Analysis via Uniform Error Bounds

Hermann Karcher, Bonn. Version May 2001

The standard analysis route proceeds as follows:

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

I propose the following reorganization of the material:

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

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

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

1. Derivatives of Polynomials

The elementary formula

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

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

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

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

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

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

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

$$\begin{aligned}&x \in [a-r, a+r], R := |a| + r \Rightarrow \\ &\left|(x^{k-1} + x^{k-2}a + \ldots + a^{k-1}) - ka^{k-1}\right| \leq k(k-1)/2 R^{k-2} \cdot |x-a|\end{aligned}$$

hence

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

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

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

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

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

2. Differentiation Rules, Derivatives of Rational Functions

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

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

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

Composition of polynomials with this one extra function gets us to all rational functions. Of course, the proof of the chain rule must pay attention to the distances from zeros which occur in the denominators.

In a similar way we handle the square root function. Here extra attention is needed for the domain of the function: As long as we only know rational numbers, the domain is rather thin, it contains only the squares of rational numbers (which, however, are still a dense subset). The following computation remains valid as more numbers become known.

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

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

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

All known proofs of the differentiation rules have the following property:
Given two differentiable functions f, g we abbreviate their tangent functions as
$l_f(x) := f(a) + f'(a)(x-a),\ l_g(x) := g(a) + g'(a)(x-a)$.
Then, if we assume that the differences $f - l_f,\ g - l_g$ are "small", then also the differences
$(\alpha \cdot f + \beta \cdot g) - (\alpha \cdot l_f + \beta \cdot l_g), \quad f \cdot g - l_f \cdot l_g, \quad f \circ g - l_f \circ l_g$
are in the same sense "small".
To prove something we have to specify "small". Because of the functions known so far we may say: $f - l_f$ is "**small**" means, there exists an interval $[a-r,\ a+r]$ and a constant K such that

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

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

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

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

3. Monotonicity Theorem and Related Results

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

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

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

Archimedes Principle

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

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

Monotonicity Theorem

Main assumption:

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

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

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

Then f is nondecreasing:

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

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

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

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

then the strategy of the following proof also works.

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

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

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

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

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

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

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

Immediate consequences are:

Generalized monotonicity:

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

Explicit bounds:

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

Multiplicative version:

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

Iterated application to second derivatives:

$$\begin{aligned} |f''| \leq B,\ a < x\ &\Rightarrow -B(x-a) \leq f'(x) - f'(a) \leq B(x-a) \\ &\Rightarrow \frac{-B}{2}(x-a)^2 \leq f(x) - f(a) - f'(a)(x-a) \leq \frac{B}{2}(x-a)^2. \end{aligned}$$

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

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

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

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

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

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

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

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

4. Completeness and Limits of Sequences of Functions

At most occasions I have prefered to begin the discussion of completeness with nested intervals; immediate applications are Leibniz series like the just mentioned Taylor polynomials. Thus we have interesting limit functions together with their derivatives right from the start.

Cauchy sequences are the second step; the most important applications are sequences or series which are dominated by the geometric series (i.e. power series and the contraction lemma). Existence of *sup* and *inf* for bounded nonempty sets gives the most compact formulation for dealing with completeness. I have used the standard text book arguments, only the applications change since the results of the earlier sections are of significant technical help.
Note. Since so much emphasis was placed on computable error bounds I mention that the theorem "*Monotone increasing, bounded sequences converge*" is a very significant exception. In my opinion this intuitively desirable theorem is a major justification of the standard limit definition: Because the convergence speed of monotone increasing, bounded sequences can be slowed down arbitrarily (namely by repeating the elements of the sequence more and more often) one **cannot** prove the monotone sequence theorem with any version of a limit definition which requires more explicit error control than the standard definition. By contrast, in numerical analysis one tries to use sequences which converge at least as **fast** as some geometric sequence, i.e. one has more explicit control like $|a_{n+p} - a_n| \le C \cdot q^n$.

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

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

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

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

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

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

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

As mentioned, the Contraction Lemma is also proved with explicit error bounds: Functions $f : M \to M$ with $|f'| \le q < 1$ are contracting, $|f(x) - f(y)| \le q \cdot |x-y|$. And for contracting maps sequences generated by iteration, $a_{n+1} := f(a_n)$, are geometrically dominated and hence Cauchy:

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

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

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

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

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

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

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

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

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

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

and Archimedes Principle extends this uniform estimate to the limit function

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

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

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

we use the Monotonicity Theorem to get uniform tangent approximations

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

and Archimedes Principle again extends these uniform bounds to the limit

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

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

5. Integrals, Riemann Sums, Antiderivatives

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

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

Riemann Sum

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

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

Error Bound

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

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

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

hence

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

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

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

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

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

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

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

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

Also other analogies between sums and integrals need to be discussed.
I describe a construction of an antiderivative which is independent of the definition of the integral (its proof, based on uniformity, works in Banach spaces).

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

$$s_n(t_j) := f(t_j),\ j = 0, \dots, n,\ (a = t_0 < t_1 < \dots < t_n = b)$$

$$s_n(x) := \frac{f(t_{j-1}) \cdot (t_j - x) + f(t_j) \cdot (x - t_{j-1})}{t_j - t_{j-1}} \quad \text{for } x \in [t_{j-1}, t_j]$$

$$|f(x) - s_n(x)| \le L \cdot \max |t_j - t_{j-1}| =: r_n$$

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

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

we have the (uniform) Cauchy property:

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

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

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

hence

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

Archimedes Principle gives the final result:

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

which says that S_∞ is differentiable with derivative f, as claimed.

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

6. Continuity, Theorems and Examples

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

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

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

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

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

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

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

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

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

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

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

Theorem: Derivative bounds are Lipschitz bounds.

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

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

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

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

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

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

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

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

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

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