CONSTRUCTIVE ANALYSIS WITH WITNESSES

HELMUT SCHWICHTENBERG

Contents

Date: April 18, 2012.

We are interested in *exact real numbers*, as opposed to floating point numbers. The final goal is to develop the basics of real analysis in such a way that from a proof of an existence formula one can extract a program. For instance, from a proof of the intermediate value theorem we want to extract a program that, given an arbitrary error bound 2^{-k}, computes a rational x where the given function is zero up to the error bound.

Why should we be interested in logic in a study of constructive analysis? There are at least two reasons.

(1) Obviously we need to be aware of the difference of the classical and the constructive existential quantifier, and try to prove the stronger statements involving the latter whenever possible. Then one is forced to give "constructive" proofs, whose algorithmic content can be "seen" and then used as a basis to formulate a program for computing the solution. This was the point of view in Bishop's classic textbook [3] (and its successor [5]), and more explicitely carried through in Andersson's Master's thesis [2] (based on Palmgren's [18]), with Mathematica as the target programming language.
(2) However, one can go one step further and automatize the step from the (formalized) constructive proof to the corresponding program. This can be done by means of the so-called realizability interpretation, whose existence was clear from the beginnings of constructive logic. The desire to have "mathematics as a numerical language" in this sense was clearly expressed by Bishop in his article [4] (with just that title). There are now many implementations of these ideas, for instance Nuprl, Coq, Minlog and recently also Isabelle, to mention only a few.

What are the requirements on a constructive logic that should guide us in our design?

- It should be as close as possible to the mathematical arguments we want to use. Variables should carry (functional) types, with free algebras (e.g., natural numbers) as base types. Over these, inductive definitions and the corresponding introduction and elimination axioms should be allowed.
- The constants of the language should denote computable functionals in the Scott-Ersov sense, and hence the higher-order quantifiers should range over their (mathematically correct) domain, the partial continuous functionals.
- The language of the logic should be strong (in the sense of being expressive), but the existence axioms used should be weak.
- Type parameters (ML style) should be allowed, but quantification over types should be disallowed in order to keep the theory predicative. Similarly, predicate variables should be allowed as placeholders for properties, but quantification over them should be disallowed, again to ensure predicativity.

On the technical side, since we need to actually construct formal proofs, we want to have some machine support in building them. In particular, to simplify equational reasoning, the system should identify terms with the same "normal form", and we should be able to add rewrite rules used to generate normal forms. Decidable predicates should be implemented via boolean valued functions, so that the rewrite mechanism applies to them as well.

Compared with the literature, the novel aspect of the present work is the development of elementary constructive analysis in such a way that witnesses have as low a type level as possible. This clearly is important for the complexity of the extracted programs. Here are some examples.

(1) A continuous function on the reals is determined by its values on the rationals, and hence can be represented by a type-one (rather than type-two) object.
(2) In the proof that the range of a continuous function on a compact intervall has a supremum, Brouwer's notion of a *totally bounded* set of reals (which has type-level two) is replaced by the notion of being *located above* (which has type-level one).
(3) The Cauchy-Euler construction of approximate solutions to ordinary differential equations can be seen as a type-level one process.

Acknowledgement. Part of the material in these notes was the subject of two proseminars at the Mathematics department, University of Munich, in the Sommersemester 2004 and 2005. I would like to thank the participating students for their very useful contributions. Also, at many points I have used arguments form Otto Forster's well-known textbook [13].

1. Real numbers

1.1. **Approximation of $\sqrt{2}$.** To motivate real numbers, we show that there is a Cauchy sequence of rational numbers that does not converge to a rational number. First we show

Lemma (Irrationality of $\sqrt{2}$). *There is no rational number b with $b^2 = 2$.*

Proof. Assume $b = \frac{n}{m} \in \mathbb{Q}$ such that $n^2 = 2m^2$. The number of prime factors 2 in n^2 ist even; however, it is odd in $2m^2$. This contradicts the uniqueness of prime factorization of natural numbers. □

Theorem (Approximation of $\sqrt{a}$)**.** *Let $a > 0$ and $a_0 > 0$ be given. Define the sequence a_n recursively by*

$$a_{n+1} := \frac{1}{2}\Big(a_n + \frac{a}{a_n}\Big).$$

Then

(a) $(a_n)_{n\in\mathbb{N}}$ *is a Cauchy sequence.*
(b) *If $\lim_{n\to\infty} a_n = c$, then $c^2 = a$.*

Proof. By induction on n one can see easily that $a_n > 0$ for all $n \in \mathbb{N}$. Moreover,

$$a_{n+1}^2 \geq a \quad \text{for all } n; \tag{1}$$

this follows from

$$a_{n+1}^2 - a = \frac{1}{4}\Big(a_n^2 + 2a + \frac{a^2}{a_n^2}\Big) - a = \frac{1}{4}\Big(a_n^2 - 2a + \frac{a^2}{a_n^2}\Big) = \frac{1}{4}\Big(a_n - \frac{a}{a_n}\Big)^2 \geq 0.$$

Next

$$a_{n+2} \leq a_{n+1} \quad \text{for all } n, \tag{2}$$

since

$$a_{n+1} - a_{n+2} = a_{n+1} - \frac{1}{2}\Big(a_{n+1} + \frac{a}{a_{n+1}}\Big) = \frac{1}{2a_{n+1}}\Big(a_{n+1}^2 - a\Big) \geq 0.$$

Let

$$b_n := \frac{a}{a_n}.$$

Then $b_{n+1}^2 \leq a$ for all n, since by (1) we have $\frac{1}{a_{n+1}^2} \leq \frac{1}{a}$, hence also

$$b_{n+1}^2 = \frac{a^2}{a_{n+1}^2} \leq \frac{a^2}{a} = a.$$

From (2) we obtain $b_{n+1} \leq b_{n+2}$ for all n. Next we have

$$b_{n+1} \leq a_{m+1} \quad \text{for all } n, m \in \mathbb{N}. \tag{3}$$

To see this, observe that – say for $n \geq m$ – we have $b_{n+1} \leq a_{n+1}$ (this follows from (1) by multiplying with $1/a_{n+1}$), and $a_{n+1} \leq a_{m+1}$ by (2).

We now show

$$a_{n+1} - b_{n+1} \leq \frac{1}{2^n}(a_1 - b_1), \tag{4}$$

by induction on n. Basis: for $n = 0$ both sides are equal. Step:

$$\begin{aligned} a_{n+2} - b_{n+2} \le a_{n+2} - b_{n+1} &= \frac{1}{2}(a_{n+1} + b_{n+1}) - b_{n+1} \\ &= \frac{1}{2}(a_{n+1} - b_{n+1}) \le \frac{1}{2^{n+1}}(a_1 - b_1) \quad \text{by IH.} \end{aligned}$$

$(a_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, since for $n \le m$ by (2), (3) and (4)

$$|a_{n+1} - a_{m+1}| = a_{n+1} - a_{m+1} \le a_{n+1} - b_{n+1} \le \frac{1}{2^n}(a_1 - b_1).$$

Now assume $\lim a_n = c$. Then also $\lim b_n = c$, for

$$|c - b_{n+1}| \le |c - a_{n+1}| + |a_{n+1} - b_{n+1}|$$

and both summands can be made arbitrarily small for large n, by (4). Hence

$$c^2 = (\lim b_n)^2 = \lim b_n^2 \le a \le \lim a_n^2 = (\lim a_n)^2 = c^2$$

because of $b_{n+1}^2 \le a \le a_{n+1}^2$, and therefore $c^2 = a$. □

1.2. **Reals, equality of reals.** We shall view a real as a Cauchy sequence of rationals with a separately given modulus.

Definition. A real number x is a pair $((a_n)_{n\in\mathbb{N}}, M)$ with $a_n \in \mathbb{Q}$ and $M \colon \mathbb{N} \to \mathbb{N}$ such that $(a_n)_n$ is a Cauchy sequence with modulus M, that is

$$|a_n - a_m| \le 2^{-k} \quad \text{for } n, m \ge M(k).$$

and M is weakly increasing. M is called Cauchy modulus of x.

We shall loosely speak of a real $(a_n)_n$ if the Cauchy modulus M is clear from the context or inessential. Every rational a is tacitly understood as the real represented by the constant sequence $a_n = a$ with the constant modulus $M(k) = 0$.

Definition. Two reals $x := ((a_n)_n, M)$, $y := ((b_n)_n, N)$ are called *equivalent* (or *equal* and written $x = y$, if the context makes clear what is meant), if

$$|a_{M(k+1)} - b_{N(k+1)}| \le 2^{-k} \quad \text{for all } k \in \mathbb{N}$$

We want to show that this is an equivalence relation. Reflexivity and symmetry are clear. For transitivity we use the following lemma:

Lemma (RealEqChar). *For reals $x := ((a_n)_n, M)$, $y := ((b_n)_n, N)$ the following are equivalent:*

(a) *$x = y$;*
(b) $\forall_k \exists_q \forall_{n \ge q}\, |a_n - b_n| \le 2^{-k}$.

Proof. (a) → (b). For $n \geq M(k+2), N(k+2)$ we have

$$|a_n - b_n| \leq |a_n - a_{M(k+2)}| + |a_{M(k+2)} - b_{N(k+2)}| + |b_{N(k+2)} - b_n| \\ \leq 2^{-k-2} + 2^{-k-1} + 2^{-k-2}.$$

(b) → (a). Let $l \in \mathbb{N}$ and $n \geq q, M(k+1), N(k+1)$ with q provided by (b). Then

$$|a_{M(k+1)} - b_{N(k+1)}| \leq |a_{M(k+1)} - a_n| + |a_n - b_n| + |b_n - b_{N(k+1)}| \\ \leq 2^{-k-1} + 2^{-l} + 2^{-k-1}.$$

The claim follows, because this holds for every $l \in \mathbb{N}$. □

Lemma. *Equality between reals is transitive.*

Proof. Let $(a_n)_n$, $(b_n)_n$, $(c_n)_n$ be the Cauchy sequences for x, y, z. Assume $x = y$, $y = z$ and pick p, q according to the lemma above. Then $|a_n - c_n| \leq |a_n - b_n| + |b_n - c_n| \leq 2^{-k-1} + 2^{-k-1}$ for $n \geq p, q$. □

1.3. **The Archimedian axiom.** For every function on the reals we certainly want compatibility with equality. This however is not always the case; here is an important example.

Lemma (RealBound). *For every real $x := ((a_n)_n, M)$ we can find an upper bound 2^{k_x} on the elements of the Cauchy sequence: $|a_n| \leq 2^{k_x}$ for all n.*

Proof. Let k_x be such that $\max\{\, |a_n| \mid n \leq M(0)\,\} + 1 \leq 2^{k_x}$, hence $|a_n| \leq 2^{k_x}$ for all n. □

Clearly this assignment of k_x to x is not compatible with equality.

1.4. **Nonnegative and positive reals.** A real $x := ((a_n)_n, M)$ is called *nonnegative* (written $x \in \mathbb{R}^{0+}$) if

$$-2^{-k} \leq a_{M(k)} \quad \text{for all } k \in \mathbb{N}.$$

It is *k-positive* (written $x \in_k \mathbb{R}^+$, or $x \in \mathbb{R}^+$ if k is not needed) if

$$2^{-k} \leq a_{M(k+1)}.$$

We want to show that both properties are compatible with equality. First we prove a useful characterization of nonnegative reals.

Lemma (RealNNegChar). *For a real $x := ((a_n)_n, M)$ the following are equivalent:*
(a) *$x \in \mathbb{R}^{0+}$;*
(b) *$\forall_k \exists_p \forall_{n \geq p} -2^{-k} \leq a_n$.*

Proof. (a) ⇒ (b). For $n \geq M(k+1)$ we have

$$\begin{aligned} -2^{-k} &\leq -2^{-k-1} + a_{M(k+1)} \\ &= -2^{-k-1} + (a_{M(k+1)} - a_n) + a_n \\ &\leq -2^{-k-1} + 2^{-k-1} + a_n. \end{aligned}$$

(b) ⇒ (a). Let $l \in \mathbb{N}$ and $n \geq p, M(k)$ with p provided by (b) (for l). Then

$$\begin{aligned} -2^{-k} - 2^{-l} &\leq -2^{-k} + a_n \\ &= -2^{-k} + (a_n - a_{M(k)}) + a_{M(k)} \\ &\leq -2^{-k} + 2^{-k} + a_{M(k)}. \end{aligned}$$

The claim follows, because this holds for every l. □

Lemma (RealNNegCompat)**.** *If $x \in \mathbb{R}^{0+}$ and $x = y$, then also $y \in \mathbb{R}^{0+}$.*

Proof. Let $x := ((a_n)_n, M)$ and $y := ((b_n)_n, N)$. Assume $x \in \mathbb{R}^{0+}$ and $x = y$, and let k be given. Pick p according to the lemma above and q according to the characterization of equality of reals in 1.2 (both for $k+1$). Then for $n \geq p, q$

$$-2^{-k} \leq -2^{-k-1} + a_n \leq (b_n - a_n) + a_n.$$

Hence $y \in \mathbb{R}^{0+}$ by definition. □

We now show compatibility of positivity with equality. Again we need a lemma:

Lemma (RealPosChar)**.** *For a real $x := ((a_n)_n, M)$ the following are equivalent:*

(a) $\exists_k\ x \in_k \mathbb{R}^+$.

(b) $\exists_{l,p} \forall_{n \geq p}\ 2^{-l} \leq a_n$.

For $\forall_{n \geq p}\ 2^{-l} \leq a_n$ write $x \in_{l,p} \mathbb{R}^+$ or $0 <_{l,p} x$.

Proof. (a) ⇒ (b). Assume $x \in_k \mathbb{R}^+$, that is $2^{-k} \leq a_{M(k+1)}$. Then

$$2^{-k-1} \leq -2^{-k-1} + a_{M(k+1)} = -2^{-k-1} + (a_{M(k+1)} - a_n) + a_n \leq a_n$$

for $M(k+1) \leq n$. Hence we can take $l := k+1$ and $p := M(k+1)$.

(b) ⇒ (a).

$$\begin{aligned} 2^{-l-1} &< -2^{-l-2} + 2^{-l} \\ &\leq -2^{-l-2} + a_n && \text{fr } p \leq n \\ &\leq (a_{M(l+2)} - a_n) + a_n && \text{fr } M(l+2) \leq n. \end{aligned}$$

Hence we can take $k := l + 1$; then x is k-positive. □

Lemma. *Positivity of reals is compatible with equality.*

Proof. Assume $0 <_{k,p} x$ and $x = y$, so in particular we have a q such that $\forall n.\ q \leq n \to |a_n - b_n| \leq 2^{-k-1}$. Then for $\max(p, q) \leq n$

$$2^{-k-1} = -2^{-k-1} + 2^k \leq (b_n - a_n) + a_n = b_n,$$

hence $0 <_{k+1,\max(p,q)} y$. □

1.5. **Arithmetical functions.** Given real numbers $x := ((a_n)_n, M)$ and $y := ((b_n)_n, N)$, we define $x + y$, $-x$, $|x|$, $x \cdot y$, and $\frac{1}{x}$ (the latter only provided that $|x| \in_l \mathbb{R}^+$) as represented by the respective sequence (c_n) of rationals with modulus L:

	c_n	$L(k)$
$x + y$	$a_n + b_n$	$\max\bigl(M(k+1), N(k+1)\bigr)$
$-x$	$-a_n$	$M(k)$
$\lvert x\rvert$	$\lvert a_n\rvert$	$M(k)$
$x \cdot y$	$a_n \cdot b_n$	$\max\bigl(M(k{+}1{+}k_{\lvert y\rvert}), N(k{+}1{+}k_{\lvert x\rvert})\bigr)$
$\frac{1}{x}$ for $\lvert x\rvert \in_l \mathbb{R}^+$	$\begin{cases} \frac{1}{a_n} & \text{if } a_n \neq 0 \\ 0 & \text{if } a_n = 0 \end{cases}$	$M(2(l+1)+k)$,

where 2^{k_x} is the upper bound provided by 1.3.

Lemma. *For reals x, y also $x+y$, $-x$, $|x|$, $x\cdot y$ and (provided that $|x| \in_l \mathbb{R}^+$) also $1/x$ are reals.*

Proof. We restrict ourselves to the cases $x \cdot y$ and $1/x$.

$$\begin{aligned} |a_n b_n - a_m b_m| &= |a_n(b_n - b_m) + (a_n - a_m)b_m| \\ &\leq |b_n - b_m| \cdot |a_n| + |a_n - a_m| \cdot |b_m| \\ &\leq |b_n - b_m| \cdot 2^{k_{|x|}} + |a_n - a_m| \cdot 2^{k_{|y|}} \leq 2^{-k} \end{aligned}$$

for $n, m \geq \max\bigl(M(k + 1 + k_{|y|}), N(k + 1 + k_{|x|})\bigr)$.

For $1/x$ assume $|x| \in_l \mathbb{R}^+$. Then by the (proof of our) characterization of positivity in 1.4, $2^{-l-1} \leq |a_n|$ for $M(l + 1) \leq n$. Hence

$$\begin{aligned} \Bigl|\frac{1}{a_n} - \frac{1}{a_m}\Bigr| &= \frac{|a_m - a_n|}{|a_n a_m|} \\ &\leq 2^{2(l+1)}|a_m - a_n| \quad \text{fr } n, m \geq M(l + 1) \\ &\leq 2^{-k} \quad \text{fr } n, m \geq M(2(l + 1) + k). \end{aligned}$$

The claim now follows from the assumption that M is weakly increasing. □

Lemma. *The functions $x + y$, $-x$, $|x|$, $x \cdot y$ and (provided that $|x| \in_l \mathbb{R}^+$) also $1/x$ are compatible with equality.*

Proof. Routine. □

Lemma. *For reals x, y, z*

$$x + (y + z) = (x + y) + z \qquad x \cdot (y \cdot z) = (x \cdot y) \cdot z$$
$$x + 0 = x \qquad x \cdot 1 = x$$
$$x + (-x) = 0 \qquad 0 < |x| \to x \cdot \frac{1}{x} = 1$$
$$x + y = y + x \qquad x \cdot y = y \cdot x$$
$$x \cdot (y + z) = x \cdot y + x \cdot z$$

Proof. Routine. □

Lemma (ProdIsOne). *For reals x, y from $x \cdot y = 1$ we can infer $0 < |x|$.*

Proof. Pick k such that $|b_n| \le 2^k$ for all n. Pick q such that $q \le n$ implies $1/2 \le a_n \cdot b_n$. Then for $q \le n$, $1/2 \le |a_n| \cdot 2^k$, and hence $2^{-k-1} \le |a_n|$. □

Lemma. *For reals x, y,*
(a) $x, y \in \mathbb{R}^{0+} \to x + y, x \cdot y \in \mathbb{R}^{0+}$,
(b) $x, y \in \mathbb{R}^{+} \to x + y, x \cdot y \in \mathbb{R}^{+}$,
(c) $x \in \mathbb{R}^{0+} \to -x \in \mathbb{R}^{0+} \to x = 0$.

Proof. (a), (b). Routine. (c). Let k be given. Pick p such that $-2^{-k} \le a_n$ and $-2^{-k} \le -a_n$ for $n \ge p$. Then $|a_n| \le 2^{-k}$. □

1.6. **Comparison of reals.** We write $x \le y$ for $y - x \in \mathbb{R}^{0+}$ and $x < y$ for $y - x \in \mathbb{R}^+$. Unwinding the definitions yields that $x \le y$ is to say that for every k, $a_{L(k)} \le b_{L(k)} + 2^{-k}$ with $L(k) := \max(M(k), N(k))$, or equivalently (using the characterization of reals in 1.4) that for every k there exists p such that $a_n \le b_n + 2^{-k}$ for all $n \ge p$. Furthermore, $x < y$ is a shorthand for the presence of k with $a_{L(k+1)} + 2^{-k} \le b_{L(k+1)}$ with L the maximum of M and N, or equivalently (using the characterization of reals in 1.4) for the presence of k, q with $a_n + 2^{-k} \le b_n$ for all $n \ge q$; we then write $x <_k y$ (or $x <_{k,q} y$) whenever we want to call these witnesses.

Lemma (RealApprox). $\forall_{x,k} \exists_a \, |a - x| \le 2^{-k}$.

Proof. Let $x = ((a_n), M)$. Given k, pick $a_{M(k)}$. We show $|a_{M(k)} - x| \le 2^{-k}$, that is $|a_{M(k)} - a_{M(l)}| \le 2^{-k} + 2^{-l}$ for every l. But this follows from

$$|a_{M(k)} - a_{M(l)}| \le |a_{M(k)} - a_{M(k+l)}| + |a_{M(k+l)} - a_{M(l)}| \le 2^{-k} + 2^{-l}. \quad \square$$

Lemma (PlusPos)**.** *$0 \le x$ and $0 <_k y$ imply $0 <_{k+1} x + y$.*

Proof. From $0 \le x$ we have $\forall_l \exists_p \forall_{n \ge p} \; -2^{-l} \le a_n$. From $0 <_k y$ we have some q such that $\forall_{n \ge q} \; 2^{-k} \le b_n$. Pick p for $k+1$. Then $p, q \le n$ implies $0 \le a_n + 2^{-k-1}$ and $2^{-k-1} \le b_n - 2^{-k-1}$, hence $2^{-k-1} \le a_n + b_n$. □

Lemma. *For reals x, y, z,*

$$x \le x \qquad x \not< x$$
$$x \le y \to y \le x \to x = y \qquad x < y \to y < z \to x < z$$
$$x \le y \to y \le z \to x \le z \qquad x < y \to x + z < y + z$$
$$x \le y \to x + z \le y + z \qquad x < y \to 0 < z \to x \cdot z < y \cdot z$$
$$x \le y \to 0 \le z \to x \cdot z \le y \cdot z$$

Proof. From 1.5. □

Lemma. *$x \le y \to y <_k z \to x <_{k+1} z$.*

Proof. This follows from the next to last lemma. □

As is to be expected in view of the existential and universal character of the predicates $<$ and $\le$ on the reals, we have:

Lemma. *$x \le y \leftrightarrow y \not< x$.*

Proof. $\to$. Assume $x \le y$ and $y < x$. By the previous lemma we obtain $x < x$, a contradiction.

$\leftarrow$. It clearly suffices to show $0 \not< z \to z \le 0$, for a real z given by $(c_n)_n$. Assume $0 \not< z$. We must show $\forall_k \exists_p \forall_n (p \le n \to c_n \le 2^{-k})$. Let k be given. By assumption $0 \not< z$, hence $\neg \exists_l (2^{-l} \le c_{M(l+1)})$. For $l := k+1$ this implies $c_{M(k+2)} < 2^{-k-1}$, hence $c_n \le c_{M(k+2)} + 2^{-k-2} < 2^{-k}$ for $M(k+2) \le n$. □

Constructively, we cannot compare two reals, but we can compare every real with a nontrivial interval.

Lemma (ApproxSplit)**.** *Let x, y, z be given and assume $x < y$. Then either $z \le y$ or $x \le z$.*

Proof. Let $x := ((a_n)_n, M)$, $y := ((b_n)_n, N)$, $z := ((c_n)_n, L)$. Assume $x <_k y$, that is (by definition) $1/2^k \le b_p - a_p$ for $p := \max(M(k+2), N(k+2))$. Let $q := \max(p, L(k+2))$ and $d := (b_p - a_p)/4$.

Case $c_q \le \frac{a_p + b_p}{2}$. We show $z \le y$. It suffices to prove $c_n \le b_n$ for $n \ge q$. This follows from

$$c_n \le c_q + \frac{1}{2^{k+2}} \le \frac{a_p + b_p}{2} + \frac{b_p - a_p}{4} = b_p - \frac{b_p - a_p}{4} \le b_p - \frac{1}{2^{k+2}} \le b_n.$$

Case $c_q \not\leq \frac{a_p+b_p}{2}$. We show $x \leq z$. This follows from $a_n \leq c_n$ for $n \geq q$:

$$a_n \leq a_p + \frac{1}{2^{k+2}} \leq a_p + \frac{b_p - a_p}{4} \leq \frac{a_p + b_p}{2} - \frac{b_p - a_p}{4} \leq c_q - \frac{1}{2^{k+2}} \leq c_n. \quad \square$$

Notice that the boolean object determining whether $z \leq y$ or $x \leq z$ depends on the representation of x, y and z. In particular this assignment is not compatible with our equality relation.

One might think that the non-available comparison of two reals could be circumvented by using a maximum function. Indeed, such a function can easily be defined (component wise), and it has the expected properties $x, y \leq \max(x, y)$ and $x, y \leq z \to \max(x, y) \leq z$. However, what is missing is the knowledge that $\max(x, y)$ equals one of its arguments, i.e., we do not have $\max(x, y) = x \vee \max(x, y) = y$.

However, in many cases it is sufficient to pick the up to ε largest real out of finitely many given ones. This is indeed possible. We give the proof for two reals; it can be easily generalized.

Lemma (Maximum of two reals)**.** *Let* $x := ((a_n)_n, M)$ *and* $y := ((b_n)_n, N)$ *be reals, and* $k \in \mathbb{N}$. *Then for* $\varepsilon := 2^{-k}$ *either* $x \leq y + \varepsilon$ *or else* $y \leq x + \varepsilon$.

Proof. Let $p := \max\bigl(M(k+1), N(k+1)\bigr)$.

Case $a_p \leq b_p$. Then for $p \leq n$

$$a_n \leq a_p + \frac{\varepsilon}{2} \leq b_p + \frac{\varepsilon}{2} \leq b_n + \varepsilon.$$

This holds for all n, therefore $x \leq y + \varepsilon$.

Case $b_p < a_p$. Then for $p \leq n$

$$b_n \leq b_p + \frac{\varepsilon}{2} < a_p + \frac{\varepsilon}{2} \leq a_n + \varepsilon.$$

This holds for all n, therefore $y \leq x + \varepsilon$. $\square$

1.7. **Non-countability.** Recall that every rational a is tacitly understood as the real represented by the constant sequence $a_n = a$ with the constant modulus $M(k) = 0$.

Lemma ($\mathbb{Q}$ is dense in $\mathbb{R}$)**.** *For any two reals* $x < y$ *there is a rational* a *such that* $x < a < y$.

Proof. Let $z := (x + y)/2$ be given by $(c_n)_n$. Then for some k we have $x <_k z <_k y$. Let $a := c_{M(k+1)}$, with M the Cauchy modulus of z. $\square$

Notice that a depends on the representations of x and y.

Theorem (Cantor)**.** *Let a sequence (x_n) of reals be given. Then we can find a real y with $0 \leq y \leq 1$ that is apart from every x_n, in the sense that $x_n < y \vee y < x_n$.*

Proof. We construct sequences $(a_n)_n$, $(b_n)_n$ of rationals such that for all n

$$0 = a_0 \leq a_1 \leq \cdots \leq a_n < b_n \leq \cdots \leq b_1 \leq b_0 = 1, \tag{5}$$

$$x_n < a_{n+1} \vee b_{n+1} < x_n, \tag{6}$$

$$b_n - a_n \leq 2^{-n}. \tag{7}$$

Let $a_0, \dots, a_n$ and $b_0, \dots, b_n$ be already constructed such that (5)-(7) hold (as far as they are defined). Now compare the real x_n with $a_n < b_n$.

Case 1. $x_n < b_n$. Let $b_{n+1} := b_n$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, we can find a rational a_{n+1} such that

$$\max\bigl(x_n, a_n, b_n - 2^{-n-1}\bigr) < a_{n+1} < b_{n+1} = b_n.$$

Case 2. $a_n < x_n$. Let $a_{n+1} := a_n$, and find a rational b_{n+1} such that

$$a_n = a_{n+1} < b_{n+1} < \min\bigl(x_n, b_n, a_n + 2^{-n-1}\bigr).$$

Clearly (5)-(7) continue to hold for $n+1$ (as far as defined). Now $y := (a_n)_n$ is a Cauchy sequence, since for $m \geq n$ we have $|a_m - a_n| = a_m - a_n \leq b_n - a_n \leq 2^{-n}$. Similarly $z := (b_n)_n$ is a Cauchy sequence. $y = z$ follows from (7), and from (6) together with (5) we obtain $x_n < y \vee z < x_n$. □

1.8. **Cleaning of reals.** After some computations involving real numbers it is to be expected that the rational numbers occurring in the Cauchy sequences may become rather complex. Hence under computational aspects it is necessary to be able to *clean up* a real, as follows.

Proof. Let $c_n := \lfloor a_{M(n)} \cdot 2^n \rfloor$ and $b_n := c_n \cdot 2^{-n}$, hence

$$\frac{c_n}{2^n} \leq a_{M(n)} < \frac{c_n}{2^n} + \frac{1}{2^n} \quad \text{with } c_n \in \mathbb{Z}.$$

Then for $m \leq n$

$$\begin{aligned}
|b_m - b_n| &= |c_m \cdot 2^{-m} - c_n \cdot 2^{-n}| \\
&\leq |c_m \cdot 2^{-m} - a_{M(m)}| + |a_{M(m)} - a_{M(n)}| + |a_{M(n)} - c_n \cdot 2^{-n}| \\
&\leq 2^{-m} + 2^{-m} + 2^{-n} \\
&< 2^{-m+2},
\end{aligned}$$

hence $|b_m - b_n| \leq 2^{-k}$ for $n \geq m \geq k + 2 =: N(k)$, so $(b_n)_n$ is a Cauchy sequence with modulus N.

To prove that x is equivalent to $y := ((b_n)_n, N)$, observe

$$\begin{aligned}|a_n - b_n| &\le |a_n - a_{M(n)}| + |a_{M(n)} - c_n \cdot 2^{-n}| \\ &\le 2^{-k-1} + 2^{-n} \quad \text{for } n, M(n) \ge M(k+1) \\ &\le 2^{-k} \quad \text{if in addition } n \ge k+1.\end{aligned}$$

Hence $|a_n - b_n| \le 2^{-k}$ for $n \ge \max(k+1, M(k+1))$, and therefore $x = y$. □

2. Sequences and series of real numbers

2.1. Completeness.

Definition. A sequence $(x_n)_{n\in\mathbb{N}}$ of reals is a Cauchy *sequence* with modulus $M : \mathbb{N} \to \mathbb{N}$ whenever $|x_n - x_m| \le 2^{-k}$ for $n, m \ge M(k)$, and *converges* with modulus $M : \mathbb{N} \to \mathbb{N}$ to a real y, its *limit*, whenever $|x_n - y| \le 2^{-k}$ for $n \ge M(k)$.

Clearly the limit of a convergent sequence of reals is uniquely determined.

Lemma (RatCauchyConvMod)**.** *Every modulated Cauchy sequence of rationals converges with the same modulus to the real number it represents.*

Proof. Let $x := ((a_n)_n, M)$ be a real. We must show $|a_n - x| \le 2^{-k}$ for $n \ge M(k)$. Fix $n \ge M(k)$. It suffices to show $|a_n - a_m| \le 2^{-k}$ for $m \ge M(k)$. But this holds by assumption. □

By the triangle inequality, every convergent sequence of reals with modulus M is a Cauchy sequence with modulus $k \mapsto M(k+1)$. We now prove the reverse implication.

Theorem (Sequential Completeness)**.** *For every Cauchy sequence of reals we can find a real to which it converges.*

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence of reals with modulus M; say x_n is $((a_{nk})_k, N_n)$. Note first that, for each $n \in \mathbb{N}$ and every p, by the lemma above we have $|x_n - a_{nl}| \le 2^{-p}$ for all $l \ge N_n(p)$. Next, set

$$b_n := a_{nN_n(n)}$$

for every $n \in \mathbb{N}$, so that

$$|x_n - b_n| \le 2^{-n} \quad \text{for all } n \in \mathbb{N}$$

by the particular case $l = N_n(n)$ of the foregoing consideration. Then

$$|b_m - b_n| \le |b_m - x_m| + |x_m - x_n| + |x_n - b_n| \le 2^{-m} + 2^{-q-1} + 2^{-n} \le 2^{-q}$$

for all $m, n \geq \max(M(q+1), q+2)$, which is to say that $y := (b_n)_n$ is a Cauchy sequence with modulus $L(q) := \max(M(q+1), q+2)$. Moreover, again by the lemma above

$$|x_n - y| \leq |x_n - b_n| + |b_n - y| \leq 2^{-n} + 2^{-q-1} \leq 2^{-q}$$

for all $n \geq L(q+1)$. In other words: (x_n) converges to y with modulus $q \mapsto L(q+1)$. □

One can even say that (x_n) converges to y with the same modulus that (x_n) has as a Cauchy sequence. More generally, the lemma above holds for Cauchy sequences of reals as well.

Lemma (RealCauchyConvMod)**.** *Every modulated Cauchy sequence of reals converges with the same modulus to its limit.*

Proof. Let $(x_n)_n$ be a Cauchy sequence of reals with modulus M, that is

$$|x_n - x_m| \leq 2^{-k} \quad \text{for } n, m \geq M(k).$$

Let y be the limit of $(x_n)_n$, that is

$$|x_n - y| \leq 2^{-l} \quad \text{for } n \geq L(l).$$

We shall prove

$$|x_n - y| \leq 2^{-k} \quad \text{for } n \geq M(k).$$

Fix $n \geq M(k)$, and let $l \in \mathbb{N}$. Then

$$\begin{aligned} |x_n - y| &\leq |x_n - x_m| + |x_m - y| \quad \text{for } m \geq M(k), L(l) \\ &\leq 2^{-k} + 2^{-l}. \end{aligned}$$

The claim follows, because this holds for every l. □

It will be useful to have a criterion for convergence of a sequence of reals, in terms of their approximations.

Lemma. *For reals x_n, x represented by $(a_{nk})_k, (b_k)_k$, we can infer that $(x_n)_n$ converges to x, i.e.,*

$$\forall_p \exists_q \forall_{n \geq q} \, |x_n - x| \leq 2^{-p}$$

from

$$\forall_p \exists_q \forall_{n,k \geq q} \, |a_{nk} - b_k| \leq 2^{-p}.$$

Proof. Given p, we have to find q such that $|x_n - x| \leq 2^{-p}$ for $n > q$. By the characterization of non-negative reals in 1.4 it suffices to have $|a_{nk} - b_k| \leq 2^{-p} + 2^{-l}$ for $k \geq r$ with r depending on l. But by assumption we even have $|a_{nk} - b_k| \leq 2^{-p}$ for $k \geq q$. □

2.2. **Limits and inequalities.** We show that limits interact nicely with non-strict inequalities.

Lemma (RealNNegLim)**.** *Let* $(x_n)_{n\in\mathbb{N}}$ *be a convergent sequence of reals and* x *its limit. Then* $0 \le x_n$ *for all* n *implies* $0 \le x$.

Proof. By assumption $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence of reals, say with modulus M. Let x_n be $((a_{nk})_k, N_n)$. Assume $0 \le x_n$ for all n, that is

$$-2^{-k'} \le a_{nN_n(k')} \quad \text{for all } n, k' \in \mathbb{N}.$$

By the theorem above, $b_n := a_{nN_n(n)}$ is a Cauchy sequence with modulus $L(k) := \max(M(k+1), k+2)$ representing x. We must show $0 \le x$, that is

$$-2^{-k} \le b_{L(k)} \quad \text{for all } k \in \mathbb{N}.$$

For $k' := L(k)$ and $n := L(k)$ we obtain

$$-2^{-k} \le -2^{-L(k)} \le a_{L(k)N_{L(k)}(L(k))} = b_{L(k)} \quad \text{for all } k \in \mathbb{N}.$$

Note that $k < L(k)$ by definition of L. □

2.3. **Series.** Series are special sequences. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of reals, and define

$$s_n := \sum_{k=0}^{n} x_k.$$

We call s_n a *partial sum* of the sequence (x_n). The sequence

$$(s_n)_{n\in\mathbb{N}} = \Big(\sum_{k=0}^{n} x_k\Big)_{n\in\mathbb{N}} =: \sum_{k=0}^{\infty} x_k$$

is called the *series* determined by the sequence $(x_n)_{n\in\mathbb{N}}$. We say that the series $\sum_{k=0}^{\infty} x_k$ converges if and only if the sequence (s_n) converges. Its limit is somewhat sloppily denoted by $\sum_{k=0}^{\infty} x_k$ as well.

Example. Consider the series $\sum_{k=1}^{\infty} \frac{1}{k(k+1)}$. Its partial sums are

$$s_n := \sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1};$$

this can be proved by induction on n, as follows. For $n = 0$ the claim clearly holds, and in the induction step $n \mapsto n+1$ we have

$$\begin{aligned}\sum_{k=1}^{n+1} \frac{1}{k(k+1)} &= \sum_{k=1}^{n} \frac{1}{k(k+1)} + \frac{1}{(n+1)(n+2)} \\ &= \frac{n}{n+1} + \frac{1}{(n+1)(n+2)} \\ &= \frac{n(n+2)+1}{(n+1)(n+2)} \\ &= \frac{(n+1)^2}{(n+1)(n+2)} \\ &= \frac{n+1}{n+2}.\end{aligned}$$

Because of $\lim_{n\to\infty} \frac{n}{n+1} = 1$ we obtain $\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1$.

Theorem (Infinite geometric series)**.** *For* $|x| < 1$ *we have*

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}.$$

Proof. Let $|x| < 1$. The n-th partial sum is

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

which can be proved easily by induction. Hence

$$\lim_{n\to\infty} \frac{1-x^{n+1}}{1-x} = \frac{1}{1-x}\big(1 - \lim_{n\to\infty} x^{n+1}\big) = \frac{1}{1-x},$$

since $\lim x^{n+1} = 0$ for $|x| < 1$. □

For instance, $\sum_{n=-k}^{\infty} a_n 2^{-n}$ with $a_n \in \{-1, 0, 1\}$ converges, because

$$\Big|\sum_{n=m+1}^{l} a_n 2^{-n}\Big| \le \sum_{n=m+1}^{l} 2^{-n} < \sum_{n=m+1}^{\infty} 2^{-n} = 2^{-m}.$$

We show that every real x can be written in this form.

2.4. **Signed digit representation of reals.**

Theorem (Signed digit representation of reals). *Every real x can be represented in the form*

$$\sum_{n=-k}^{\infty} a_n 2^{-n} \quad \text{with } a_n \in \{-1, 0, 1\}. \tag{8}$$

Proof. By 1.3 we can find k such that $-2^{k+1} \leq x \leq 2^{k+1}$. We recursively construct $a_{-k}, a_{-k+1}, \dots, a_m, \dots$ such that

$$-2^{-m} \leq x - \sum_{n=-k}^{m} a_n 2^{-n} \leq 2^{-m} \quad \text{for } m \geq -k-1.$$

For $m = -k-1$ this holds by the choice of k. Now assume the claim holds for m; we need to construct a_{m+1} such that it holds for $m+1$ as well. Let $y := x - \sum_{n=-k}^{m} a_n 2^{-n}$, hence $-2^{-m} \leq y \leq 2^{-m}$. By comparing y first with $-2^{-m-1} < 0$ and then $0 < 2^{-m-1}$ we can define a_{m+1} such that

$$a_{m+1} = \begin{cases} -1 & \text{if } y \leq 0 \\ 0 & \text{if } -2^{-m-1} \leq y \leq 2^{-m-1} \\ 1 & \text{if } 0 \leq y. \end{cases}$$

Then in each of the three cases

$$(a_{m+1} - 1)2^{-m-1} \leq y \leq (a_{m+1} + 1)2^{-m-1},$$

hence

$$-2^{-m-1} \leq y - a_{m+1}2^{-m-1} \leq 2^{-m-1},$$

which was to be shown. □

2.5. **Convergence tests.** We now consider some of the standard convergence tests for series.

Theorem (Cauchy convergence test). *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of reals. The series $\sum_{n=0}^{\infty} a_n$ converges if and only if for every $k \in \mathbb{N}$ there is an $N \in \mathbb{N}$ such that for all $n \geq m \geq N$*

$$\Big|\sum_{i=m}^{n} a_i\Big| \leq 2^{-k}.$$

Proof. The condition expresses that the sequence of partial sums is a Cauchy sequence. □

It follows that the convergence of series does not depend on a possible change of finitely many of its members. Howver, the limit of the series may well change.

Theorem. *A necessary (but not sufficient) condition for the convergence of a series* $\sum_{n=0}^{\infty} a_n$ *is* $\lim_{n\to\infty} a_n = 0$.

Proof. Assume $\sum_{n=0}^{\infty} a_n$ is convergent. We must show $\lim_{n\to\infty} a_n = 0$, which means

$$\forall_{k\in\mathbb{N}}\exists_{N\in\mathbb{N}}\forall_{n\geq N}|a_n| \leq 2^{-k}.$$

So let $k \in \mathbb{N}$. Then there is an $N \in \mathbb{N}$ such that for all $n \geq m \geq N$

$$\Big|\sum_{k=m}^{n} a_k\Big| \leq 2^{-k}.$$

In particular then $|a_n| \leq 2^{-k}$ for $n \geq N$. □

Example. The *harmonic series* $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges to $+\infty$. This can be seen by grouping its members together:

$$1 + \frac{1}{2} + \underbrace{\Big(\frac{1}{3} + \frac{1}{4}\Big)}_{\geq \frac{2}{4} = \frac{1}{2}} + \underbrace{\Big(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\Big)}_{\geq \frac{4}{8} = \frac{1}{2}} + \dots.$$

More precisely, one first shows that for all $p \in \mathbb{N}$

$$\sum_{k=2^p+1}^{2^{p+1}} \frac{1}{k} \geq 2^p \cdot \frac{1}{2^{p+1}} = \frac{1}{2}.$$

This implies

$$\sum_{k=1}^{2^{n+1}} \frac{1}{k} = 1 + \sum_{p=0}^{n} \sum_{k=2^p+1}^{2^{p+1}} \frac{1}{k} \geq 1 + \sum_{p=0}^{n} \frac{1}{2},$$

which implies the claim. The harmonic series is an example that the condition $\lim_{n\to\infty} a_n = 0$ does *not* ensure convergence of the series $\sum_{n=0}^{\infty} a_n$.

Theorem 2.1 (Leibniz test for alternating series)**.** *Let* $(x_n)_{n\in\mathbb{N}}$ *be a decreasing sequence of non-negative reals with* $\lim_{n\to\infty} x_n = 0$. *Then the series*

$$\sum_{n=0}^{\infty} (-1)^n x_n$$

converges.

Proof. Because of $\lim_{n\to\infty} x_n = 0$ it suffices to show

$$\forall_m \forall_n (0 \le (-1)^n \sum_{k=n}^{n+m} (-1)^k x_k \le x_n).$$

The proof is by induction on m. For $m = 0$ the claim is $0 \le (-1)^{2n} x_n = x_n$, and in the step $m \mapsto m+1$ we have

$$\begin{aligned}(-1)^n \sum_{k=n}^{n+m+1} (-1)^k x_k &= (-1)^n((-1)^n x_n + \sum_{k=n+1}^{n+m+1} (-1)^k x_k) \\ &= x_n - (-1)^{n+1} \sum_{k=n+1}^{n+1+m} (-1)^k x_k,\end{aligned}$$

and by induction hypothesis $0 \le \sum_{k=n+1}^{n+1+m} (-1)^k x_k \le x_{n+1}$. □

For example, the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ converges by the Leibniz test.

Definition. A series $\sum_{n=0}^{\infty} x_n$ is *absolutely convergent* if $\sum_{n=0}^{\infty} |x_n|$ converges.

Clearly every absolutely convergent series is convergent. The converse does not hold generally, by the example above.

Theorem 2.2 (Comparison test)**.** *Let $\sum_{n=0}^{\infty} y_n$ be a convergent series with non-negative y_n. If $|x_n| \le y_n$ for all $n \in \mathbb{N}$, then $\sum_{n=0}^{\infty} x_n$ is absolutely convergent.*

Proof. We have to show that $\sum_{n=0}^{\infty} |x_n|$ converges. Let $k \in \mathbb{N}$. Since $\sum_{n=0}^{\infty} y_n$ converges, we have an $N \in \mathbb{N}$ such that for all $n \ge m \ge N$

$$\sum_{k=m}^{n} y_k \le 2^{-k}.$$

But then also

$$\sum_{k=m}^{n} |x_k| \le \sum_{k=m}^{n} y_k \le 2^{-k}.$$ □

Example. The series $\sum_{n=1}^{\infty} \frac{1}{n^k}$ converges for every $k \ge 2$. To see this, recall that $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ converges, hence also $\sum_{n=1}^{\infty} \frac{2}{n(n+1)}$. Because of $k \ge 2$ we have for all $n \ge 1$

$$\frac{1}{n^k} \le \frac{1}{n^2} \le \frac{2}{n(n+1)}.$$

Hence by the comparison test $\sum_{n=1}^{\infty} \frac{1}{n^k}$ converges as well.

Theorem 2.3 (Ratio test)**.** *Assume*

$$|x_{n+1}| \leq q|x_n| \quad \textit{for all } n \geq n_0$$

with $0 \leq q < 1$. *Then the series* $\sum_{n=0}^{\infty} x_n$ *is absolutely convergent.*

Proof. Since the convergence of series does not depend on a possible change of finitely many of its members, we may assume $n_0 = 0$. By assumption we have for all n

$$|x_n| \leq q^n |x_0|;$$

this can be seen easily by induction. The geometric series $\sum_{n=0}^{\infty} q^n$ converges (because of $0 \leq q < 1$), hence also $\sum_{n=0}^{\infty} q^n |x_0|$. From the comparison test we can conclude the absolute convergence of $\sum_{n=0}^{\infty} x_n$. □

Example. The series $\sum_{n=1}^{\infty} \frac{n^2}{2^n}$ converges. To see this, observe that for all $n \geq 3$

$$\frac{(n+1)^2 \cdot 2^n}{2^{n+1} \cdot n^2} = \frac{1}{2}\Big(1 + \frac{1}{n}\Big)^2 \leq \frac{1}{2} \cdot \frac{16}{9} = \frac{8}{9} < 1.$$

Hence the series converges by the ratio test.

2.6. **Reordering.** Let $\sum_{n=0}^{\infty} x_n$ be a series. If $\tau \colon \mathbb{N} \to \mathbb{N}$ is a bijective map, then the series $\sum_{n=0}^{\infty} x_{\tau(n)}$ is a *reordering* of $\sum_{n=0}^{\infty} x_n$.

Theorem (Reordering theorem)**.** *Let* $\sum_{n=0}^{\infty} x_n$ *be absolutely convergent with limit* x. *Then every reordering of it converges to* x *as well.*

Proof. (See [13]). We must show

$$\lim_{m \to \infty} \sum_{k=0}^{m} x_{\tau(k)} = x.$$

Let $k \in \mathbb{N}$. Because of the absolute convergence of $\sum_{n=0}^{\infty} x_n$ we have an n_0 such that

$$\sum_{k=n_0}^{\infty} |x_k| \leq 2^{-k-1}.$$

Hence

$$\Big|x - \sum_{k=0}^{n_0-1} x_k\Big| = \Big|\sum_{k=n_0}^{\infty} x_k\Big| \leq \sum_{k=n_0}^{\infty} |x_k| \leq 2^{-k-1}.$$

Now choose N such that $\{\tau(0), \tau(1), \dots, \tau(N)\} \supseteq \{0, 1, \dots, n_0 - 1\}$. Then for all $m \geq N$

$$\Big|\sum_{k=0}^{m} x_{\tau(k)} - x\Big| \leq \Big|\sum_{k=0}^{m} x_{\tau(k)} - \sum_{k=0}^{n_0-1} x_k\Big| + \Big|\sum_{k=0}^{n_0-1} x_k - x\Big|$$

$$\leq \sum_{k=n_0}^{\infty} |x_k| + 2^{-k-1} \leq 2^{-k}. \qquad \square$$

2.7. **The exponential series.**

Theorem. *The* exponential series

$$\exp(x) := \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

is absolutely convergent, for every real x.

Proof.

$$\Big|\frac{x^{n+1}}{(n+1)!}\Big| \leq \frac{1}{2}\Big|\frac{x^n}{n!}\Big|$$

is equivalent to $2|x| \leq n+1$. Hence the series converges absolutely by the ratio test. $\square$

The Euler number e is defined as

$$e := \exp(1) = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

Theorem (Estimate of the rest)**.**

$$\Big|\sum_{n=N+1}^{\infty} \frac{x^n}{n!}\Big| \leq 2\frac{|x|^{N+1}}{(N+1)!} \quad \textit{for } |x| \leq 1 + \tfrac{N}{2}.$$

Proof.

$$\begin{aligned}
&\Big|\sum_{n=N+1}^{\infty} \frac{x^n}{n!}\Big| \\
&\leq \sum_{n=N+1}^{\infty} \frac{|x^n|}{n!} \\
&= \frac{|x|^{N+1}}{(N+1)!}\Big(1 + \frac{|x|}{N+2} + \dots + \frac{|x|^k}{(N+2)\dots(N+k+1)} + \dots\Big).
\end{aligned}$$

For $\frac{|x|}{N+2} \leq \frac{1}{2}$ or $|x| \leq 1+\frac{N}{2}$ we can estimate this series against the geometric series, since

$$\frac{|x|^k}{(N+2)\dots(N+k+1)} \leq \Big(\frac{|x|}{N+2}\Big)^k \leq \frac{1}{2^k}.$$

Hence for $|x| \leq 1 + \frac{N}{2}$

$$\Big|\sum_{n=N+1}^{\infty} \frac{x^n}{n!}\Big| \leq \frac{|x|^{N+1}}{(N+1)!}\sum_{k=0}^{\infty}\frac{1}{2^k} = 2\frac{|x|^{N+1}}{(N+1)!}. \qquad \square$$

Theorem (Cauchy product)**.** *Assume* $\sum_{n=0}^{\infty} x_n$ *and* $\sum_{n=0}^{\infty} y_n$ *are absolutely convergent, and define*

$$z_n := \sum_{k=0}^{n} x_{n-k} y_k.$$

Then $\sum_{n=0}^{\infty} z_n$ *is absolutely convergent as well, and*

$$\sum_{n=0}^{\infty} z_n = \Big(\sum_{n=0}^{\infty} x_n\Big) \cdot \Big(\sum_{n=0}^{\infty} y_n\Big).$$

Proof. (See [13]). Define

$$Z_n := \sum_{k=0}^{n} z_k.$$

We first show

$$\lim_{n\to\infty} Z_n = \sum_{k=0}^{\infty} z_k = \Big(\sum_{n=0}^{\infty} x_n\Big) \cdot \Big(\sum_{n=0}^{\infty} y_n\Big).$$

For

$$Z_n^* := \Big(\sum_{k=0}^{n} x_k\Big) \cdot \Big(\sum_{k=0}^{n} y_k\Big),$$

we clearly have

$$\lim_{n\to\infty} Z_n^* = \Big(\sum_{n=0}^{\infty} x_n\Big) \cdot \Big(\sum_{n=0}^{\infty} y_n\Big).$$

Hence it suffices to show

$$\lim_{n\to\infty} (Z_n^* - Z_n) = 0.$$

To prove this, consider

$$P_n^* := \Big(\sum_{k=0}^{n} |x_k|\Big) \cdot \Big(\sum_{k=0}^{n} |y_k|\Big).$$

Since by assumption both $\sum_{n=0}^{\infty} x_n$ and $\sum_{n=0}^{\infty} y_n$ are absolutely convergent, $(P_n^*)_{n\in\mathbb{N}}$ converges.

Now let $k \in \mathbb{N}$. From the convergence of $(P_n^*)_{n\in\mathbb{N}}$ we obtain an N such that for all $n \geq m \geq N$

$$P_n^* - P_m^* = \sum_{\substack{i,j\leq n\\ m<\max(i,j)}} |x_i||x_j| \leq 2^{-k}.$$

Hence for $n \geq 2N$

$$\begin{aligned} |Z_n^* - Z_n| = \Big|\sum_{\substack{i,j\leq n\\ n<i+j}} x_i x_j\Big| \leq \sum_{\substack{i,j\leq n\\ n<i+j}} |x_i||x_j| &\leq \sum_{\substack{i,j\leq n\\ N<\max(i,j)}} |x_i||x_j| \\ &= P_n^* - P_N^* \leq 2^{-k}. \end{aligned}$$

It remains to show that $\sum_{n=0}^\infty z_n$ is absolutely convergent. This follows from the comparison test and the previous arguments, applied to the series $\sum_{n=0}^\infty |x_n|$ und $\sum_{n=0}^\infty |y_n|$ instead of $\sum_{n=0}^\infty x_n$ and $\sum_{n=0}^\infty y_n$. For then

$$\sum_{n=0}^\infty \sum_{k=0}^n |x_{n-k}||y_k|$$

converges to $(\sum_{n=0}^\infty |x_n|) \cdot (\sum_{n=0}^\infty |y_n|)$. Because of

$$|z_n| = \Big|\sum_{k=0}^n x_{n-k} y_k\Big| \leq \sum_{k=0}^n |x_{n-k}||y_k|$$

the comparison test implies the absolute convergence of $\sum_{n=0}^\infty z_n$. □

If instead of the absolute convergence of $\sum_{n=0}^\infty x_n$ and $\sum_{n=0}^\infty y_n$ we only assume ordinary convergence, $\sum_{n=0}^\infty z_n$ in general will not converge.

Theorem (Functional equation for the exponential function)**.**

$$\exp(x+y) = \exp(x)\exp(y) \quad \textit{for all } x, y \in \mathbb{R}.$$

Proof. Applying the Cauchy product to the absolutely convergent series

$$\exp(x) = \sum_{n=0}^\infty \frac{x^n}{n!} \quad \text{and} \quad \exp(y) = \sum_{n=0}^\infty \frac{y^n}{n!}$$

gives

$$\begin{aligned} \exp(x)\exp(y) &= \sum_{n=0}^\infty \sum_{k=0}^n \frac{x^{n-k}}{(n-k)!}\frac{y^k}{k!} \\ &= \sum_{n=0}^\infty \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k \end{aligned}$$

$$= \sum_{n=0}^{\infty} \frac{(x+y)^n}{n!} \quad \text{by the Binomial theorem}$$
$$= \exp(x+y). \qquad \square$$

Corollary. (a) $\exp(x) > 0$ *for all* $x \in \mathbb{R}$.
(b) $\exp(-x) = \exp(x)^{-1}$ *for all* $x \in \mathbb{R}$.
(c) $\exp(n) = e^n$ *for every integer* $n \in \mathbb{Z}$.

Proof. First notice

$$\exp(x)\exp(-x) = \exp(x-x) = \exp(0) = 1.$$

(a). For $x \geq 0$ we clearly have $\exp(x) \geq 1$. For $-\frac{1}{2} \leq x \leq \frac{1}{2}$,

$$\begin{aligned}
\sum_{n=0}^{\infty} \frac{x^n}{n!} &= 1 + x + \sum_{n=2}^{\infty} \frac{x^n}{n!} \\
&\geq 1 + x - \Big|\sum_{n=2}^{\infty} \frac{x^n}{n!}\Big| \\
&\geq 1 + x - 2 \cdot \frac{|x|^2}{2!} \quad \text{by the theorem on page 22} \\
&\geq 1 - |x| - |x|^2 \\
&\geq 1 - \frac{1}{2} - \frac{1}{4} = \frac{1}{4},
\end{aligned}$$

and for $x \leq 0$, $\exp(-x)\exp(x) = 1$ and $\exp(-x) \geq 1$ together imply $\exp(x) > 0$.

(b) is now immediate, and for (c) we use induction on n. Clearly $\exp(0) = 1 = e^0$; for $n \mapsto n+1$

$$\exp(n+1) = \exp(n)\exp(1) = e^n \cdot e = e^{n+1} \quad \text{by induction hypothesis,}$$

and for $n < 0$

$$\exp(n) = \frac{1}{\exp(-n)} = \frac{1}{e^{-n}} = e^n. \qquad \square$$

3. The exponential function for complex numbers

Later we shall define the sine and cosine functions by means of the complex exponential function, using the Euler equation

$$e^{ix} = \cos x + i \sin x.$$

As a preparation we introduce the complex numbers and prove their fundamental properties.

On the set $\mathbb{R} \times \mathbb{R}$ we define addition and multiplication by

$$(x_1, y_1) + (x_2, y_2) := (x_1 + x_2, y_1 + y_2),$$
$$(x_1, y_1) \cdot (x_2, y_2) := (x_1 x_2 - y_1 y_2, x_1 y_2 + y_1 x_2).$$

One can check easily that all the field axioms are satisfied if one defines $(0, 0)$ as zero and $(1, 0)$ als one. This field is called the field $\mathbb{C}$ of *complex numbers*. Because of

$$(x_1, 0) + (x_2, 0) = (x_1 + x_2, 0),$$
$$(x_1, 0) \cdot (x_2, 0) = (x_1 x_2, 0)$$

a real number x can be identified with the complex number $(x, 0)$; in this sense we have $\mathbb{R} \subseteq \mathbb{C}$.

Defining

$$i := (0, 1),$$

we obtain

$$i^2 = (0, 1)(0, 1) = (-1, 0) = -1;$$

therefore in the field of complex numbers there is an element whose square is the negative of the unit element. Every complex number $z = (x, y)$ can – using the above identification – be written in the form

$$z = x + iy.$$

x is called the *real part* $\Re(z)$ and y the *imaginary part* $\Im(z)$ of z. Clearly two complex numbers are equal if and only if they have the same real and imaginary parts.

Every complex number $z = x + iy$ can be viewed as point in the Gaußian plane. The real part x is the projection of z to the x-axis and the imaginary part y the projection to the y-axis.

For every complex number $z = x + iy$ we define the *conjugated complex number* $\overline{z}$ durch $\overline{z} := x - iy$. In the Gaußian plane the conjugated complex number is obtained by mirroring at the x-axis. One can check easily that for all $z, z_1, z_2 \in \mathbb{C}$

$$\overline{\overline{z}} = z, \quad \overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}, \quad \overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}.$$

Moreover for every $z \in \mathbb{C}$ we clearly have

$$\Re(z) = \frac{1}{2}(z + \overline{z}) \quad \text{and} \quad \Im(z) = \frac{1}{2i}(z - \overline{z}).$$

The *modulus* $|z|$ of a complex number z is defined by means of conjugated complex numbers; this will be useful for some of our later calculations. Let

$z = x + iy$ with $x, y \in \mathbb{R}$. Then

$$z\overline{z} = (x + iy)(x - iy) = x^2 + y^2 \geq 0,$$

and we can define

$$|z| := \sqrt{z\overline{z}}.$$

Because of $|z| = \sqrt{x^2 + y^2}$ we can view $|z|$ as the distance of the point z in the Gaußian plane from the orign. Observe that for $z \in \mathbb{R}$ the modulus as defined for real numbers coincides with the modulus for complex numbers as we just defined it. Also we clearly have $|z| = |\overline{z}|$.

Theorem. *For all $z, z_1, z_2 \in \mathbb{C}$ we have*
(a) $|z| \geq 0$*, and* $|z| = 0$ *iff* $z = 0$*.*
(b) $|z_1 z_2| = |z_1||z_2|$*,*
(c) $|z_1 + z_2| \leq |z_1| + |z_2|$ *(triangle inequality).*

Proof. (a) is clear.
(b) $|z_1 z_2|^2 = z_1 z_2 \overline{z_1 z_2} = z_1 z_2 \overline{z_1}\,\overline{z_2} = z_1 \overline{z_1} z_2 \overline{z_2} = |z_1|^2 |z_2|^2$.
(c)

$$\begin{aligned}
|z_1 + z_2|^2 &= (z_1 + z_2)(\overline{z_1} + \overline{z_2}) \\
&= z_1\overline{z_1} + z_1\overline{z_2} + z_2\overline{z_1} + z_2\overline{z_2} \\
&= |z_1|^2 + z_1\overline{z_2} + \overline{z_1\overline{z_2}} + |z_2|^2 \\
&= |z_1|^2 + 2\Re(z_1\overline{z_2}) + |z_2|^2 \\
&\leq (|z_1| + |z_2|)^2,
\end{aligned}$$

because of

$$\Re(z_1\overline{z_2}) \leq |z_1\overline{z_2}| = |z_1||\overline{z_2}| = |z_1||z_2|. \qquad \square$$

Remark. A field with a modulus function satisfying the three properties of the theorem above is called a *valued field*. $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ are valued fields.

The notions and results above concerning the convergence of sequences and series can be carried over routinely from reals to complex numbers.

Definition. A sequence $(c_n)_{n\in\mathbb{N}}$ of complex numbers is a Cauchy *sequence* with modulus $M : \mathbb{N} \to \mathbb{N}$ whenever $|c_m - c_n| \leq 2^{-k}$ for $m, n \geq M(k)$, and *converges* with modulus $M : \mathbb{N} \to \mathbb{N}$ to a complex number z, its *limit*, whenever $|c_n - z| \leq 2^{-k}$ for $n \geq M(k)$.

One can see easily that a sequence $(c_n)_{n\in\mathbb{N}}$ of complex numbers is a Cauchy sequence if and only if the two sequences of reals $(\Re(c_n))_{n\in\mathbb{N}}$ and

$(\Im(c_n))_{n\in\mathbb{N}}$ are, and that it converges if and only if the two sequences of reals $(\Re(c_n))_{n\in\mathbb{N}}$ and $(\Im(c_n))_{n\in\mathbb{N}}$ converge. In this case we have

$$\lim_{n\to\infty} c_n = \lim_{n\to\infty} \Re(c_n) + i \lim_{n\to\infty} \Im(c_n).$$

Theorem. *In $\mathbb{C}$ every Cauchy sequence converges.*

Proof. The two sequences $(\Re(c_n))_{n\in\mathbb{N}}$ and $(\Im(c_n))_{n\in\mathbb{N}}$ are Cauchy sequences, hence converge in the reals. This implies the claim. □

The treatment of the exponential series

$$\exp(z) := \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

can be carried over without any difficulty to the complex numbers. This also applies to the estimate of the rest, and the functional equation. As a consequence, we have $\exp(z) \neq 0$ for all $z \in \mathbb{C}$, because of $\exp(z)\exp(-z) = \exp(z-z) = \exp(0) = 1$. Notice also that

$$\exp(\overline{z}) = \overline{\exp(z)} \quad (z \in \mathbb{C}); \tag{9}$$

this follows from

$$\exp(\overline{z}) = \sum_{n=0}^{\infty} \frac{(\overline{z})^n}{n!} = \lim_{n\to\infty} \sum_{k=0}^{n} \overline{\left(\frac{z^k}{k!}\right)} = \lim_{n\to\infty} \overline{\left(\sum_{k=0}^{n} \frac{z^k}{k!}\right)} = \overline{\exp(z)}.$$

4. Continuous functions

For $x, y \in \mathbb{R}$ the *finite intervals* are defined by

$$[x,y] := \{\, z \in \mathbb{R} \mid x \le z \le y \,\},$$
$$(x,y) := \{\, z \in \mathbb{R} \mid x < z < y \,\},$$
$$(x,y] := \{\, z \in \mathbb{R} \mid x < z \le y \,\},$$
$$[x,y) := \{\, z \in \mathbb{R} \mid x \le z < y \,\}.$$

The interval $[x,y]$ is *closed*, (x,y) is *open*, $(x,y]$ is *half-open on the left*, $[x,y)$ is *half-open on the right*. We also allow the *infinite intervals*

$$[x,\infty) := \{\, z \in \mathbb{R} \mid x \le z \,\},$$
$$(x,\infty) := \{\, z \in \mathbb{R} \mid x < z \,\},$$
$$(\infty,y] := \{\, z \in \mathbb{R} \mid z \le y \,\},$$
$$(\infty,y) := \{\, z \in \mathbb{R} \mid z < y \,\}.$$

An inhabited, closed finite interval is called a *compact interval*. We use I, J to denote compact intervals with rational end points.

4.1. **Suprema and infima.**

Definition. Let S be a set of reals. A real y is an *upper bound* of S if $x \leq y$ for all $x \in S$. A real y is a *supremum* of S if y is an upper bound of S, and in addition for every rational $a < y$ there is real $x \in S$ such that $a \leq x$. The set S is *order located above* if for every $a < b$, either $x \leq b$ for all $x \in S$ or else $a \leq x$ for some $x \in S$.

Every set S can have at most one supremum. To see this, assume that y, z are suprema of S. It is enough to show $y \leq z$, and for this it suffices to show $z \not< y$. So assume $z < y$. Then $z < a < y$ for some rational a, hence $a \leq x$ for some $x \in S$, contradicting the assumption that z is an upper bound of S. If the supremum of S exists, it is denoted by $\sup S$.

Theorem (Least-upper-bound principle)**.** *Assume that S is an inhabited set of reals that is bounded above. Then S has a supremum if and only if it is order located above.*

Proof. If $\sup S$ exists and $a < b$, then either $\sup S < b$ or else $a < \sup S$. In the former case $x \leq b$ for all $x \in S$, and in the latter case clearly $a \leq x$ for some $x \in S$. Hence S is order located above.

For the converse it is useful to consider

$$\Pi_S(a,b): \qquad \text{both } y \leq b \text{ for all } y \in S \text{ and } a < x \text{ for some } x \in S$$

as a property of any pair a, b of rational numbers with $a < b$. By assumption we have $a, b \in \mathbb{Q}$ with $a < b$ such that $\Pi_S(a,b)$. We construct two sequences $(c_n)_n$ and $(d_n)_n$ of rationals such that for all n

$$a = c_0 \leq c_1 \leq \cdots \leq c_n < d_n \leq \cdots \leq d_1 \leq d_0 = b, \tag{10}$$

$$\Pi_S(c_n, d_n), \tag{11}$$

$$d_n - c_n \leq \left(\frac{2}{3}\right)^n (b - a). \tag{12}$$

Let $c_0, \dots, c_n$ and $d_0, \dots, d_n$ be already constructed such that (10)-(12) hold. Let $c = c_n + \frac{1}{3}(d_n - c_n)$ and $d = c_n + \frac{2}{3}(d_n - c_n)$. Since S is order located above, either $s \leq d$ for all $s \in S$ or else $c < r$ for some $r \in S$. In the first case let $c_{n+1} := c_n$ and $d_{n+1} := d$, and in the second case let $c_{n+1} := c$ and $d_{n+1} := d_n$. Then clearly $\Pi_S(c_{n+1}, d_{n+1})$, (10) and (12) continue to hold for $n + 1$, and the real number $x = y$ given by the modulated Cauchy sequences of rationals $(c_n)_n$ and $(d_n)_n$ is the least upper bound of S. □

4.2. **Continuous functions.**

Definition. A *continuous function* $f\colon I \to \mathbb{R}$ on a compact interval I with rational end points is given by

(a) an approximating map $h_f\colon (I \cap \mathbb{Q}) \times \mathbb{N} \to \mathbb{Q}$ and a map $\alpha_f\colon \mathbb{Z} \to \mathbb{N}$ such that $(h_f(a,n))_n$ is a Cauchy sequence with (uniform) modulus α_f;

(b) a modulus $\omega_f\colon \mathbb{Z} \to \mathbb{N}$ of (uniform) continuity, which satisfies

$$|a-b| \le 2^{-\omega_f(k)+1} \to |h_f(a,n) - h_f(b,n)| \le 2^{-k} \quad \text{for } n \ge \alpha_f(k);$$

(c) a lower bound N_f and an upper bound M_f for all $h_f(a,n)$.

α_f and ω_f are required to be weakly increasing. A function $f\colon J \to \mathbb{R}$ on an arbitrary interval J is continuous if it is continuous on every compact subinterval of J with rational end points.

Notice that a continuous function is given by objects of type level ≤ 1 only. This is due to the fact that it suffices to define its values on rational numbers.

The lower and upper bound of the values of the approximating map have been included to ease the definition of composition of continuous functions; however, they also have an effect on computational efficiency.

Instead of making the lower and upper bound part of the definition of a continuous function, a substitute could have been defined, making use of the fact that every continuous function on a compact interval comes with a modulus of uniform continuity. This can be seen as follows. Let $f\colon [a,b] \to \mathbb{R}$ be continuous, given by h_f, α_f and ω_f. Then for all $n \ge n_0 := \alpha_f(0)$ and rationals $c \in I$,

$$|h_f(c,n)| \le M := |h_f(a,n_0)| + N + 1,$$

where $(c-a)2^{\omega_f(0)-1} \le N \in \mathbb{N}$. For the proof recall that for $c,d \in [a,b]$

$$|c-d| \le 2^{-\omega_f(0)+1} \to |h_f(c,n_0) - h_f(d,n_0)| \le 1.$$

Let $\varepsilon := (c-a)/N$ and $a_i := a + i\varepsilon$ for $i = 0, \dots, N$. Then $a_N = a + N\varepsilon = c$ and

$$\bigl|h_f(c,n_0) - h_f(a,n_0)\bigr| \le \sum_{i=0}^{N-1} \bigl|h_f(a_{i+1},n_0) - h_f(a_i,n_0)\bigr| \le N,$$

hence $|h_f(c,n_0)| \le |h_f(a,n_0)| + N$ and therefore $|h_f(c,n)| \le |h_f(a,n_0)| + N + 1$ for $n \ge n_0 := \alpha_f(0)$.

An example is the squaring function $\mathsf{sq}\colon \mathbb{R} \to \mathbb{R}$, which is given on every compact interval with rational end points $0 \le c < d$ by

(a) the approximating map $h_{\mathsf{sq}}(a,n) := a^2$ and modulus $\alpha_{\mathsf{sq}}(k) := 0$;
(b) the modulus $k \mapsto k+p+1$ of uniform continuity, where p is such that $|a+b| \le 2^p$ for $a,b \in [c,d]$, because

$$|a-b| \le 2^{-k-p} \to |a^2-b^2| = |(a-b)(a+b)| \le 2^{-k};$$

(c) the lower bound $N_{\mathsf{sq}} = c^2$ and upper bound $M_{\mathsf{sq}} = d^2$,

Similarly all polynomials with rational coefficients on finite intervals can be viewed as continuous functions in our sense.

Another example is the inverse function $\mathsf{inv}\colon (0,\infty) \to \mathbb{R}$, given on every compact interval $[2^{-l}, d]$ by

(a) the approximating map $h_{\mathsf{inv}}(a,n) := \frac{1}{a}$ and modulus $\alpha_{\mathsf{inv}}(k) := 0$;
(b) the modulus $k \mapsto k+2l+1$ of uniform continuity, for

$$|a-b| \le 2^{-k-2l} \to \Big|\frac{1}{a}-\frac{1}{b}\Big| = \Big|\frac{b-a}{ab}\Big| \le 2^{-k},$$

because $ab \ge 2^{-2l}$;
(c) the lower bound $N_{\mathsf{inv}} = 1/d$ and upper bound ${}_{\mathsf{inv}}M = 2^l$.

Yet another example is the *square root function*; it differs from the previous ones in that the values on rational numbers will not be rationals any more. Given $a > 0$ and – for definiteness – $a_0 := 1$, recall from 1.1 that we can approximate $\sqrt{a}$ by

$$a_{n+1} := \frac{1}{2}\Big(a_n + \frac{a}{a_n}\Big).$$

One can verify easily that $\min(a,1) \le a_n \le \max(a,1)$, for all n. Hence the square root function on $[c,d]$ $(0 < c < d)$ is given by

(a) the approximating map $h_{\sqrt{\cdot}}(a,n) := a_n$, and a modulus $\alpha_{\sqrt{\cdot}}$, which can easily be computed from the fact (established in the proof of the theorem in 1.1) that $|a_{n+1} - a_{m+1}| \le (a_1 - b_1)/2^n$ for $n \le m$, with $a_1 = (1+a)/2$ and $b_1 = a/a_1$;
(b) the modulus of uniform continuity can be obtained from

$$\big|\sqrt{a} - \sqrt{b}\big| \le \frac{1}{\sqrt{a}+\sqrt{b}}|a-b|,$$

because $\sqrt{a}+\sqrt{b} \ge 2\min(c,1)$;
(c) the lower bound $N_{\sqrt{\cdot}} = \min(c,1)$ and upper bound $M_{\sqrt{\cdot}} = \max(d,1)$.

Our final example is the *exponential function*. On $[c,d]$ $(c < d)$ it is given by

(a) the approximating map

$$h_{\exp}(a,n) := \sum_{k=0}^{n} \frac{a^k}{k!},$$

and a uniform Cauchy modulus $\alpha_{\exp}$, which can easily be computed from the theorem in 2.7:

$$\Big|\sum_{k=0}^{n} \frac{a^k}{k!} - \sum_{k=0}^{m} \frac{a^k}{k!}\Big| = \Big|\sum_{k=n+1}^{m} \frac{a^k}{k!}\Big| \le \sum_{k=n+1}^{\infty} \frac{|a|^k}{k!} \le 2\frac{|a|^{n+1}}{(n+1)!}$$

for $|a| \le 1 + \frac{n}{2}$ and $n \le m$;

(b) the modulus of uniform continuity, which can be obtained from

$$\Big|\sum_{k=0}^{n} \frac{a^k}{k!} - \sum_{k=0}^{n} \frac{b^k}{k!}\Big| = \Big|\sum_{k=1}^{n} \frac{a^k - b^k}{k!}\Big| = |a-b| \sum_{k=1}^{n} \frac{1}{k!}\Big|\sum_{l=0}^{k-1} a^{k-1-l}b^l\Big|$$

$$\le |a-b| \sum_{k=1}^{n} \frac{kM^{k-1}}{k!} = |a-b| \sum_{k=0}^{n-1} \frac{M^k}{k!} < |a-b|\exp(M),$$

where $M = \max(|c|,|d|)$;

(c) the lower bound $N_{\exp} := 0$ and upper bound

$$M_{\exp} = \sum_{k=0}^{L} \frac{d^k}{k!} + 2\frac{|d|^{L+1}}{(L+1)!} \quad \text{with } L := 2\lceil |d| \rceil;$$

these can easily be verified, using the theorem in 2.7.

4.3. **Application of a continuous function to a real.** Since the approximating map operates on rationals only, we need to define separately what it means to apply a continuous function in our sense to a real.

Definition. *Application* of a continuous function $f\colon I \to \mathbb{R}$ (given by h_f, α_f, ω_f) to a real $x := ((a_n)_n, M)$ in I is defined to be

$$(h_f(a_n,n))_n$$

with modulus $\max(\alpha_f(k+2), M(\omega_f(k+1)-1))$. This is a modulus, for

$$\begin{aligned}
&|h_f(a_m,m) - h_f(a_n,n)| \\
&\le |h_f(a_m,m) - h_f(a_m,p)| + |h_f(a_m,p) - h_f(a_n,p)| + |h_f(a_n,p) - h_f(a_n,n)| \\
&\le 2^{-k-2} + 2^{-k-1} + 2^{-k-2}
\end{aligned}$$

if $m, n \geq M(\omega_f(k+1)-1)$ and $p \geq \alpha_f(k+1)$ (for the middle term), and moreover $m, n, p \geq \alpha_f(k+2)$ (for the first and last term). We denote this real by $f(x)$. The set of all such reals is called the *range* of f.

We want to show that application is compatible with equality.

Lemma (ContAppComp)**.** *Let f be a continuous function and x, y be reals in its domain. Then*

$$x = y \to f(x) = f(y).$$

Proof. Let $x := ((a_n)_n, M)$ and $y := ((b_n)_n, N)$, and assume $x = y$. We must show that $(h_f(a_n, n))_n = (h_f(b_n, n))_n$. This follows from

$$|h_f(a_n, n) - h_f(b_n, n)| \leq 2^{-k}$$

if $n \geq \alpha_f(k)$ and in addition $n \geq p$ with p provided by the characterization of equality of reals in 1.2 (for $\omega_f(k) - 1$). □

Next we show that indeed a continuous function f has ω_f as a modulus of uniform continuity.

Lemma (ContMod)**.** *Let f be continuous and x, y be reals in its domain. Then*

$$|x - y| \leq 2^{-\omega_f(k)} \to |f(x) - f(y)| \leq 2^{-k}.$$

Proof. Let $x := ((a_n)_n, M)$ and $y := ((b_n)_n, N)$. Assume $|a_n - b_n| \leq 2^{-\omega_f(k)+1}$ for $n \geq p$. Then for $n \geq p, \alpha_f(k)$

$$|h_f(a_n, n) - h_f(b_n, n)| \leq 2^{-k},$$

that is $|f(x) - f(y) \leq 2^{-k}$. □

4.4. **Continuous functions and limits.** We show that continuous functions commute with limits.

Lemma (ContLim)**.** *Let $(x_n)_n$ be a sequence of reals which converges to y. Assume $x_n, y \in I$ and let $f\colon I \to \mathbb{R}$ be continuous. Then $(f(x_n))_n$ converges to $f(y)$.*

Proof. For a given k, pick p such that for all n

$$p \leq n \to |x_n - y| \leq 2^{-\omega_f(k)}.$$

Then by the previous lemma

$$p \leq n \to |f(x_n) - f(y)| \leq 2^{-k}.$$

Hence $(f(x_n))_n$ converges to $f(y)$. □

Lemma (ContRat). *Assume that $f, g\colon I \to \mathbb{R}$ are continuous and coincide on all rationals $a \in I$. Then $f = g$.*

Proof. Let $x = ((a_n)_n, M)$. By ContLim, $(f(a_n))_n$ converges to $f(x)$ and $(g(a_n))_n$ to $g(x)$. Now $f(a_n) = g(a_n)$ implies $f(x) = g(x)$. □

4.5. **Composition of continuous functions.** We define the *composition* of two continuous functions.

Definition. Assume that $f\colon I \to \mathbb{R}$ and $g\colon J \to \mathbb{R}$ are continuous, with I, J compact intervals with rational end points, and $N_f, M_f \in J$. Then the function $g \circ f\colon I \to \mathbb{R}$ is defined by

(a) the approximating map

$$h_{g\circ f}\colon (I \cap \mathbb{Q}) \times \mathbb{N} \to \mathbb{Q}, \quad h_{g\circ f}(a, n) := h_g(h_f(a, n), n)$$

with modulus $\alpha_{g\circ f}(k) := \max\bigl(\alpha_g(k + 2), \alpha_f(\omega_g(k + 1) - 1)\bigr)$;

(b) the modulus $\omega_{g\circ f}(k) := \omega_f(\omega_g(k) - 1) + 1$ of (uniform) continuity for $g \circ f$.

(c) $N_{g\circ f} = N_g$ and $M_{g\circ f} = M_g$.

For arbitrary continuous f, g, we are given for every compact subinterval with rational end points of the domain interval of f a lower bound N and an upper bound M on the range of the approximating map for f on this subinterval. Take the instantiation of g to $[N, M]$, and form the composition of the two functions as described above. This defines $g \circ f$.

We need to show that this indeed defines a continuous function.

Lemma. *Under the assumptions of the definition above we have*

(a) *$\alpha_{g\circ f}$ is a Cauchy modulus;*

(b) *$\omega_{g\circ f}$ is a modulus of uniform continuity for $g \circ f$.*

Proof. (a). $(h_f(a, n))_n$ is a real with modulus α_f. This real is between M_f and N_f, hence in J. By 4.3, application of g to this real gives the Cauchy sequence $(h_g(h_f(a, n), n))_n$ with Cauchy modulus

$$\max(\alpha_g(k + 2), \alpha_f(\omega_g(k + 1) - 1)) = \alpha_{g\circ f}(k).$$

(b). Assume $|a - b| \le 2^{-\omega_f(\omega_g(k)-1)+1}$. Then $|h_f(a, n) - h_f(b, n)| \le 2^{-\omega_g(k)+1}$ provided $n \ge \alpha_f(\omega_g(k) - 1)$, and therefore $|h_g(h_f(a, n), n) - h_g(h_f(b, n), n)| \le 2^{-k}$ provided also $n \ge \alpha_g(k)$. Both conditions on n hold for $n \ge \alpha_{g\circ f}(k)$, since α_g, α_f and ω_g are weakly increasing. □

For example, assume that $f\colon I \to \mathbb{R}$ is continuous and that for some l, $2^{-l} \leq h_f(c, n) \in \mathbb{Q}$ for all n and all rationals $c \in I$. Then we can form the inverse $1/f$ as a continuous map $I \to \mathbb{R}$, by composing $\mathsf{inv} \circ f$.

4.6. **Properties of continuous functions.** The supremum of the range of a continuous function on a compact interval can be shown to exist constructively.[1] We prove that the range is order located above, which entails (by the least-upper-bound principle) that it has a supremum.

Lemma. *Let $f\colon I \to \mathbb{R}$ be continuous, I compact with rational end points. Then the range of f is order located above.*

Proof. Let h_f, α_f and ω_f be the data for f. Given a, b with $a < b$, fix k such that $2^{-k} \leq \frac{1}{3}(b-a)$. Take a partition $a_0, \dots, a_l$ of I of mesh $\leq 2^{-\omega_f(k)+2}$. Then for every $c \in I$ there is an i such that $|c - a_i| \leq 2^{-\omega_f(k)+1}$. Let $n_k := \alpha_f(k)$ and consider all finitely many

$$h(a_i, n_k) \quad \text{for } i = 0, \dots, l.$$

Let $h(a_j, n_k)$ be the maximum of all those.

Case $h(a_j, n_k) \leq a + \frac{1}{3}(b-a)$. We show that $f(x) \leq b$ for all x. Let $x = ((b_n)_n, M)$. Then for $n \geq n_k$

$$\begin{aligned} h_f(b_n, n) &\leq h_f(b_n, n_k) + 2^{-k} \\ &\leq h_f(a_i, n_k) + 2^{-k+1} \quad \text{for } i \text{ such that } |b_n - a_i| \leq 2^{-\omega_f(k)+1} \\ &\leq h_f(a_j, n_k) + 2^{-k+1} \\ &\leq b. \end{aligned}$$

Case $a + \frac{1}{3}(b-a) < h(a_j, n_k)$. We show $a \leq f(x)$ for $x := a_j$. Then $f(x)$ is given by the Cauchy sequence $(h_f(a_j, n))_n$. We have for $n \geq n_k$

$$h_f(a_j, n) \geq h_f(a_j, n_k) - 2^{-k} \geq a + \frac{1}{3}(b-a) - 2^{-k} \geq a.$$

Hence $a \leq f(x)$. □

Corollary. *Let $f\colon I \to \mathbb{R}$ be continuous, I compact with rational end points. Then the range of f has a supremum, denoted $\|f\|_I$.*

Proof. The range of f is bounded above, and by the last lemma it is order located above. Hence by the least-upper-bound principle it has a supremum. □

[1] This is proved in [5], using the notion of a "totally bounded" set. However, the latter is a type-level 2 concept, which we wish to avoid.

4.7. **Intermediate value theorem.** We next supply the standard constructive versions of the *intermediate value theorem.*

Theorem (Approximate intermediate value theorem)**.** *Let $a < b$ be rational numbers. For every continuous function $f\colon [a,b] \to \mathbb{R}$ with $f(a) \le 0 \le f(b)$, and every k, we can find $c \in [a,b]$ such that $|f(c)| \le 2^{-k}$.*

Proof. In the sequel we repeatedly invoke the approximate splitting principle from 1.6. Given k, let $\varepsilon := 2^{-k}$. We compare $f(a)$ and $f(b)$ with $-\varepsilon < -\frac{\varepsilon}{2}$ and $\frac{\varepsilon}{2} < \varepsilon$, respectively. If $-\varepsilon < f(a)$ or $f(b) < \varepsilon$, then $|f(c)| < \varepsilon$ for $c = a$ or $c = b$; whence we may assume that

$$f(a) < -\frac{\varepsilon}{2} \quad \text{and} \quad \frac{\varepsilon}{2} < f(b).$$

Now pick l so that, for all $x, y \in [a,b]$, if $|x-y| \le 2^{-l}$, then $|f(x)-f(y)| \le \varepsilon$, and divide $[a,b]$ into $a = a_0 < a_1 < \cdots < a_m = b$ such that $|a_{i-1} - a_i| \le 2^{-l}$. Compare every $f(a_i)$ with $-\frac{\varepsilon}{2} < \frac{\varepsilon}{2}$. By assumption $f(a_0) < -\frac{\varepsilon}{2}$ and $\frac{\varepsilon}{2} < f(a_m)$; whence we can find j minimal such that

$$f(a_j) < \frac{\varepsilon}{2} \quad \text{and} \quad -\frac{\varepsilon}{2} < f(a_{j+1}).$$

Finally, compare $f(a_j)$ with $-\varepsilon < -\frac{\varepsilon}{2}$ and $f(a_{j+1})$ with $\frac{\varepsilon}{2} < \varepsilon$. If $-\varepsilon < f(a_j)$, we have $|f(a_j)| < \varepsilon$. If $f(a_{j+1}) < \varepsilon$, we have $|f(a_{j+1})| < \varepsilon$. If both $f(a_j) < -\frac{\varepsilon}{2}$ and $\frac{\varepsilon}{2} < f(a_{j+1})$, then we would have $|f(a_{j+1}) - f(a_j)| > \varepsilon$, contradicting $|a_{j+1} - a_j| \le 2^{-l}$. □

Alternative Proof. We give a different proof, which more directly makes use of the fact that our continuous functions come with witnessing data.

We may assume $f(a) < -2^{-k-1}$ and $2^{-k-1} < f(b)$ (see above). Divide $[a,b]$ into $a = a_0 < a_1 < \cdots < a_m = b$ such that $|a_{i-1} - a_i| \le 2^{-\omega_f(k+1)}$. Consider all finitely many

$$h(a_i, n_0) \quad \text{for } i = 1, \dots, m,$$

with $n_0 := \alpha_f(k+1)$. Pick j such that $h(a_{j-1}, n_0) \le 0 \le h(a_j, n_0)$; this can be done because $f(a) < -2^{-k-1}$ and $2^{-k-1} < f(b)$. We show $|f(a_j)| \le 2^{-k}$; for this it clearly suffices to show $|h(a_j, n)| \le 2^{-k}$ for $n \ge n_0$. Now

$$|h(a_j, n)| \le \bigl|h(a_j, n) - h(a_j, n_0)\bigr| + \bigl|h(a_j, n_0)\bigr| \le 2^{-k-1} + 2^{-k-1},$$

where the first estimate holds by the choice of n_0, and the second one follows from the choice of a_j and $|h(a_{i-1}, n) - h(a_i, n)| \le 2^{-k-1}$. □

A problem with both of these proofs is that the algorithms they provide are rather bad: in each case one has to partition the interval into as many pieces as the modulus of the continuous function requires for the given error bound, and then for each of these (many) pieces perform certain operations. This problem seems to be unavoidable, since our continuous function may be rather flat. However, we can do somewhat better if we assume a uniform *modulus of increase* (or lower bound on the slope) of f, that is, some $l \in \mathbb{N}$ such that for all $c, d \in \mathbb{Q}$ and all $k \in \mathbb{Z}$

$$2^{-k} \leq d - c \to f(c) <_{k+l} f(d).$$

We begin with an auxiliary lemma, which from a "correct" interval $c < d$ (that is, $f(c) \leq 0 \leq f(d)$ and $2^{-k} \leq d - c$) constructs a new one $c_1 < d_1$ with $d_1 - c_1 = \frac{2}{3}(d - c)$.

Lemma (IVTAux)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be continuous, and with a uniform modulus l of increase. Assume $a \leq c < d \leq b$, say $2^{-k} < d - c$, and $f(c) \leq 0 \leq f(d)$. Then we can construct c_1, d_1 with $d_1 - c_1 = \frac{2}{3}(d - c)$, such that again $a \leq c \leq c_1 < d_1 \leq d \leq b$ and $f(c_1) \leq 0 \leq f(d_1)$.*

Proof. Let $c_0 = \frac{2c+d}{3}$ and $d_0 = \frac{c+2d}{3}$. From $2^{-k} < d - c$ we obtain $2^{-k-2} \leq d_0 - c_0$, so $f(c_0) <_{k+2+l} f(d_0)$. Now compare 0 with this proper interval, using ApproxSplit. In the first case we have $0 \leq f(d_0)$; then let $c_1 = c$ and $d_1 = d_0$. In the second case we have $f(c_0) \leq 0$; then let $c_1 = c_0$ and $d_1 = d$. □

Theorem (IVT)**.** *If $f\colon [a,b] \to \mathbb{R}$ is continuous with $f(a) \leq 0 \leq f(b)$, and with a uniform modulus of increase, then we can find $x \in [a,b]$ such that $f(x) = 0$.*

Proof. Iterating the construction in the auxiliary lemma IVTAux above, we construct two sequences $(c_n)_n$ and $(d_n)_n$ of rationals such that for all n

$$\begin{aligned} &a = c_0 \leq c_1 \leq \cdots \leq c_n < d_n \leq \cdots \leq d_1 \leq d_0 = b, \\ &f(c_n) \leq 0 \leq f(d_n), \\ &d_n - c_n = \big(2/3\big)^n (b - a). \end{aligned}$$

Let x, y be given by the Cauchy sequences $(c_n)_n$ and $(d_n)_n$ with the obvious modulus. As f is continuous, $f(x) = 0 = f(y)$ for the real number $x = y$. □

Remark. The proposition can also be proved for locally nonconstant functions. A function $f\colon [a,b] \to \mathbb{R}$ is *locally nonconstant* whenever if $a \leq a' < b' \leq b$ and c is an arbitrary real, then $f(x) \neq c$ for some real $x \in [a', b']$. Note that if f is continuous, then there also is a rational with that property.

Strictly monotonous functions are clearly locally nonconstant, and so are nonconstant real polynomials.

From the Intermediate Value Theorem we obtain

Theorem (Inv). *Let $f\colon [a,b] \to \mathbb{R}$ be continuous with a uniform modulus of increase, and assume $f(a) \le a' < b' \le f(b)$. We can find a continuous $g\colon [a',b'] \to \mathbb{R}$ such that $f(g(y)) = y$ for every $y \in [a',b']$ and $g(f(x)) = x$ for every $x \in [a,b]$ such that $a' \le f(x) \le b'$.*

Proof. Let $f\colon [a,b] \to \mathbb{R}$ be continuous with a uniform modulus of increase, that is, some $l \in \mathbb{Z}$ such that for all $c, d \in [a,b]$ and all $k \in \mathbb{Z}$

$$2^{-k} \le d - c \to f(c) <_{k+l} f(d).$$

Let $f(a) \le a' < b' \le f(b)$. We construct a continuous $g\colon [a',b'] \to \mathbb{R}$.

Let $u \in [a',b']$ be rational. Using $f(a) - u \le a' - u \le 0$ and $0 \le b' - u \le f(b) - u$, the IVT gives us an x such that $f(x) - u = 0$, as a Cauchy sequence (c_n). Let $h_g(u,n) := c_n$. Define the modulus α_g such that for $n \ge \alpha_g(k)$, $(2/3)^n(b-a) \le 2^{-\omega_f(k+l+2)}$. For the uniform modulus ω_g of continuity assume $a' \le u < v \le b'$ and $k \in \mathbb{Z}$. We claim that with $\omega_g(k) := k + l + 2$ (l from the hypothesis on the slope) we can prove the required property

$$|u - v| \le 2^{-\omega_g(k)+1} \to |h_g(u,n) - h_g(v,n)| \le 2^{-k} \quad (n \ge \alpha_g(k)).$$

Let $a' \le u < v \le b'$ and $n \ge \alpha_g(k)$. For $c_n^{(u)} := h_g(u,n)$ and $c_n^{(v)} := h_g(v,n)$ assume that $|c_n^{(u)} - c_n^{(v)}| > 2^{-k}$; we must show $|u - v| > 2^{-\omega_g(k)+1}$.

By the proof of the Intermediate Value Theorem we have

$$d_n^{(u)} - c_n^{(u)} \le (2/3)^n(b-a) \le 2^{-\omega_f(k+l+2)} \quad \text{for } n \ge \alpha_g(k).$$

Using $f(c_n^{(u)}) - u \le 0 \le f(d_n^{(u)}) - u$, the fact that a continuous function f has ω_f as a modulus of uniform continuity gives us

$$|f(c_n^{(u)}) - u| \le |(f(d_n^{(u)}) - u) - (f(c_n^{(u)}) - u)| = |f(d_n^{(u)}) - f(c_n^{(u)})| \le 2^{-k-l-2}$$

and similarly $|f(c_n^{(v)}) - v| \le 2^{-k-l-2}$. Hence, using $|f(c_n^{(u)}) - f(c_n^{(v)})| \ge 2^{-k-l}$ (which follows from $|c_n^{(u)} - c_n^{(v)}| > 2^{-k}$ by the hypothesis on the slope),

$$|u - v| \ge |f(c_n^{(u)}) - f(c_n^{(v)})| - |f(c_n^{(u)}) - u| - |f(c_n^{(v)}) - v| \ge 2^{-k-l-1}.$$

Now $f(g(u)) = u$ follows from

$$|f(g(u)) - u| = |h_f(c_n, n) - u| \le |h_f(c_n, n) - h_f(c_n, m)| + |h_f(c_n, m) - u|,$$

which is $\le 2^{-k}$ for $n, m \ge \alpha_f(k+1)$. Since continuous functions are determined by their values on the rationals, we have $f(g(y)) = y$ for $y \in [a',b']$.

For every $x \in [a,b]$ with $a' \leq f(x) \leq b'$, from $g(f(x)) < x$ we obtain the contradiction $f(x) = f(g(f(x))) < f(x)$ by the hypothesis on the slope, and similarly for $>$. Using $u \not< v \leftrightarrow v \leq u$ we obtain $g(f(x)) = x$. □

As an example, consider the squaring function $f\colon [1,2] \to [1,4]$, given by the approximating map $h_f(a,n) := a^2$, constant Cauchy modulus $\alpha_f(k) := 1$, and modulus $\omega_f(k) := k+1$ of uniform continuity. The modulus of oncrease is $l := 0$, because for all $c, d \in [1,2]$

$$2^{-k} \leq d - c \to c^2 <_k d^2.$$

Then $h_g(u,n) := c_n^{(u)}$, as constructed in the IVT for $x^2 - u$, iterating IVTAux. The Cauchy modulus α_g is such that $(2/3)^n \leq 2^{-k+3}$ for $n \geq \alpha_g(k)$, and the modulus of uniform continuity is $\omega_f(k) := k+2$.

4.8. **Continuity for functions of more than one variable.** Without loss of generality we restrict ourselves to functions of two real variables.

Definition. A *continuous function* $f\colon I_1 \times I_2 \to \mathbb{R}$ for compact intervals I_1, I_2 with rational end points is given by

(a) an approximating map $h_f\colon (I_1 \cap \mathbb{Q}) \times (I_2 \cap \mathbb{Q}) \times \mathbb{N} \to \mathbb{Q}$ and a map $\alpha_f\colon \mathbb{Z} \to \mathbb{N}$ such that $(h_f(a,b,n))_n$ is a Cauchy sequence with (uniform) modulus α_f;
(b) a modulus $\omega_f\colon \mathbb{Z} \to \mathbb{N}$ of (uniform) continuity, which satisfies

$$|a - a'|, |b - b'| \leq 2^{-\omega_f(k)+1} \to |h_f(a,b,n) - h_f(a',b',n)| \leq 2^{-k}$$

for $n \geq \alpha_f(k)$;
(c) a lower bound N_f and an upper bound M_f for all $h_f(a,b,n)$.

α_f and ω_f are required to be weakly increasing. A function $f\colon D \to \mathbb{R}$ on an arbitrary domain $D \subseteq \mathbb{R}^2$ is continuous if it is continuous on every $I_1 \times I_2 \subseteq D$, where I_1, I_2 are compact intervals with rational end points.

An example is the exponential function of a complex variable. Continuity of a function $f\colon D \to \mathbb{C}$ for some domain $D \subseteq \mathbb{C}$ is treated as continuity of the two real valued functions $\Re(f(z))$ and $\Im(f(z))$, and the latter as binary real valued functions, e.g., $\Re(f(x+iy))$. The example above of the continuity of the real exponential function can easily be modified to yield the continuity the exponential function of a complex variable, in the sense described.

5. Differentiation

5.1. Derivatives.

Definition. Let $f, g\colon I \to \mathbb{R}$ be continuous. g is called *derivative* of f with modulus $\delta_f\colon \mathbb{Z} \to \mathbb{N}$ of differentiability if for $x, y \in I$ with $x < y$,

$$y \le x + 2^{-\delta_f(k)} \to \bigl|f(y) - f(x) - g(x)(y - x)\bigr| \le 2^{-k}(y - x).$$

f is said to be *differentiable* on I and g is called a *derivative* of f on I.

To say that g is a derivative of f we write

$$g = f', \quad g = Df, \quad \text{or} \quad g(x) = \frac{df(x)}{dx}.$$

If f has two derivatives, then clearly they are equal functions.

For example, a constant function has derivative 0, and the identity function has the constant 1 function as derivative.

5.2. **Bounds on the slope.** We show that a bound on the derivative of f serves as a Lipschitz constant of f:

Lemma. *Let $f\colon I \to \mathbb{R}$ be continuous with derivative f'. Assume that f' is bounded on I by M. Then for $x, y \in I$ with $x < y$,*

$$\bigl|f(y) - f(x)\bigr| \le M(y - x).$$

Proof. Given $k \in \mathbb{N}$, it suffices to prove

$$\bigl|f(y) - f(x)\bigr| \le M(y - x) + 2^{-k}.$$

Choose l such that $2^{-l}(y - x) \le 2^{-k}$, and let $x = x_0 < x_1 < \cdots < x_n = y$ such that $x_{i+1} \le x_i + 2^{-\delta_f(l)}$, where δ_f is the modulus of differentiability of f. Then

$$\begin{aligned}
&\bigl|f(y) - f(x)\bigr| \\
&= \Bigl|\sum_{i=0}^{n-1} f(x_{i+1}) - f(x_i)\Bigr| \\
&= \Bigl|\sum_{i=0}^{n-1}\bigl(f'(x_i)(x_{i+1} - x_i)\bigr) + \sum_{i=0}^{n-1}\bigl(f(x_{i+1}) - f(x_i) - f'(x_i)(x_{i+1} - x_i)\bigr)\Bigr| \\
&\le M(y - x) + 2^{-l}(y - x).
\end{aligned}$$

□

Corollary (DerivZero)**.** *Let $f\colon I \to \mathbb{R}$ be continuous with derivative $f' = 0$. Then f is a constant.*

Proof. The lemma yields $f(x) = f(y)$ for $x, y \in I$, $x < y$. So $f(a)$ is constant for all rationals $a \in I$, hence also for all $x \in I$. □

5.3. **Properties of derivatives.**

Lemma. *Let $f, g\colon I \to \mathbb{R}$ be continuous with derivatives f', g' of moduli δ_f, δ_g. Then*

$$(f+g)' := f' + g'$$

is a derivative of $f+g$ with modulus

$$\delta_{f+g}(k) := \max\bigl(\delta_f(k+1), \delta_g(k+1)\bigr).$$

Proof. Let $x < y \le x + 2^{-m}$. Then

$$\begin{aligned}
&\bigl|f(y) + g(y) - f(x) - g(x) - (f'(x) - g'(x))(y-x)\bigr| \\
&\le \bigl|f(y) - f(x) - f'(x)(y-x)\bigr| + \bigl|g(y) - g(x) - g'(x)(y-x)\bigr| \\
&\le 2^{-k-1}(y-x) + 2^{-k-1}(y-x).
\end{aligned}$$

for $m \ge \delta_f(k+1), \delta_g(k+1)$. □

Lemma. *Let $f, g\colon I \to \mathbb{R}$ be continuous with derivatives f', g' of moduli δ_f, δ_g. Then*

$$(fg)' := f'g + fg'$$

is a derivative of fg with modulus

$$\delta_{fg}(k) := \max\bigl(\omega_g(r+k+1), \delta_f(k+q+2), \delta_g(k+p+2)\bigr),$$

where $2^r, 2^p, 2^q$ are upper bounds for f', f, g in I, respectively.

Proof. Let $x < y \le x + 2^{-m}$. Then, using the lemma in 5.2

$$\begin{aligned}
&\bigl|f(y)g(y) - f(x)g(x) - f'(x)g(x)(y-x) - f(x)g'(x)(y-x)\bigr| \\
&= \bigl|\bigl(f(y) - f(x)\bigr)g(y) + f(x)\bigl(g(y) - g(x)\bigr) - \\
&\qquad f'(x)g(x)(y-x) - f(x)g'(x)(y-x)\bigr| \\
&= \bigl|\bigl(f(y) - f(x)\bigr)\bigl(g(y) - g(x)\bigr) + \bigl(f(y) - f(x) - f'(x)(y-x)\bigr)g(x) + \\
&\qquad f(x)\bigl(g(y) - g(x) - g'(x)(y-x)\bigr)\bigr| \\
&\le 2^r(y-x)2^{-r-k-1} + 2^{-k-q-2}(y-x)\bigl|g(x)\bigr| + \bigl|f(x)\bigr|(y-x)2^{-k-p-2} \\
&\le 2^{-k}(y-x)
\end{aligned}$$

for $m \ge \omega_g(r+k+1), \delta_f(k+q+2), \delta_g(k+p+2)$. □

Lemma. *Let $g\colon I \to \mathbb{R}$ be continuous with derivative g' of modulus δ_g, and $|g'(x)| \le 2^q$ for all $x \in I$. Moreover assume that $2^{-p} \le |g(x)|$ for all $x \in I$. Then*

$$\Big(\frac{1}{g}\Big)' := -\frac{g'}{g^2}$$

is a derivative of $\frac{1}{g}$ with modulus

$$\delta_{\frac{1}{g}}(k) := \max\big(\delta_g(k + 2p + 1), \omega_{\frac{1}{g}}(p + q + k + 1)\big).$$

Proof. Let $x < y \le x + 2^{-m}$. Then

$$\begin{aligned}
&\Big|\frac{1}{g(y)} - \frac{1}{g(x)} - \frac{g'(x)}{g(x)^2}(y - x)\Big| \\
&= \Big|\frac{1}{g(y)g(x)}\big(g(y) - g(x) - g'(x)(y - x)\big) + \frac{g'(x)}{g(x)}(y - x)\Big(\frac{1}{g(y)} - \frac{1}{g(x)}\Big)\Big| \\
&\le 2^p \cdot 2^p \cdot 2^{-k-2p-1} \cdot (y - x) + 2^p \cdot 2^q \cdot (y - x) \cdot 2^{-p-q-k-1} \\
&\le 2^{-k}(y - x)
\end{aligned}$$

for $m \ge \delta_g(k + 2p + 1), \omega_{\frac{1}{g}}(p + q + k + 1)$. □

Notice that the well-known *quotient rule* can now be derived easily: under the appropriate assumptions we have

$$\Big(\frac{f}{g}\Big)' = f\frac{-g'}{g^2} + \frac{1}{g}f' = \frac{f'g - fg'}{g^2}.$$

Theorem (Chain Rule)**.** *Let $f\colon I \to J$ and $g\colon J \to \mathbb{R}$ be continuous with derivatives f', g' of moduli δ_f, δ_g. Then*

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

is a derivative of $g \circ f$ with modulus

$$\delta_{g\circ f}(k) = \max\big(\delta_g(k + 1 + r), \delta_f(k + 1 + q)\big),$$

where $2^r, 2^q$ are upper bounds for f', g' in I, J, respectively.

Proof. Let $x < y \le x + 2^{-m}$. Then, using the lemma in 5.2

$$\begin{aligned}
&\big|g(f(y)) - g(f(x)) - g'(f(x))f'(x)(y-x)\big| \\
&\le \big|g(f(y)) - g(f(x)) - g'(f(x))\big(f(y) - f(x)\big)\big| + \\
&\qquad \big|g'(f(x))\big| \cdot \big|f(y) - f(x) - f'(x)(y-x)\big| \\
&\le 2^{-k-1-r}\big|f(y) - f(x)\big| + 2^q\big|f(y) - f(x) - f'(x)(y-x)\big| \\
&\le 2^{-k-1}(y-x) + 2^q\big|f(y) - f(x) - f'(x)(y-x)\big| \\
&\le 2^{-k-1}(y-x) + 2^q 2^{-k-1-q}(y-x) \\
&= 2^{-k}(y-x)
\end{aligned}$$

for $m \ge \delta_g(k+1+r), \delta_f(k+1+q)$. □

5.4. **Rolle's Lemma, mean value theorem.**

Lemma (Rolle)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be continuous with derivative f', and assume $f(a) = f(b)$. Then for every $k \in \mathbb{N}$ we can find $c \in [a,b]$ such that $|f'(c)| \le 2^{-k}$.*

Proof. Let δ_f be the modulus of differentiability of f, and let $a = a_0 < a_1 < \dots < a_n = b$ such that $a_{i+1} \le a_i + 2^{-\delta_f(k+2)}$. Compare all $|f'(a_i)|$ with $2^{-k-1} < 2^{-k}$. If some is $< 2^{-k}$, we are done. Otherwise we argue as in the lemma in 5.2.

$$\begin{aligned}
&f(b) - f(a) \\
&= \sum_{i=0}^{n-1}\big(f(a_{i+1}) - f(a_i)\big) \\
&= \sum_{i=0}^{n-1}\big(f'(a_i)(a_{i+1} - a_i)\big) + \sum_{i=0}^{n-1}\big(f(a_{i+1}) - f(a_i) - f'(a_i)(a_{i+1} - a_i)\big) \\
&\ge 2^{-k-1}(b-a) - 2^{-k-2}(b-a) > 0.
\end{aligned}$$

This contradiction proves the claim. □

Theorem (Mean value theorem)**.** *Let $f\colon [a,b] \to \mathbb{R}$ be continuous with derivative f'. Then for every $k \in \mathbb{N}$ we can find $c \in [a,b]$ such that*

$$\big|f(b) - f(a) - f'(c)(b-a)\big| \le 2^{-k}(b-a).$$

Proof. Let $2^{-l} \le 2^{-k}(b-a)$ and define a continuous $h\colon [a,b] \to \mathbb{R}$ by

$$h(x) := (x-a)\big(f(b) - f(a)\big) - f(x)(b-a).$$

Then $h(a) = h(b) = -f(a)(b-a)$. Hence by Rolle's lemma we can find c in $[a, b]$ such that

$$\left|h'(c)\right| = \left|f(b) - f(a) - f'(c)(b-a)\right| \le 2^{-l}.$$

This proves the claim. □

6. Integration

To begin with, we define the integral of a continuous function on a compact interval with rational end points only. The reason for this restriction is that we need to establish $\int_a^x f(t)\,dt$ as a continuous function of x. Later we shall extend the definition of the integral to compact intervals whose end points are apart.

6.1. Riemannian sums.

Definition. Let a, b be rationals with $a < b$. A list $P = a_0, \dots, a_n$ of rationals is a *partition* of the interval $[a, b]$, if $a = a_0 \le a_1 \le \dots \le a_n = b$. $\max\{\, a_{i+1} - a_i \mid i < n \,\}$ is the *mesh* of P. A partition $Q = a'_0, \dots, a'_m$ of $[a, b]$ is a *refinement* of P, if

$$\forall_{i \le n} \exists_{j \le m} a'_j = a_i.$$

If $f \colon [a, b] \to \mathbb{R}$ is a continuous function given by h_f, α_f and ω_f, and $P = a_0, a_1, \dots, a_n$ a partition of $[a, b]$, then an arbitrary sum of the form

$$\sum_{i=0}^{n-1} h_f(e_i, n) \cdot (a_{i+1} - a_i)$$

with $e_i \in [a_i, a_{i+1}]$ is denoted by $S(f, P)$. In particular for $a_i = a + \frac{i}{n}(b-a)$

$$S(f, n) := S(f, a, b, n) := \frac{b-a}{n} \sum_{i=0}^{n-1} h_f(a_i, n)$$

is one of the numbers $S(f, P)$.

Theorem. *Assume that $f \colon [a, b] \to \mathbb{R}$ is continuous with modulus ω_f of (uniform) continuity. Then*

$$\bigl(S(f, n)\bigr)_{n \in \mathbb{N}}$$

is a Cauchy sequence of rationals with modulus

$$M(p) = \max(2^{\omega_f(p+q+1)}(b-a), \alpha_f(p+q+2)),$$

where q is such that $b - a \leq 2^q$; we denote this real by

$$\int_a^b f(x)\,dx.$$

Moreover, if P is a partition of mesh $\leq 2^{-\omega_f(l)}$, then

$$\left| S(f,P) - \int_a^b f(x)\,dx \right| \leq 2^{-l}(b-a).$$

Proof. Let k, l be given and $P = a_0, \ldots, a_n$, $Q = b_0, \ldots, b_m$ partitions of $[a,b]$ with mesh $\leq 2^{-\omega_f(k+1)}$ or $\leq 2^{-\omega_f(l+1)}$, respectively. Let $R = c_0, \ldots, c_r$ be the common refinement of P and Q, obtained by arranging $a_0, \ldots, a_n, b_0, \ldots, b_m$ into a monotone sequence (here we make use of the assumption that a_i, b_i are rational numbers). Let $d_j \in [c_j, c_{j+1}]$ for $j < r$. For every $i < n$ denote by $\sum_i$ the summation over all indices j such that $a_i \leq c_j < a_{i+1}$. Then

$$\begin{aligned}
&|S(f,P) - S(f,R)| \\
&= \Big| \sum_{i=0}^{n-1} h_f(e_i,n) \cdot (a_{i+1} - a_i) - \sum_{i=0}^{r-1} h_f(d_i,r) \cdot (c_{i+1} - c_i) \Big| \\
&= \Big| \sum_{i=0}^{n-1} h_f(e_i,n) \sum_i (c_{j+1} - c_j) - \sum_{i=0}^{n-1} \sum_i h_f(d_j,r) \cdot (c_{j+1} - c_j) \Big| \\
&\leq \sum_{i=0}^{n-1} \sum_i |h_f(e_i,n) - h_f(d_j,r)|(c_{j+1} - c_j) \\
&\leq \sum_{i=0}^{n-1} \sum_i \big(|h_f(e_i,n) - h_f(e_i,r)| + |h_f(e_i,r) - h_f(d_j,r)| \big)(c_{j+1} - c_j) \\
&\leq \sum_{i=0}^{n-1} \sum_i 2^{-k}(c_{j+1} - c_j) \quad \text{for } n \geq \alpha_f(k+1) \\
&= 2^{-k}(b-a)
\end{aligned}$$

Similarly, for $n \geq \alpha_f(l+1)$

$$|S(f,Q) - S(f,R)| \leq 2^{-l}(b-a),$$

hence

$$|S(f,P) - S(f,Q)| \leq (2^{-k} + 2^{-l})(b-a).$$

In particular

$$|S(f,m) - S(f,n)| \leq 2^{-k+1}(b-a)$$

for $m, n \geq 2^{\omega_f(k)}(b-a), \alpha_f(k+1)$. Hence $(S(f,n))_n$ is a Cauchy sequence.

Moreover we have for $n \geq 2^{\omega_f(l+1)}(b-a), \alpha_f(l+1)$

$$|S(f,P) - S(f,n)| \leq (2^{-k} + 2^{-l})(b-a).$$

Now let $n \to \infty$ and $k \to \infty$. Then we obtain

$$|S(f,P) - \int_a^b f(x)\,dx| \leq 2^{-l}(b-a),$$

as was to be shown. □

Remark. We will also need to consider $S(f,n)$ in case $b < a$. Then we can use the same definition, and by the same argument we see that $(S(f,n))_n$ is a Cauchy sequence; its limit is denoted by $\int_a^b f(t)\,dt$. One can see easily that $\int_a^b f(t)\,dt = -\int_b^a f(t)\,dt$.

Immediately from the definition we obtain:

Corollary. *Assume that* $f\colon [a,b] \to \mathbb{R}$ *is continuous and* $c \in [a,b]$*. Then* $\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$*.*

Corollary. *Assume that* $f, g\colon [a,b] \to \mathbb{R}$ *are continuous.*

(a) *If* $f \leq g$*, then* $\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx$*.*
(b) $\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx$*.*
(c) $\int_a^b \big(f(x) + g(x)\big)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$*.*
(d) $\int_a^b \big(c \cdot f(x)\big)\,dx = c \cdot \int_a^b f(x)\,dx$*.*
(e) $\int_a^b c\,dx = c \cdot (b-a)$*.*

6.2. **Integration and differentiation.** Up to now we have considered the integral with respect to a fixed integration interval. Now we view the upper bound of this interval as variable and study the function obtained in this way. It is called the "undetermined integral".

Given $a_0 < c < b_0$ and a continuous $f\colon [a_0, b_0] \to \mathbb{R}$, we first need to establish $F(x) := \int_c^x f(t)\,dt$ as a continuous function. This means that we have to come up with h_F, α_F and ω_F; as lower bound we can take $N_F := (b_0 - a_0)N_f$ and as upper bound $M_F := (b_0 - a_0)M_f$. Let

$$h_F(a,n) := S(f,c,a,n).$$

By the theorem above we know that $(h_F(a,n))_n$ is a Cauchy sequence with modulus $k \mapsto 2^{\omega_f(k+1)}$. It remains to provide a modulus ω_F of (uniform) continuity. To this end, we may assume $c < a < b$. Divide the intervals $[c,a]$

and $[c,b]$ in n pieces each, and let $a_i := c + \frac{i}{n}(a-c)$ and $b_i := c + \frac{i}{n}(b-c)$. Then

$$\begin{aligned}
&\bigl|h_F(a,n) - h_F(b,n)\bigr| \\
&= \bigl|S(f,c,a,n) - S(f,c,b,n)\bigr| \\
&= \frac{1}{n}\bigl|(a-c)\sum_{i=0}^{n-1} h_f(a_i,n) - (b-c)\sum_{i=0}^{n-1} h_f(b_i,n)\bigr| \\
&\le \frac{1}{n}\sum_{i=0}^{n-1}\bigl((a-c)\bigl|h_f(a_i,n) - h_f(b_i,n)\bigr| + |a-b|\cdot|h_f(b_i,n)|\bigr) \\
&\le \frac{1}{n}\sum_{i=0}^{n-1}\bigl((a-c)\cdot 2^{-k-1} + |a-b|\cdot 2^p\bigr) \\
&\le 2^{-k}
\end{aligned}$$

provided $|a_i - b_i| \le 2^{-\omega_f(k+1)+1}$ and $|a-b| \le 2^{-p-k-1}$, where p is such that $h_f(b_i,n) \le 2^p$. So let

$$\alpha_F(k) := \max\bigl(\alpha_f(0), 2^{\omega_f(k+1)}\bigr), \quad \omega_F(k) := \max\bigl(p+k, \omega_f(k+1)\bigr).$$

Proposition. *Let $f\colon [a,b] \to \mathbb{R}$ be continuous with modulus ω_f of (uniform) continuity. Fix $c \in [a,b]$ and let*

$$F(x) := \int_c^x f(t)\,dt.$$

be the continuous function just described. Then this function $F\colon [a,b] \to \mathbb{R}$ has f as derivative, with modulus ω_f. Morover, if G is any differentiable function on $[a,b]$ with $G' = f$, then the difference $F - G$ is a constant.

Proof.

$$\begin{aligned}
\bigl|F(y) - F(x) - f(x)(y-x)\bigr| &= \Bigl|\int_c^y f(t)\,dt - \int_c^x f(t)\,dt - f(x)(y-x)\Bigr| \\
&= \Bigl|\int_x^y f(t)\,dt - \int_x^y f(x)\,dt\Bigr| \\
&\le \int_x^y \bigl|f(t) - f(x)\bigr|\,dt \\
&\le \int_x^y 2^{-k}\,dt = 2^{-k}(y-x)
\end{aligned}$$

for $y \le x + 2^{-\omega_f(k)}$; this was to be shown. Now let G be any differentiable function on $[a,b]$ with $G' = f$. Then $(F-G)' = F' - G' = f - f = 0$, hence $F - G$ is a constant, by corollary DerivZero in 5.2. $\square$

Theorem (Fundamental theorem of calculus). *Let $f\colon I \to \mathbb{R}$ be continuous and $F\colon I \to \mathbb{R}$ such that $F' = f$. Then for all $a, b \in I$*

$$\int_a^b f(t)\,dt = F(b) - F(a).$$

Proof. For $x \in I$ define

$$F_0(x) := \int_a^x f(t)\,dt.$$

By the proposition we have $F_0' = f$. Clearly

$$F_0(a) = 0 \qquad \text{and} \qquad F_0(b) = \int_a^b f(t)\,dt.$$

Hence for any $F\colon I \to \mathbb{R}$ such that $F' = f$, by corollary DerivZero in 5.2 the function $F - F_0$ is a constant. Therefore

$$F(b) - F(a) = F_0(b) - F_0(a) = F_0(b) = \int_a^b f(t)\,dt. \qquad \square$$

It is common to use the notation

$$F(x)\Big|_a^b \quad \text{or} \quad \Big[F(x)\Big]_a^b \quad \text{for} \quad F(b) - F(a).$$

The formula from the fundamental theorem of calculus can then be written as

$$\int_a^b f(x)\,dx = F(x)\Big|_a^b \quad \text{or} \quad \int_a^b f(x)\,dx = \Big[F(x)\Big]_a^b.$$

Let $f\colon I \to \mathbb{R}$ be continuous. For arbitrary reals $x, y \in I$ we define

$$\int_x^y f(t)\,dt := F(y) - F(x), \tag{13}$$

where F is the function from the proposition (which has f as derivative). Clearly this definition does not depend on the choice of the constant c implicit in the function F.

6.3. **Substitution rule, partial integration.**

Theorem (Substitution rule)**.** *Let $f\colon I \to \mathbb{R}$ be continuous and $\varphi\colon [a,b] \to \mathbb{R}$ differentiable such that $\varphi([a,b]) \subseteq I$. Then*

$$\int_a^b f(\varphi(t))\varphi'(t)\,dt = \int_{\varphi(a)}^{\varphi(b)} f(x)\,dx.$$

Remark. With the symbolic notation

$$d\varphi(t) := \varphi'(t)\,dt$$

the above formula can be written as

$$\int_a^b f(\varphi(t))\,d\varphi(t) = \int_{\varphi(a)}^{\varphi(b)} f(x)\,dx.$$

This is easy to remember, for one only has to replace x by $\varphi(t)$. The integration bounds can be inferred as well: if t ranges from a to b, then x $(= \varphi(t))$ ranges from $\varphi(a)$ to $\varphi(b)$.

Proof. Let $F\colon I \to \mathbb{R}$ be such that $F' = f$. For $F \circ \varphi\colon [a,b] \to \mathbb{R}$ we have by the chain rule

$$(F \circ \varphi)'(t) = F'(\varphi(t))\varphi'(t) = f(\varphi(t))\varphi'(t).$$

Hence by the fundamental theorem of calculus

$$\int_a^b f(\varphi(t))\varphi'(t)\,dt = F(\varphi(b)) - F(\varphi(a)) = \int_{\varphi(a)}^{\varphi(b)} f(x)\,dx. \qquad \square$$

Theorem (Partial integration)**.** *Let $f, g\colon [a,b] \to \mathbb{R}$ be differentiable functions. Then*

$$\int_a^b f(x)g'(x)\,dx = f(x)g(x)\Big|_a^b - \int_a^b g(x)f'(x)\,dx.$$

Remark. A short notation for this formula is

$$\int f\,dg = fg - \int g\,df.$$

Proof. For $F := fg$ we have by the product rule

$$F'(x) = f'(x)g(x) + f(x)g'(x),$$

whence by the fundamental theorem of calculus

$$\int_a^b f'(x)g(x)\,dx + \int_a^b f(x)g'(x)\,dx = F(x)\Big|_a^b = f(x)g(x)\Big|_a^b. \qquad \square$$

6.4. **Intermediate value theorem of integral calculus.**

Theorem (Intermediate value theorem of integral calculus)**.** *Let $f, \varphi\colon I \to \mathbb{R}$ be continuous and f locally nonconstant. Assume that we have rationals $a \le c < d \le b$ in I such that*

$$f(c) \le f(t) \le f(d) \quad (t \in [a,b]).$$

Assume further $\varphi \ge 0$ and $0 < \int_a^b \varphi(t)\,dt$. Then we can find $x \in [c,d]$ such that

$$\int_a^b f(t)\varphi(t)\,dt = f(x)\int_a^b \varphi(t)\,dt.$$

Proof. By assumption

$$f(c)\int_a^b \varphi(t)\,dt \le \int_a^b f(t)\varphi(t)\,dt \le f(d)\int_a^b \varphi(t)\,dt,$$

whence we have $y \in [f(c), f(d)]$ such that

$$\int_a^b f(t)\varphi(t)\,dt = y\int_a^b \varphi(t)\,dt.$$

By the intermediate value theorem we obtain an $x \in [c,d]$ such that $f(x) = y$, as required. □

6.5. **Inverse of the exponential function.** We use the machinery developed in this section to define the inverse of the exponential function. To motivate the definition, suppose we already have a differentiable function $\ln\colon (0,\infty) \to \mathbb{R}$ such that $\exp(\ln(x)) = x$ for $x > 0$. Then the chain rule entails

$$\frac{d}{dx}\exp(\ln(x)) = \exp(\ln(x)) \cdot \frac{d}{dx}\ln(x) = 1,$$

hence

$$\frac{d}{dx}\ln(x) = \frac{1}{x}.$$

Because of $\exp(\ln(1)) = 1$ we must also have $\ln(1) = 0$.

Therefore we define

$$\ln(x) := \int_1^x \frac{dt}{t} \quad (x > 0). \tag{14}$$

Because of $\exp(x) > 0$ the composite function $\ln \circ \exp$ is continuous on $\mathbb{R}$. Its derivative is

$$\frac{d}{dx}\ln(\exp(x)) = \frac{1}{\exp(x)} \cdot \exp(x) = 1 = \frac{d}{dx}x.$$

By corollary DerivZero in 5.2 the function $\ln(\exp(x)) - x$ is a constant, and because of $\ln(\exp(0)) = \ln(1) = 0$ this constant must be 0. Hence

$$\ln(\exp(x)) = x \quad (x \in \mathbb{R}). \tag{15}$$

Now fix $x > 0$ and let $y := \exp(\ln(x))$. Then

$$\ln(y) = \ln(\exp(\ln(x))) = \ln(x)$$

by (15), hence

$$0 = \ln(y) - \ln(x) = \int_x^y \frac{dt}{t}.$$

Assuming $x < y$ clearly leads to a contradiction, hence $x \geq y$. Similarly we obtain $y \geq x$ and therefore $x = y$. Hence

$$\exp(\ln(x)) = x \quad (x > 0). \tag{16}$$

To prove the familiar functional equation for the logarithm, fix $y > 0$ and consider $\ln(xy) - \ln(y)$ $(x > 0)$. Then

$$\frac{d}{dx}\big(\ln(xy) - \ln(y)\big) = \frac{1}{xy} \cdot y - 0 = \frac{1}{x} = \frac{d}{dx}\ln(x).$$

By corollary DerivZero in 5.2 $x \mapsto \ln(xy) - \ln(y) - \ln(x)$ is a constant, which must be 0 since this expression vanishes at $x = 1$. Hence

$$\ln(xy) = \ln(y) + \ln(x) \quad (x, y > 0). \tag{17}$$

Now we can define general exponentiation by

$$x^y := \exp(y \cdot \ln(x)) \quad (x > 0)$$

and derive easily all its usual properties.

7. Taylor series

We now study more systematically the development of functions in power series.

Theorem (Taylor formula)**.** *Let $f\colon I \to \mathbb{R}$ be $(n+1)$-times differentiable. Then for all $a, x \in I$*

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_{n+1}(x)$$

with

$$R_{n+1}(x) = \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\, dt.$$

Proof. By induction on n. Basis $n = 0$. By the fundamental theorem of calculus

$$f(x) = f(a) + \int_a^x f'(t)\,dt.$$

Step $n \to n+1$. By induction hypothesis

$$\begin{aligned} R_n(x) &= \frac{1}{(n-1)!}\int_a^x (x-t)^{n-1} f^{(n)}(t)\,dt \\ &= -\int_a^x f^{(n)}(t)\Big(\frac{d}{dt}\frac{(x-t)^n}{n!}\Big)\,dt \\ &= -f^{(n)}(t)\frac{(x-t)^n}{n!}\Big|_{t=a}^{t=x} + \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)\,dt \\ &= \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\,dt. \end{aligned}$$ □

Corollary. *Let $f\colon I \to \mathbb{R}$ be $(n+1)$-times differentiable with $f^{(n+1)}(x) = 0$ for all $x \in I$. Then f is a polynomial of degree $\le n$.*

Proof. In this case we have $R_{n+1}(x) = 0$. □

Theorem (Lagrange)**.** *Let $f\colon I \to \mathbb{R}$ be $(n+1)$-times differentiable and $a, x \in I$. Assume that $f^{(n+1)}$ is locally nonconstant and that we have rationals c, d with $a \le c < d \le x$ such that*

$$f^{(n+1)}(c) \le f^{(n+1)}(t) \le f^{(n+1)}(d) \quad (t \in [a, x]).$$

Then we can find $\xi \in [a, b]$ such that

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}.$$

Proof. By the intermediate value theorem of integral calculus we can construct $\xi \in [c, d]$ such that

$$\begin{aligned} R_{n+1}(x) &= \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\,dt \\ &= f^{(n+1)}(\xi)\int_a^x \frac{(x-t)^n}{n!}\,dt \\ &= -f^{(n+1)}(\xi)\frac{(x-t)^{n+1}}{(n+1)!}\Big|_{t=a}^{t=x} \\ &= \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}. \end{aligned}$$ □

8. Sequences of functions

8.1. **Uniform convergence.** We define the notion of uniform convergence of a sequence of continuous functions $f_n \colon I \to \mathbb{R}$ to a continuous function $f \colon I \to \mathbb{R}$. The definition is in terms of witnesses for the given continuous functions f_n, in order to ensure that from a proof of uniform convergence we can extract the right data.

Definition. Let $f_n, f \colon I \to \mathbb{R}$ be continuous, with approximating maps h_n, h and Cauchy moduli α_n, α. The sequence $(f_n)_{n\in\mathbb{N}}$ is *uniformly convergent* to f if

$$\forall_p \exists_q \forall_{n\geq q} \forall_{a\in I} \big(\big| h_n(a, \alpha_n(p+2)) - h(a, \alpha(p+2)) \big| \leq 2^{-p} \big).$$

The next lemma gives a useful characterization of uniform convergence.

Lemma (UnifConvChar)**.** *Let $f_n, f \colon I \to \mathbb{R}$ be continuous, with approximating maps h_n, h. Then the following are equivalent.*

(a) *The sequence $(f_n)_{n\in\mathbb{N}}$ is uniformly convergent to f.*
(b) $\forall_p \exists_{q_1} \forall_{n\geq q_1} \exists_{q_2} \forall_{k\geq q_2} \forall_{a\in I} (|h_n(a,k) - h(a,k)| \leq 2^{-p})$.

Proof. (a) $\Rightarrow$ (b). Given p, pick q_1 by (a) for $p+1$. Given $n \geq q_1$,

$$\begin{aligned} \big| h_n(a,k) - h(a,k) \big| \leq{} & \big| h_n(a,k) - h_n(a, \alpha_n(p+3)) \big| + \\ & \big| h_n(a, \alpha_n(p+3)) - h(a, \alpha(p+3)) \big| + \\ & \big| h(a, \alpha(p+3)) - h(a,k) \big| \\ \leq{} & 2^{-p-3} + 2^{-p-1} + 2^{-p-3} \end{aligned}$$

if $k \geq q_2 := \alpha_n(p+3), \alpha(p+3)$. Then the first and last term are $\leq 2^{-p-3}$, and the middle term is $\leq 2^{-p-1}$ by the choice of q_1.

(b) $\Rightarrow$ (a).

$$\begin{aligned} \big| h_n(a, \alpha_n(p+2)) - h(a, \alpha(p+2)) \big| \leq{} & \big| h_n(a, \alpha_n(p+2)) - h_n(a,k) \big| + \\ & \big| h_n(a,k) - h(a,k) \big| + \\ & \big| h(a,k) - h(a, \alpha(p+2)) \big| \\ \leq{} & 2^{-p-2} + 2^{-p-1} + 2^{-p-2} \end{aligned}$$

if $k \geq \alpha_n(p+2), \alpha(p+2)$ (for the first and last term) and in addition $n, k \geq p$ with p provided for $p+1$ by (b). $\square$

We now show that a uniformly convergent sequence indeed is uniformly convergent in the usual sense.

Lemma. *Let $f_n, f\colon I \to \mathbb{R}$ be continuous. Assume that the sequence $(f_n)_{n\in\mathbb{N}}$ is uniformly convergent to f. Then*

$$\forall_p \exists_q \forall_{n\geq q} \forall_x \bigl(\bigl|f_n(x) - f(x)\bigr| \leq 2^{-p}\bigr).$$

Proof. Let $x = ((a_k)_k, M)$, and let h_n, h be approximating maps for f_n, f, respectively. By lemma UnifConvChar

$$\forall_p \exists_{q_1} \forall_{n\geq q_1} \exists_{q_2} \forall_{k\geq q_2} \forall_{a\in I} (|h_n(a,k) - h(a,k)| \leq 2^{-p}),$$

whence the claim. □

The next lemma gives a useful criterion as to when and how we can construct the limit function. It will be used below.

Lemma (UnifConvLim)**.** *Let $f_n\colon I \to \mathbb{R}$ be continuous functions, given by approximating functions h_n and moduli α of Cauchyness and ω of (uniform) continuity, where the latter two are* independent *of n. Assume we have a weakly increasing modulus $\delta\colon \mathbb{N} \to \mathbb{N}$ of* uniform convergence *satisfying*

$$\bigl|h_n(a,k) - h_m(a,k)\bigr| \leq 2^{-p}$$

for $n, m \geq \delta(p)$ and $k \geq \alpha(p)$, and all $a \in I$. Then $(f_n)_{n\in\mathbb{N}}$ uniformly converges to the continuous function $f\colon I \to \mathbb{R}$ given by

$$h_f(a,k) := h_k(a,k), \quad \alpha_f(p) := \max\bigl(\delta(p+1), \alpha(p+1)\bigr), \quad \omega_f := \omega.$$

Proof. It is easy to see that this function $f\colon I \to \mathbb{R}$ given by h_f, α_f and ω_f is indeed continuous: α_f is a Cauchy modulus, because

$$\begin{aligned} |h_k(a,k) - h_l(a,l)| &\leq |h_k(a,k) - h_l(a,k)| + |h_l(a,k) - h_l(a,l)| \\ &\leq 2^{-p-1} + 2^{-p-1} \end{aligned}$$

for $k, l \geq \alpha_f(p)$, and ω is a modulus of (uniform) continuity, because

$$|a - b| \leq 2^{-\omega(p)+1} \to |h_k(a,k) - h_k(b,k)| \leq 2^{-p}$$

for $k \geq \alpha_f(p)$. Moreover, for a given p we may pick $p := \max(\delta(p), \alpha(p))$. Then for $n, k \geq p$ and all a, $|h_n(a,k) - h_k(a,k)| \leq 2^{-p}$. By lemma UnifConvChar, this implies that $(f_i)_{i\in\mathbb{N}}$ uniformly converges to f. □

8.2. **Integration, differentiation and limits.** We show that for a uniformly convergent sequence of continuous functions, integration and limits can be exchanged.

Theorem (IntLimit)**.** *Let $f_n, f\colon [a,b] \to \mathbb{R}$ be continuous, and assume that for f_n the moduli of Cauchyness and of (uniform) continuity are independent of n. Assume that the sequence $(f_n)_{n\in\mathbb{N}}$ is uniformly convergent to f. Then*

$$\lim_{n\to\infty} \int_a^b f_n(t)\,dt = \int_a^b f(t)\,dt.$$

Proof. Let $a_{nk} := S(f_n, k)$, $a_k := S(f, k)$,

$$\int_a^b f_n(t)\,dt = (a_{nk})_k =: x_n \quad \text{and} \quad \int_a^b f(t)\,dt = (a_k)_k =: x.$$

We show that $(x_n)_n$ converges to x, that is $|x_n - x| \le 2^{-p}$ for $n \ge M(p)$. Observe that for any k,

$$|x_n - x| \le |x_n - a_{nk}| + |a_{nk} - a_k| + |a_k - x|.$$

Recall that by definition

$$\begin{aligned}&\int_a^b f_n(t)\,dt = (S(f_n, n), M_n) \quad \text{with}\\ &M_n(p) = \max(2^{\omega_{f_n}(p+q+1)}(b-a), \alpha_{f_n}(p+q+2)),\end{aligned}$$

where q is such that $b - a \le 2^q$. In our case, the moduli α_{f_n} of Cauchyness and ω_{f_n} of (uniform) continuity are independent of n, say α and ω. So instead of M_n we can take $M(p) := \max(2^{\omega_f(p+q+1)}(b-a), \alpha_f(p+q+2))$.

We now estimate each of the three parts of $|x_n - a_{nk}| + |a_{nk} - a_k| + |a_k - x|$ separately.

First, $|x_n - a_{nk}| \le 2^{-p-2}$ for $k \ge M(p+2)$; here we need lemma RatCauchyConvMod in 2.1.

Second, for a given l such that $b - a \le 2^l$, by lemma UnifConvChar in 8.1 we can pick q_1 such that for all $n \ge q_1$ we can pick q_2 such that for all $k \ge q_2$ we have $|h_n(a_i, k) - h(a_i, k)| \le 2^{-p-1-l}$. Hence

$$\begin{aligned}|a_{nk} - a_k| &\le \frac{b-a}{k} \sum_{i=0}^{k-1} \bigl|h_n(a_i, k) - h(a_i, k)\bigr|\\ &\le \frac{b-a}{k} k 2^{-p-1-l} \le 2^{-p-1}.\end{aligned}$$

Third, $|a_k - x| \le 2^{-p-2}$ for $k \ge \alpha'(p+2)$ with α' the Cauchy modulus of $(a_k)_k$; here again we need lemma RatCauchyConvMod in 2.1.

Finally, for $n \geq q_1$ and $k \geq \max(M(p+2), \alpha'(p+2), q_2)$ (with q_2 depending on n) we have all three estimates simultaneously and hence

$$\begin{aligned} |x_n - x| &\leq |x_n - a_{nk}| + |a_{nk} - a_k| + |a_k - x| \\ &\leq 2^{-p-2} + 2^{-p-1} + 2^{-p-2} = 2^{-p}. \end{aligned}$$

Therefore it suffices to take $n \geq q_1$. □

The final theorem gives a sufficient criterium as to when differentiation and limits can be exchanged.

Theorem (DiffLimit). *Let the continuous functions $f_n \colon [a,b] \to \mathbb{R}$ be uniformly convergent to a continuous $f \colon [a,b] \to \mathbb{R}$. Assume that each f_n is differentiable with derivative f'_n, and assume that for f'_n the moduli of Cauchyness and of (uniform) continuity are independent of n. Moreover assume that the sequence $(f'_n)_{n \in \mathbb{N}}$ is uniformly convergent to a continuous $f^* \colon [a,b] \to \mathbb{R}$. Then f is differentiable with derivative f^*.*

Proof. By the fundamental theorem of calculus

$$f_n(c) = f_n(a) + \int_a^c f'_n(t)\,dt \quad (c \in [a,b]).$$

By the theorem above, $\int_a^c f'_n(t)\,dt$ converges for $n \to \infty$ to $\int_a^c f^*(t)\,dt$. whence

$$f(c) = f(a) + \int_a^c f^*(t)\,dt \quad (c \in [a,b]).$$

Now let $x = (c_k)_k$ be a real in $[a,b]$. Then

$$f(x) = f(a) + \int_a^x f^*(t)\,dt \quad (x \in [a,b]).$$

By 6.2, f is differentiable with derivative f^*. □

9. Trigonometric functions

9.1. **Euler's formula.** For all $x \in \mathbb{R}$ let

$$\cos x := \Re(e^{ix}), \quad \sin x := \Im(e^{ix}),$$

hence $e^{ix} = \cos x + i \sin x$ (*Euler's formula*indexEuler's formula).

Notice that for all $x \in \mathbb{R}$ we have $|e^{ix}| = 1$, because

$$|e^{ix}|^2 = e^{ix}\overline{e^{ix}} = e^{ix}e^{-ix} = e^0 = 1.$$

Therefore e^{ix} is a point on the unit circle of the Gaußian plane and $\cos x$, $\sin x$ are the projections of this point to the x- and y-axis. Immediately from the definitions we have

$$\cos x = \frac{1}{2}(e^{ix} + e^{-ix}) \quad \text{and} \quad \sin x = \frac{1}{2i}(e^{ix} - e^{-ix}),$$
$$\cos(-x) = \cos x \quad \text{and} \quad \sin(-x) = -\sin x,$$
$$\cos^2 x + \sin^2 x = 1.$$

Theorem. *The functions* $\cos x$ *and* $\sin x$ *are continuous on all of* $\mathbb{R}$.

Proof. Omitted. □

9.2. **Addition theorems.**

Theorem (Addition Theorems)**.** *For all* $x, y \in \mathbb{R}$ *we have*

$$\cos(x + y) = \cos x \cos y - \sin x \sin y,$$
$$\sin(x + y) = \sin x \cos y + \cos x \sin y.$$

Proof. From the functional equation of the exponential function

$$e^{i(x+y)} = e^{ix}e^{iy}$$

we obtain by Euler's formula

$$\begin{aligned}
&\cos(x + y) + i\sin(x + y) \\
&= (\cos x + i\sin x)(\cos y + i\sin y) \\
&= (\cos x \cos y - \sin x \sin y) + i(\sin x \cos y + \cos x \sin y).
\end{aligned}$$

Comparing the real and imaginary parts gives the claim. □

Corollary. *For all* $x, y \in \mathbb{R}$ *we have*

$$\sin x - \sin y = 2\cos\frac{x+y}{2}\sin\frac{x-y}{2},$$
$$\cos x - \cos y = -2\sin\frac{x+y}{2}\sin\frac{x-y}{2}.$$

Proof. Let $u := \frac{x+y}{2}$ and $v := \frac{x-y}{2}$; that $x = u + v$ and $y = u - v$. The addition theorem for sin entails

$$\begin{aligned}
\sin x - \sin y &= \sin(u + v) - \sin(u - v) \\
&= \sin u \cos v + \cos u \sin v - \sin u \cos(-v) - \cos u \sin(-v) \\
&= \sin u \cos v + \cos u \sin v - \sin u \cos v + \cos u \sin v \\
&= 2\cos u \sin v.
\end{aligned}$$

The second equation is proved similarly. □

Theorem. *For all $x \in \mathbb{R}$ we have*

$$\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - + \dots,$$

$$\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - + \dots.$$

Both series converge absolutely.

Proof. Absolute convergence follows from the absolute convergence of the exponential series. Using

$$i^n = \begin{cases} 1, & \text{if } n = 4m; \\ i, & \text{if } n = 4m+1; \\ -1, & \text{if } n = 4m+2; \\ -i, & \text{if } n = 4m+3 \end{cases} \qquad (m \in \mathbb{Z}).$$

we obtain for all $x \in \mathbb{R}$

$$\begin{aligned} e^{ix} &= \sum_{n=0}^{\infty} i^n \frac{x^n}{n!} \\ &= 1 + ix - \frac{x^2}{2!} - i\frac{x^3}{3!} + \frac{x^4}{4!} + i\frac{x^5}{5!} - \frac{x^6}{6!} - i\frac{x^7}{7!} + \frac{x^8}{8!} + \dots \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} + i \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} \end{aligned}$$

Comparing the real and imaginary parts gives the claim. □

9.3. **Estimate of the rest.**

Theorem (Estimate of the rest)**.** *For all $x \in \mathbb{R}$ we have*

$$\cos x = \sum_{k=0}^{n} (-1)^k \frac{x^{2k}}{(2k)!} + r_{2n+2}(x),$$

$$\sin x = \sum_{k=0}^{n} (-1)^k \frac{x^{2k+1}}{(2k+1)!} + r_{2n+3}(x),$$

where

$$|r_{2n+2}(x)| \le \frac{|x|^{2n+2}}{(2n+2)!} \quad \textit{for } |x| \le 2n+3,$$

$$|r_{2n+3}(x)| \leq \frac{|x|^{2n+3}}{(2n+3)!} \quad \textit{for } |x| \leq 2n+4.$$

Remark. These estimates are valid for *all* $x \in \mathbb{R}$; this can be proved by means of the Taylor formula.

Proof. For all $x \in \mathbb{R}$ we have

$$r_{2n+2}(x) = \pm \frac{x^{2n+2}}{(2n+2)!}\big(1 - \frac{x^2}{(2n+3)(2n+4)} + \dots \\ \pm \frac{x^{2k}}{(2n+3)\cdots(2n+2k+2)} \mp \dots\big).$$

For $k \geq 1$ let

$$a_k := \frac{x^{2k}}{(2n+3)(2n+4)\cdots(2n+2k+2)}.$$

Then for all $|x| \leq 2n+3$

$$1 \geq a_1 \geq a_2 \geq \dots \geq a_k \geq 0.$$

As in the proof of the Leibniz test we obtain

$$a_1 - a_2 + a_3 - + \dots \mp a_k \geq 0 \quad \text{and} \quad 1 - a_1 + a_2 - a_3 + - \dots \pm a_k \geq 0,$$

hence

$$|r_{2n+2}(x)| = \frac{|x|^{2n+2}}{(2n+2)!}(1 - a_1 + a_2 - a_3 + - \dots \pm a_k \mp \dots) \leq \frac{|x|^{2n+2}}{(2n+2)!}.$$

The second estimate is proved similarly. □

Corollary.

$$\lim_{\substack{x \to 0 \\ x \neq 0}} \frac{\sin x}{x} = 1.$$

Proof. We use the 3rd order rest, i.e.,

$$\sin x = x + r_3(x)$$

with

$$|r_3(x)| \leq \frac{|x|^3}{3!} \quad \text{for } |x| \leq 4.$$

This gives for $0 < |x| \leq 3$

$$\Big|\frac{\sin x}{x} - 1\Big| = \frac{|r_3(x)|}{|x|} \leq \frac{|x|^2}{6}$$

and hence the claim. □

9.4. Definition of π.

Theorem. *The function* cos *has exactly one zero in the interval* $[0, 2]$.

For the proof we need three auxiliary lemmata.

Lemma. $\cos 2 \le -\frac{1}{3}$.

Proof. We use the 4th order rest, i.e.,

$$\cos x = 1 - \frac{x^2}{2} + r_4(x)$$

with

$$|r_4(x)| \le \frac{|x|^4}{4!} \quad \text{for } |x| \le 5.$$

Hence

$$\cos 2 = 1 - 2 + r_4(2)$$

with

$$|r_4(2)| \le \frac{16}{24} = \frac{2}{3}$$

and hence the claim $\cos 2 \le -\frac{1}{3}$. □

Lemma. $\sin x > 0$ *for all* $x \in]0, 2]$.

Proof. Because of

$$\sin x = x + r_3(x)$$

with

$$|r_3(x)| \le \frac{|x|^3}{3!} \quad \text{for } |x| \le 4$$

we have for $x \in]0, 2]$

$$\sin x \ge x - \frac{x^3}{6} = \frac{x}{6}(6 - x^2) > 0.$$

This proves the claim. □

Lemma. *The function* cos *is strictly decreasing in the interval* $[0, 2]$.

Proof. Let $0 \le x < x' \le 2$. From the Corollary of the addition theorem we have

$$\cos x' - \cos x = -2 \sin \frac{x' + x}{2} \sin \frac{x' - x}{2}.$$

$0 < \frac{x'-x}{2} \le 1$ implies $\sin \frac{x'-x}{2} > 0$ by the second auxiliary lemma, and $0 < \frac{x'+x}{2} \le 2$ implies $\sin \frac{x'+x}{2} > 0$ again by the second auxiliary lemma, hence $\cos x' - \cos x < 0$. □

Proof of the theorem. $\cos 0 = 1$, $\cos 2 \leq -\frac{1}{3}$ and the fact that cos is strictly decreasing in the interval $[0, 2]$ imply the claim. □

We now define the real number $\pi/2$ as the (uniquely determined) zero of the function cos in the interval $[0, 2]$.

Theorem (Special values of the exponential function)**.**

$$e^{i\frac{\pi}{2}} = i, \quad e^{i\pi} = -1, \quad e^{i\frac{3\pi}{2}} = -i, \quad e^{2\pi i} = 1.$$

Proof. Because of $\cos^2 x + \sin^2 x = 1$ and the definition of $\frac{\pi}{2}$ we have $\sin\frac{\pi}{2} = \pm 1$, hence by the second auxiliary lemma above $\sin\frac{\pi}{2} = 1$. This implies

$$e^{i\frac{\pi}{2}} = \cos\frac{\pi}{2} + i\sin\frac{\pi}{2} = i.$$

Hence by the functional equation

$$\begin{aligned} e^{i\pi} &= e^{i\frac{\pi}{2}}e^{i\frac{\pi}{2}} = i^2 = -1 \\ e^{i\frac{3\pi}{2}} &= e^{i\pi}e^{i\frac{\pi}{2}} = -i \\ e^{2\pi i} &= e^{i\pi}e^{i\pi} = 1. \end{aligned}$$

□

Therefore

x	0	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π
$\sin x$	0	1	0	-1	0
$\cos x$	1	0	-1	0	1

Corollary. *For all $x \in \mathbb{R}$ we have*

(a) $\cos(x + 2\pi) = \cos x$, $\sin(x + 2\pi) = \sin x$
(b) $\cos(x + \pi) = -\cos x$, $\sin(x + \pi) = -\sin x$
(c) $\cos x = \sin(\frac{\pi}{2} - x)$, $\sin x = \cos(\frac{\pi}{2} - x)$

Proof. (a) $e^{i(x+2\pi)} = e^{ix}e^{2\pi i} = e^{ix}$.
(b) $e^{i(x+\pi)} = e^{ix}e^{i\pi} = -e^{ix}$.
(c) $e^{ix} = e^{ix - i\frac{\pi}{2} + i\frac{\pi}{2}} = e^{i(x-\frac{\pi}{2})}e^{i\frac{\pi}{2}} = ie^{i(x-\frac{\pi}{2})}$. Hence

$$\begin{aligned} \cos x &= -\sin\bigl(x - \frac{\pi}{2}\bigr) = \sin(\frac{\pi}{2} - x), \\ \sin x &= \cos(\frac{\pi}{2} - x) = \cos\bigl(x - \frac{\pi}{2}\bigr). \end{aligned}$$

□

Corollary (Zeros of sine and cosine)**.** *In the interval* $[0, 2\pi[$ *the function* cos *has exactly the zeros* $\frac{\pi}{2}$ *and* $\frac{3\pi}{2}$*, and* sin *has exactly the zeros* 0 *and* π*.*

Proof. 1. $\cos\frac{\pi}{2} = 0$ by definition, hence also $\cos\frac{3\pi}{2} = -\cos\frac{\pi}{2} = 0$. Moreover, cos is strictly decreasing in $[0, \frac{\pi}{2}]$, and $\cos(\frac{\pi}{2}+x) = -\cos(\frac{\pi}{2}-x)$. Hence cos is strictly decreasing in $[\frac{\pi}{2}, \pi]$ as well. Therefore $\frac{\pi}{2}$ is the unique zero of cos in $[0, \pi]$. Furthermore $\cos(\pi + x) = \cos(-\pi - x) = \cos(\pi - x)$; hence cos has exactly one zero in $[\pi, 2\pi]$, namely $\frac{3\pi}{2}$.

2. Because of $\sin x = \cos(\frac{\pi}{2} - x) = \cos(x - \frac{\pi}{2})$ the claim follows from the fist part. □

Corollary. *In the interval* $[0, 2\pi[$, *the function* cos *assumes the value* 1 *exactly in the point* 0.

Proof. We have just shown that cos is strictly decreasing in $[0, \pi]$, and that $\cos(\pi + x) = \cos(\pi - x)$. Because of $\cos 0 = 1$ the claim follows. □

We can now define the tangens function for $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$ by

$$\tan x := \frac{\sin x}{\cos x}.$$

9.5. **The inverse functions arcsin, arccos, arctan.** The inverse functions arccos for cos, arcsin for sin and arctan for tan may now be defined similarly to how we defined the logarithm as the inverse of the exponential function, i.e., by means of integrals. We carry this out for the sine function. To motivate the definition of the inverse arcsin of the sine function, suppose we already have a differentiable function $\arcsin\colon (-1, 1) \to (-\pi/2, \pi/2)$ such that $\sin(\arcsin x) = x$ for $-1 < x < 1$. Then the chain rule entails

$$\frac{d}{dx}\sin(\arcsin x) = \cos(\arcsin x) \cdot \frac{d}{dx}\arcsin x = 1,$$

hence

$$\frac{d}{dx}\arcsin x = \frac{1}{\cos(\arcsin x)} = \frac{1}{\sqrt{1-\sin^2(\arcsin x)}} = \frac{1}{\sqrt{1-x^2}}.$$

Because of $\sin(\arcsin(0)) = 0$ we must also have $\arcsin(0) = 0$.

Therefore we define

$$\arcsin x := \int_0^x \frac{dt}{\sqrt{1-t^2}} \quad (-1 < x < 1). \tag{18}$$

Because of $-1 < \sin x < 1$ the composite function $\arcsin \circ \sin$ is continuous on $(-\pi/2, \pi/2)$. Its derivative is

$$\frac{d}{dx}\arcsin(\sin x) = \frac{1}{\sqrt{1-\sin^2 x}} \cdot \cos x = 1 = \frac{d}{dx}x.$$

By corollary DerivZero in 5.2 the function $\arcsin(\sin x) - x$ is a constant, and because of $\arcsin(\sin(0)) = \arcsin(0) = 0$ this constant must be 0. Hence

$$\arcsin(\sin x) = x \quad (-\pi/2 < x < \pi/2). \tag{19}$$

Now fix $x > 0$ and let $y := \sin(\arcsin x)$. Then

$$\arcsin y = \arcsin(\sin(\arcsin x)) = \arcsin x$$

by (19), hence

$$0 = \arcsin y - \arcsin x = \int_x^y \frac{dt}{\sqrt{1-t^2}}.$$

Assuming $x < y$ clearly leads to a contradiction, hence $x \geq y$. Similarly we obtain $y \geq x$ and therefore $x = y$. Hence

$$\sin(\arcsin x) = x \quad (-1 < x < 1). \tag{20}$$

Similarly we can introduce $\arccos\colon (-1, 1) \to (0, \pi)$ by

$$\arccos x := \frac{\pi}{2} - \int_0^\pi \frac{dt}{\sqrt{1-t^2}}. \tag{21}$$

For the inverse $\arctan\colon \mathbb{R} \to (-1, 1)$ of the tangens function $\tan x := \frac{\sin x}{\cos x}$ first recall that by the quotient rule

$$\frac{d}{dx} \tan x = \frac{1}{\cos^2 x}.$$

To motivate the definition of the inverse arctan of the tangens function, suppose we already have a differentiable function $\arctan\colon \mathbb{R} \to (-1, 1)$ such that $\tan(\arctan x) = x$ for $x \in \mathbb{R}$. Then the chain rule entails

$$\frac{d}{dx} \tan(\arctan x) = \frac{1}{\cos^2(\arctan x)} \cdot \frac{d}{dx} \arctan x = 1,$$

hence

$$\frac{d}{dx} \arctan x = \cos^2(\arctan x).$$

Now let $y := \arctan x$, hence $x = \tan y$. Then

$$x^2 = \tan^2 y = \frac{\sin^2 y}{\cos^2 y} = \frac{1 - \cos^2 y}{\cos^2 y} = \frac{1}{\cos^2 y} - 1,$$

hence

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

and therefore

$$\frac{d}{dx} \arctan x = \frac{1}{1+x^2}.$$

Because of $\tan(0) = 0$ we must also have $\arctan(0) = 0$. So we define

$$\arctan x := \int_0^x \frac{dt}{1+t^2} \quad (x \in \mathbb{R}). \tag{22}$$

Clearly the composite function $\arctan \circ \tan$ is continuous on $(-\pi/2, \pi/2)$. Its derivative is

$$\frac{d}{dx} \arctan(\tan x) = \frac{1}{1+\tan^2 x} \cdot \frac{1}{\cos^2 x} = 1 = \frac{d}{dx} x.$$

By corollary DerivZero in 5.2 the function $\arctan(\tan x) - x$ is a constant, and because of $\arctan(\tan(0)) = \arctan(0) = 0$ this constant must be 0. Hence

$$\arctan(\tan x) = x \quad (-\pi/2 < x < \pi/2). \tag{23}$$

Now fix $x > 0$ and let $y := \tan(\arctan x)$. Then

$$\arctan y = \arctan(\tan(\arctan x)) = \arctan x$$

by (23), hence

$$0 = \arctan y - \arctan x = \int_x^y \frac{dt}{1+t^2}.$$

Assuming $x < y$ clearly leads to a contradiction, hence $x \geq y$. Similarly we obtain $y \geq x$ and therefore $x = y$. Hence

$$\tan(\arctan x) = x \quad (x \in \mathbb{R}). \tag{24}$$

9.6. **Polar coordinates.**

Theorem (Polar coordinates)**.** *Every complex number $z \neq 0$ can be written uniquely in the form*

$$z = re^{i\varphi} \quad \text{with } r = |z| \text{ and } \varphi \in [0, 2\pi).$$

Proof. Let $\xi := \frac{z}{|z|}$, $x := \Re(\xi)$ and $y := \Im(\xi)$. Because of $|\xi| = 1$ we have $x^2 + y^2 = 1$, hence $x, y \in [-1.1]$. Let $\alpha \in [0, \pi]$ we the unique real such that $\cos\alpha = x$. Because of $y^2 = 1 - x^2 = 1 - \cos^2\alpha = \sin^2\alpha$ we have

$$y = \pm \sin\alpha.$$

Let

$$\varphi := \begin{cases} \alpha, & \text{if } y = \sin\alpha; \\ 2\pi - \alpha, & \text{if } y = -\sin\alpha. \end{cases}$$

Then in any case $\sin\varphi = y$ (we may assume here $|y| \geq \frac{1}{3}$, otherwise we work with x instead), and hence

$$e^{i\varphi} = \cos\varphi + i\sin\varphi = x + iy = \xi = \frac{z}{|z|}.$$

For uniqueness, assume $e^{i\varphi_1} = e^{i\varphi_2}$ with $0 \leq \varphi_1 \leq \varphi_2 < 2\pi$. Then $e^{i(\varphi_2-\varphi_1)} = 1$ with $\varphi_2 - \varphi_1 \in [0, 2\pi)$, hence $\varphi_2 - \varphi_1 = 0$. □

Remark. The product of two complex numbers can now simply be written as

$$r_1 e^{i\varphi_1} \cdot r_2 e^{i\varphi_2} = r_1 r_2 e^{i(\varphi_1+\varphi_2)}.$$

Corollary (nth root of unity)**.** *Let n be a natural number ≥ 2. The equation $z^n = 1$ has exactly n complex roots, namely*

$$e^{i\frac{2k\pi}{n}} \quad \textit{for } k = 0, \dots, n-1.$$

Proof. First notice that

$$\big(e^{i\frac{2k\pi}{n}}\big)^n = e^{2k\pi i} = 1.$$

Now let $z \in \mathbb{C}$ with $z^n = 1$. Then $|z| = 1$, hence by the theorem z can be written uniquely in the form $e^{i\varphi}$ with $\varphi \in [0, 2\pi)$. By asumption $(e^{i\varphi})^n = e^{in\varphi} = 1$, hence $n\varphi = 2k\pi$ for some $k \in \mathbb{Z}$, hence $\varphi = \frac{2k\pi}{n}$. Because of $\varphi \in [0, 2\pi)$ we obtain $0 \leq k < n$. □

10. Metric Spaces

We now generalize our treatment of the reals to metric spaces. Just as the rationals $\mathbb{Q}$ were taken as basic for $\mathbb{R}$, we assume that our starting point is a countable set Q of *approximations*. On these we assume a given *metric*, i.e., a map $d\colon Q \times Q \to \mathbb{R}$ such that for all $a, b, c \in Q$

(a) $d(a, b) = 0$ iff $a = b$,
(b) $d(a, b) = d(b, a)$ (symmetry), and
(c) $d(a, c) \leq d(a, b) + d(b, c)$ (triangle inequality).

Definition. A *metric* on a set X is a map $d\colon X \times X \to \mathbb{R}$ such that for all $x, y, z \in X$

(a) $d(x, y) = 0$ iff $x = y$,
(b) $d(x, y) = d(y, x)$ (symmetry), and
(c) $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

A *metric space* is a pair (X, d) consisting of a set X and a metric d on X. The real $d(x, y)$ is called *distance* of the points x and y w.r.t. the metric d.

We assume that every metric space considered is *separable* in the sense that it comes with a countable set Q of *approximations*, which is *dense* in X, i.e., such that $\forall_{x \in X} \forall_k \exists_{a \in Q}(d(x,a) \le 2^{-k})$.

Remark. The axioms entail $d(x,y) \ge 0$ for all $x, y \in X$. This follows from the triangle inequality applied to x, y, x:

$$0 = d(x,x) \le d(x,y) + d(y,x) = 2d(x,y).$$

A set S of elements of a metric space (X,d) is *located* if for every approximation $c \in Q$ the set $\{ d(x,c) \mid x \in S \}$ is order located below. This is to say that for every $c \in Q$ and all $a, b \in \mathbb{Q}$ with $a < b$, either $a \le d(x,c)$ for all $x \in S$ or else $d(x,c) \le b$ for some $x \in S$.

Lemma (Bishop's Lemma [6, p.92])**.** *Let A be a complete located inhabited subset of a metric space X and c an approximation of a point in X. Then we can find a point $y \in A$ such that $d(c,y) > 0$ entails $d(c,A) > 0$.*

11. Ordinary differential equations

11.1. **The Cauchy-Euler approximation method.** We consider a differential equation

$$y' = f(x,y), \tag{25}$$

where $f\colon D \to \mathbb{R}$ is a continuous function on some subset D of $\mathbb{R}^2$. A solution of (25) on an interval I is a function $\varphi\colon I \to \mathbb{R}$ with a continuous derivative φ' such that for all $x \in I$

$$(x, \varphi(x)) \in D \text{ (hence } f(x,\varphi(x)) \text{ is defined) and } \varphi'(x) = f(x, \varphi(x)).$$

We want to construct approximate solutions to (25). Let $f\colon D \to \mathbb{R}$ be continuous, and consider an interval I. A function $\varphi\colon I \to \mathbb{R}$ is an *approximate solution up to the error* 2^{-k} of (25) if

(a) φ is *admissible*, i.e., $(x, \varphi(x)) \in D$ for $x \in I$.
(b) φ is continuous.
(c) φ has a piecewise continuous derivative on I.
(d) $|\varphi'(x) - f(x, \varphi(x))| \le 2^{-k}$ for all $x \in I$ where $\varphi'(x)$ is defined.

Notice that we only required the differential equation (25) to be satisfied up to the error 2^{-k}. Later we shall see that under certain conditions which guarantee a unique exact solution, every approximate solution differs from the exact one by a constant multiple of its error.

Theorem (Cauchy-Euler approximation)**.** *Let* $f\colon D \to \mathbb{R}$ *be continuous, and* $(a_0, b_0) \in D$ *such that the rectangle* R *given by* $|x - a_0| \le a$*,* $|y - b_0| \le b$ *is in* D*. Assume* $|f(x, y)| \le M$ *for* $(x, y) \in R$*, and let* $h := \min(a, b/M)$*. Then for every* $k \in \mathbb{N}$ *we can construct an approximate solution* $\varphi\colon [a_0 - h, a_0 + h] \to \mathbb{R}$ *of* (25) *up to the error* 2^{-k} *such that* $\varphi(a_0) = b_0$*.*

Proof. By definition of h, the rectangle

$$S\colon\ |x - a_0| \le h,\ |y - b_0| \le Mh$$

is in D. f is continuous, hence comes with a modulus of (uniform) continuity; so for our given k we have an l such that

$$|f(\tilde{x}, \tilde{y}) - f(x, y)| \le 2^{-k-1} \tag{26}$$

for $(\tilde{x}, \tilde{y}), (x, y)$ in D and $|\tilde{x} - x|, |\tilde{y} - y| \le 2^{-l}$.

We now divide the interval $[a_0, a_0 + h]$ such that

(a) $a_0 < a_1 < \cdots < a_{n-1} < a_n = a_0 + h$
(b) $a_i - a_{i-1} \le \min(2^{-l}, 2^{-l}/M)$ for $i = 1, \dots, n$

and construct an approximate solution on $[a_0, a_0 + h]$; similarly this can be done on $[a_0 - h, a_0]$.

The idea is to start at (a_0, b_0) and draw a line with slope $f(a_0, b_0)$ until it intersects $x = a_1$, say at (a_1, b_1), then starting from (a_1, b_1) draw a line with slope $f(a_1, b_1)$ until it intersects $x = a_2$, say at (a_2, b_2), etc. Since we want an approximate solution which maps rationals to rationals, we approximate the slopes $f(a_{i-1}, b_{i-1})$ by rationals s_{i-1}.

More precisely, we recursively define for $i = 1, \dots, n$

$$\begin{aligned} &\varphi(x) = b_{i-1} + (x - a_{i-1})s_{i-1} \quad \text{for } a_{i-1} \le x \le a_i, \\ &b_i := \varphi(a_i), \\ &-M \le s_{i-1} \le M \quad \text{such that } |s_{i-1} - f(a_{i-1}, b_{i-1})| \le 2^{-k-1}. \end{aligned}$$

Clearly φ is continuous, admissible and has piecewise derivatives

$$\varphi'(x) = s_{i-1} \quad \text{for } a_{i-1} < x < a_i.$$

Now for $a_{i-1} < x < a_i$ we have $|x - a_{i-1}| \le 2^{-l}$ and

$$|\varphi(x) - b_{i-1}| \le |x - a_{i-1}| \cdot |s_{i-1}| \le \frac{2^{-l}}{M} \cdot M = 2^{-l},$$

hence by (26)

$$\begin{aligned}|\varphi'(x) - f(x,\varphi(x))| &= |s_{i-1} - f(x,\varphi(x))| \\ &\le |s_{i-1} - f(a_{i-1},b_{i-1})| + |f(a_{i-1},b_{i-1}) - f(x,\varphi(x))| \\ &\le 2^{-k-1} + 2^{-k-1} = 2^k.\end{aligned}$$

Hence φ is an approximate solution up to the error 2^{-k}. □

The approximate solutions we have constructed are *rational polygons*, i.e., piecewise differentiable continuous functions with rational corners and rational slopes.

Lemma (Rational polygons). *Given a rational polygon φ on $[a,b]$ and $c \in [a,b]$. Then one of the following alternatives will hold.*

(a) $0 \le \varphi(x)$ *for* $a \le x \le b$, *or*
(b) $\varphi(x) \le 0$ *for* $a \le x \le b$, *or*
(c) *there is a* $d < c$ *such that* $\varphi(d) = 0$ *and either* $0 \le \varphi(x)$ *for* $d \le x \le c$, *or else* $\varphi(x) \le 0$ *for* $d \le x \le c$.

Proof. Let $a = a_0 < a_1 < \cdots < a_n = b$ be the exception points for φ. We can locate c in $a = a_0 < a_1 < \cdots < a_n = b$. Pick i maximal such that $a_{i-1} \le c$. Compare $\varphi(a_0)$, $\varphi(a_1)$, $\varphi(a_{i-1})$ and $\varphi(c)$. If all have the same sign, we are done. Otherwise pick j maximal such that $\varphi(a_j)$ and $\varphi(a_{j+1})$ (or $\varphi(c)$, respectively) change sign. Then we are done as well. □

11.2. **The fundamental inequality.** We need an additional restriction in order to estimate the difference of approximate solutions. A function $f\colon D \to \mathbb{R}$, $D \subseteq \mathbb{R}^2$ is said to satisfy a *Lipschitz condition* w.r.t. its second argument for the constant $L > 0$, if for every $(x,y_1),(x,y_2) \in D$

$$|f(x,y_1) - f(x,y_2)| \le L|y_1 - y_2|.$$

Theorem (Fundamental inequality). *Let $f\colon D \to \mathbb{R}$ be continuous, and satisfy a Lipschitz condition w.r.t. its second argument for the constant $L > 0$. Let*

$$\varphi,\psi\colon [a,b] \to \mathbb{R}$$

be approximate solutions up to the error $2^{-k}, 2^{-l}$ of (25). Assume $\varphi \le \psi$ on $[a,b]$, or else that φ and ψ are rational polygons. Then

$$\big|\psi(x) - \varphi(x)\big| \le e^{L(x-a)}\big|\psi(a) - \varphi(a)\big| + \frac{2^{-k} + 2^{-l}}{L}\big(e^{L(x-a)} - 1\big)$$

for all $x \in [a,b]$. This is the fundamental inequality.

Proof. *Part* 1. $\varphi \leq \psi$ in $[a, b]$. Let $\sigma := \psi - \varphi$. By definition we have for all points except finitely many

$$\left|\varphi'(x) - f(x, \varphi(x))\right| \leq 2^{-k} \quad \text{and} \quad \left|\psi'(x) - f(x, \psi(x))\right| \leq 2^{-l},$$

hence with $\varepsilon := 2^{-k} + 2^{-l}$

$$\psi'(x) - \varphi'(x) \leq \left|f(x, \psi(x)) - f(x, \varphi(x))\right| + \varepsilon \leq L\bigl(\psi(x) - \varphi(x)\bigr) + \varepsilon$$

by the Lipschitz condition, hence

$$\sigma'(x) \leq L\sigma(x) + \varepsilon.$$

Therefore

$$\int_a^x e^{-Lt}\bigl(\sigma'(t) - L\sigma(t)\bigr)\,dt \leq \varepsilon \int_a^x e^{-Lt}\,dt.$$

The integrand on the left-hand side has finitely many discontinuities but a continuous indefinite integral, so

$$\begin{aligned} &\bigl[e^{-Lt}\sigma(t)\bigr]_a^x \leq \varepsilon\bigl[-\frac{1}{L}e^{-Lt}\bigr]_a^x \\ &e^{-Lx}\sigma(x) - e^{-La}\sigma(a) \leq \frac{\varepsilon}{L}\Bigl(e^{-La} - e^{-Lx}\Bigr) \\ &\sigma(x) \leq e^{L(x-a)}\sigma(a) + \frac{\varepsilon}{L}\bigl(e^{L(x-a)} - 1\bigr), \end{aligned}$$

which is the required inequality.

Part 2. φ and ψ are rational polygons. It suffices to prove the claim for a rational c. By the lemma on rational polygons in 11.1 we can assume $\varphi(d) = \psi(d)$ for some $d < c$, and $\varphi(x) \leq \psi(x)$ for $d \leq x \leq c$. Then by the first part

$$\psi(c) - \varphi(c) \leq \frac{2^{-k} + 2^{-l}}{L}\bigl(e^{L(c-d)} - 1\bigr),$$

which is an even stronger inequality. □

11.3. **Uniqueness.** To prove uniqueness of solutions we again need the Lipschitz condition.

Theorem (Uniqueness). *Let $f\colon D \to \mathbb{R}$ be continuous, and satisfy a Lipschitz condition w.r.t. its second argument. Let*

$$\varphi, \psi\colon I \to \mathbb{R}$$

be two (exact) solutions of (25). *If $\varphi(a) = \psi(a)$ for some $a \in I$, then $\varphi(x) = \psi(x)$ for all $x \in I$.*

Proof. We show $\varphi(x) = \psi(x)$ for $a \le x \le a + \frac{1}{2L}$, $x \in I$. Similarly this can be shown for $a - \frac{1}{2L} \le x \le a$; hence the claim follows.

Integrating the two equations

$$\varphi'(x) = f(x, \varphi(x)) \quad \text{and} \quad \psi'(x) = f(x, \psi(x))$$

we obtain from $\varphi(a) = \psi(a)$

$$\varphi(x) - \psi(x) = \int_a^x \bigl(f(t, \varphi(t)) - f(t, \psi(t))\bigr)\, dt$$

and hence

$$|\varphi(x) - \psi(x)| \le \int_a^x \bigl|f(t, \varphi(t)) - f(t, \psi(t))\bigr|\, dt \le L \int_a^x \bigl|\varphi(t) - \psi(t)\bigr|\, dt.$$

Let M be the supremum of the range of $|\varphi - \psi|$ on $[a, a + \frac{1}{2L}]$. Then for $a \le x \le a + \frac{1}{2L}$

$$|\varphi(x) - \psi(x)| \le L(x - a)M \le \frac{1}{2}M,$$

hence $M = 0$ and therefore $\varphi = \psi$ on $[a, a + \frac{1}{2L}]$. □

The example

$$y' = y^{1/3}, \quad y(0) = y_0. \tag{27}$$

shows that the Lipschitz condition is indeed necessary for uniqueness: for $y_0 = 0$ we have two solutions $\varphi(x) = 0$ and $\varphi(x) = (\frac{2}{3}x)^{3/2}$.

11.4. **Construction of an exact solution.** To prove the existence of an exact solution we again assume the Lipschitz condition.

Theorem (Exact solutions)**.** *Let $f\colon D \to \mathbb{R}$ be continuous, and satisfy a Lipschitz condition w.r.t. its second argument. Let $(a_0, b_0) \in D$ such that the rectangle R given by $|x - a_0| \le a$, $|y - b_0| \le b$ is in D. Assume $|f(x, y)| \le M$ for $(x, y) \in R$, and let $h := \min(a, b/M)$. Then we can construct an exact solution $\varphi\colon [a_0 - h, a_0 + h] \to \mathbb{R}$ of* (25) *such that $\varphi(a_0) = b_0$.*

Proof. By the Cauchy-Euler approximation theorem in 11.1 we have an approximate solution φ_n up to the error 2^{-n}, which is a rational polygon, over $I := [a_0 - h, a_0 + h]$.

By lemma UnifConvLim in 8.1 the sequence $(\varphi_n)_{n \in \mathbb{N}}$ uniformly converges over I to a continuous function φ. Hence the sequence $(f(x, \varphi_n(x)))_n$ of continuous functions uniformly converges over I to the continuous function $f(x, \varphi(x))$.[2] Therefore, by theorem IntLimit in 8.2

$$\lim_{n\to\infty} \int_a^b f(t, \varphi_n(t))\,dt = \int_a^b f(t, \varphi(t))\,dt.$$

We now prove that φ is an exact solution. By the choice of φ_n

$$\left|\varphi_n'(x) - f(x, \varphi_n(x))\right| \le 2^{-n}$$

for all $x \in I$ where $\varphi_n'(x)$ is defined. Integrating this inequality from a_0 to x gives

$$\Big|\int_{a_0}^x \big[\varphi_n'(t) - f(t, \varphi_n(t))\big]\,dt\Big| \le 2^{-n}(x - a_0) \le 2^{-n}h.$$

Since φ_n is continuous, by the fundamental theorem of calculus

$$\Big|\varphi_n(x) - \varphi_n(a_0) - \int_{a_0}^x f(t, \varphi_n(t))\,dt\Big| \le 2^{-n}h.$$

Approaching the limit for $n \to \infty$ gives

$$\varphi(x) - \varphi(a_0) - \int_{a_0}^x f(t, \varphi(t))\,dt = 0.$$

Differentiation yields $\varphi'(x) = f(x, \varphi(x))$, and $\varphi_n(a_0) = b_0$ entails $\varphi(a_0) = b_0$. □

For the construction above of an exact solution we have made use of the Lipschitz condition. However, it is well known that classically one has Peano's existence theorem, which does not require a Lipschitz condition.

Following [1, 8] we now want to argue that Peano's existence theorem entails that for every real x we can decide whether $x \ge 0$ or $x \le 0$, hence we cannot expect to be able to prove it constructively.

Consider again the initial value problem (27). First note that for $0 < a < y < b$ the continuous function $f(x, y) := y^{1/3}$ satisfies a Lipschitz condition w.r.t. its second argument (and similarly for $b < y < a < 0$). To see this, by the lemma in refSS:BoundSlope it suffices to find a bound on the derivative of $y^{1/3}$. But this is easy, since $\frac{d}{dy}y^{1/3} = \frac{1}{3}y^{-2/3} < \frac{1}{3}a^{-2/3}$ for $0 < a < y < b$.

[2] This is best proved by a slightly more general setup, where metric spaces (e.g., $\mathbb{R}^2$) are considered. One shows that if φ_n, ψ_n are uniformly convergent to φ, ψ, respectively, then (φ_n, ψ_n) is uniformly convergent to (φ, ψ), and if φ_n is uniformly convergent to φ, then $f(\varphi_n(x))$ is uniformly convergent to $f(\varphi(x))$.

Therefore in case $y_0 > 0$ the solution $\varphi_+(x) := (\frac{2}{3}x + y_0^{2/3})^{3/2}$ is unique, and similarly in case $y_0 < 0$ the solution $\varphi_-(x) := -(\frac{2}{3}x + |y_0|^{2/3})^{3/2}$ is unique. Pick $|y_0|$ small enough such that $(\frac{2}{3} - |y_0|^{2/3})^{3/2} > \frac{1}{2}$.

Now suppose that (27) has a solution φ, for a given real y_0. Compare $\varphi(1)$ with $[-1/2, 1/2]$. If $\varphi(1) < 1/2$, then $y_0 \not> 0$, hence $y_0 \le 0$. If $-1/2 < \varphi(1)$, then $y_0 \not< 0$, hence $y_0 \ge 0$.

We finally show that an approximate solution of (25) up to the error 2^{-k} differs from the exact solution by a constant multiple of 2^{-k}.

Theorem. *Let $f\colon D \to \mathbb{R}$ be continuous, and satisfy a Lipschitz condition w.r.t. its second argument. Let $(a_0, b_0) \in D$ such that the rectangle R given by $|x - a_0| \le a$, $|y - b_0| \le b$ is in D. Assume $|f(x, y)| \le M$ for $(x, y) \in R$, and let $h := \min(a, b/M)$. Assume further that we have an exact solution $\varphi\colon [a_0-h, a_0+h] \to \mathbb{R}$ of (25) such that $\varphi(a_0) = b_0$, that ψ is an approximate solution up to the error 2^{-k} such that $\psi(a_0) = b_0$, and that $\varphi \le \psi$ or $\psi \le \varphi$. Then there is a constant N independent of k such that $|\varphi(x)-\psi(x)| \le 2^{-k}N$ for $|x - a_0| \le h$.*

Proof. By the fundamental inequality

$$|\varphi(x) - \psi(x)| \le \frac{2^{-k}}{L}(e^{Lh} - 1).$$

Hence we can define $N := (e^{Lh} - 1)/L$. □

12. Notes

There are many approaches to exact real number computation in the literature. One of those - using Möbius (or linear fractional) transformations - has been put forward by Edalat; a good survey can be found in [9] (see also [10]). Exact real numbers based on the so-called redundant b-adic notation have been treated by Wiedmer [20] and by Böhm and Cartwright [7], and based on continued fractions by Gosper (1972), Vuillemin [19], Nielsen and Kornerup [17] and also by Geuvers and Niqui [14].

Generally, one can see these approaches as either using Cauchy sequences with (fixed or separately given) modulus, or else Dedekind cuts. We prefer Cauchy sequences over Dedekind cuts, since the latter are given by sets, and hence we would need additional enumerating devices in order to compute approximations of a real number presented as a Dedekind cut. There is not much of a difference between fixed or separate moduli for Cauchy sequences: one can always transform one form into the other. However, in order to keep the standard series representations of particular reals (like e) we prefer to work with separate moduli.

Another treatment (including an implementation in Mathematica) has been given by Andersson in his Master's thesis [2] (based on Palmgren's [18]). He treats trigonometric functions, and includes Picard's and Euler's methods to constructively prove the existence of solutions for differential equations.

Some authors (in particular the so-called Russian school) restrict attention to computable real numbers. We do not want to make this restriction, since it makes sense, also constructively, to speak about arbitrary sequences. This view of higher type computability is the basis of Scott/Ershov domain theory, and we would like to adopt it here.

However, the domain theoretic setting for dealing with exact real numbers (cf. Escardó, Edalat and [11]) is usually done in such a way that continuous functions are viewed as objects of the function domain, and hence are objects of type level 2. This clearly is one type level higher than necessary, since a continuous function is determined by its values on the rational numbers already. In particular from the point of view of program extraction it seems crucial to place objects as basic as continuous functions at the lowest possible type level. Therefore we propose a special concept of continuous functions, as type 1 objects.

Some of the (rather standard) calculus material, for instance in the section on sequences and series of real numbers, is taken form Forster's text [12]. The section on ordinary differential equations is based on Chapter 1 of Hurewicz's textbook [15], adapted to our constructive setting. I have also made use of a note of Bridges [8].

References

1. Oliver Aberth, *Computable analysis and differential equations*, in A. Kino and Vesley [16], pp. 47–52.
2. Patrik Andersson, *Exact real arithmetic with automatic error estimates in a computer algebra system*, Master's thesis, Mathematics department, Uppsala University, 2001.
3. Errett Bishop, *Foundations of constructive analysis*, McGraw-Hill, New York, 1967.
4. ———, *Mathematics as a numerical language*, in A. Kino and Vesley [16], pp. 53–71.
5. Errett Bishop and Douglas Bridges, *Constructive analysis*, Grundlehren der mathematischen Wissenschaften, vol. 279, Springer Verlag, Berlin, Heidelberg, New York, 1985.
6. Errett Bishop and Douglas Bridges, *Constructive analysis*, Grundlehren der mathematischen Wissenschaften, vol. 279, Springer Verlag, Berlin, Heidelberg, New York, Berlin, 1985.

7. Hans Boehm and Robert Cartwright, *Exact real arithmetic. formulating real numbers as functions*, Research Topics in Functional Programming, Addison Wesley, 1990, pp. 43–64.
8. Douglas Bridges, *A Note on the Existence of Solutions for ODEs*, Unpublished note, June 2003.
9. Abbas Edalat, *Exact real number computation using linear fractional transformations*, Tech. report, Imperial College, Dept. of Computing, 2003.
10. Abbas Edalat and Reinhold Heckmann, *Computing with real numbers: (i) LFT approach to real computations, (ii) domain-theoretic model of computational geometry*, Proc APPSEM Summer School in Portugal (G. Barthe et al., eds.), LNCS, Springer Verlag, Berlin, Heidelberg, New York, 2001.
11. Abbas Edalat and Dirk Pattinson, *Initial value problems in domain theory*, to appear: Computation in Analysis.
12. Otto Forster, *Analysis 1*, Vieweg, 1999, 5th.
13. ______, *Analysis 1*, 7th ed., Vieweg, 2004.
14. Herman Geuvers and Milad Niqui, *Constructive reals in coq: Axioms and categoricity*, Proc. Types 2000 (P. Callaghan, Z. Luo, J. McKinna, and R. Pollack, eds.), LNCS, vol. 2277, Springer Verlag, Berlin, Heidelberg, New York, 2000, pp. 79–95.
15. Witold Hurewicz, *Lectures on ordinary differential equations*, MIT Press, Cambridge, Mass., 1958.
16. Akiko Kino, John Myhill, and Richard E. Vesley (eds.), *Intuitioninism and proof theory*, Studies in Logic and the Foundations of Mathematics, North-Holland, Amsterdam, 1970.
17. Asger Munk Nielsen and Peter Kornerup, *MSB-first digit serial arithmetic*, Journal of Univ. Comp. Science **1** (1995), no. 7, 523–543.
18. Erik Palmgren, *Constructive nonstandard analysis*, Méthods et analyse non standard (A. Petry, ed.), vol. 9, Cahiers du Centre de Logique, 1996, pp. 69–97.
19. Jean Vuillemin, *Exact real computer arithmetic with continued fractions*, IEEE Transactions on Computers **39** (1990), no. 8, 1087–1105.
20. Edwin Wiedmer, *Computing with infinite objects*, Theoretical Computer Science **10** (1980), 133–155.

Index