We have $a_k \leq r \leq b_k$ because both sequences are monotone. In other words, the number r lies in all the nested intervals, so that these are convergent as in the first formulation of completeness.
For the other direction we construct for a given Cauchy sequence $\{c_k\}$ convergent nested intervalls $\{[a_m, b_m]\}$, so that each interval contains infinitely many of the c_k. We need several steps:

1.) The whole Cauchy sequence $\{c_k\}$ lies in some interval $[a_1, b_1]$.
Proof. The definition gives: For $n = 2$ we have k_2 such that $k, \ell \geq k_2 \Rightarrow |c_k - c_\ell| \leq \frac{1}{2}$.
This implies for all ℓ: $a_1 := \min\{c_k; k \leq k_2\} - \frac{1}{2} \leq c_\ell \leq \max\{c_k; k \leq k_2\} + \frac{1}{2} =: b_1$.

2.) Consider $d := \frac{1}{2}(a_1 + b_1)$. Either we have in $[a_1, d]$ infinitely many elements of the Cauchy sequence, then put $a_2 = a_1, b_2 = d$. Otherwise put $a_2 = d, b_2 = b_1$. Repeat this procedure.

3.) Let $[a_m, b_m]$ the last already chosen interval that contains infinitely many elements of the Cauchy sequence. Consider $d = \frac{1}{2}(a_m + b_m)$. Either $[a_m, d]$ contains infinitely many elements of the Cauchy sequence, then put $a_{m+1} = a_m, b_{m+1} = d$. Otherwise put $a_{m+1} = d, b_{m+1} = b_m$, thus choosing the next interval.

4.) Since every interval is one half of the preceeding one we see that $\{[a_m, b_m]\}$ are convergent nested intervals and they have, by the first formulation of completeness, some limit r.

5.) We need to show that $\{c_k\}$ converges to r, or that $\{c_k - r\}$ converges to 0. Let $\epsilon > 0$ be given. First choose m so that $|b_m - a_m| \leq \epsilon/2$. Next, by definition of a Cauchy sequence, choose k_ϵ so that $k, \ell \geq k_\epsilon \Rightarrow |c_k - c_\ell| \leq \epsilon/2$. In the interval $[a_m, b_m]$ are infinitely many of the c_ℓ, pick one, $c_{\ell*}$, with $\ell* \geq k_\epsilon$. Of course $r \in [a_m, b_m]$. Finally we have with the triangle inequality $k \geq k_\epsilon \Rightarrow |c_k - r| \leq |c_k - c_{\ell*}| + |c_{\ell*} - r| \leq \epsilon/2 + |b_m - a_m| \leq \epsilon$.

Example. Let $\{a_k\}$ be some "geometrically majorized" sequence, that is we have for all k

$$(q1) \qquad |a_{k+2} - a_{k+1}| \leq q \cdot |a_{k+1} - a_k|, \; q < 1.$$

Claim. This sequence $\{a_k\}$ is Cauchy.
Note that we can allow $a_k \in \mathbb{R}^n$ or even in infinite dimensional complete spaces.

Proof. By applying the assumption k times we arrive at an estimate that also deserves the name "geometrically majorized":

$$(q2) \qquad |a_{k+2} - a_{k+1}| \leq q^k \cdot |a_1 - a_0|.$$

From $(q2)$ and the triangle inequality we conclude (assuming $\ell > k$):

$$|a_\ell - a_k| \leq \sum_{m=k}^{\ell-1} |a_{m+1} - a_m| = |a_1 - a_0| \cdot \sum_{m=k}^{\ell-1} q^{m-1} \; \leq \frac{|a_1 - a_0|}{1-q} \cdot q^{k-1}.$$

This majorization by a known null sequence is the Cauchy property of $\{a_k\}$ – so that a limit in $\mathbb{R}$ exists. This simple argument has a frequently quoted consequence, the

Contraction Lemma. Let f be a map with expansion bound $q < 1$, so that we have $|f(y) - f(x)| \leq q \cdot |y - x|$. Choose some a_0 in the domain of f. The previous claim shows: The definition $a_{k+1} := f(a_k), \; k \in \mathbb{N}$, gives a geometrically majorized Cauchy sequence. The limit $a := \lim_{k\to\infty} a_k$ together with the expansion bound q gives a (moreover unique)

Fixed Point $a = f(a)$.

Since the contraction lemma is a very effective tool for proving existence of solutions of nonlinear equations $a = f(a)$ it has also the more prestigious name *Banach Fixed Point Theorem.*

Example. $f(x) := 1/(2+x),\ x \in [0, \infty),\ |f'(x)| \leq 1/4$ satisfies the assumptions of the contraction lemma. Choose $a_0 = 0$. The recursive sequence $a_{k+1} := f(a_k), k \in \mathbb{N}$ runs through the finite values of the convergent continued fraction

$$0,\ \frac{1}{2},\ \frac{2}{5} = \frac{1}{2+\frac{1}{2}},\ \frac{5}{12} = \frac{1}{2+\frac{1}{2+1/2}},\ f(\frac{5}{12}),\ \ldots,\frac{1}{2+\frac{1}{2+1/2\ldots}}\ ,$$

which converges to the fixed point $a = f(a) = 1/(2+a)$ or $a = \sqrt{2} - 1$.

The third formulation of completeness goes back to Dedekind. It is based on the following

Definition. Consider a nonempty subset $\emptyset \neq A \subset \mathbb{R}$. A number s is called **smallest upper bound** or **least upper bound** or **Supremum** of A, notation: $s = \sup A$, if the following holds:

1.) s is an upper bound of A (i.e. $a \in A \Rightarrow a \leq s$).
2.) There is no smaller upper bound (i.e. $r < s \Rightarrow$ there exists $a \in A$ with $r < a$).

Third Formulation of the Axiom of Completeness

Every non-empty and bounded from above subset A of $\mathbb{R}$ has a smallest upper bound or least upper bound, also called supremum of A, and denoted $\sup A$.

Remark. The theorem of Bolzano-Weierstraß follows trivially. In the opposite direction, the proof of this third formulation is similar to the proof of this theorem. The supremum formulation of completeness is often considered as the most elegant one. However it requires better technical skills in using analysis arguments than the other two formulations.

Exercise in Argumentation. Assume completeness in its third formulation. Consider nested intervals $\{[a_k, b_k]\}$ with $\lim(b_k - a_k) = 0$. We show convergence, i.e. we find a number $c \in [a_k, b_k]$ for all k. The non-empty set $\{a_k\}$ which is bounded above by b_1 has by assumption a least upper bound c, $a_k \leq c$. Since each b_k is an upper bound of the set $\{a_k\}$, we conclude for all k: $c \leq b_k$, hence $c \in [a_k, b_k]$. Therefore c is limit of the nested intervals.

Application. Definition of arclength for expansion-bounded curves $p : [a, b] \to \mathbb{R}^n$.

Assumption: $|p'| \leq L$, hence $|p(y) - p(x)| \leq L \cdot |y - x|$ ("expansion-bounded").

Consider an arbitrary subdivision $a = x_0 < x_1 < \ldots < x_n = b$ of the interval $[a, b]$ of definition of p.

First we get a bound for all sums of secant-lengths which is independent of the chosen subdivision:

$$\sum_{k=1}^{N} |p(x_k) - p(x_{k-1})| \leq \sum_{k=1}^{N} L \cdot |x_k - x_{k-1}| = L \cdot |b - a|.$$

The sums of secant lengths are therefore a non-empty bounded set of real numbers and we can use the third completeness formulation **to define:**

$$\text{Length}\,(p\big|_{[a,b]}) := \sup\{\sum_{k=1}^{N} |p(x_k) - p(x_{k-1})|; a = x_0 < \ldots < x_N = b\}.$$

This intuitive definition does not lead to formulas for computation of arc lengths. We find these in the chapter on integration.

Summary on the Success of Completeness. The real numbers allowed us to solve the existence problems from the beginning of the chapter by just quoting the completeness axiom. By using uniform error bounds we were also able to extend knowlege on approximating functions (like "differentiability") to the limit function by just quoting the Archimedes Axiom (or trick). With the exception of the fact that we did not create an intuitive understanding of the real numbers in their entirety, all our arguments were rather simple. They consisted in quoting those two axioms. I cannot emphasize enough that this simplicity is the result of our usage of **uniform error estimates** which, in turn, were consequences of the monotonicity theorem. – The interplay between the real numbers and the notion of continuity, treated in the final chapter, can only be handled with new kinds of arguments, indirect arguments as we will see.

Complex Numbers, Functions and Differentiation

After having introduced the real numbers it is appropriate to review the definitions and proofs of the theory of differentiation. I want to use the complex numbers in this repetition so that things do not look completely familiar and we have to rethink our arguments. Between 1500-1800 mathematicians have used the "imaginary unit" $\mathbf{i}^2 = -1$. The name *imaginary number'* may indicate that they were not totally comfortable with this tool. This changed after 1800 when Gauss and others explained and visualized them as *two-dimensional numbers*. Now electrical engineering, quantum theory and large portions of mathematics cannot do without them.
The complex numbers will be defined such that the Euclidean plane is made into a complete number field. Multiplication is compatible with Euclidean geometry in that $|z \cdot w| = |z| \cdot |w|$ holds. The complex rational functions are shown to be differentiable with the same arguments as the real rational functions. We need different visualizations for the "two-dimensional" complex functions than were used for the "one-dimensional" real functions. There is a "mixed chain rule" for situations were real and complex differentiation play together as in

$$c : I \to \mathbb{C}, \quad f : \mathbb{C} \to \mathbb{C} \;\Rightarrow\; (f \circ c)' = (f' \circ c) \cdot c'.$$

We will prove for (not necessarily uniformly) differentiable functions a replacement of the (one-dimensional) monotonicity theorem, the (D⇒L)–Theorem , which says that a **D**erivative bound is a **L**ipschitz bound:

$$|f'| \le L \Rightarrow |f(w) - f(z)| \le L \cdot |w - z|.$$

The first goal is to make the Euclidean plane into a number field. Let $\mathbf{e}, \mathbf{i}$ be our preferred orthonormal basis. We identify the real numbers with the first coordinate axis, $\mathbf{e} \cdot \mathbb{R}$, but instead of $r \cdot \mathbf{e}$ we only write r. The additive group of the $\mathbb{R}$-vectorspace $\mathbb{C} := \mathrm{span}_{\mathbb{R}}\{\mathbf{e}, \mathbf{i}\}$ is the additive group of the number field to be constructed:
$(a \cdot \mathbf{e} + b \cdot \mathbf{i}) + (x \cdot \mathbf{e} + y \cdot \mathbf{i}) = (a + x) \cdot \mathbf{e} + (b + y) \cdot \mathbf{i}$.
Moreover, we know the multiplication $\mathbb{R} \times \mathbb{C} \to \mathbb{C}$ (rsp. $\mathbb{C} \times \mathbb{R} \to \mathbb{C}$) already, namely as the multiplication by scalars in the $\mathbb{R}$-vectorspace $\mathbb{C}$:
$(r \cdot \mathbf{e}) \cdot (x \cdot \mathbf{e} + y \cdot \mathbf{i}) = r \cdot (x \cdot \mathbf{e} + y \cdot \mathbf{i}) = rx \cdot \mathbf{e} + ry \cdot \mathbf{i} = (x \cdot \mathbf{e} + y \cdot \mathbf{i}) \cdot (r \cdot \mathbf{e})$.
In particular: $\mathbf{e} \cdot \mathbf{i} = \mathbf{i} = \mathbf{i} \cdot \mathbf{e}$.
For making $\mathbb{C}$ into a commutative ring, we are only missing *one* multiplication, the one with the famous history:

$$\mathbf{i} \cdot \mathbf{i} := \; -\mathbf{e}.$$

We write down the trivial extension to linear combinations to get the desired *multiplication formula*:

$$(a \cdot \mathbf{e} + b \cdot \mathbf{i}) \cdot (x \cdot \mathbf{e} + y \cdot \mathbf{i}) := (ax - by) \cdot \mathbf{e} + (ay + bx) \cdot \mathbf{i}\,.$$

This multiplication has, because of $(a\cdot\mathbf{e}+b\cdot\mathbf{i})\cdot(a\cdot\mathbf{e}-b\cdot\mathbf{i})=(a^2+b^2)\cdot\mathbf{e}$,

an inverse: $$(a\cdot\mathbf{e}+b\cdot\mathbf{i})^{-1}=(a\cdot\mathbf{e}-b\cdot\mathbf{i})/(a^2+b^2).$$

This makes $\mathbb{C}$, or the Euclidean Plane, into a commutative field.

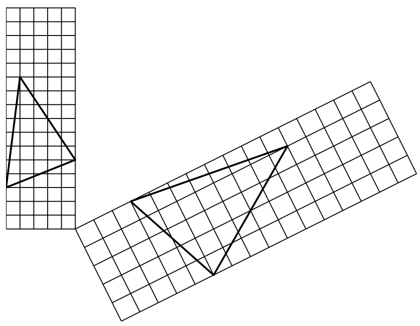
Compatibility of the Complex Numbers with the Euclidean Geometry

To get used to complex numbers as two-dimensional numbers it is important that also the multiplication is compatible with the geometry.
(i) Addition acts as translation, $\text{Trans}_a : z \to a+z$, with $a, z \in \mathbb{C}$.

Definition, for each $z = x + \mathbf{i}y$ its complex conjugate is: $\overline{z} := x - y\cdot\mathbf{i}$.
Definition of the Euclidean length of $z = x + y\cdot\mathbf{i}$: $|z| := \sqrt{x^2+y^2} = \sqrt{z\overline{z}}$.

(ii) **Theorem** (Compatibility of Euclidean lengths with multiplication)
For $c := a + b\cdot\mathbf{i}$ and $z := x + y\cdot\mathbf{i}$ holds:
1.) $\overline{c\cdot z} = \overline{z}\cdot\overline{c}$, $|c\cdot z| = |c|\cdot|z|$.
2.) For $c \neq 0$ the map $z \mapsto c\cdot z$ is a similarity transformation: Every triangle $\{z_1, z_2, z_3\}$ is mapped to the triangle $\{c\cdot z_1, c\cdot z_2, c\cdot z_3\}$. The image edgelengths are $|c\cdot z_j - c\cdot z_k| = |c|\cdot|z_j - z_k|$, i.e. all are changed by the same factor $|c|$. If $c \neq 1$ then 0 is the only fixed point. The map $z \mapsto c\cdot z$ is, therefore, a rotation around 0 composed with a scaling by the factor $|c|$.
Proof.
1.) $\overline{c\cdot z} = (ax-by) - (ay+bx)\cdot\mathbf{i} = (a - b\cdot\mathbf{i})\cdot(x - y\cdot\mathbf{i}) = \overline{c}\cdot\overline{z}$.
$|c\cdot z|^2 = (c\cdot z)\cdot\overline{(c\cdot z)} = c\cdot(z\cdot\overline{z})\cdot\overline{c} = |c|^2\cdot|z|^2$.

2.) $|(c\cdot z_j) - (c\cdot z_k)| = |c(z_j - z_k)| =$
$|c|\cdot|z_j - z_k|$,
$j, k \in \{1, 2, 3\}$.



Multiplication by a complex number $c \neq 0$ is an angle preserving map, $z \mapsto cz$.

(iii) **Theorem** (Compatibility of inversion Inv: $z \mapsto 1/z$ with straight lines and circles)
The inversion map $z \mapsto 1/z$ acts as follows:
1. Lines through $0 \mapsto$ lines through 0, (namely $r \to r\cdot(c/|c|)$ goes to $r \to 1/r\cdot(\overline{c}/|c|)$,)
2. Lines not through $0 \mapsto$ circles through 0, circles through $0 \mapsto$ lines not through 0,
3. Circles not through $0 \mapsto$ circles not through 0.

Proof. Circles with midpoint m and radius r we describe as $\{z\ ;\ |z-m|=r\}$ or better $K_r(m) := \{z\ ;\ (z-m)(\bar{z}-\bar{m}) = r^2\}$. If $m\bar{m} = r^2$ holds then $0 \in K_r(m)$.

The image of such a circle under the inversion map Inv: $z \mapsto w = 1/z$ is

$\mathrm{Inv}(K_r(m)) = \{w = 1/z\ ;\ (z-m)(\bar{z}-\bar{m}) - r^2 = 0 = (1-mw)(1-\bar{m}\bar{w}) - r^2 w\bar{w}\ \}$,

which is, if $m\bar{m} = r^2$, the following line not passing through 0:

$\mathrm{Inv}(K_r(m)) = \{w\ ;\ 1 = mw + \bar{m}\bar{w}\}$,

and is, if $m\bar{m} \neq r^2$, $M := \bar{m}/(m\bar{m}-r^2)$, $R^2 := M\bar{M} - 1/(m\bar{m}-r^2)$, the circle:

$\mathrm{Inv}(K_r(m)) = \{w\ ;\ (w-M)(\bar{w}-\bar{M}) = R^2\ \}$.

The images of straight lines are easily determined. Note $\mathrm{Inv} \circ \mathrm{Inv} = \mathrm{id}$.

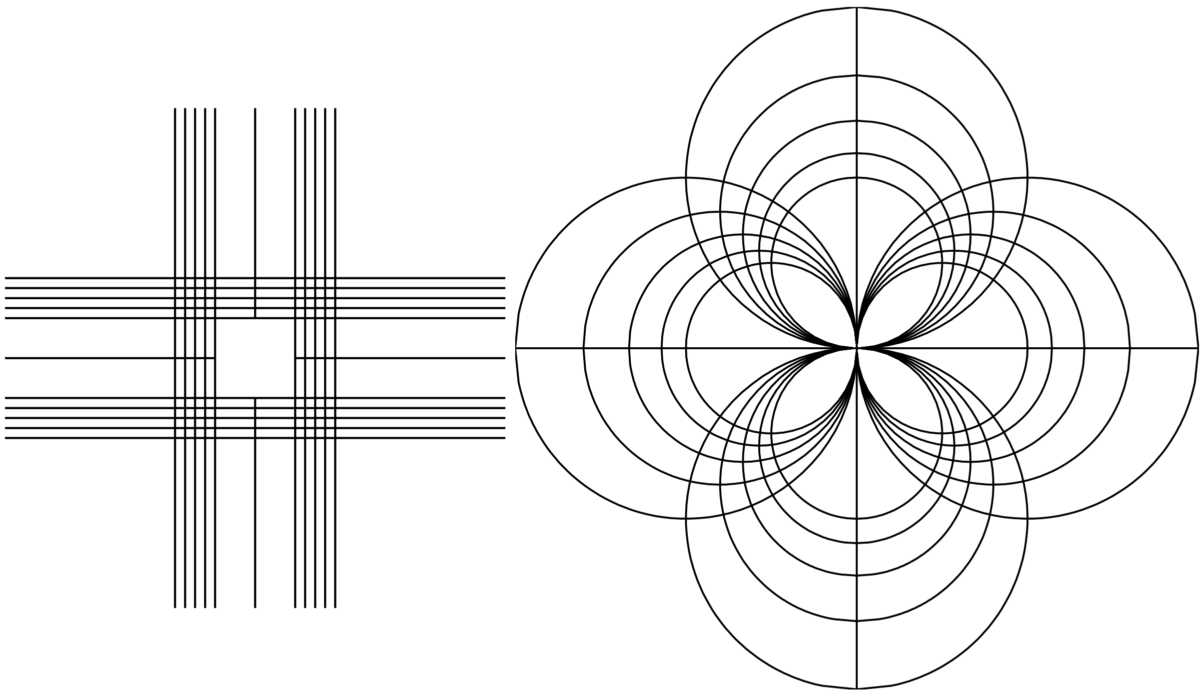


Image under $z \mapsto 1/z$ of four half-strips parallel to the axes (and not drawn to ∞).
The parallel lines are mapped to circles, which are tangent to the axes at 0.
The middle square in the domain is mapped to the outside of the figure in the range.

Extension. This preservation of circles carries over to maps (called Möbius transformations) $z \mapsto (az+b)/(cz+d) \neq const$, which can be written if $c \neq 0$, $ad-bc \neq 0$ as

$$z \mapsto \frac{(az+b)}{(cz+d)} = \frac{a}{c} - \frac{(ad-bc)/c^2}{(z+d/c)},$$

which is a composition of two translations (by a/c and d/c), one rotation with scaling (from the factor $-(ad-bc)/c^2$) and one inversion. All these are maps which preserve circles.

Theorem: **Completeness of $\mathbb{C}$**

The complex numbers are complete,

that is, every Cauchy sequence $\{z_k\}$, $z_k \in \mathbb{C}$, converges.

Proof. $\{z_k = x_k + y_k \cdot \mathbf{i}\}$ is a complex Cauchy sequence if and only if $\{x_k\}$ and $\{y_k\}$ are two real Cauchy sequences, because of

$$|z_\ell - z_k|^2 = |x_k - x_\ell|^2 + |y_k - y_\ell|^2.$$

We are pleased that no new efforts were needed, the work with the reals $\mathbb{R}$ was sufficient. Approximation plus quoting completeness will also in the complex case prove to be an efficient method to construct new interesting functions.

Rational Functions.

The necessary computations can hardly be distinguished from the corresponding real ones, building visual intuitions will require more work. First we repeat for polynomials what we did in the real case, starting with

$$|a|, |z| \leq R \Rightarrow |z^n - a^n - n \cdot a^{n-1} \cdot (z-a)| \leq \frac{n \cdot (n-1)}{2} \cdot R^{n-2} \cdot |z-a|^2.$$

In Words: The power functions $P : z \mapsto P(z) := z^n$ can be approximated by the linear functions $z \to a^n + n \cdot a^{n-1} \cdot (z-a)$ with uniform quadratic errors. Therefore we also call in the complex case the polynomial $P' : z \mapsto P'(z) := nz^{n-1}$ the "derivative" of the power function P.

In the same way as in $\mathbb{R}$ these linear approximations with uniform quadratic errors extend to polynomials (note that the error constant K is explicitly computed from the coefficients and the "interval", better a circle):

$$|a|, |z| \leq R,\ P(z) := \sum_{k=0}^{n} a_k z^k,\ \ P'(z) = \sum_{k=1}^{n} a_k k z^{k-1},\ \ K := \sum_{k=2}^{n} |a_k| \frac{k(k-1)}{2} R^{k-2}$$
$$\Rightarrow \qquad |P(z) - P(a) - P'(a)(z-a)| \leq K \cdot |z-a|^2.$$

The function $f : z \mapsto f(z) := 1/z$ is at $a \neq 0$ differentiable with $f'a) = -1/a^2$ and also uniformly differentiable, if we keep some distance away from 0, e.g. $|a| \geq r > 0$:

$$0 < r \leq |a|,\ \ |z-a| \leq r/2 \Rightarrow |\frac{1}{z} - \frac{1}{a} + \frac{z-a}{a^2}| \leq \frac{2}{|a|^2}|z-a|^2 \leq \frac{2}{r^2}|z-a|^2.$$

Before we try to get an intuitive understanding of the derivatives of these simple functions we need to ask: what is a promising way to visualize complex functions? In the real case

one almost always chooses to draw the graph of f, but in the complex case this graph $\{(z, f(z)) \in \mathbb{C} \times \mathbb{C}\}$ cannot be used because $\mathbb{C} \times \mathbb{C}$ is 4-dimensional. What can one do?
In many areas of mathematics the words “function” and “map, mapping” are used interchangeably. In the one-dimensional real case this did not become popular, because 1-dimensional images are not visually interesting. However, for complex functions one obtains useful visualizations if one considers them as maps: Decorate the domain of the function with some texture or draw some grid into it. Then draw the image of the domain with its decoration. For each point we can “see” the image point with the help of the decoration. For the same reason textures are used in computer graphics. Also, we can use and compare maps of the earth which are made with quite different methods.
In our first example we use a square grid in the domain of $f(z) = z^2$ as the texture. It will always be the case that the strongest local deformation of the domain occurs near zeros of the derivative. Grid squares farther away are mapped to only slightly deformed squares, and the deviation from true squares gets less under subdivision of the grid. The grid curves in the image are parabolas, as follows from

$$z = x + i \cdot b \to u(x) + i \cdot v(x) := (x^2 - b^2) + i \cdot 2b \cdot x, \quad u = v^2/(4b^2) - b^2.$$

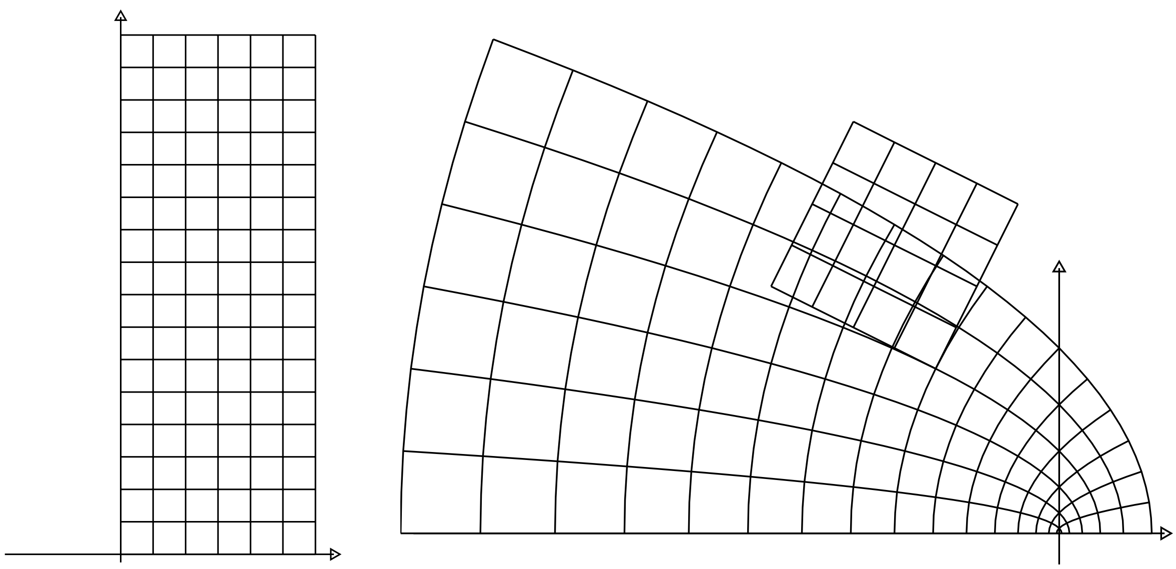


Image under $z \to z^2$, together with a linear approximation
The sizes of coordinate meshes in the image at $f(a)$ are given by $|f'(a)|$, $f'(0) = 0$.

We show the strong deformation of the parameter grid near zeros of the derivative again; further away from these points the square grid meshes stay approximately square. The second example is $z \to z - z^3/3$. The zeros of the derivative are at ± 1.

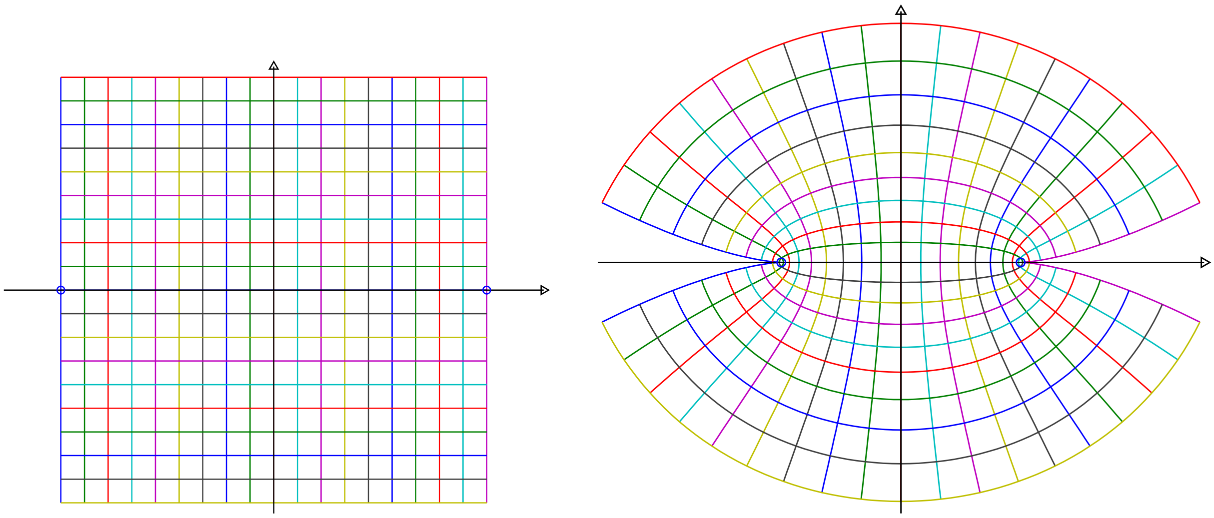


Image under $z \to z - z^3/3$. The zeros of f' at ± 1 are marked in domain and range.

It is instructive, to map also other decorations of the domain, for example polar coordinate grids. We postpone this, until some more theory suggests to us, what features of the images to concentrate on.
As in the real case we need *Differentiation rules*, after having done the simplest examples directly from the definition. The differentiation rules look the same as in the real case and their proofs can be taken over without changes. In addition we have a new conceptional problem: the grid curves in our images are curves in the Euclidean plane and we know already how to differentiate them. The comparison of curves in the domain grid with their images in the image grid will be used to better understand the complex function (or map). Therefore we need a rule, which uses the derivative of the *complex* function to transform curve tangents in the domain to curve tangents in the range. I call this rule *mixed chain rule* because it applies to situations where both, real and complex differentiability, occur. Because of this mixture of concepts I show the proof, although the details are not much different from other chain rule proofs. First the reformulation of the already known rules:

Theorem. The Differentiation Rules.
Let f, g be complex functions, about which we assume: there are other complex functions f', g' and constants K_f, K_g such that the following assumptions hold (one may choose the constants to be dependent on a or hold uniformly as long as one makes the same choice in assumption and claim):

Assumptions:
$$|z-a| \le r \Rightarrow |f(z) - f(a) - f'(a)\cdot(z-a)| \le K_f \cdot |z-a|^2,$$
$$|z-b| \le r \Rightarrow |g(z) - g(b) - g'(b)\cdot(z-b)| \le K_g \cdot |z-b|^2.$$

Then we have the **Differentiation Rules:**
$$(\alpha f + \beta g)' = \alpha \cdot f' + \beta \cdot g' \qquad (a = b),$$
$$(f \cdot g)' = f' \cdot g + f \cdot g' \qquad (a = b),$$
$$(f \circ g)' = (f' \circ g) \cdot g' \qquad (a = g(b)).$$

These rules imply that the complex rational functions $z \to P(z)/Q(z)$ are, except at zeros of denominators, complex differentiable – at this point formally as described by the inequlities of the definition, a visualization will follow below. The proofs show that the functions F formed from f and g, namely $\alpha \cdot f + \beta \cdot g$, $f \cdot g$, $f \circ g$, are approximated by the linear functions $z \mapsto F(a) + F'(a) \cdot (z - a)$ ("tangents") with quadratic errors as was assumed to be the case for f and g. Later results show – in strong contrast to the real case – that complex differentiable functions which are approximated by their tangents with *coarser errors* do not exist!

We turn to the mixed chain rule and repeat the question: Given a (real) differentiable curve $c : I \mapsto \mathbb{C}$ and a complex differentiable map (function) f how can the (real) derivative of the image curve $\hat{c} = f \circ c : I \to \mathbb{C}$ be computed from the derivatives of f and c? Recall for the differentiable curve $c : I \to \mathbb{C}$ that there is another curve (or velocity or derivative) $c' : I \to \mathbb{C}$ and constants r_c, K_c (usually dependent on t_0) such that

$$|t - t_0| \leq r_c \Rightarrow |c(t) - c(t_0) - c'(t_0)(t - t_0)| \leq K_c \cdot |t - t_0|^2.$$

Theorem. **Mixed Chain Rule**

Let c be a differentiable curve as above, put $a := c(t_0)$ and $\hat{c} := f \circ c$, where f is complex differentiable at a, that is:

Assumption: $|z - a| \leq r_f \Rightarrow |f(z) - f(a) - f'(a) \cdot (z - a)| \leq K_f \cdot |z - a|^2.$

Claim: $(f \circ c)'(t_0) = f'(c(t_0)) \cdot c'(t_0),$

which means: the linear approximation $z \to f(a) + f'(a) \cdot (z - a)$ of f maps the tangent $t \to c(t_0) + c'(t_0) \cdot (t - t_0)$ of c to the tangent $t \to \hat{c}(t_0) + \hat{c}'(t_0) \cdot (t - t_0)$ of $\hat{c}$.
And expressed as inequality:

$$|t - t_0| \leq r_{\hat{c}} \Rightarrow |f \circ c(t) - f \circ c(t_0) - f'(c(t_0)) \cdot c'(t_0) \cdot (t - t_0)| \leq K_{\hat{c}} \cdot |t - t_0|^2.$$

Proof. The argument follows the familiar pattern, except that we use two different differentiability concepts. The triangle inequality and the differentiability assumptions imply Lipschitz bounds ℓ, L, for c: $\ell := |c'(t_0)| + K_c \cdot r_c$ in the interval $[t_0 - r_c, t_0 + r_c]$
and for f: $L := |f'(a)| + K_f \cdot r_f$ in the disk $\{z; |z - a| \leq r_f\}$.
First we decrease r_c to $r_1 := \min(r_c,\ r_f/\ell)$, so that the r_1-interval around t_0 is mapped by c into the r_f-disk around $a = c(t_0)$, where we have our assumptions on f:

$$|t - t_0| \leq r_1 \Rightarrow |c(t) - c(t_0)| \leq \ell \cdot r_1 \leq r_f.$$

(This is an important step, which should not be overlooked in chain rule proofs.)
On the r_1-interval we have the Lipschitz bound $L \cdot l$ for $\hat{c}$, which is only slightly bigger than the product $|f'(a)| \cdot |c'(t_0)|$, in agreement with the chain rule to be proved. In this proof the main step consists in inserting the values of c (note $c(t_0) = a$) into the differentiability inequality for f:

$$|t - t_0| \leq r_1 \Rightarrow |c(t) - c(t_0)| \leq r_f \Rightarrow |f(c(t)) - f(a) - f'(a) \cdot (c(t) - c(t_0))| \leq K_f \cdot |c(t) - c(t_0)|^2.$$

Here the not yet final term $f'(a) \cdot (c(t) - c(t_0))$ is approximately the desired final term $f'(a) \cdot c'(t_0) \cdot (t - t_0)$ with an error $\leq |f'(a)| \cdot K_c \cdot |t - t_0|^2$.

The triangle inequality gives
(Recall: $|A - B| \leq F_1, |B - C| \leq F_2 \Rightarrow |A - C| = |(A - B) - (C - B)| \leq F_1 + F_2$.)
$|t-t_0| \leq r_1 \Rightarrow |f(c(t))-f(a)-f'(a)\cdot c'(t_0)\cdot(t-t_0)| \leq K_f \cdot|c(t)-c(t_0)|^2+|f'(a)|\cdot K_c\cdot|t-t_0|^2$.
To unite the errors recall $K_f \cdot |c(t) - c(t_0)|^2 \leq K_f \cdot l^2 \cdot |t - t_0|^2$.
Now define $K_{\hat{c}} := K_f \cdot \ell^2 + |f'(a)| \cdot K_c$, and $r_{\hat{c}} := r_1$ and we have achieved our goal:
With the constants $r_{\hat{c}}, K_{\hat{c}}$ we have:

$$|t - t_0| \leq r_{\hat{c}} \Rightarrow |\hat{c}(t) - \hat{c}(t_0) - \hat{c}'(t_0) \cdot (t - t_0)| =$$
$$= |f(c(t)) - f(c(t_0)) - f'(c(t_0)) \cdot c'(t_0) \cdot (t - t_0)| \leq K_{\hat{c}} \cdot |t - t_0|^2.$$

This mixed chain rule implies an essential property of complex differentiable functions (maps) which has immediate implications about how our image figures look like and which is also responsible for the surprisingly far reaching consequences of complex differentiability (as opposed to real differentiability).

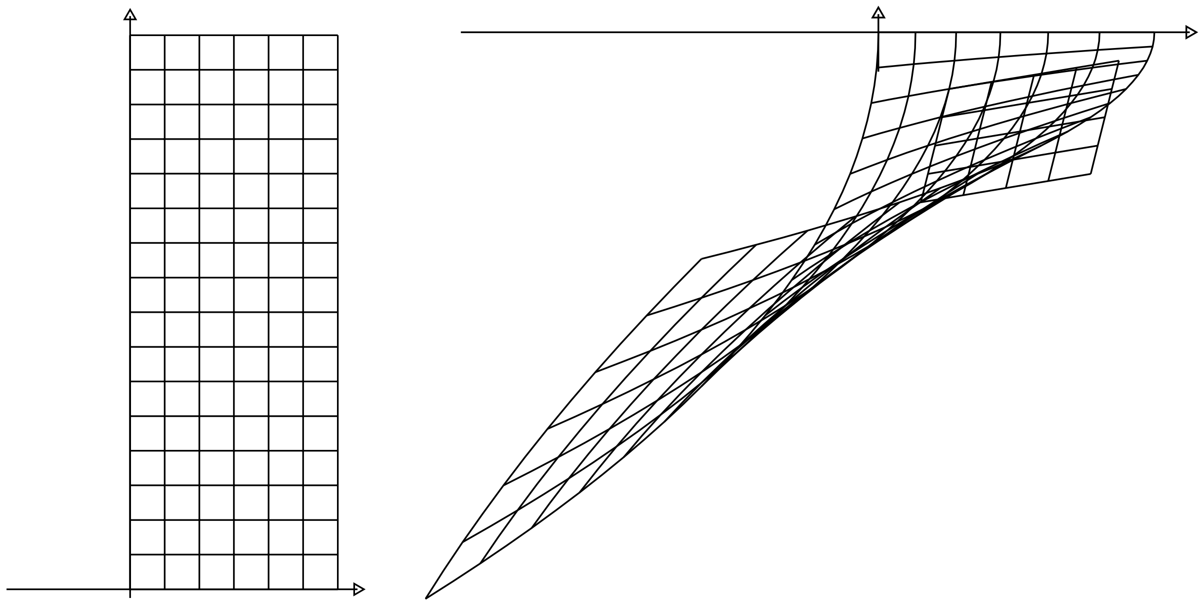
Theorem on the Preservation of Angles (Conformality) of complex differentiable maps.

Let f be complex differentiable. The map f is at all points a where $f'(a) \neq 0$ angle preserving.

Proof. The linear approximations $z \mapsto f(a) + f'(a) \cdot (z - a)$ are (see the first result in this chapter) *rotations composed with scaling* and therefore angle preserving. The mixed chain rule implies: Let two differentiable curves c_1, c_2 intersect at a, then the angle between their tangents is the same as the angle between the tangents of the image curves $f \circ c_1$, $f \circ c_2$ at $f(a)$; this property is called *angle preservation* or *conformality* of the map f.

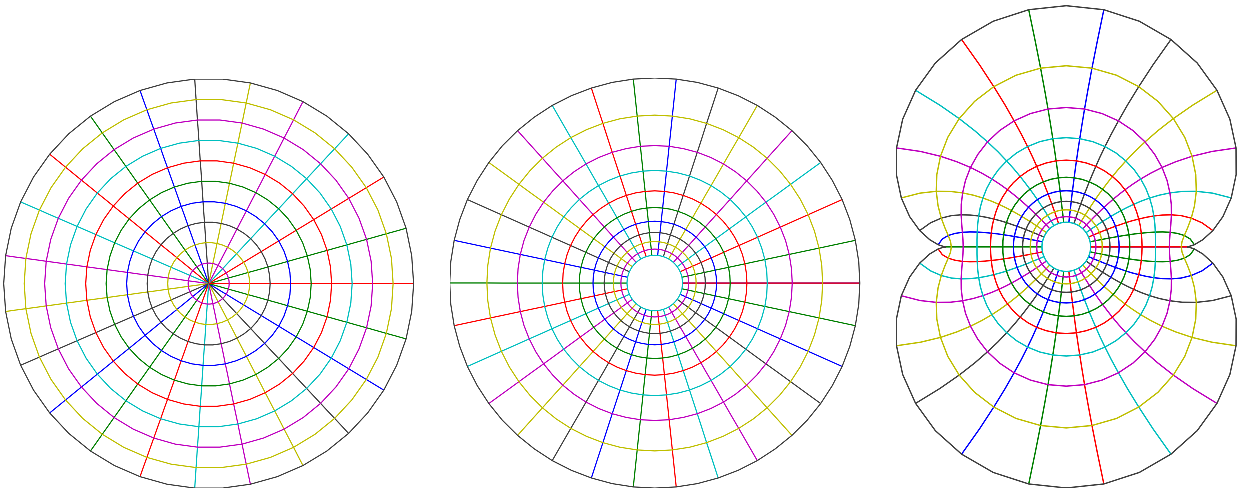
Visualization of Complex Functions. Because conformality is so important, we use for the visualization of complex maps f (as in all previous illustrations) orthogonal grid curves in the domain. It is easy to see in the image grid whether the image curves are also orthogonal or not. It is even better to start from square grids (or at least infinitesimally square grids) in the domain. Conformality implies that also their diagonals intersect orthogonally in domain and range. Such grids are perceived by our eyes as made up of slightly bent squares, and rather small deviations from this "squareness" are easily noticed. (View the image below of $z \to z + a \cdot \bar{z}^2$ with small a.)
The proposed visualization of complex functions by the pair of domain and image grids (chosen infinitesimally square) shows very clearly whether or not a function is complex differentiable.
After having agreed on how to visualize functions we ask: "How can we add a visualization of the derivative to our figures?" Derivatives at a point are defined as linear approximations $z \to f(a) + f'(a) \cdot (z - a)$ and are therefore themselves maps. We represent them in the same way, except that we only take a small square in the domain and superimpose its image onto the image grid of f. This image is again a square if f is conformal at a, see the visualization of $z \mapsto z^2$ three pages above.

Summary of what we see in the following picture of a non-conformal map: The two grid curves of the image grid of f through $f(a)$ have as their tangents two grid lines of the superimposed image parallelgram (not a square!) obtained from the linear approximation of f (at the point a).

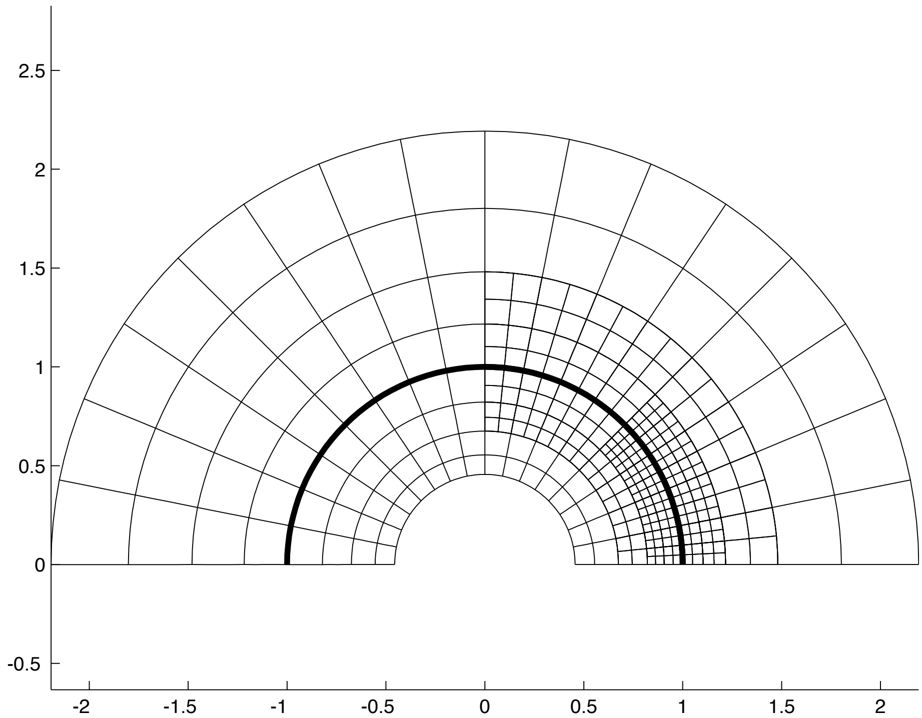


Obviously not a conformal image under $z \to \bar{z} + z^2/4$.
The image grid is not "infinitesimally square" and the image has a "fold edge".

We know now what to concentrate on when viewing visualizations of complex functions. Therefore we understand why standard polar coordinates are not well suited to represent conformal maps: Although the parameter lines (radial rays and concentric circles) are orthogonal, the shape of the (curved) parameter rectangles changes drastically. Their ratio of edgelengths is neither bounded from below, nor from above. Therefore it is difficult to judge whether one is viewing a conformal map or not. Is it possible to make better polar coordinates? We want to keep the radial rays but want to adjust the distances between the circles. Begin with a pair of circles with radii 1 and R. A conformal stretching from 0 by the factor R maps the initial circles to circles of radii R, R^2. Each radial line is mapped onto itsself. One can repeat this process. The next pictures show the result: One can indeed make "infinitesimally square", or conformal, polar coordinates. They are particularly well compatible with inversion, the map $z \to 1/z$ sends such a conformal polar grid onto itsself, exchanging the interior and exterior of the unit circle. Also the behavior of $z \to z^2$ in disks around the zero $z = 0$ of the derivative is clearly represented: Circles around 0 of radius $= r$ are mapped to circles around 0 of radius $= r^2$, radial rays are mapped to radial rays in such a way that the angle of two rays at 0 is *doubled*, since points $c, |c| = 1$ on the unit circle double their arc length distance from 1 under multiplication by c. Note how the little squares double their edgelengths under the map.



Ordinary and conformal ("localy square") polar coordinate grids. Image of the locally square grid under $z \mapsto z(1 - z^2/3)$ is locally square. The zeros of f' dominate the image.



Conformal polar coordinates serve to visualize $z \mapsto z^2$ and $z \mapsto 1/\bar{z}$.

The pieces of the figure covered by the smallest squares are sent by the map $z \mapsto z^2$ to the pieces covered by the middles-sized squares, and these in turn are sent to twice as large (curved) squares again. This shows that the angles of the rays at zero are doubled. – The inversion $z \mapsto 1/z$ interchanges the interior and the exterior of the unit disk, and also the upper and lower halfplane. But the conformal polar grid as a whole is mapped to itsself. See p.72

Before we can start with the construction of new functions, we need a replacements for the monotonicity theorem which can't even be formulated in the complex case. The theorem that derivative bounds are Lipschitz bounds – quoted as the (D⇒L)–Theorem – is not quite as intuitive to use but serves the same purpose. Since completeness has been introduced, we prove this theorm under weaker assumptions than the earlier monotonicity theorem.

Derivative Bounds are Lipschitz Bounds for Complex Functions.
Quoted as (D⇒L)–Theorem

Assumption. Let f be complex differentiable along the segment between z and w and satisfy there $|f'| \leq L$. Since we plan to use completeness we no longer assume uniform tangent approximations but only: For each a between z and w there exist constants $r_a > 0$, $K_a > 0$ (which are explicitly allowed to depend on a) so that:

$$|c-a| \leq r_a \Rightarrow |f(c) - f(a) - f'(a) \cdot (c-a)| \leq K_a \cdot |c-a|^2.$$

Claim. $$|f'| \leq L \Rightarrow |f(z) - f(w)| \leq L \cdot |z-w|.$$

Proof. Because of the non-uniform assumptions we need an indirect argument. Note that later – when other "small" errors than the presently used quadratic errors are considered – this same argument will work.
Assume the claim were wrong, i.e. we had some z, w such that

$$\begin{array}{lcll} |f(z)-f(w)| & > & L \cdot |z-w|, & \text{and hence for some suitable } n \\ |f(z)-f(w)| & \geq & (L+1/n) \cdot |z-w|. & \end{array}$$

Put $z_0 = z$, $w_0 = w$ and consider the midpoint $m := \frac{1}{2}(z_0 + w_0)$. Then at least one of the following two inequalities holds:
(*) $\quad |f(z_0) - f(m)| \geq (L+1/n) \cdot |z_0 - m|, \qquad |f(m) - f(w_0)| \geq (L+1/n) \cdot |m - w_0|,$
since otherwise the triangle inequality would give

$$|f(z_0) - f(w_0)| \leq |f(z_0) - f(m)| + |f(m) - f(w_0)| < (L+1/n) \cdot (|z_0 - w_0|),$$

contradicting the initial assumption.
If the first inequality (*) holds put $z_1 = z_0$, $w_1 = m$, otherwise put $z_1 = m$, $w_1 = w_0$.
This procedure is repeated. Assume $z_0, \ldots, z_k$, $w_0, \ldots, w_k$ are already defined such that

$$|f(z_k) - f(w_k)| \geq (L+1/n) \cdot |z_k - w_k|, \quad |z_k - w_k| = 2^{-k} \cdot |z_0 - w_0|.$$

Put again $m := \frac{1}{2}(z_k + w_k)$. In at least one of the two halfs we have the inequality with the same factor $(L+1/n)$, therefore we can define z_{k+1}, w_{k+1} such that

$$|f(z_{k+1} - f(w_{k+1})| \geq (L+1/n) \cdot |z_{k+1} - w_{k+1}|, \quad |z_{k+1} - w_{k+1}| = |z_k - w_k|/2.$$

Therefore $\{z_k\}$ and $\{w_k\}$ are geometrically majorized Cauchy sequences, because of completeness they converge, moreover to the *same* limit c. The complex number c is by construction for each k on the segment between z_k and w_k ($c = z_k$ or $c = w_k$ are allowed).
Therefore we have **at least one** of the two inequalities

(+) $|f(z_k) - f(c)| \geq (L+1/n) \cdot |z_k - c| > 0$ or $|f(w_k) - f(c)| \geq (L+1/n) \cdot |w_k - c| > 0.$

These inequalities lead to a contradiction, if we use the differentiability assumption on f at c (a key point to make the argument work):
Choose k large enough so that $|z_k - w_k| < \min(r_c, 1/(2nK_c))$ holds. Then we have for all z on the segment between z_k and w_k that $|z - c| \leq r_c$. From this we conclude:

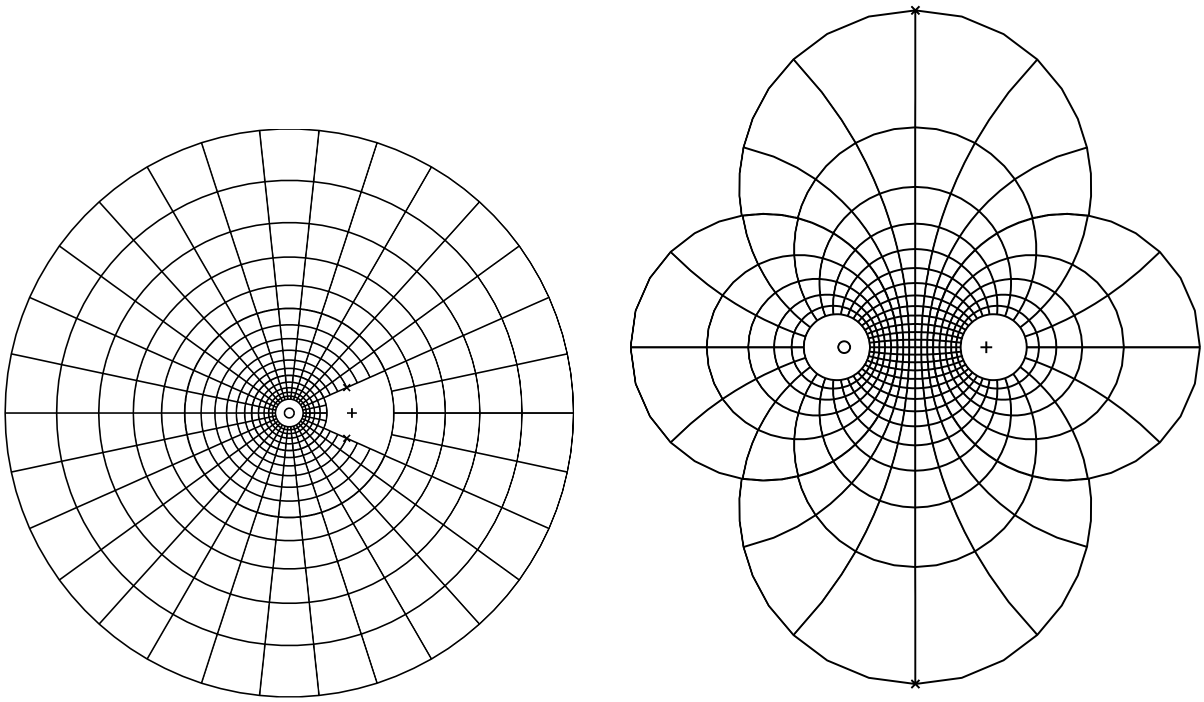
$$|f(z) - f(c) - f'(c) \cdot (z - c)| \leq K_c \cdot |z - c|^2.$$

Finally use the bound $|f'(c)| \leq L$ and $|z - c| \leq 1/(2nK_c)$ to get:

$$|f(z) - f(c)| \leq (L + 1/2n) \cdot |z - c|,$$

contradicting one of the two inequalities (+) derived above for $z = z_k$ or for $z = w_k$.

Remark. The structure of this proof: "With the help of completeness an indirect argument allows to localize a point, where a contradiction to the assumptions exists", can be found in many analysis proofs. Other famous examples are the boundedness theorem for continuous functions or the equivalence theorem of different continuity definitions or, later, the Cauchy integral theorem. Note, that for this argument to work, the precise specification of a "small error" (in the continuity and differentiability definitions) is irrelevant!



Using conformal polar coordinates with the map $f(z) := (z + 1)/(z - 1)$. This map sends straight lines and circles to straight lines and circles. Therefore one may extrapolate what happens outside the shown decoration of domain and range. Because of $f(f(z)) = z$ one may view each of the figures as domain, the other as image. f interchanges 0 and -1 (marked o), maps the interior (resp. exterior) of the unit circle to the left (resp. right) halfplane, and vice versa. The hole around $+1$ (marked $+$) in one figure is mapped onto the exterior of the other figure. This suggests to talk about a "point ∞" which gets interchanged with $+1$. The right figure shows the image of circular neighborhoods of 0 rsp. ∞ from the left figure mapped to circular neighborhoods of finite points.

Complex Power Series

The completeness of the complex numbers and the (D⇒L)–Theorem allow to define limit functions of complex valued function sequences. After introducing completeness we constructed only three (particularly important) functions. With the tool of power series we obtain a large class of limit functions which share many properties with the polynomials. Their treatment also practises the simplest and most important majorization technique, namely comparisons with the geometric series. Power series are an important class of functions, every complex differentiable function (in particular all functions that have individual names) can be represented as power series.

Definition. For a given sequence of numbers $a_k \in \mathbb{C}$ consider the

sequence of polynomials: $$P_n(z) := \sum_{k=0}^{n} a_k \cdot z^k.$$

Both, the sequence of polynomials $\{z \mapsto P_n(z)\}$ and also (in case of convergence) their limit function $z \mapsto P_\infty(z)$ are called *Power Series*; usually they are written as $\sum_{k=0}^{\infty} a_k z^k$.

Theorem: Majorisation by a point of covergence.
Assumption. For some $z_0 \neq 0$ the sequence $\{P_n(z_0)\}$ converges. Put $r := |z_0|$.
This convergence implies that $\{b_k := P_k(z_0) - P_{k-1}(z_0) = a_k z_0^k\}$ is a null sequence, hence bounded (recall: choose $\epsilon := 1$ then find n_1 such that $m, n \geq n_1 \Rightarrow |b_m - b_n| \leq 1$; then we have the bound $|b_k| \leq \max_{j=1\ldots n_1}(|b_j| + 1)$).
Therefore a **weaker assumption** is popular:
There exist $M, r > 0$ such that $|a_k| \cdot r^k \leq M$ for all $k \in \mathbb{N}$.
Claim. Let $0 < q < 1$ and $|z| \leq q \cdot r$. Then the sequence $\{P_n(z)\}$ is q-geometrically majorized and thus for all z with $|z| < |z_0|$ convergent.

Proof. $|P_n(z) - P_{n-1}(z)| = |a_n \cdot z^n| \leq |a_n r^n q^n| \leq M \cdot q^n$, so that $\{P_n(z)\}$ is Cauchy:

$$|P_{n+m}(z) - P_n(z)| \leq \frac{M}{1-q} \cdot q^{n+1}.$$

Contraposition.
Should $\{P_n(z_1)\}$ *not* converge, then $\{P_n(z)\}$ does *not* converge for any z with $|z| > |z_1|$.

This result allows to almost completely describe the
Domain of Convergence of Power Series:
1.) There are power series, which do not converge for any $z \neq 0$, example $a_k := k^k$.
2.) There are power series, which converge for all $z \in \mathbb{C}$.
Examples: the already known power series for exp, sin, cos converge for all $z \in \mathbb{R}$, and the theorem implies that they converge for all $z \in \mathbb{C}$.

3.) For all other power series there exists $R > 0$, so that for $|z| < R$ the sequence $\{P_n(z)\}$ converges and for $|z| > R$ diverges. (For $|z| = R$ we do not make a general statement.) This number R is called **radius of convergence** of the power series. Clearly:

$$R = \sup\{|z|; P_n(z) \text{ converges}\}.$$

Please note right from the start: In the open disk $|z| < R$, although one does have convergence, one does not have any *quantitative control;* however, in every smaller disk $|z| \le q \cdot R$, $(0 < q < 1)$ majorisation by the geometric series succeeds.

Before we do further work I show two tricks which will be our main tools for handling power series. The first one is used to deal with differentiated power series. The second trick is used if the above first result cannot be applied because there is no convergence point on the boundary of the disk of convergence (e.g. the derivative of the geometric series). For the first trick we differentiate the summation formula of the *finite* geometric series and get the desired n-independent bounds by dropping the negative terms:

Assumption: $0 \le x < 1$,

Claim (trick 1, convenient majorisations)**:**

$$\sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x} \le \frac{1}{1-x},$$

$$\left(\sum_{k=0}^{n} x^k\right)' = \sum_{k=1}^{n} kx^{k-1} = \frac{-(n+1)x^n}{1-x} + \frac{1-x^{n+1}}{(1-x)^2} \le \frac{1}{(1-x)^2},$$

$$\left(\sum_{k=0}^{n} x^k\right)'' = \sum_{k=2}^{n} k(k-1)x^{k-2} = \frac{-(n+1)nx^{n-1}}{1-x} - \frac{2(n+1)x^n}{(1-x)^2} + 2\frac{1-x^{n+1}}{(1-x)^3} \le \frac{2}{(1-x)^3}.$$

Assumption:

Let the radius of convergence of the power series $\{\sum_{k=0}^{n} a_k \cdot z^k\}$ be $R \in (0,\ \infty)$.

Claim (trick 2, exploit $q < \sqrt{q} < 1$) **:**

Since $\sqrt{q} \cdot R$ is a convergence point, there exists $M > 0$ such that $|a_k|(\sqrt{q}R)^k \le M$. Therefore $|z| \le q \cdot R \Rightarrow |a_k \cdot z^k| \le M \cdot (\sqrt{q})^k$, $|a_k \cdot kz^{k-1}| \le \frac{M}{qR} \cdot k(\sqrt{q})^k$, etc, which gives with trick 1 convenient bounds independent of n:

$$\sum_{k=0}^{n} |a_k \cdot z^k| \le M \sum_{k=0}^{n} (\sqrt{q})^k \le \frac{M}{1-\sqrt{q}},$$

$$\sum_{k=1}^{n} |a_k k \cdot z^{k-1}| \le \frac{M}{\sqrt{q}R} \sum_{k=0}^{n} k(\sqrt{q})^{k-1} \le \frac{M}{\sqrt{q}R} \frac{1}{(1-\sqrt{q})^2},$$

$$\sum_{k=2}^{n} |a_k k(k-1) \cdot z^{k-2}| \le \frac{M}{qR^2} \frac{2}{(1-\sqrt{q})^3}.$$

One can determine the radius of convergence from the coefficients a_k of the power series by two famous criteria. I will not use them much but their proofs demonstrate the importance of trick 2 above: if one has $|z| \leq q \cdot R$ (R the radius of convergence) one might hope to, but cannot, use majorization by a geometric series with the same q. One has to be content with a slightly larger factor, like $\sqrt{q},\ q < \sqrt{q} < 1$ as in trick 2.
The first, and simpler, of the two criteria gives only a *sufficient* condition:

Quotient Criterium:
If $R := \lim_{k\to\infty} |a_k/a_{k+1}|$ exists, then R is the radius of convergence.

Example: This provides a simple test for the convergence of the Taylor series of exp.

Proof. I only treat the case $0 < R < \infty$. For $q \in (0,1)$ arbitrary we will show that the radius of convergence is between $q \cdot R$ and R/q. Letting q go to 1 finishes the proof.
Because of the assumed convergence of $\{|a_k/a_{k+1}|\}$ we can find for every $q \in (0,1)$ some k_q such that:

$$k \geq k_q \Rightarrow \sqrt{q} \cdot R \leq \left|\frac{a_k}{a_{k+1}}\right| \leq \frac{1}{\sqrt{q}} \cdot R.$$

This implies for all $k \geq k_q$ a simple but crucial estimate:

$$|z| \leq q \cdot R \ \Rightarrow\ |P_{k+1}(z) - P_k(z)| = |a_{k+1} \cdot z^{k+1}| \leq q \cdot R \cdot |a_{k+1}| \cdot |z|^k \leq$$
$$\leq \sqrt{q} \cdot |a_k| \cdot |z|^k = \sqrt{q} \cdot |P_k(z) - P_{k-1}(z)|.$$

In the disk $|z| \leq q \cdot R$ the power series is $\sqrt{q}$-*geometrically majorized* and therefore convergent, i. e. $q \cdot R$ is smaller or equal to the radius of convergence. (The initial portion of the power series, that is for $k < k_q$, is irrelevant for the convergence.) Similarly we conclude with $|a_{k+1}| \cdot R \geq \sqrt{q} \cdot |a_k|$:

$$|z| \geq R/q \Rightarrow |P_{k+1}(z) - P_k(z)| \geq (1/\sqrt{q}) \cdot |P_k(z) - P_{k-1}(z)|,$$

hence divergence of $\{P_k(z)\}$, or: R/q is greater or equal to the radius of convergence.

The second criterion is necessary and sufficient:

Root Test. Put $s := \limsup \sqrt[k]{|a_k|} := \lim_{j\to\infty}(\sup_{k\geq j} \sqrt[k]{|a_k|})$. Then:
1.) The power series converges only for $z = 0$ if and only if $s = \infty$ holds.
2.) The power series converges for all $z \in \mathbb{C}$ if and only if $s = 0$.
3.) If $0 < s < \infty$, then we have for the radius of convergence R:

$$R = \frac{1}{s} = 1/\limsup \sqrt[k]{|a_k|}.$$

Proof. I only treat Nr. 3.).
First, we show $R \cdot s \leq 1$ with the only proof in this chapter, that does not use the geometric series:

For each z, for which $\{P_k(z)\}$ converges, the sequence $\{a_k z^k\}$ is bounded, say $|a_k \cdot z^k| \leq M$ for all $k \in \mathbb{N}$. Because the k^{th} root is a monotone function that lies below its tangent at 1 we have

$$\sqrt[k]{|a_k|} \cdot |z| \leq \sqrt[k]{M} \leq 1 + \frac{M-1}{k}, \quad \text{implying (at the last step with Archimedes' trick)}$$
$$s \cdot |z| = \limsup \sqrt[k]{|a_k|} \cdot |z| \leq \sup_{k \geq j}(1 + \frac{M}{k}) = 1 + \frac{M}{j}, \quad \Rightarrow \quad s \cdot |z| \leq 1.$$

This holds for every $|z| < R$ (:= radius of convergence) and therefore proves $s \cdot R \leq 1$.
Second, we will show for each $q \in (0,1)$ that $s \cdot R \geq q$. Both inequalities (plus Archimedes' trick) give $s \cdot R = 1$.
Since $0 < s = \limsup \sqrt[k]{|a_k|} < \infty$ we find for each $q \in (0,1)$ some k_q so that:

$$k \geq k_q \Rightarrow \sup \sqrt[k]{|a_k|} \leq s/\sqrt{q}\ , \text{ hence } \ |a_k| \cdot (\sqrt{q}/s)^k \leq 1.$$

Consequently for all $k \geq k_q$ and all $|z| \leq q/s$ holds

$$|P_{k+1}(z) - P_k(z)| = |a_{k+1} \cdot z^{k+1}| \leq |a_{k+1} \cdot (q/s)^{k+1}| \leq \sqrt{q}^{k+1}.$$

Thus, in the disk $|z| \leq q/s$ we have the power series $\sqrt{q}$-geometrically majorized (at least for $k \geq k_q$), hence proved convergence; i.e. q/s is less or equal to the radius of convergence R. Therefore we also showed the left of the two inequalities $q \leq s \cdot R \leq 1$, hence $R = 1/s$.

Exercise. Treat the cases 1.) and 2.) of the root test.

Now we know enough about how power series converge, but we haven't said anything about *properties of the limit functions*. In particular:

"Are they differentiable? If so, how does one find their derivative?"

We will answer this by geometrically majorizing the differentiated power series $\{P'_n(z)\}$, $\{P''_n(z)\}$, with the technical tool of trick 2.

Exercise. There are other ways to geometrically majorize the differentiated power series. Recall: $(n+1) \cdot q^n \leq 1 + q + \ldots + q^n \leq 1/(1-q)$ and conclude (one may replace q by $q^{1/2}$, $q^{1/3}$ etc. and square or cube etc. the inequality):

$$\begin{array}{rcl} (n+1) \cdot q^n = (n+1) \cdot (q^{1/2})^n \cdot (q^{1/2})^n & \leq & (q^{1/2})^n/(1-q^{1/2}), \\ (n+1)^2 \cdot q^n = (n+1)^2 \cdot (q^{2/3})^n \cdot (q^{1/3})^n & \leq & (q^{1/3})^n/(1-q^{1/3})^2, \\ (n+1)^k \cdot (q^{k/(k+1)})^n \cdot (q^{1/(k+1)})^n & \leq & (q^{1/(k+1)})^n/(1-q^{1/(k+1)})^k. \end{array}$$

Corollary: for each fixed k the sequence $\{a_n := n^k \cdot q^n\}$ converges to 0.

Theorem. Geometric Majorization of Differentiated Power Series

Assumption. The power series with the coefficients a_k has a radius of convergence $r > 0$. Choose $q \in (0,1)$.

Claim. The differentiated polynomial squences $\{P_n'(z)\}$, $\{P_n''(z)\}$ are in the disk $|z| \le q \cdot r$ geometrically majorized, hence convergent. They are, depending on the choice of q, but independent of n, *bounded* in the disk $|z| \le q \cdot r$ (For bounds see the proof).

Corollary. All the differentiated power series have the same radius of convergence.

Proof. By assumption the power series $\{P_n(z)\}$ converges if $|z| \le \sqrt{q} \cdot r$, consequently $a_k \cdot (\sqrt{q}r)^k$ is bounded: $|a_k| \cdot \sqrt{q}^k \cdot r^k \le M$ (trick 2 above). This implies in the smaller disk $|z| \le q \cdot r$ a geometric majorization of the power series and its derivatives:

First:

$|P_n(z) - P_{n-1}(z)| = |a_n \cdot z^n| \le |a_n \cdot q^n \cdot r^n| \le M \cdot \sqrt{q}^n$.

Because of trick 1 above

$\sum \sqrt{q}^k \le 1/(1-\sqrt{q}),\ \sum k\sqrt{q}^{k-1} \le 1/(1-\sqrt{q})^2,\ \sum k(k-1)\sqrt{q}^{k-2} \le 2/(1-\sqrt{q})^3$

we have for all $n \in \mathbb{N}$ and all $|z| \le q \cdot r$ the bounds

$$\begin{aligned} |P_n(z)| &\le M/(1-\sqrt{q}) =: C_q, \\ |P_n'(z)| &\le \frac{M}{\sqrt{q}r} \cdot 1/(1-\sqrt{q})^2 =: L_q, \\ |P_n''(z)| &\le \frac{M}{qr^2} \cdot 2/(1-\sqrt{q})^3 =: B_q. \end{aligned}$$

The bound L_q and Archimedes' trick give the same Lipschitz bound for the limit function. The bound B_q controls the quadratic deviation of all the $P_n(z)$ from their tangents. The following theorem extends this control to the limit function:

Theorem. Derivative of the Limit Function

The differentiated power series $\{P_n'(z)\}$ converges in the disk $|z| < r$ to the derivative of the limit function P_∞ of the power series $\{P_n(z)\}$. In every smaller disk $D_{qr} = \{z\ ;\ |z| \le q \cdot r\}$ one has *uniform* quadratic approximation of the tangents:

$$z, a \in D_{qr} \Rightarrow |P_\infty(z) - P_\infty(a) - \lim P_n'(a)(z-a)| \le \frac{B_q}{2}|z-a|^2.$$

Consequently P_∞ is differentiable and the differentiation rule is: $(P_\infty)' := \lim P_n'$.

Proof. We first recall why L_q is a Lipschitz bound for the limit function. From $|P_n'(z)| \le L_q$ we have for the polynomial P_n from the (D⇒L)–Theorem that we proved for polynomials:

$$z, w \in D_{q\cdot r} \Rightarrow |P_n(w) - P_n(z)| \le L_q \cdot |w - z|.$$

This inequality holds for every n. The Archimedes' trick implies the same Lipschitz bound for the limit function P_∞:

$$z, w \in D_{q\cdot r} \Rightarrow |P_\infty(w) - P_\infty(z)| \le L_q \cdot |w - z|.$$

We repeat this argument for the tangent approximation. We give the limit function of the power series $\{P_n'(z)\}$ the name $Q_\infty(z)$. Now $|P_n''(z)| \le B_q$ implies via the (D⇒L)–Theorem

$$z, w \in D_{q\cdot r} \Rightarrow |P_n(w) - P_n(z) - P_n'(z)\cdot(w-z)| \le B_q \cdot |w-z|^2.$$

For the limit function holds via the triangle inequality:

$$z, w \in D_{q\cdot r} \Rightarrow |P_\infty(w) - P_\infty(z) - Q_\infty(z)\cdot(w-z)| \le B_q \cdot |w-z|^2 + \textit{sum of three sequences converging to } 0.$$

Archimedes' trick gets rid of the three sequences:

$$z, w \in D_{q\cdot r} \Rightarrow |P_\infty(w) - P_\infty(z) - Q_\infty(z)\cdot(w-z)| \le B_q \cdot |w-z|^2.$$

This inequlity says, that P_∞ satisfies the definition for uniform differentiability with derivative $(P_\infty)' = Q_\infty$. In other words: "One differentiates power series (in $D_{q\cdot r}$) term by term like polynomials". Power series are automatically Taylor series:

$$\left(\sum_{k=0}^{\infty} a_k\, z^k\right)' = \sum_{k=1}^{\infty} a_k \cdot k \cdot z^{k-1}, \qquad a_k = \frac{P_\infty^{(k)}(0)}{k!},\ P_\infty(z) = \sum_{k=0}^{\infty} \frac{P_\infty^{(k)}(0)}{k!}\, z^k.$$

Remark. Every convergent power series is derivative of some other convergent power series:

$$\left(\sum_{k=0}^{\infty} a_k \cdot \frac{z^{k+1}}{k+1}\right)' = \sum_{k=0}^{\infty} a_k\, z^k.$$

Example. Up to now we could not write the function $f(z) = 1/z$ as derivative of some other function; however, in a subdisk of its domain of definition it could be written as a geometric series. In this disk we now have an antiderivative:

$$|z-1| < 1 \Rightarrow \frac{1}{z} = \frac{1}{1-(1-z)} = \sum_{k=0}^{\infty}(1-z)^k = \left(-\sum_{k=0}^{\infty} \frac{(1-z)^{k+1}}{k+1}\right)'.$$

In larger disks we have a similar albeit less esthetic result:

$$|z-R| < R \Rightarrow \frac{1}{z} = \frac{1}{R-(R-z)} = \frac{1}{R}\cdot\sum_{k=0}^{\infty}(1-\frac{z}{R})^k =$$

$$= \left(\sum_{k=0}^{\infty}\left[\frac{-1}{k+1}(1-\frac{z}{R})^{k+1} + (\frac{1}{k+1}(1-\frac{1}{R})^{k+1}\right]\right)'.$$

The limit function of the last power series we denote by $L_R(z)$. We have $L_R(1) = 0$ and $L_R'(z) = 1/z$ in the disk $|z-R| < R$. Can we identify $L_R(z)$ with the logarithm function?

About the Exponential Function. For the following kown *real* power series we define complex extensions:

$$\exp(z) := \sum_{k=0}^{\infty} \frac{z^k}{k!}, \quad \sin(z) := \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k+1}}{(2k+1)!}, \quad \cos(z) := \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k}}{(2k)!}.$$

exp satisfies the same differential equation and functional equation that it satisfies as real function:

$$\exp' = \exp \quad \text{and} \quad \exp(a) \cdot \exp(z) = \exp(a+z).$$

Proof. We find the derivative by termwise differentiation. The quotient

$$q(z) := \exp(a) \cdot \exp(z) / \exp(a+z)$$

satisfies $q(0) = 1$ and the quotient rule implies $q'(z) = 0$. The (D⇒L)–Theorem gives $q(z) = 1$ for all z. (Without this theorem this conclusion is decidedly more lengthy.)

By comparing the power series one finds Euler's formulas:

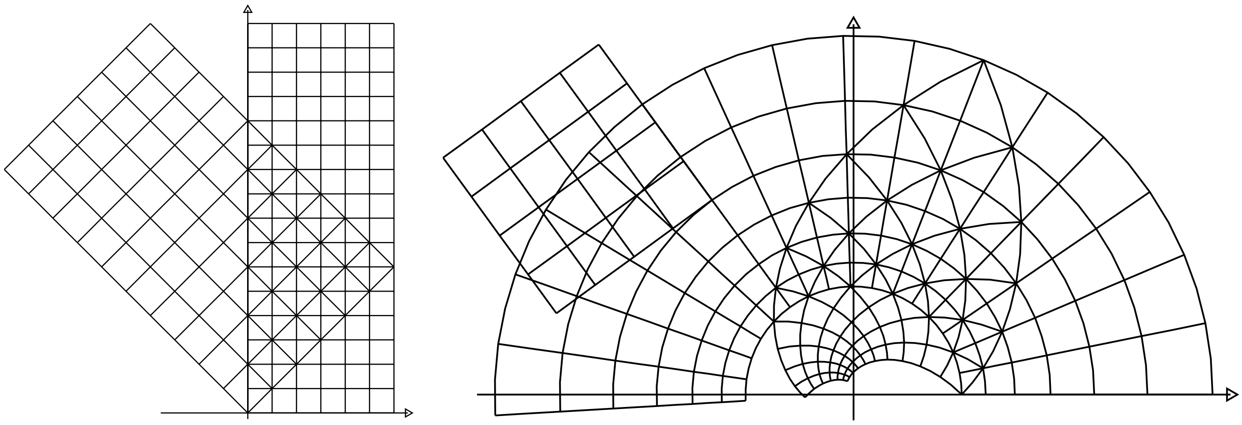
$$\begin{aligned} \exp(iz) &= \cos(z) + i \cdot \sin(z)\,, \quad \exp(2\pi i) = \exp(0) = 1, \\ \cos z &= (\exp(iz) + \exp(-iz))/2, \\ \sin z &= (\exp(iz) - \exp(-iz))/2i. \end{aligned}$$

The addition theorem for exp implies (via insertion) the addition theorems of sin and cos. The argument that we used in the real domain does not work since for complex numbers we do not have $a^2 + b^2 = 0 \Rightarrow a = 0 = b$. One example:

$$\begin{aligned} \cos(z+w) &= \frac{1}{2}(\exp(iz+iw) + \exp(-iz-iw)) = \\ &= \frac{1}{2}(\exp(iz) \cdot \exp(iw) + \exp(-iz) \cdot \exp(-iw)) \\ &= +\frac{1}{2}(\exp(iz) + \exp(-iz)) \cdot \frac{1}{2}(\exp(iw) + \exp(-iw)) \\ &\quad - \frac{1}{2i}(\exp(iz) - \exp(-iz)) \cdot \frac{1}{2i}(\exp(iw) - \exp(-iw)) \\ &= \cos z \cdot \cos w - \sin z \cdot \sin w. \end{aligned}$$

From $\exp(2\pi i) = \exp(0) = 1$ we therefore get that cos and sin are also in the complex domain 2π-periodic:

$$\cos(z + 2\pi) = \cos(z), \quad \sin(z + 2\pi) = \sin(z).$$



Angle preserving (= conformal) polar coordinates are obtained as **image of the Exponential function.** The square net with its orthogonal 45°-diagonals (left) is mapped to the orthogonal net (right) with orthogonal diagonal curves (spirals). Also, a linear approximation is shown as superimposed square.

Inverse of the complex function exp. Since exp is $2\pi i$-periodic, there is *no* unique inverse function. Let $\ell(z)$ be *some* (inner) inverse, i.e. $\exp(\ell(z)) = z$.
If we knew differentiability of ℓ, the chain rule would give:

$$1 = \exp'(\ell(z)) \cdot \ell'(z) = \ z \cdot \ell'(z) = 1, \ \Rightarrow \ \ell'(z) = 1/z.$$

At least in disks $|z - R| < R$ we just constructed power series $L_R(z)$ with derivative $L_R'(z) = \ 1/z$ and with $L_R(1) = 0$. Are they inverses of exp? We insert them into the exponential function, $h(z) := \exp(L(z))/z$, and check:

$$h(1) = 1 \text{ and } h'(z) = \exp'(L(z)) \cdot L'(z)/z - \exp(L(z))/z^2 = 0.$$

The (D⇒L)–Theorem implies that h is the constant 1, hence $\exp(L_R(z)) = z$ is proved at least in $|z - R| < R$!
Finally we observe, that the union of the expanding disks $\{z; |z - R| < R\}$ is the right half plane $\{z : \text{Re } z > 0\}$; moreover, any two of the functions $L_R(z)$ agree on the intersection of their disks of convergence (which is the smaller disk):

$$r \le R, \ |z - r| < r \ \Rightarrow \ h(z) := L_R(z) - L_r(z) \ \Rightarrow \ h(1) = 0, \ h'(z) = 0.$$

The inner inverse function of exp, which was above only defined on the right halfplane (the extension to the complex plane minus the negative real axis has to wait until the chapter on integrals), is called

principal value of the complex logarithm function, Log (z),

$$\text{Log } (1) = 0 \ , \ \exp(\text{Log } z) = z \ , \ \text{Log }'(z) = 1/z.$$

As application we define *powers with complex exponents*

$$z^\alpha := \exp(\alpha \cdot \text{Log } z)\ ,\ (\alpha \in \mathbb{C}, \text{Re } z > 0).$$

The chain rule generalizes the differentiation rule for powers from integer exponents to complex exponents:

$$(z^\alpha)' = \alpha/z \cdot \exp(\alpha \cdot \text{Log } z) = \alpha \cdot z^{\alpha-1}.$$

Problem. The function $z \to (1+z)^\alpha$ and its Taylor series at $z = 0$ (radius of convergence 1) have at $z = 0$ the value 1 and both have the same growth rate $f'/f = \alpha/(1+z)$. Hence their quotient is 1. Determine the coefficients of this power series and check its growth rate.

Non approximating Taylor series. We pointed out already that Taylor polynomials of a function f can only be expected to be convergent approximations if the derivatives $f^{(k)}(x)$ do not grow too fast. This has to be emphasized by an example:

Definition: $\quad f(0) := 0, \quad x \in \mathbb{R} \setminus \{0\}\ :\ f(x) := \exp(-1/x^2).$

Claim: $\quad f^{(k)}(0) := 0, \quad x \in \mathbb{R} \setminus \{0\}\ :\ f'(x) := \dfrac{2}{x^3}\exp(-1/x^2).$

Proof. From our (real) approximation of exp by the compounded interest functions, that is from $0 \le x \Rightarrow \exp(-x) \le 1/(1+x/n)^n$, we conclude

$$0 \le f(x) \le \frac{1}{(1+1/(x^2 n))^n} \le n^n \cdot x^{2n}, \qquad |f'(x)| \le 2n^n \cdot |x|^{2n-3}.$$

Already with $n = 1$ the first inequality implies $f'(0) = 0$. The second shows with $n = 3$ that:

$$|f'(x) - f'(0) - 0 \cdot x| \le 27|x|^3, \text{ hence } f''(0) = 0.$$

We continue by induction. Each differentiation of $\exp(-1/x^2)$ gives a factor $2/x^3$; differentiation of the rational factors increases the denominator exponents only by 1. Therefore constants C_k exist such that:

$$|x| \le 1 \Rightarrow |f^{(k)}(x)| \le C_k n^n \cdot |x|^{2n-3k}.$$

The choice $2n \ge 3k+2$ implies

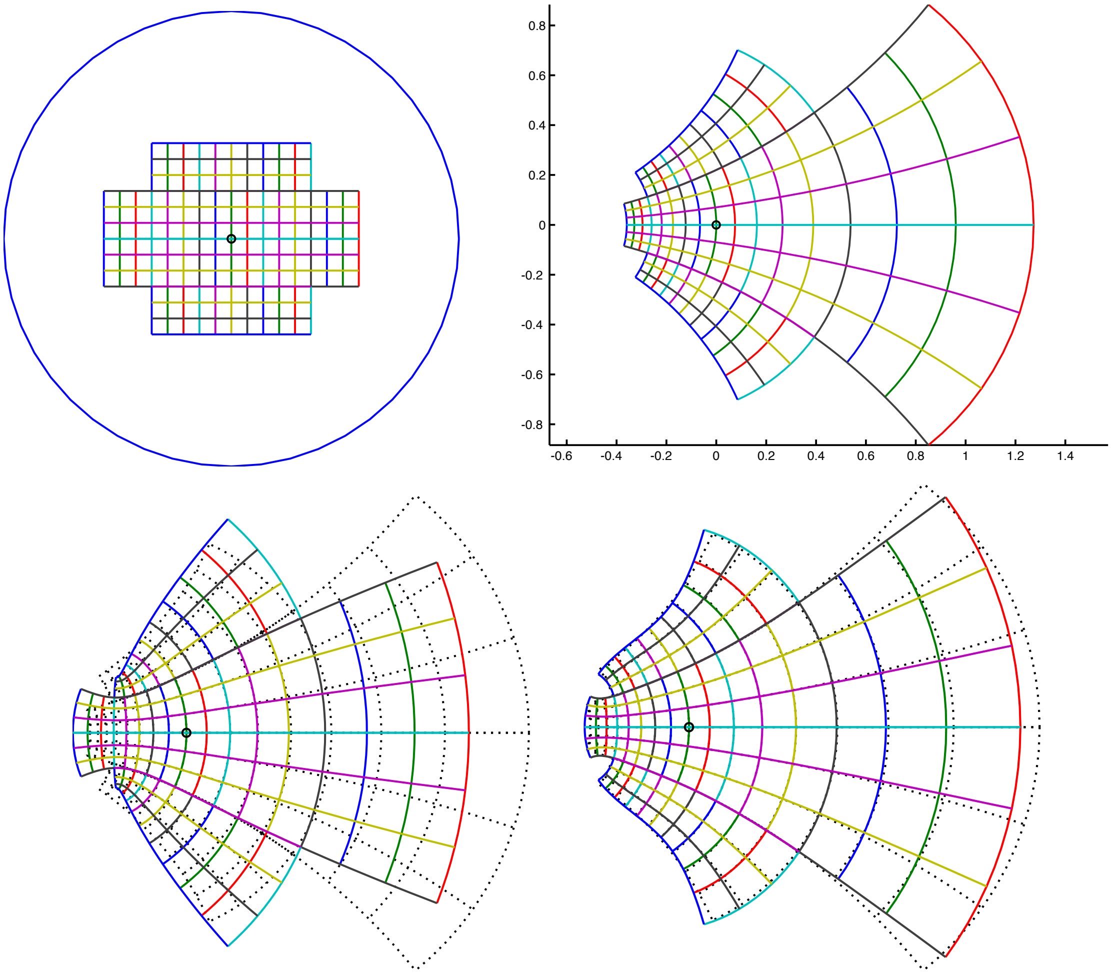
$$|x| \le 1 \Rightarrow |f^{(k)}(x) - 0 - 0 \cdot x| \le C_k n^n |x|^2, \text{ hence } f^{(k)}(0) = 0.$$

All Taylor polynomials of this function f are therefore 0. The Taylor series converges phantastically, but their limit function 0 has little in common with f.

So much right now for power series. They allowed to extend all functions that were treated in school to *complex differentiable functions*. This has also extended our intuition of functions to 2 dimensions. The theory of these functions holds further surprises. – The following problem leads to a natural question.

Problem. Assume we are given a function f with $f'' = -f,\ f^2 + f'^2 = 1$. Define a function g by $g(x) := f(x/3) \cdot (3 - 4f(x/3)^2)$.

Show: $g'' = -g$. For $f(0) = 0$, we find $f'(0) = g'(0)$ and our *real* proof of the addition theorems would show $f = g$. We do not yet have a theorem that allows the same conclusion in the complex case, or do we?



Taylor approximation of the geometric series $z/(1-z) = z + z^2 + z^3 + \dots$
The left top figure shows inside the disk of convergence a twelve-gon. The right top figure is the image under $z \to z/(1-z)$; since this is a Möbius-transformation, all curves in the image net are circles, see p. 63. In the two bottom figures we see the image of the twelve-gon under the third and the fifth Taylor polynomial; the dotted image under $z \to z/(1-z)$ is added for comparison. The base point for the Taylor expansion is marked (o). – The pictures show that, also in this 2-dimensional situation, illustrations of the Taylor approximations are quite suggestive; of course one cannot use the graph representation, the preferred visualization in the 1-dimensional case.

Riemann Sums, Integrals, Antiderivatives

Tangents of certain curves were already considered more than 2000 years ago and for the notion "slope", say of a mountain path, one has some intuitive understanding before any definition. By contrast, integrals are a completely new concept. Integrals, for example, allow to reconstruct from the velocity of a motion the travelled path. If the motion would consist of "jumps", such a reconstruction would be a finite sum. For a given smooth velocity one needs a generalization which may be called a "continuous summation". Our definition, therefore, starts from Riemann sums. We show that integration is an inverse of differentiation and that integrals can be computed from antiderivatives of the integrand.

The concept of antiderivative is simple: A function F is called antiderivative of f, if $F' = f$ holds. We use this notion for real and complex functions and also for curves $f : I \to \mathbb{R}^3$ (one may interprete f as velocity of the movement F). We know already that convergent power series have antiderivatives. And, towards integrals, we know that $F(b) - F(a)$ is not very different from the values of Riemann sums of f on the interval $[a, b]$. These facts will be extended to the integral calculus. Our first step is an existence result which gives us *new functions* in a different way than with power series:

Existence Theorem for Antiderivatives. L-Lipschitz-bounded maps $f : I \to \mathbb{R}$ (or $f : I \to \mathbb{R}^2, \mathbb{R}^d$) have antiderivatives $F : I \to \mathbb{R}$ (or $\mathbb{R}^2, \mathbb{R}^d$), i.e. $F' = f$.

Remark. We will see that this theorem implies (the remaining part of) the fundamental theorem of calculus.

Proof. The Lipschitz assumption: $x, y \in I \Rightarrow |f(y) - f(x)| \leq L \cdot |y - x|$ implies that f can be uniformly approximated by piecewise linear functions as follows:
Subdivide the interval $I = [x_0, x_n]$ into n equal parts $x_0 < x_1 < \ldots < x_n$, with distances $x_j - x_{j-1} = (x_n - x_0)/n = |I|/n$ and define the "**S**ecant**a**pproximations" sa_n of f:

$$x \in [x_{j-1}, x_j] \Rightarrow sa_n(x) := \frac{(x_j - x) \cdot f(x_{j-1}) + (x - x_{j-1}) \cdot f(x_j)}{x_j - x_{j-1}}.$$

All slopes of sa_n are secant slopes of f, hence bounded by L:

$$x, y \in I \to |sa_n(y) - sa_n(x)| \leq L \cdot |y - x|.$$

The Difference $f - sa_n$ has the Lipschitz bound $2L$; moreover, f and sa_n agree on all the division points x_j, $f(x_j) = sa_n(x_j)$ $(j = 0, \ldots, n)$, finally $\min_j |x - x_j| \leq |x_n - x_0|/2n$. This implies for the maximal difference between f and sa_n:

$$x \in I \Rightarrow |f(x) - sa_n(x)| \leq 2L \cdot \min_j |x - x_j| \leq \frac{L \cdot |I|}{n}.$$

We express this property by:

"f can be approximated uniformly by piecewise linear functions".

The functions sa_n have explicit piecewise **quadratic** antiderivatives $SQ_n,\ SQ_n' = sa_n$; we write them first for each subinterval

$$x \in [x_{j-1}, x_j] \Rightarrow SQ_n(x) := C_{j-1} + \frac{1}{2}\frac{(x-x_{j-1})^2 \cdot f(x_j) - (x_j - x)^2 \cdot f(x_{j-1})}{x_j - x_{j-1}}.$$

Finally we put $C_0 = 0$ and we choose C_1 so that the two definitions of $SQ_n(x_1)$ on $[x_0, x_1]$ and on $[x_1, x_2]$ agree. Now repeat: Let the constants $C_0, C_1, \ldots, C_{j-1}$ be determined so that the left-sided and right-sided function values at the division points $x_0, ..., x_{j-1}$ agree; then choose C_j so that the definitions of $SQ_n(x_j)$ on $[x_{j-1}, x_j]$ and on $[x_j, x_{j+1}]$ agree. Now we have

$$SQ_n(x_0) = 0\ ,\ x \in I \Rightarrow SQ_n'(x) = sa_n(x).$$

The sequences $\{SQ_n(x)\}$, $x \in I$, are Cauchy: we have $|sa_n(x) - sa_m(x)| \le \left(\frac{1}{n} + \frac{1}{m}\right) \cdot L \cdot |I|$ already and can conclude with the monotonicity theorem:

$$\begin{aligned} x \in I \Rightarrow |SQ_n(x) - SQ_m(x)| &\le \max |\big(SQ_n - SQ_m\big)'(x)| \cdot |x - x_0| \\ &\le \left(\frac{1}{n} + \frac{1}{m}\right) \cdot L \cdot |I|^2. \end{aligned}$$

By completeness the Cauchy sequences $\{SQ_n(x)\}$ converge to give the values of a *limit function* $x \mapsto SQ_\infty(x)$. Apply the monotonicity theorem again (for piecewise quadratic functions with $|SQ_n'(x) - sa_n(a)| \le L|x-a|$) to get:

$$a, x \in I \Rightarrow |SQ_n(x) - SQ_n(a) - sa_n(a) \cdot (x-a)| \le L \cdot |x-a|^2.$$

Finally by Archimedes' trick (without quoting this old reference one says: taking $\lim_{n\to\infty}$)

$$x, y \in I \Rightarrow |SQ_\infty(y) - SQ_\infty(x) - f(x) \cdot (y-x)| \le L \cdot |y-x|^2.$$

This shows that the limit function SQ_∞ satisfies in I the inequality needed to apply the differentiability definition with derivative $SQ_\infty'(x) = f(x)$.

SQ_∞ is the desired antiderivative of f.

The next task is, to define the integral of a function f in terms of its Riemann sums. We need to make precise the following loose description: All Riemann sums of f based on sufficiently fine subdivisions of the integration interval are so close to each other that *exactly one number* can be considered as their "generalized limit".

Definition of the Integral. Let $f : I \to \mathbb{R}^1$ be a function. Let $x_0 < x_1 < \ldots < x_n$ be a subdivision $\mathfrak{S}$ of the integration interval $I = [a, b] = [x_0, x_n]$. We define a *measure of smallness* for the subdivision:

$$\delta(\mathfrak{S}) := \max\{|x_j - x_{j-1}|;\ j = 1, \ldots, n\}.$$

Finally let $W_{\mathfrak{S}} \subset \mathbb{R}$ be the nonempty set of values of Riemann sums of f for the subdivision $\mathfrak{S}$, i.e.

$$w \in W_{\mathfrak{S}} \Leftrightarrow w = \sum_{j=1}^{n} f(\tau_j) \cdot (x_j - x_{j-1}) \text{ for some } \tau_j \in [x_{j-1}, x_j].$$

The function f is called **Riemann-integrable** on $I = [a, b]$ if the following holds:

$$\lim_{n\to\infty} \sup\{\bigcup_{\mathfrak{S}} W_{\mathfrak{S}}\ ;\ \delta(\mathfrak{S}) \le 1/n\} =: \limsup_{\delta(\mathfrak{S})\to 0} W_{\mathfrak{S}}$$
$$= \liminf_{\delta(\mathfrak{S})\to 0} W_{\mathfrak{S}} := \lim_{n\to\infty} \inf\{\bigcup_{\mathfrak{S}} W_U\ ;\ \delta(U) \le 1/n\}.$$

This condition says: there is exactly one real number, namely

$$\limsup_{\delta(\mathfrak{S})\to 0} W_{\mathfrak{S}} = \liminf_{\delta(\mathfrak{S})\to 0} W_{\mathfrak{S}},$$

which is approximated by Riemann sums for finer and finer subdivisions.
We call it "integral of f on $[a, b]$" and write $\int_I f$ or $\int_a^b f(x)dx$:

Integral of f on $[a, b]$: $$\int_I f := \int_a^b f(x)dx := \limsup_{\delta(\mathfrak{S})\to 0} W_{\mathfrak{S}} = \liminf_{\delta(\mathfrak{S})\to 0} W_{\mathfrak{S}}.$$

A function $f = (f^1, f^2, \ldots, f^d) : I \to \mathbb{R}^d$ is called integrable, if all component functions f^k are integrable. We integrate component wise:

$$\int_I f := (\int_I f^1, \int_I f^2, \ldots, \int_I f^d).$$

The vector-valued $\int_I f$ can also be approximated by Riemann sums, because the Pythagorean theorem allows to combine all the component deviations:

For every error bound $1/n$ there exists a smallness measure δ_n so that:

Let $\mathfrak{S}$ be a subdivision with smallness measure $\delta(\mathfrak{S}) \le \delta_n$

and let $\mathfrak{RS}(f) := \sum_k f(\tau_k)(t_k - t_{k-1})$ be a Riemann sum for the subdivision $\mathfrak{S}$,

then: $$|\int_I f - \mathfrak{RS}(f)| \le \frac{1}{n}.$$

Our existence theorem for antiderivatives and our control of Riemann sums of f by an antiderivative F of f (generalized in the next chapter from Lipschitz to continuous f) gives the

Fundamental Theorem of Calculus.
L-Lipschitz-bounded $f : I = [a,b] \to \mathbb{R}^d$ are Riemann integrable. f has an antiderivative $F : I \to \mathbb{R}^d,\ F' = f$, and

$$\int_I f := \int_a^b f(x)dx = F(b) - F(a).$$

Proof. Our previous existence theorem provides the antiderivative F of f. With the Lipschitz bound L for f we proved in this existence theorem:

$$x, y \in I \Rightarrow |F(y) - F(x) - f(x) \cdot (y - x)| \leq L \cdot |y - x|^2.$$

Choose $n \in \mathbb{N}$ arbitrarily. We have for every subdivision $\mathfrak{S}$ with smallness measure $\delta(\mathfrak{S}) \leq 1/n$ and *every* Riemann sum $\mathfrak{RS}(f) = \sum_{j=1}^{n} f(\tau_j) \cdot (x_j - x_{j-1})$ for this subdivision by the last theorem in the chapter on the monotonicity theorem

$$|F(b) - F(a) - \mathfrak{RS}(f)| \leq \frac{L \cdot |I|}{n}.$$

This bounds the difference between $F(b) - F(a)$ and arbitrary Riemann sums for subdivisions with smallness measure $\leq 1/n$ by a sequence that converges to 0; the condition for integrability is satisfied and $F(b) - F(a) = \limsup_{\delta(\mathfrak{S})\to 0} W_{\mathfrak{S}} = \liminf_{\delta(\mathfrak{S})\to 0} W_{\mathfrak{S}}$ holds.

Comment. The differential calculus to determine tangents and the integral calculus of generalized summation were developed independently of each other. It was a great discovery that integration can be viewed as an inverse of differentiation. The supply of functions with *known* antiderivatives was large right from the beginning, the fundamental theorem could therefore be used to explicitly compute many integrals. This was, of course, a great success because the definition of the integral provides only approximations, not methods of computation.

A somewhat **different formulation of the Fundamental Theorem** is also important:
If a function $F : I \to \mathbb{R}^d$ has a Lipschitz-bounded derivative $f = F' : I \to \mathbb{R}^d$ and known initial value $F(a)$, then F can be reconstructed by integrating f:

$$F(x) - F(a) = \int_a^x f(t)dt.$$

The assumption "Lipschitz-bounded" will be weakened to "continuous" in the next chapter; but recall that the fundamental theorem is almost 200 years older than continuity.

For Interpretations of Integrals it is often important to think of integrals as sums. Here is a list of formal similarities:

1.) $\sum_k (\alpha \cdot a_k + \beta \cdot b_k) \;=\; \alpha \cdot \sum_k a_k + \beta \cdot \sum_k b_k,$ linear

$\int (\alpha \cdot f(x) + \beta \cdot g(x))dx = \alpha \cdot \int f(x)dx + \beta \cdot \int g(x)dx.$ linear

2.) $\sum_{k=1}^{m} a_k + \sum_{k=m+1}^{n} a_k = \sum_{1}^{n} a_k,$ associative

$\int_a^b f + \int_b^c f \;=\; \int_a^c f,$ interval-additive

3.) $a_k = b_{k+1} - b_k \Rightarrow \sum_{k=1}^{n} a_k = b_{n+1} - b_1,$ collapsing sum

$f = F' \Rightarrow \int_a^b f(x)dx = F(b) - F(a),$ Fundamental Theorem

4.) $a_k \le b_k \Rightarrow \sum a_k \le \sum b_k,$ Monotonicity

$f \le g \;\Rightarrow \int_a^b f \;\le \int_a^b g,$ Monotonicity

5.) $|\sum a_k| \;\le\; \sum |a_k|,$ generalized triangle inequality

$|\int_a^b f(x)dx| \le \int_a^b |f(x)|dx,$ continuous triangle inequality.

Proof of the continuous triangle inequality for functions $f : I \to \mathbb{R}^d$ with respect to any norm $|.|$ on $\mathbb{R}^d$. Choose — for each $n \in \mathbb{N}$ — the smallness measure of subdivisions $\mathfrak{S}$ of the interval $I = [a, b]$ so small that for all such subdivisions the difference between Riemann sums and integral is at most $1/n$, on both sides of the inequality. For the Riemann sums we have the generalized triangle inequality:

$$\Big|\sum_{j=1}^{n} f(\tau_j) \cdot (x_j - x_{j-1})\Big| \le \sum_{j=1}^{n} |f(\tau_j)| \cdot (x_j - x_{j-1}),$$

this implies for the integrals (the error $2/n$ is then removed with the Archimedes trick):

$$\Big|\int_a^b f(x)dx\Big| \le \int_a^b |f(x)|dx + \frac{2}{n} \quad \text{for all} \quad n \in \mathbb{N}.$$

Numerical Computations of Integrals. The above analogy between sums amd integrals is also emphasized by formulas for numerical evaluation; the numerics therefore helps the intuition. In accordance with the definition of the integral one subdivides the integration interval I into small subintervals. The main idea is that the integrals of *differentiable* functions f can be approximated on small intervals by really simple formulas so well that the summation over all the subintervals still leads to a good approximation for $\int_I f$. The two simplest approximations are as follows
(a) the integral of the secant between the endpoints is computed
(b) the integral of the tangent at the midpoint is computed:

$$\text{a) Secant Trapezoid:} \qquad \int_a^b f(x)dx \sim \frac{f(a)+f(b)}{2}(b-a).$$

$$\text{b) Tangent Trapezoid:} \qquad \int_a^b f(x)dx \sim f(\frac{a+b}{2})(b-a).$$

How good are these approximations? As a consequence of the Monotonicity Theorem we proved that bounds for the second derivative allow to bound the difference of f and either its tangent or its secant as follows:

$$\begin{array}{ll}
\text{Tangent at } c: & T_c(x) := f(c) + f'(c)(x-c), \\
\text{Secant between } a,b: & S_{ab}(x) := \big(f(a)(b-x) + f(b)(x-a)\big)/(b-a). \\
\text{Assumption:} & a \le c, x \le b, \quad 0 \le f'' \le B. \\
\text{Monotonicity Theorem} \Rightarrow & 0 \le f(x) - T_c(x) \le \frac{1}{2}B(x-c)^2, \\
 & 0 \le S_{ab}(x) - f(x) \le \frac{1}{2}B(x-a)(b-x).
\end{array}$$

The *error estimate* of the two trapezoid rules is obtained by integrating these two inequalities and using the monotonicity of the integral:

$$\begin{array}{ll}
\text{Assumption:} & 0 \le f'' \le B, \quad c := \frac{a+b}{2}. \\
\text{Tangent Trapezoid:} & 0 \le \int_a^b f(x)dx - f(c)(b-a) \le \frac{B}{24}(b-a)^3, \\
\text{Secant Trapezoid:} & 0 \le \frac{f(a)+f(b)}{2}(b-a) - \int_a^b f(x)dx \le \frac{B}{12}(b-a)^3.
\end{array}$$

The assumption $|f''| \le B$ implies the corresponding absolute value inequalities. It is useful to remember though: As long as the second derivative does not change sign, one of the trapezoid rules gives too large an approximation the other is too small. This leads to the question:
Is it possible to average the two trapezoid rules to get a better rule? If our error estimates

are realistic then the tangent trapezoid is twice as good as the secant trapezoid rule. This agrees with Archimedes' result that the area between parabola and secant is twice as large as the area between parabola and tangent. Since the errors also have opposite sign we try a 1:2 average and find the:

Simpson rule: $$\int_a^b f(x)dx \sim \big(\frac{f(a)+f(b)}{2} + 2\cdot f(\frac{a+b}{2})\big)\cdot\frac{(b-a)}{3}.$$

For the quadratic parabola the Simpson rule gets the correct result:

$$\int_a^b x^2 dx = \frac{1}{3}(b^3-a^3) = \big(\frac{a^2+b^2}{2} + 2(\frac{a+b}{2})^2\ \big)\cdot\frac{(b-a)}{3}.$$

In fact, the Simpson rule computes for every *cubic* polynomial P the correct integral. Namely, with c being the interval midpoint we write $x^3 = (x-c+c)^3 = (x-c)^3+$ quadratic remainder. The quadratic remainder is integrated correctly by the Simpson rule and the integral of $(x-c)^3$ is 0, also the value of the Simpson approximation. Therefore we have the interpretation:

The Simpson rule computes the integral of the cubic polynomial P that agrees with f in three values $f(a), f(\frac{a+b}{2}), f(b)$ and the derivative $f'(\frac{a+b}{2})$.

An error bound for the Simpson rule first shows that a bound for the fourth derivative of f, say $|f^{(4)}| \le C$, controls the difference between f and P and the integral of this difference bounds the error of the Simpson rule.

The proof of this frequently used and rather precise error estimate is a more complicated example for arguing with the monotonicity theorem. I simplify the notation by assuming $a=-b$. For the function $h := f - P$ we have the following

Assumption: $$h(\pm a) = 0 = h(0),\ h'(0) = 0,\ |h^{(4)}| \le C,$$

Claim: $$-a \le x \le a \ \Rightarrow\ |h(x)| \le \frac{C}{24}x^2(a^2-x^2),$$

Simpson error: $$\left|\int_a^b f(x)dx - \frac{1}{6}\left(f(a) + 4f(\frac{a+b}{2}) + f(b)\right)(b-a))\right| \le \frac{C}{2880}(b-a)^5.$$

Proof. As with other applications of the monotonicity theorem the main result is the special case where some derivative, here the fourth, does not change sign:

Claim0: $$h(\pm a) = 0 = h(0),\ h'(0) = 0,\ 0 \le h^{(4)} \ \Rightarrow\ h \le 0.$$

We interprete this: If $f^{(4)} \ge 0$ then the Simpson approximation is larger than the integral of f. Next, if Claim0 is proved and $|f^{(4)}| \le C$ is assumed then Claim0 can be used for $g_+(x) := +h(x) - \frac{C}{24}x^2(a^2-x^2),\ g_-(x) := -h(x) - \frac{C}{24}x^2(a^2-x^2)$ since $0 \le g_\pm^{(4)}$.

Of course $g_\pm \le 0$ implies the Claim for $|h|$. Then the Simpson-error is obtained from the antiderivative $(a^2x^3/3 - x^5/5)' = x^2(a^2-x^2)$.

Proof of Claim0. Our early applications of the monotonicity theorem include: $0 \le h^{(4)}$ implies that h'' lies above each tangent and below each secant. From this we conclude by contradiction $h''(0) \le 0$: Assume $h''(0) > 0$; the tangent at 0 of h'' is at least on one of the intervals $[-a, 0]$ (case $h'''(0) \le 0$) or $[0, a]$ (if $h'''(0) \ge 0$) strictly positive. The same is true for h'' because it is above this tangent. Consequently is h' on at least one of these intervals strictly increasing and, because of $h'(0) = 0$ we have $h' < 0$ on $[-a, 0)$ OR $h' > 0$ on $(0, a]$. This implies that h were strictly monotone on at least one of these intervals, contradicting $h(-a) = h(0) = h(a)$, thus proving $h''(0) \le 0$.
We give two versions for the end of the proof. The first and simpler version uses from the next chapter that h assumes its maximum; the second version only uses what is presently proved.
If the function h had positive values then it had a positive maximum $h(c)$. Necessary conditions for a maximum are $h'(c) = 0, h''(c) \le 0$. Since h'' is below each secant (recall $h^{(4)} \ge 0$), we have $h'' \le 0$ between 0 and c. Therefore h had to be below its horizontal tangent at 0, contradicting $h(c) > 0$.
Now the same conclusion without the Maximum Theorem. Assume that h had some positive value $h(x) > 0$, for example $0 < x < a$. Then h' would have to have some positive values on $[0, x]$ and some negative values on $[x, a]$. For h' to grow from $h'(0) = 0$ to positive values we need h'' positive somewhere in $[0, x]$, e.g. $h''(c) > 0, c \in (0, x)$. Later h' decreases from positive values to negative values which is impossible unless h'' is negative somewhere, e.g. $h''(d) < 0, d \in (c, a)$. As in the first version h'' is below the secant on $[0, d\,]$ with values ≤ 0 at the endpoints – contradicting $h''(c) > 0$.

Integrals and Area below Graphs of Functions. The definition of area is built up over many steps and cannot be summarized here. But an early result in that construction is: *compact sets have area*. Another important property of area is its *monotonicity*: if area(X) and area(Y) are defined and $X \subset Y$ then area$(X) \le$ area(Y).
These properties suffice to see that

$$\text{area}(\{(x, y); a \le x \le b, 0 \le y \le f(x)\}) = \int_a^b f(x)\, dx,$$

or, the area under the graph of f is computed by $\int_a^b f(x)\, dx$.
First, we interprete each term $f(\tau_j) \cdot (x_j - x_{j-1})$ of a Riemann sum of f as the area of a rectangle which has as one edge the interval $[x_{j-1}, x_j]$ on the x-axis and which has height $f(\tau_j)$. Therefore every Riemann sum is the area of a union of rectangles. Then, as the measure of smallness of the subdivisions decreases to 0, the Riemann sums converge to the integral. Finally, the set under the graph of f is compact and therefore has an area; and by monotonicity, the integral is the only candidate number for the area, hence equals the area. – In the 1-dimensional case we can fully explain the construction:

Computation of Arc Length by Integrals: $\text{length}\big(p([a, x])\big) = \int_a^x |p'(t)| dt$.
The arc length of Lipschitz-bounded curves $p : [a, b] \to \mathbb{R}^d$ was earlier (in case $d = 2$) defined as:

length($p([a,b])$) :=
$\sup\{\sum_{k=1}^{N}|p(x_k)-p(x_{k-1})|$; with $a=x_0<x_1<\ldots<x_N=b$ a subdivision of $[a,b]$ $\}$.

We assume in addition that $p(\ )$ is a differentiable curve with (presently) Lipschitz-bounded derivative $p' : [a,b] \to \mathbb{R}^d$. The Fundamental Theorem of Calculus and the continuous triangle inequality imply:

$$|p(x_k)-p(x_{k-1})| = \Big|\int_{x_{k-1}}^{x_k} p'(x)dx\Big| \le \int_{x_{k-1}}^{x_k} |p'(x)|dx,$$

hence

$$\sum_{k=1}^{N}|p(x_k)-p(x_{k-1})| \le \int_a^b |p'(x)|dx.$$

This shows that the integral is an upper bound for the length of any inscribed secant polygon. The sup-definition of arc length now gives:

$$\text{length}(p[a,b]) \le \int_a^b |p'(x)|dx.$$

Next, we view the length and the integral as functions of the right endpoint of the interval:

$$L(t) := \text{length }(p([a,t])), \quad F(t) := \int_a^t |p'(x)|dx,$$

and we aim to show the equality of these functions. Since $L(a)=F(a)=0$ it would suffice to prove $L'(t)=F'(t)$. We now write inequalities for difference quotients (the first one says that the secant is shorter than the length, the second is the inequality proved above):

$$\frac{|p(t_2)-p(t_1)|}{t_2-t_1} \le \frac{L(t_2)-L(t_1)}{t_2-t_1} = \frac{\text{length }(p([t_1,t_2]))}{t_2-t_1} \le \frac{1}{t_2-t_1}\int_{t_1}^{t_2}|p'(t)|dt.$$

Since $p(\ )$ is differentiable we can take the limit $t_2 \to t_1$ explicitly at the two outermost terms of the chain of inequalities:

$$|p'(t_1)| \le \lim_{t_2\to t_1}\frac{L(t_2)-L(t_1)}{t_2-t_1} \le |p'(t_1)| = F'(t_1).$$

This shows that $L(\ )$ is differentiable with derivative $L'(t_1)=|p'(t_1)|=F'(t_1)$ for all $t_1\in[a,b]$. So finally $F=L$ holds and we have proved:

$$\text{length}(p([a,x]) = \int_a^x |p'(t)|dt.$$

We repeat this result with a different wording. View the parameter t of the curve $t\to p(t)$ as "time": a moving point p is at time t at position $p(t)$. This implies the interpretation:

$|p'(t)|$ is the (absolute value of the) velocity at time t of the moving point. And the length $L(t)$ is the distance travelled until time t. In formulas:

$$\text{Travelled distance} = L(t) = \int_0^t |p'(\tau)|d\tau = \text{ Integral of the velocity.}$$

Let us look with this interpretation at the Riemann sums of the integral. Each term $|p'(\tau_j)| \cdot |t_j - t_{j-1}$ is the distance that would be travelled with the (intermediate) velocity $|p'(\tau_j)|$ between time t_{j-1} and time t_j. *Approximately* this is the length of the secant $|p(t_j) - p(t_{j-1})|$. But the approximation is, percentage-wise, the better the shorter the time interval is. The Riemann sum is the sum of these approximately travelled distances and the deviation from the arc length gets, percentage-wise, smaller as the smallness measure of the subdivision decreases.

The daily language cannot really express what the integral achieves in this situation. But our comparison of Riemann sums and integrals gives, I believe, a sufficiently descriptive mental image so that the following colloquial words make sense:

> The integral $\int_a^t |p'(\tau)|d\tau$ "sums continuously" the infinitesimal distances, which are travelled as time progresses by the motion $t \to p(t)$ with instantaneous velocity $|p'(t)|$.

Mean Value Property of Integrals

First recall that the center of mass S of points P_k with position $p_k \in \mathbb{R}^3$ and masses m_k is computed as the mean value:

$$\text{Center of Mass } S = \sum_{k=1}^{N} \frac{m_k}{M} \cdot p_k, \quad \text{Total Mass } M = \sum m_k \; .$$

For some integrable function $f : [a, b] \to \mathbb{R}^3$ we want to interprete the integral $\frac{1}{b-a} \int_a^b f(t)dt$ as a mean value. Consider the Riemann sums $\sum_{k=1}^N f(\tau_k) \cdot (t_k - t_{k-1})/(b - a)$. Clearly, these can be viewed as mean values of the $f(\tau_k)$ with masses $(t_k - t_{k-1})/(b - a)$.

Each refinement of the subdivision on which the Riemann sums are based uses more function values $f(\tau_k)$ (with smaller masses) for this averaging. The colloquial formulation:

$$\int_a^b f(t) \frac{dt}{b-a} \text{ is a continuous average of the values } f(t)$$

is a useful verbal rendition of the underlying mathematics.

Continuity and Uniform Convergence

Towards the end of the 19th century continuity was established as a fundamental notion of analysis, roughly 200 years after the begin of differentiation and integration. I cannot retrace the reasons for this development here, I have to contend myself with simpler motivations for introducing *continuity* than the historically correct ones. In the earlier chapters we saw already how important convergent sequences are in analysis. It is therefore a plausible question, that will lead to continuity, to ask: *What are the functions that map every sequence that is convergent in the domain of the function to a convergent sequence in the range?* The main theorems on continuous functions cannot be proved without using completeness. First we discuss the theorems that follow easily from sequence-continuity. Then ϵ-δ-continuity is introduced and the remaining main theorems are proved. Crucial are the convergence properties of sequences of continuous functions. We close with famous examples by Cantor, Weierstraß and Hilbert.

Definition of continuity. A function f with domain D is called *sequence-continuous at* $a \in D$, if every sequence $\{a_k\}$ in D with $\lim a_k = a$ is mapped to a sequence that converges to $f(a)$, $\lim f(a_k) = f(a)$.
The function f is called continuous if it is continuous at every $a \in D$.

We use this definition for real functions, complex functions, curves $c : \mathbb{R} \to \mathbb{R}^d$, maps $F : \mathbb{R}^d \to \mathbb{R}^e$ and as we progress in increasingly general situations. Maps with a Lipschitz bound $|F(x) - F(y)| \le L \cdot |x - y|$ are, trivially, sequence-continuous since $\lim a_k = a$ and $|F(a_k) - F(a)| \le L \cdot |a_k - a|$ imply $\lim F(a_k) = F(a)$.
The definition implies directly:

Theorem. Sums, products and compositions of continuous functions are continuous.

We will get to a more precise intuition how to imagine continuous functions once we establish powerful function construction tools at the end of this section.

The first three theorems deal with properties so plausible that they are often used without even realizing that they need to be proved. First the

Intermediate Value Theorem. Let $f : [a, b] \to \mathbb{R}$ be (sequence-)continuous and $w \in \mathbb{R}$ with $f(a) < w < f(b)$. Then there is at least one $c \in (a, b)$ such that $f(c) = w$.
(It is enough to prove this in case $w = 0$, otherwise change f to $f - w$.)

Proof. We construct convergent nested intervals, beginning with $a_1 := a$, $b_1 := b$. Assume we had the first n intervals $[a_1, b_1] \supset [a_2, b_2] \supset \cdots \supset [a_n, b_n]$ defined in such a way that $f(a_n) \le 0$, $f(b_n) \ge 0$ and $|b_n - a_n| = |b_{n-1} - a_{n-1}|/2$.
Consider the midpoint $m := (a_n + b_n)/2$. If $f(m) \ge 0$, put $a_{n+1} = a_n$, $b_{n+1} = m$; otherwise

put $a_{n+1} = m,\ b_{n+1} = b_n$. We obtain the next interval. Let c be the limit point of these convergent nested intervals, $\lim a_n = c = \lim b_n$. Sequence-continuity of f together with $f(a_n) \le 0,\ f(b_n) \ge 0$ implies

$$0 \ge \lim f(a_n) = f(c) = \lim f(b_n) \ge 0, \text{ i.e. } f(c) = 0.$$

Problem. Use another formulation of completeness to prove this result.

Functions with a uniform Lipschitz bound L on bounded intervals $[a, b]$ (e.g. polynomials) have values $f(x) \in [f((a+b)/2) - L \cdot (b-a)/2,\ f((a+b)/2) + L \cdot (b-a)/2]$, i.e. are bounded.
The rational function $f(x) = 1/x$ is continuous in $(0, 1)$ but not bounded. Such boundedness failure can only happen near a boundary of the interval because we have the

Boundedness Theorem. Let $[a, b]$ be finite and $f : [a, b] \to \mathbb{R}$ be continuous.
Then f is bounded.

Proof. Assume that f is unbounded on $[a, b] =: [a_1, b_1]$. We first construct convergent nested intervals $\{[a_n, b_n]\}$ so that f is unbounded on all of them. These constructions are always by induction. So we assume that we have the first n intervals already and $m := (a_n + b_n)/2$ is the last midpoint. Since f is unbounded on $[a_n, b_n]$, f has to be unbounded at least on one of $[a_n, m],\ [m, b_n]$. We choose $[a_{n+1}, b_{n+1}]$ in two steps. First we take the interval on which f is unbounded (or one of them); then we make the interval smaller so that the maximum of f at the two endpoints is $\ge n$. Because we used interval halfing, these nested intervals are convergent, say to c. Because of continuity we should have $\lim f(a_n) = f(c),\ \lim f(b_n) = f(c)$, but this is a contradiction because at least one of the sequences $\{f(a_n)\}, \{f(b_n)\}$ is unbounded.

The following refinement is quoted frequently

Maximal Value Theorem. Let $f : [a, b] \to \mathbb{R}$ be continuous. Then there is at least one $c \in [a, b]$ such that $x \in [a, b]$ implies $f(x) \le f(c)$.
One quotes this as: f assumes at some $c \in [a, b]$ its maximum $f(c)$.

Proof. Since f was already proved to be bounded on $[a, b] =: [a_1, b_1]$, we can define:

$$S := \sup\{f(x); x \in [a_1, b_1]\}, \quad S < \infty.$$

Again we construct convergent nested intervals $\{[a_n, b_n]\}$, now with $S = \sup f([a_n, b_n])$. Assume we have the first n such intervals and $m := (a_n + b_n)/2$. On one of the two subintervals $[a_n, m],\ [m, b_n]$ the supremum of f must still be S. Choose it preliminarily as $[a_{n+1}, b_{n+1}]$. In this interval find x with $f(x) \ge S - 1/(n+1)$. On one of the two subintervals $[a_{n+1}, x],\ [x, b_{n+1}]$ the supremum of f must still be S, make it our choice of $[a_{n+1}, b_{n+1}]$. Let c be the limit of these nested intervals. By (sequence-)continuity we have

$\lim f(a_n) = f(c) = \lim f(b_n) \leq S$, hence of course $f(c) = \lim \max\{f(a_n), f(b_n)\}$. But we also have $S - 1/n \leq \max\{f(a_n), f(b_n)\}$, hence $f(c) = S$.

The definition of *sequence-continuous* says that we know the value $f(a)$ at a if we know the values of f at convergent (to a) neighboring points a_n, namlely $f(a) = \lim f(a_n)$. We ask in the opposite direction: if one knows $f(a)$, what can one conclude about values at nearby points? The following theorem gives a simple answer which, however, can only be proved indirectly.

Theorem on Positive Neighbors. Let $f : (a, b) \to \mathbb{R}$ be sequence-continuous at $c \in (a, b)$ and *positive*, $f(c) > 0$. Then there is some interval $(c - \delta, c + \delta)$ on which f is positive.

Indirect Proof. First choose n_1 so large, that $(c - 1/n_1, c + 1/n_1) \subset (a, b)$. Either f is positive on this interval (and we are done) or there exists $a_1 \in (c - 1/n_1, c + 1/n_1)$ with $f(a_1) \leq 0,\ a_1 \neq c$. In the second case choose $n_2 > n_1$ so that $a_1 \notin (c - 1/n_2, c + 1/n_2)$. Again, either f is positive on this second interval or there exists $a_2 \in (c - 1/n_2, c + 1/n_2)$ with $f(a_2) \leq 0$. We can repeat this procedure and either end up with an interval on which f is positive or we find a sequence $\{a_k\}$ with $\lim a_k = c$ and $f(a_k) \leq 0$, hence $f(c) \leq 0$, a contradiction – i.e. this possibility cannot arise, we end up with a (possibly small) interval on which f is positive.

This looks like a complicated proof for an innocently looking statement. Other characterizations of continuity seem desirable which make the above statement obvious.

The following definition looks at the function in the opposite way as sequence-continuity: if one knows $f(a)$ then one can say something about the values at nearby points: For every *acceptable* deviation $\epsilon > 0$ from $f(a)$ there is an interval $[a - \delta, a + \delta]$ on which *no larger* deviations from $f(a)$ occur. In the standard more formal language:

Definition of ϵ-δ-Continuity. Let A, B be subsets of $\mathbb{R}, \mathbb{C}$ or $\mathbb{R}^d$. A map or function $f : A \to B$ is called ϵ-δ-*continuous* at $a \in A$ if the following property holds:
For every $\epsilon > 0$ (called error bound) there exists some $\delta > 0$ (called guarantee radius) so that

$$x \in A,\ |x - a| \leq \delta \ \Rightarrow\ |f(x) - f(a)| \leq \epsilon.$$

If this holds for every $a \in A$, then f is called ϵ-δ-continuous on A.
(The words "error bound, guarantee radius" are added to underscore the intentions pursued with this definition, they can be omitted.)

Theorem on Equivalent Continuity Definitions.
A function $f : (a, b) \to \mathbb{R}$ is sequence-continuous if and only if f is ϵ-δ-continuous.

Proof. a) Let f be ϵ-δ-continuous and let $\{c_k\}$ be a sequence converging to c. We want to show: The image sequence $f(c_k)$ converges to $f(c)$.
Let $\epsilon > 0$ be given, we have to find k_ϵ, such that

$$k \geq k_\epsilon \Rightarrow |f(c_k) - f(c)| \leq \epsilon.$$

The ϵ-δ-continuity of f gives for each given $\epsilon > 0$ a guarantee intervall $(c - \delta, c + \delta)$ such that $x \in (c - \delta, c + \delta) \Rightarrow |f(x) - f(c)| \le \epsilon$. The assumed convergence of $\{c_k\}$ gives us k_δ such that $k \ge k_\delta \Rightarrow |c_k - c| \le \delta$. We put $k_\epsilon := k_\delta$, and combine these two implications to give $\lim f(c_k) = f(c)$, namely:

$$k \ge k_\epsilon = k_\delta \Rightarrow |c_k - c| \le \delta \Rightarrow |f(c_k) - f(c)| \le \epsilon.$$

b) In the other direction we have to conclude from the sequence-continuity of f that for every $\epsilon > 0$ there exists some $\delta > 0$, so that

$$|x - c| < \delta \Rightarrow |f(x) - f(c)| \le \epsilon.$$

The proof, again, has to be indirect and resembles the proof of the preceeding theorem.
If, for some (error bound) ϵ^* it were NOT possible to find the needed (guarantee) interval $(c - \delta, c + \delta)$ (namely with the property $x \in (c - \delta, c + \delta) \Rightarrow |f(x) - f(c)| < \epsilon^*$) then we could define a sequence $\{a_k\}$ leading to a contradiction as follows:
Start by choosing n_1 such that $(c - 1/n_1, c + 1/n_1) \subset (a, b)$ and pick $a_1 \in (c - 1/n_1, c + 1/n_1)$ with $|f(a_1) - f(c)| > \epsilon^*$. (In other words: a_1 is a counter-example to $(c - 1/n_1, c + 1/n_1)$ being a guarantee interval.)
By induction assume that $n_1 < n_2 < \ldots < n_k$ and $a_1, a_2, \ldots, a_k$ are already chosen such that $a_j \in (c - 1/n_j, c + 1/n_j)$ and $|f(a_j) - f(c)| > \epsilon^*$. Next choose n_{k+1} large enough so that $a_k \notin (c - 1/n_{k+1}, c + 1/n_{k+1})$. Because of the indirect proof assumption no δ is good enough to prevent deviations $> \epsilon^*$. Therefore we find $a_{k+1} \in (c - 1/n_{k+1}, c + 1/n_{k+1})$ such that $|f(a_{k+1}) - f(c)| > \epsilon^*$. The sequence $\{a_k\}$ converges to c – but in contradiction to the assumed sequence-continuity we have $|f(a_{k+1}) - f(c)| > \epsilon^*$, or $f(c) \ne \lim f(a_k)$. This shows that we *cannot* construct the sequence $\{a_{k+1}\}$. Which means: after finitely many steps we find $\delta = 1/n_k > 0$ such that $|x - c| \le \delta \Rightarrow |f(x) - f(c)| \le \epsilon^*$. We proved ϵ-δ-continuity of f at c.

Problem. Use the ϵ-δ-definition to show that sum, product and composition of continuous functions are continuous. This takes more work than with the first definition. On the other hand, the theorem on positive neighbors is now straight forward: Choose $\epsilon := f(c)/2 > 0$: In the corresponding guarantee interval $(c - \delta, c + \delta)$ we have $f(c) - f(x) \le f(c)/2$ or $0 < f(c)/2 \le f(x)$, which is a positive lower bound.

The intermediate value theorem and the maximal value theorem do not get simpler proofs by using ϵ-δ-continuity. The boundedness theorem however can be proved in a completely different way. It relies on a property of finite intervals $[a, b]$ that combines well with ϵ-δ-continuity and leads to very short arguments. We first prove this interval property, the

Heine-Borel Covering Theorem. Vor every $x \in [a, b]$ we are given some $\delta_x > 0$.
The conclusion is: One can find *finitely* many $x_k \in [a, b]$ such that (abbreviating $\delta_k := \delta_{x_k}$) the interval $[a, b]$ is covered by the open intervals $(x_k - \delta_k, x_k + \delta_k)$.
This is also expressed as $[a, b] \subset \bigcup_{k=1}^{K} (x_k - \delta_k, x_k + \delta_k)$.

Indirect Proof. The argument resembles the proof of the boundedness theorem. We construct convergent nested intervals $\{[a_n, b_n]\}$, such that each $[a_n, b_n]$ is NOT covered by finitely many intervals $(x - \delta_x, x + \delta_x)$. First step: the given interval $[a, b] =: [a_1, b_1]$ is not covered by finitely many of the $(x - \delta_x, x + \delta_x)$. By induction assume that we have found the first n intervals $[a_1, b_1] \supset \ldots \supset [a_n, b_n]$, all not covered by finitely many of the $(x - \delta_x, x + \delta_x)$. Define $m := (a_n + b_n)/2$ and observe that $[a_n, m]$ or $[m, b_n]$ is not finitely covered so that one of them can be taken as $[a_{n+1}, b_{n+1}]$. Let c be the limit of these nested intervals. Clearly, the *single* interval $(c - \delta_c, c + \delta_c)$ covers those intervals $[a_n, b_n]$ for which $|b_n - a_n| = |b_1 - a_1| \cdot 2^{-n} < \delta_c$ holds. This contradiction prohibits the choice of these nested intervals, hence: $[a, b]$ is finitely covered.

Using the Heine Borel theorem we first generalize the boundedness theorem.

Definition of locally bounded. A function f is called *locally bounded*, if we have for each point x of the domain of f some $\delta_x > 0$ such that f is bounded on $(x - \delta_x, x + \delta_x)$. Continuous functions f are locally bounded because, for every c in their domain, there exists a $\delta_1(c) > 0$ such that

$$|x - c| \le \delta_1 \Rightarrow |f(x) - f(c)| \le 1 \Rightarrow |f(x)| \le |f(c)| + 1.$$

Theorem. Each function f which is locally bounded on $[a, b]$ is in fact globally bounded.

Proof. Cover $[a, b]$ by *finitely many* of the intervals $(x - \delta_x, x + \delta_x)$. The maximum of the assumed local bounds on these intervals is a global bound for f on $[a, b]$.

One can see that we have developed an easily used method of proof. In the literature these arguments are unusually standardized. One therefore has to practise in mathematical conversations to what extent variations in the wording are tolerable - learning the standard words is not enough.
We use this method to prove the

Theorem on Uniform Continuity. Let $f : [a, b] \to \mathbb{R}$ be continuous. Then f is in fact *uniformly* continuous, which means that the δ can be chosen to work for all $x \in [a, b]$:
Claim: For every $\epsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in [a, b]$ holds:

$$|x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon.$$

Proof. The pointwise continuity of f supplies for each $\epsilon > 0$ and each $x \in [a, b]$ some $\delta_x > 0$ such that

$$x, y \in [a, b] \text{ and } |x - y| < 2 \cdot \delta_x \Rightarrow |f(x) - f(y)| < \epsilon/2.$$

(Note that the extra factors $2, 1/2$ are compatible with the continuity definition.)
Finitely many of the intervals $(x - \delta_x, x + \delta_x)$ cover $[a, b]$, or $[a, b] \subset \bigcup_{k=1}^{K} (x_k - \delta_k, x_k + \delta_k)$. Put $\delta := \min_{k=1\ldots k} \delta_k$. We show that this δ is a suitable choice: For each $x \in [a, b]$ first choose x_k such that $x \in (x_k - \delta_k, x_k + \delta_k)$. This implies $|f(x) - f(x_k)| < \epsilon/2$. Assuming $|x - y| < \delta$ we have $y \in (x_k - 2\delta_k, x_k + 2\delta_k)$, hence also $|f(y) - f(x_k)| < \epsilon/2$. The triangle inequality gives $|f(x) - f(y)| < \epsilon$, finishing the proof.

Theorem on Uniform Approximation. Continuous functions $f : [a, b] \to \mathbb{R}$ can be uniformly approximated by piecewise linear functions, meaning:
For every $\epsilon > 0$ there exists a piecewise linear function ℓ such that

$$\text{for all } x \in [a, b] \text{ holds } |f(x) - \ell(x)| < \epsilon.$$

Proof. Since f is uniformly continuous on $[a, b]$ one can find for each $\epsilon > 0$ some $\delta > 0$ so that

$$x, y \in [a, b], |x - y| \leq \delta \Rightarrow |f(x) - f(y)| \leq \frac{\epsilon}{2}.$$

With this δ choose a subdivision $a = t_0 < t_1 < \ldots < t_N = b$ satisfying $|t_j - t_{j-1}| < \delta$. Next define a piecewise linear function ℓ by linearly interpolating f between t_{j-1} and t_j:

$$t \in [t_{j-1}, t_j] \Rightarrow \ell(t) := \frac{f(t_j) \cdot (t_j - t) + f(t_{j-1}) \cdot (t - t_{j-1})}{t_j - t_{j-1}}$$

The triangle inequality proves that l is ϵ-close to f:

$$t \in [a, b] \ \Rightarrow$$
$$|f(t) - \ell(t)| \leq |f(t) - f(t_j)| + |\ell(t) - \ell(t_j)| \leq \frac{\epsilon}{2} + |f(t_j) - f(t_{j-1})| \frac{t - t_{j-1}}{t_j - t_{j-1}} \leq \epsilon.$$

Problem. Give a proof of the maximal value theorem using Heine-Borel. Start as: Assume for each $x \in [a, b]$ we had $f(x) < S := \sup f([a, b])$. Continuity of f at $x \ldots$.

The next goal is to generalize the *Fundamental Theorem of Calculus* from Lipschitz continuous to continuous functions. We need to finalize the definition.

Final Definition of Differentiability. A function $f : (\alpha, \omega) \to \mathbb{R}$ is called *differentiable* at $a \in (\alpha, \omega)$ with derivative $m = f'(a)$ and tangent $T_a(x) := f(a) + m \cdot (x - a)$, if the following holds:

$$|f(x) - T_a(x)| = \Phi_a(x)|x - a| \text{ and } \Phi_a(x) \text{ is continuous at } x = a \text{ with } \Phi_a(a) = 0$$

or:

$$\left|\frac{f(x) - f(a)}{x - a} - m\right| = \Phi_a(x) \text{ and } \Phi_a(x) \text{ is continuous at } x = a \text{ with } \Phi_a(a) = 0$$

or, after plugging in the continuity definition of Φ_a :

For every $\epsilon > 0$ there is a $\delta > 0$ such that: $|x - a| \leq \delta \Rightarrow |f(x) - T_a(x)| \leq \epsilon \cdot |x - a|$.

Problem. In the chapter on complex numbers we proved that derivative bounds are Lipschitz bounds without making uniformity assumptions. Use the same strategy of proof to conclude the *Monotonicity Theorem* with the final version of differentiability in the assumption (adaption to the final definition is needed only at the end).
Let $f : (\alpha, \omega) \to \mathbb{R}$ be differentiable and continuous on $[\alpha, \omega]$, then:

$$f' \geq 0 \ \Rightarrow \ f \text{ is weakly monoton increasing, i.e. } x \leq y \Rightarrow f(x) \leq f(y).$$

Theorem on the Existence of Antiderivatives and Integrals:

Continuous functions $f : [\alpha, \omega] \to \mathbb{R}$ have antiderivatives F, i.e. $F' = f$.

Therefore they are integrable with

$$a, b \in [\alpha, \omega] \Rightarrow F(b) - F(a) = \int_a^b f(x)dx.$$

Proof. The arguments run as for Lipschitz continuous f in the chapter on integration, one only has to adapt to the ϵ-δ-deviations.
For each $1/n > 0$ let $\ell_n : [\alpha, \omega] \to \mathbb{R}$ be a uniform piecewise linear approximation satisfying $x \in [\alpha, \omega] \Rightarrow |f(x) - \ell_n(x)| \leq 1/n,\ |\ell_n(x) - \ell_m(x)| \leq (1/n + 1/m)$, see the preceeding theorem. In this application we used already that f is uniformly continuous, i.e.: For every $\epsilon > 0$ there is a $\delta > 0$ such that

$$x, y \in [\alpha, \omega],\ |x - y| \leq \delta \Rightarrow |f(x) - f(y)| \leq \epsilon \Rightarrow |\ell_n(x) - \ell_n(y)| \leq 2/n + \epsilon.$$

As before, the ℓ_n have piecewise quadratic antiderivatives $Q_n : [\alpha, \omega] \to \mathbb{R}$ satisfying $Q'_n = \ell_n,\ Q_n(0) = 0$.
Apply the monotonicity theorem for each fixed y to the function $x \mapsto Q_n(x) - Q_n(y)$; using the previous estimate we get

$$x, y \in [\alpha, \omega],\ |x - y| \leq \delta \Rightarrow |Q_n(x) - Q_n(y) - \ell_n(y)(x - y)| \leq (2/n + \epsilon) \cdot |x - y|.$$

The monotonicity theorem also shows that the $\{Q_n\}$ are a Cauchy sequence, in formulas:

$$x \in [\alpha, \omega] \Rightarrow |Q_n(x) - Q_m(x)| \leq \max(|\ell_n - \ell_m|) \cdot |x - \alpha| \leq (1/n + 1/m) \cdot |\omega - \alpha|.$$

The completeness of $\mathbb{R}$ provides the limit function Q_∞ of this Cauchy sequence. The Archimedes trick extends the tangent approximations of the Q_n to the limit function:

$$x, y \in [\alpha, \omega],\ |x - y| \leq \delta \Rightarrow |Q_\infty(x) - Q_\infty(y) - f(y)(x - y)| \leq \epsilon \cdot |x - y|.$$

Our argument found for every $\epsilon > 0$ some $\delta > 0$ such that this implication holds, in other words Q_∞ is differentiable with $Q'_\infty = f$. The limit function is an antiderivative of f.

As in the chapter on integration we can approximate Riemann sums of f - up to Archimedes errors - by $Q_\infty(b) - Q_\infty(a)$. Therefore f is integrable in terms of its antiderivative:

$$a, b \in [\alpha, \omega] \Rightarrow Q_\infty(b) - Q_\infty(a) = \int_a^b f(x)dx,\ \ Q'_\infty = f.$$

Sequences of Continuous Functions.

The previous theorems establish the main facts that are true for continuous functions - considered one at a time. The next developments are concerned with properties of convergent sequences f_n of continuous functions and with properties of the limit function. We meet a new method to construct continuous functions in the theorem on *uniformly convergent* sequences of functions. – For such sequences we ALWAYS make the following

Minimal Assumptions for Sequences of Functions $\{f_n(x)\}$.

All the f_n are continuous and the sequences of values $\{f_n(x)\}$ converge for each $x \in D \subset$ Domain (f_n). (This is called *pointwise* convergence in D.)
These assumptions allow to define a limit function:

$x \in D,\ f(x) := \lim_{n\to\infty} f_n(x)$ – which is *NOT* necessarily continuous!!

Example:
$D = [0, 1],\ f_n(x) := x^n,\ f(x) := \lim_{n\to\infty} f_n(x)$ implies $f(1) = 1,\ f(x) = 0$ für $x \in [0, 1)$.

The following theorems say under what additional assumptions one can conclude that the limit function is indeed continuous.
First we use ideas from earlier chapters: If the errors can be controlled independently of n (we also say "uniformly in n" or in the case of continuity "equicontinuous in n") then the error control extends via Archimedes to the limit function:

Theorem on Equicontinuous Sequences.
Assumptions.
For every $\epsilon > 0$ and for each $x \in D$ there is a $\delta_x > 0$ (NOT depending on n), such that:

$$x, y \in D\ ,\ |y - x| \le \delta_x \Rightarrow |f_n(y) - f_n(x)| \le \epsilon.$$

Claim: The limit function has the same continuity behavior:

$$x, y \in D\ ,\ |y - x| \le \delta_x \Rightarrow |f(y) - f(x)| \le \epsilon.$$

Proof. $r_n := |f(x) - f_n(x)|$ and $s_n := |f(y) - f_n(y)|$ are assumed to be convergent to 0. The assumed continuity behavior of the $\{f_n\}$ and the triangle inequality imply

$$x, y \in D,\ |y - x| \le \delta_x \Rightarrow |f(y) - f(x)| \le \epsilon + r_n + s_n,$$

and Archimedes' argument gets rid of r_n and s_n.

This theorem is sufficient to construct, at the end of the chapter, two famous continuous examples: Cantor's staircase and Hilbert's square filling curve.
In the earlier chapters we met situations in which such an n-independent error control was almost impossible to miss. But there are also interesting constructions with continuous functions where the error control deteriorates as the index n increases, see the third example at the end, Weierstraß' oscillator. To express for these cases the additional assumption that implies continuity we need a new notion, we need to make the "pointwise" convergence no longer depend on the point of the domain, but make it "uniform":

Definition of Uniformly Convergent Sequences of Functions.
The sequence of functions $\{f_n\}$ is called *uniformly convergent* (or closer to the assumptions: "uniformly Cauchy") on D, if for every $\epsilon > 0$ some n_ϵ exists, so that for ALL $x \in D$ holds:

$$m, n \ge n_\epsilon \Rightarrow |f_m(x) - f_n(x)| \le \epsilon.$$

One also formulates this as: The convergence (for $n \to \infty$) of $\{f_n(x)\}$ is uniform in x.
Define $||f - g|| := \sup\{|f(x) - g(x)|; x \in D\}$ to be a *distance* between the functions f, g.

Main Theorem on Uniformly Convergent Sequences of Functions.
Assume that the continuous functions $f_n : (a, b) \to \mathbb{R}$ are uniformly convergent and define $f(x) := \lim_{n\to\infty} f_n(x)$.

Claim: The limit function f is continuous.

Proof. Since the statement is different in nature from all previous claims, we also meet a new method of proof which I quote as "the $\epsilon/3$-argument". It is sufficient to prove continuity of the limit function at an arbitrary $x \in (a,b)$.
Let $\epsilon > 0$ be given. Because the sequence is uniformly convergent we can choose n_ϵ, such that

$$x \in (a,b),\ n \geq n_\epsilon \Rightarrow |f(x) - f_n(x)| \leq \frac{\epsilon}{3}.$$

Then pick any one $n^* \geq n_\epsilon$ and use the continuity of f_{n^*} at x to find $\delta_x > 0$ such that

$$x, y \in (a,b),\ |x-y| < \delta_x \Rightarrow |f_{n^*}(y) - f_{n^*}(x)| \leq \frac{\epsilon}{3}.$$

The triangle inequality gives the claim, f is continuous at x:

$$\begin{aligned} x, y \in (a,b),\ |x-y| < \delta_x \Rightarrow |f(x) - f(y)| = \\ &= |(f(x) - f_{n^*}(x)) + (f_{n^*}(x) - f_{n^*}(y)) + (f_{n^*}(y) - f(y))| \leq \\ &\leq |f(x) - f_{n^*}(x)| + |f_{n^*}(x) - f_{n^*}(y)| + |f_{n^*}(y) - f(y)| \leq \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}. \end{aligned}$$

Remark. Observe the huge difference to earlier determinations of δ. In the past, δ could mostly be determined as some explicit and reasonably simple function of ϵ. In this new case the proof gives no quantitative control of δ. If we decrease ϵ we will usually have to take a larger n^* so that δ_x has to be chosen to fit a different function f_{n^*}, a function that may have a substantially worse continuity behavior. Application of this theorem gives no quantitative control on the limit function. This fits, because the definition of continuity also does not require any quantitative control.
Since we did get quantitative control of the limit function in the case of pointwise convergent equicontinuous sequences, this looks like the stronger assumption.
We ask: Do such sequences in fact converge uniformly, at least on bounded closed intervals $[a,b]$? They do:

Assumption. Let $\{f_n : [a,b] \mapsto \mathbb{R}^n\}$ be a sequence of equicontinuous functions which converge pointwise.
Claim. The f_n converge in fact uniformly!
Proof. Given $\epsilon > 0$ we find, for each $x \in [a,b]$ by equicontinuity, a $\delta_\epsilon(x) > 0$ such that for all $n \in \mathbb{N}$

$$y \in [a,b],\ |x-y| < 2\delta_\epsilon(x) \Rightarrow |f_n(x) - f_n(y)| < \epsilon/3$$

holds. *Finitely* many of the intervals $(x - \delta_\epsilon(x),\ x + \delta_\epsilon(x))$ cover $[a,b]$. Call their centers $x_1, \ldots, x_k, \ldots, x_K$. Since the f_n converge pointwise we can find n_ϵ such that for all $k = 1, \ldots, K$ holds:

$$m, n \geq n_\epsilon \Rightarrow |f_n(x_k) - f_m(x_k)| \leq \epsilon/3.$$

For each $x \in [a,b]$ choose an x_k such that $x \in (x_k - \delta_\epsilon(x_k),\ x_k + \delta_\epsilon(x_k))$. Then we have $m, n \geq n_\epsilon \Rightarrow |f_n(x) - f_m(x) \leq |f_n(x) - f_n(x_k)| + |f_n(x_k) - f_m(x_k)| + |f_m(x_k) - f_m(x)| \leq \epsilon$, proving uniform convergence.

The two theorems on limit functions have immediate applications to integration or differentiation of the limit function of a convergent sequence of functions.

Theorem on Integration of Limit Functions. Assume that the functions $f_n : [a,b] \to \mathbb{R}$ are continuous and pointwise convergent, $\lim f_n(x) =: f(x)$. We proved already that the f_n have antiderivatives, hence are integrable. What can be said about f?

a) If the $\{f_n\}$ converge *uniformly*, then we have for the limit function

$$\lim \int_a^b f_n(x)dx = \int_a^b \lim f_n(x)dx = \int_a^b f(x)dx.$$

b) If the continuity behavior of the f_n is *uniform* in n (the f_n are equicontinuous) then

$$\lim \int_a^b f_n(x)dx = \int_a^b \lim f_n(x)dx = \int_a^b f(x)dx.$$

Proof. The assumptions a) or b) allow, to use the two theorems on continuity of limit functions. Therefore f is continuous, hence integrable.
Case a), uniform convergence, we have simple estimates for integrals:

$$||f-g|| \leq \epsilon \Rightarrow |\int_a^b f(x)dx - \int_a^b g(x)dx| \leq \int_a^b ||f-g||dx = ||f-g|| \cdot |b-a|.$$

Insert $g = f_n,\ n \geq n_\epsilon$ to obtain $\lim \int = \int \lim$, integral and uniform limit can be interchanged.

Case b). We just proved that under the assumptions b) the pointwise convergence is in fact uniform on $[a,b]$ and the statement follows from part a).

Theorem on Differentiation of limit functions of convergent function sequences. Let $f_n : [a,b] \to \mathbb{R}$ be continuously differentiable (i. e. f_n' is continuous). Assume that the sequences $\{f_n(a)\}$ and $\{f_n'(x)\}$ converge for all $x \in [a,b]$. Put $\lim f_n(a) =: f(a),\ \lim f_n'(x) =: g(x)$. These assumptions are *not* sufficient to guarantee differentiability of a limit function f of the sequence $\{f_n\}$ and to prove $f' = g$.
Assume in addition that the sequence of the *derivatives* $\{f_n'\}$ satisfies a) or b) in the previous theorem on integration of convergent sequences.
Claim: f_n converges to a differentiable limit function f and one can interchange differentiation and limit:

$$f'(x) = \Big(\lim_{n\to\infty} f_n(x)\Big)' = \lim_{n\to\infty} (f_n'(x)) = g(x).$$

Proof. Since f_n' has the antiderivative f_n, we have $\int_a^x f_n'(t)dt = f_n(x) - f_n(a)$.
Since the integrals on the left and $\{f_n(a)\}$ converge as $n \to \infty$, we conclude first that

$\{f_n(x)\}$ converges. Moreover, the left side converges to an antiderivative G of g:

$$G(x) := \int_a^x g(t)dt = \lim_{n\to\infty} \int_a^x f'_n(t)dt = f(x) - f(a).$$

Finally, $f(x) = f(a) + G(x)$ shows $f'(x) = G'(x) = g(x)$ and finishes the proof.

Remark: The arguments presented above to deal with limits of continuous functions survive many generalizations. They did not change much during the last 100 years and may have reached their final form. I expect that readers can generalize these arguments to handle limits of continuous curves $c_n : I \to \mathbb{R}^d$. Generalizations to maps $F_n : \mathbb{R}^m \to \mathbb{R}^d$ do not need much additional thought either. – By contrast, it is not obvious how to define differentiability of such maps $F : \mathbb{R}^m \to \mathbb{R}^d$. It needs even more thought to invert differentiation; attempts to define antiderivatives meet new phenomena that do not exist for 1-dimensional functions. Before dealing with higher dimensional differentiation it is therefore advantageous to have experience with continuity arguments. Therefore it is usually taken as natural to treat continuity before differentiation also for 1-dimensional functions. I have changed this order to be able to start from the knowledge the students already have. Moreover we can now, immediately after introducing continuity, construct very surprising continuous functions. In the usual order of things these functions are widely separated from the definition of continuity, because differentiability is treated after continuity, but before limits of sequences of functions.

Examples of Continuous Functions

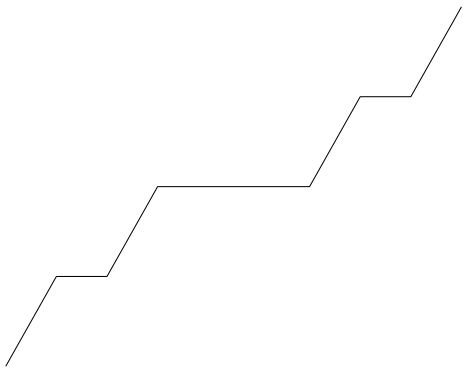
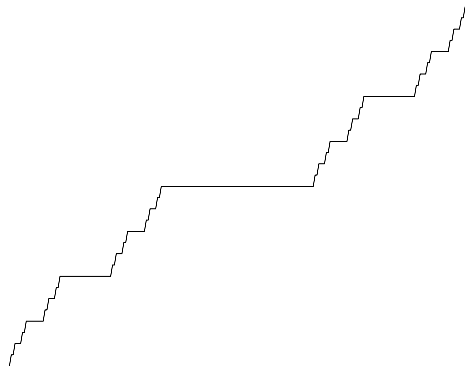
The intuitive aspect of functions, that what we imagine as their *normal* behavior, is of course strongly influenced by the functions we know best. The explicitly computable *Rational Functions* will be the first candidates for many people. They create too harmless a picture of the typical behavior of continuous functions. I will therefore present three examples, which, to an intuition educated only by rational functions, have a very unusual behavior. They are sometimes dismissed as exotic examples. However, if one adds to any of these strange functions for example a rational function, then the behavior called “strange” persists. Therefore we need to view it as the normal behavior of continuous functions. To have an adequate intuition of the basic objects in a field is important because part of the mathematical education consists in getting used to drastically abbreviated arguments that still lead to correct conclusions.
The first example is the

Cantor Staircase. Recall that a differentiable function with vanishing derivative, $f' = 0$, is necessarily constant. A plausible attempt to weaken the assumptions of this important result is refuted by the Cantor Staircase:
This functions is continuous and monotone increasing from 0 to 1 – while its derivative exists and is zero “almost everywhere” (see below)!
The function is constructed by piecewise linear (hence continuous) monotone approximations $f_k : [0,1] \to [0,1]$. As k increases the f_k get constant on more and more subintervals of $[0,1]$ and already coincide with the limit function on these constancy intervals.
The first piecewise linear function is:

$$f_1(x) := \begin{cases} \frac{3}{2} \cdot x & \text{for } 0 \le x \le \frac{1}{3} \\ \frac{1}{2} & \text{for } \frac{1}{3} \le x \le \frac{2}{3} \\ \frac{3}{2} \cdot x - \frac{1}{2} & \text{for } \frac{2}{3} \le x \le 1 \end{cases}$$

The further functions f_k are recursively defined. To get f_{k+1} from f_k, scale its graph horizontally by one third and vertically by one half; two of these decreased copies define f_{k+1} on the left and right third of $[0,1]$, in the middle third f_{k+1} is constant $= 1/2$. With formulas:

$$f_{k+1}(x) := \begin{cases} \frac{1}{2} \cdot f_k(3x) & \text{for } 0 \le x \le \frac{1}{3} \\ \frac{1}{2} & \text{for } \frac{1}{3} \le x \le \frac{2}{3} \\ \frac{1}{2} \cdot (1 + f_k(3x-2)) & \text{for } \frac{2}{3} \le x \le 1 \end{cases}$$

Two approximations of the monotone continuous Cantor staircase, $f' = 0$ almost everywhere.
The continuity of these approximations is uniform (equicontinuous) in n:

$$x, y \in [0,1], |x-y| < 3^{-n} \quad \Rightarrow \quad |f_k(x) - f_k(y)| < 2^{-n}$$

The approximations converge uniformly in x (geometrically dominated):

$$|f_k(x) - f_{k+1}(x)| \le \tfrac{1}{3} \cdot 2^{-(k+1)}.$$

The function f_k is constant on $(2^k - 1)$ disjoint subintervals having there increasing values $j \cdot 2^{-k}$, $(j = 1, \ldots, 2^k - 1)$. I call these intervals "plateau intervals". At the boundary and between these plateau intervals we have 2^k intervals of length 3^{-k} on which f_k increases from one plateau value to the next with slope $(3/2)^k$. I call these intervals "ladder intervals". The recursive definition achieves, that f_{k+1} and f_k agree on the plateau intervals of f_k. Each ladder interval of f_k is subdivided into three subintervals. Each middle third becomes a plateau interval of f_{k+1} with value in the middle between adjacent plateau values. The other subintervals are the 2^{k+1} ladder intervals of f_{k+1}; the slope of f_{k+1} is 3/2-times as large as the slopes of f_k. In particular (as claimed above):

$$x \in [0,1] \;\Rightarrow\; |f_{k+1}(x) - f_k(x)| \le \tfrac{1}{3} \cdot 2^{-(k+1)}.$$

$$x, y \in [0,1], |x-y| < 3^{-n} \quad \Rightarrow \quad |f_k(x) - f_k(y)| < 2^{-n}$$

The sequence $\{f_k\}$ of continuous functions is uniformly (in x) geometrically dominated, thus they converge to a continuous limit function. (Since the approximations are also equicontinuous (error control is uniform in k) we have a second proof of the continuity of the limit function.) This limit function agrees with f_k on all open plateau intervals of f_k. It is therefore differentiable there with derivative 0. All other points lie in 2^k intervals (ladder intervals of f_k) of total length $(2/3)^k$, which converges to 0. This is summarized in the formuation: The Cantor staircase is "almost everywhere" differentiable with derivative 0.

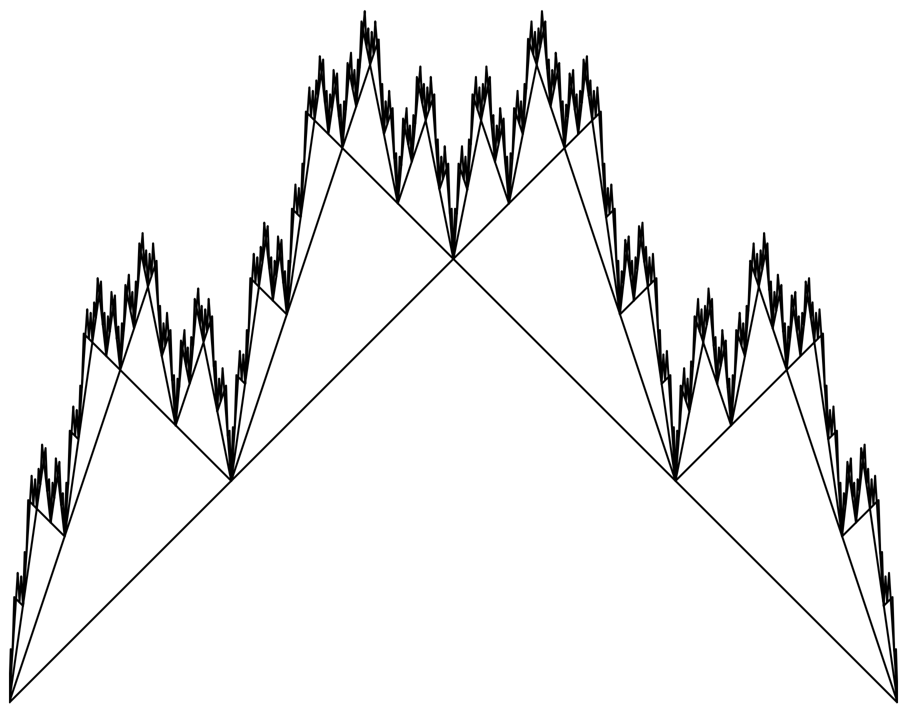
Example 2: The Weierstraß Oscillator is a continuous but nowhere differentiable function. As before it is constructed from uniformly geometrically dominated functions. Building block is the saw-tooth function (abbreviation: floor(x) resp. ceil(x) are those integers that are nearest to x from below, resp. from above):
$\text{sawtooth}(x) := \min(x - \text{floor}\,(x), \text{ceil}\,(x) - x) = \text{distance}(x, \mathbb{Z})$.

$$f_n(x) := \sum_{k=1}^{n} 2^{-k} \cdot \text{sawtooth}(4^k \cdot x).$$

Clearly we have for all x: $\qquad |f_n(x) - f_{n+1}(x)| \le 2^{-(n+1)}$,
i.e. the sequence $\{f_n\}$ is a uniformly convergent sequence of continuous functions. Therefore, it has a *continuous* limit function $W(x) := \lim f_n(x)$. We call it Weierstraß oscillator (note that in this case we do not have equicontinuous approximations). We need to argue, why W is at no point differentiable. Note first that at all rational points with denominator 4^{-n} we can compute the value of the limit function from its n-th approximation, hence:

$$\begin{aligned} x := j \cdot 4^{-n},\ \Delta x = 0.5 \cdot 4^{-n} \ \Rightarrow\ & \text{sawtooth}(4^n x) = 0,\ f_{n-1}(x) = f_n(x) = \ldots = W(x) \\ & f_n(x + \Delta x) = f_{n+1}(x + \Delta x) = \ldots = W(x + \Delta x) \\ & W(x + \Delta x) \ge W(x) + 2^{-(n+1)} \\ & \frac{W(x + \Delta x) - W(x)}{\Delta x} \ge 2^n \end{aligned}$$

Each number with denominator 4^{-n} can be written with denominator 4^{-N} for all $N \ge n$. Therefore we have from all points $j \cdot 4^{-n}$ secants with slope $\ge 2^N$ over intervals of length $0.5 \cdot 4^{-N}$, so that at these points no limit $\lim_{N\to\infty}$ of the secant slopes exists.
Finally, let $\tilde{x} \in (j \cdot 4^{-n},\ j \cdot 4^{-n} + 0.5 \cdot 4^{-n}) = (x_0, x_1)$. Then we have
$W(x_1) - W(x_0) \ge 2^{-(n+1)}$, hence $\max(|W(\tilde{x}) - W(x_0)|,\ |W(\tilde{x}) - W(x_1)|) \ge 0.5 \cdot 2^{-(n+1)}$ and $\max(\tilde{x} - x_1,\ x_2 - \tilde{x}) \le 2^{-2n-1}$. This shows that we also have from $\tilde{x}$ secants with absolute value of their slope $\ge 2^n$. Again, a finite limit of the secant slopes starting from $(\tilde{x},\ W(\tilde{x}))$ cannot exist.
In addition it is possible to show that the graph of W has infinite length over every interval of non-zero length. The popular phrase for creating an intuition of a continuous function W: "one can draw the graph of W with a pen without lifting the pen" does not make sense for the Weierstraß oscillator, because one needs to draw with infinite velocity in directions that do not exist.
Estimates for the length become simpler for the slight modification $\sum 2^{-n} \cdot \text{sawtooth}(8^n x)$.

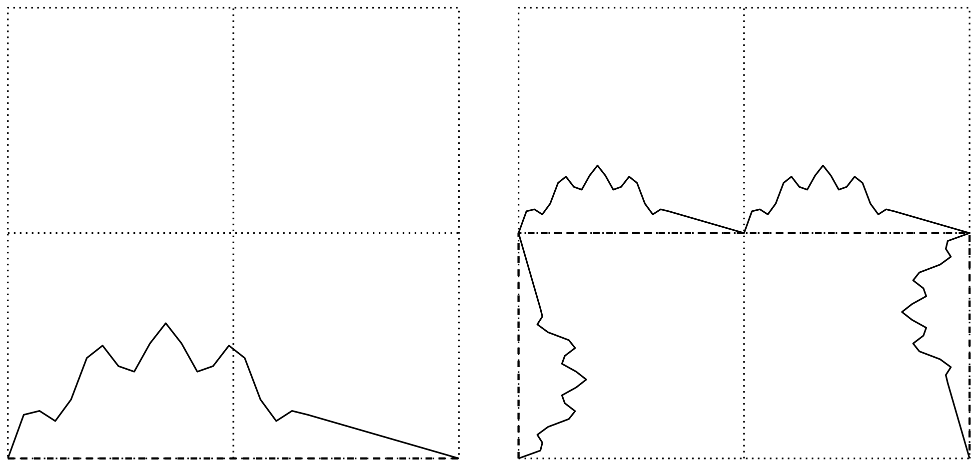


Weierstraß' continuous, nowhere differentiable function is limit
of the uniformly convergent sequence: $W(x) = \sum_{n=1}^{\infty} 2^{-n} \cdot \text{sawtooth}(4^n x)$.

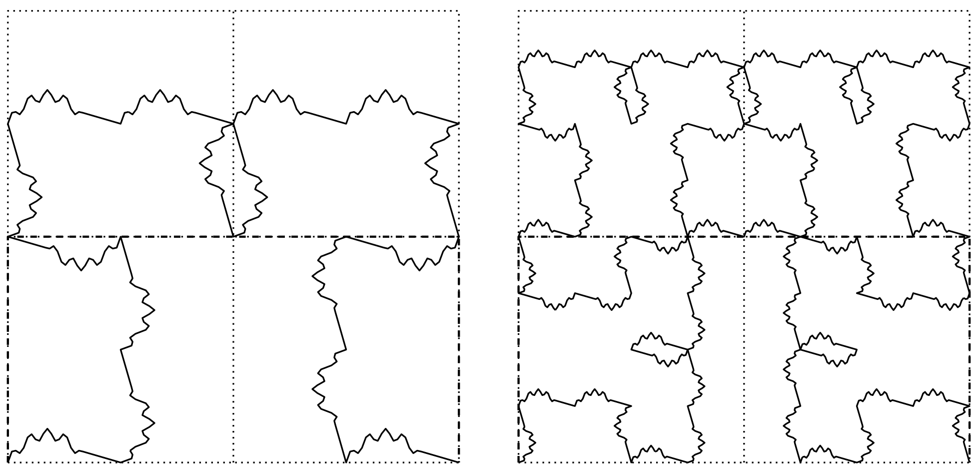
Example 3: Hilbert's Square-filling Curve. Graphs of continuous functions are "curves" which have a 1-1 projection to the x-axis. Typical continuous curves are more complicated: Hilbert's square filling curve passes through *all* points of the unit square. Again, we construct this curve as limit of uniformly convergent approximations. All one needs to understand is, how to go from one approximation to the next.
Let $c_1 : [0,1] \to [0,1] \times [0,1]$ be a continuous curve with $c_1(0) = (0,0),\ c_1(1) = (1,0)$ (it joins the two bottom vertices of the square and lies in the square).
$\tilde{c}(x) := \frac{1}{2} c_1(4 \cdot x),\ 0 \le x \le 0.25$ is a curve of half the original size, but parametrized by an interval of one quarter the original domain. Therefore we can join four of these half-sized curves to perform the recursive step to the next approximation $c_2 : [0,1] \to [0,1] \times [0,1]$.
The endpoints of the four smaller copies are
$c_2(0) = (0,0),\ c_2(0.25) = (0,0.5),\ c_2(0.5) = (0.5,0.5),\ c_2(0.75) = (1,0.5),\ c_2(1) = (1,0)$,
and, they lie each in one of the following sub-squares:
$[0,0.5] \times [0,0.5],\ [0,0.5] \times [0.5,1],\ [0.5,1] \times [0.5,1],\ [0.5,1] \times [0,0.5]$. The curve c_2 again joins the two bottom vertices of the unit square (see figures).
We have: if each point of the unit square has a distance $\le a$ to the image of c_1, then each point of the four sub-squares has a distance $\le \frac{1}{2} \cdot a$ from the corresponding subarc of c_2. It is also clear that:

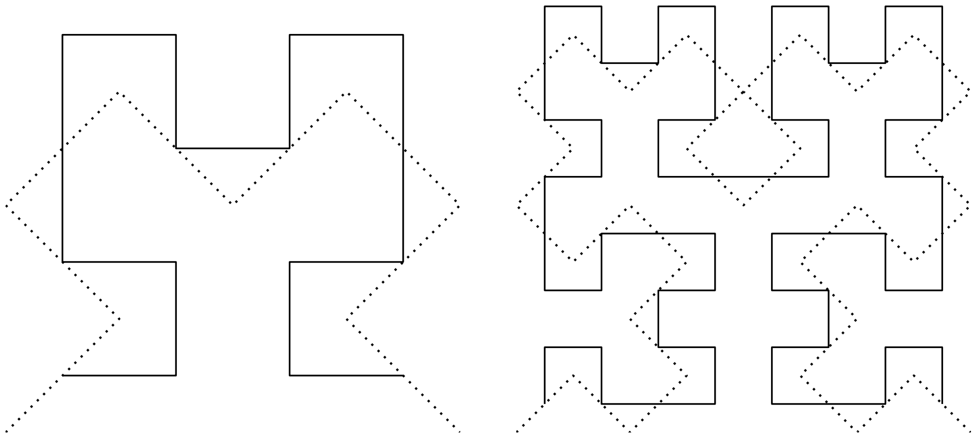
$$|c_3(s) - c_2(s)| \le \tfrac{1}{2} \cdot |c_2(s) - c_1(s)|.$$

This shows that the iteratively defined sequence $\{c_k\}$ is uniformly geometrically dominated, and therefore converges to a *continuous* limit curve. Furthermore, *each* point P of the unit square is limit of the sequence P_k of the points that are closest to P on c_k. Therefore P is on the limit curve and the whole unit square is filled by the image of the limit curve. – Pictures of the approximating curves are a bit confusing, because with each iteration the number of double points increases. Hilbert has therefore modified the construction so that all approximations are injectiv, see the following curves made of segments parallel to the axes. – The double points are simple from another point of view: They are already points on the limit curve and all points on the limit curve are limits of double points.

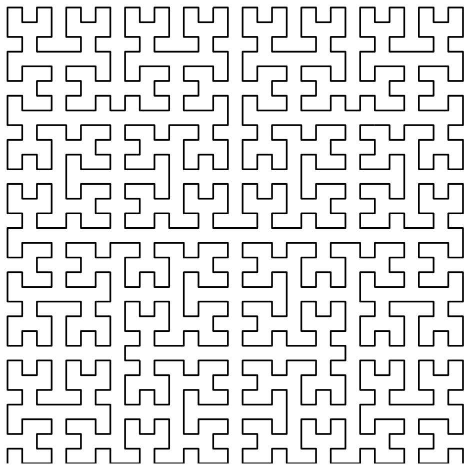


Hilbert's square filling curve. Hilbert's iteration of some arbitrary curve, which joins $(0, 0)$ and $(1, 0)$, consists of four half-sized copies which are put together along the dashed iteration guide. The procedure is repeated with the iterated curve.





Hilbert's injective approximations join the midpoints of the segments of the dotted Hilbert-iterations of the initial curve $x \to \min(x, 1-x)$.



Looking back at these examples, note that all three were made with uniformly geometrically dominated approximations. We used this fundamental tool before and we will meet it again when we come to *ordinary differential equations*.

However, since in each case the step from one approximation to the next was rather simple, we should expect that these examples are still rather mild cases of continuity and our intuition of continuity is still in its infancy.